KU LEUVEN

ARENBERG DOCTORAL SCHOOL
Faculty of Engineering Science

Proximal Algorithms for Structured Nonconvex Optimization

Andreas Themelis

Supervisors:
Prof. Panagiotis Patrinos
Prof. Alberto Bemporad

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

December 2018

Proximal Algorithms for Structured Nonconvex Optimization

Andreas THEMELIS

Examination committee:
Prof. Yves Willems, chair
Prof. Panagiotis Patrinos, supervisor
Prof. Alberto Bemporad, supervisor
Prof. Lieven De Lathauwer
Prof. Stefan Vandewalle
Prof. Russell Luke
(Universität Göttingen)
Prof. Yurii Nesterov
(UC Louvain)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

December 2018

© 2018 KU Leuven – Faculty of Engineering Science
Uitgegeven in eigen beheer, Andreas Themelis, Kasteelpark Arenberg 10, bus 2446, B-3001 Leuven (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm, electronic or any other means without written permission from the publisher.

To my parents, zia Mana, and all my family

Preface

I consider myself extremely lucky having met Prof. Panagiotis Patrinos (Panos) in IMT Lucca, and more than that for having been working with him. I am deeply and sincerely thankful for everything he so passionately taught me; his far-sighted intuition and exceptional mathematical rigor have always been inspirational and motivating. I feel privileged for having always had the chance to work on so interesting topics, knowing that Panos' ideas would be confirmed both in theory and in practice. For this, for his patience, and for the friendly working environment he always offered, I am deeply thankful.

I am also extremely thankful to Prof. Alberto Bemporad, who welcomed me in his group and walked me through my first steps into the PhD. He shared his vast experience in the field and with constant assistance he guided my first publication, knowing how to best fit my background knowledge.

I wish to express my most sincere gratitude to Prof. Russell Luke, Prof. Yurii Nesterov, Prof. Lieven de Lathauwer, Prof. Stefan Vandewalle and Prof. Mario Zanon for carefully reading my manuscript and for the many valuable suggestions that lead to this final version of the thesis.

I am very thankful to IMT Lucca and to the DYSCO group for having offered me a plesant and ideal working environment, and equally as much to KU Leuven and the STADIUS group for welcoming me first as a visitor and then as a joint PhD student. I could never thank enough Sara Olson, Daniela Giorgetti, Maria Mateos, and Serena Argentieri for their constant assistance throughout the entire PhD; their outstanding kindness and cordiality have been of huge support. Same goes for Justine, an angel in IMT who always cheered me up with her humor and post-modern spirituality.

I wish to thank my family who always supported me in every decision and since when I was a kid would always be proud about any result, no matter how bad or ridiculous. Thanks mom and dad, and thanks aunt Mana for guiding my steps into science (and sharing brick-related frustrations...), Lala Vi and brothers Leo (Sbrisjbri) Liza and Georgos, Stathoula, Pia and Dido, Baba and Dedo, you will always be my reference.

And there are those who although with different blood (racist allusion intended) I still consider family: Yamamoto Kaoru (Ass. Prof.), Shō and the バリ−ちゃんたち (とってもね); Yang Yu, Luo Bin and their handsome little toad Ziyang; Chen Liujie, wode gege Xu Ye and their little princess Zilan; and Hasegawa Mai chan, Yamato Junpei sama, ... And of course, yet unfortunately

with no racist implications, Franca and Michele, Giulia and Massimo with Bianca and Elena, Giuseppe and Marialuisa with little Gabriele, and Daniela with her German-speaking Antonio!

Overlooking her embarassing sense of duty when it comes to bringing unnecessary presents at any visit, speaking of family I cannot hide how privileged I am to have a sister like Monica. Her music always brings me to a dimension where problems cease to exist; but if I here thank Monica is for her sincere friendliness and the support she has always given me.

Well aware that infinite many words wouldn't be enough, I still wish to express my most sincere and heartfelt gratitude to Mariapaola Spadolini and Gianfranco Guidoboni for their immense generosity and delightful cordiality. Those pristine silky keys and warm notes gave me the strength to survive many challenges and difficult moments of the PhD.

Many other people took part in the chain of events that led me here since long before the starting of the PhD. Arturo Labriola and Tommaso Cobblestone Bertini, Silvia @Vivy87, Giovanni Il Giovine, Prisco Sullivan Oliva, Guido J̃ı̀(ğ)Łi, Matteo and Alessia and their struggle against cholesterol, Marika (my favorite #2) and Andrea Augusto Ronqui (too fancy a name not to write it whole), Dario and Benedetta, Marco Tombo and Riccardo, captain Philip Giof, my Mongol Rally team mate Samuele and Elisabetta (make sure to pay them a visit @ Agriturismo la Fontaccia, 25km out of Florence: agriturismolafontaccia.com!), the evergreen Vincenzo Schweppes and Stefania Procops, and the muse of acknowledgements Dario Masi (practicing Jūdō in Tōkyō) with Ilaria Fish&Chips.

I feel much obliged to acknowledge B-spline interpoler Andrea Mazzanti for guiding me into optimization back in undergrad, and Debora Mucci and Fulvio Gesmundo for easing my studies with their precious notes. Special thanks to Iro Dotsikas, Davide Poggiali, Maikol Borsetti, Eugenio Giannelli and our common friend Francesco for the great memories of that period.

The rest of the acknowledgement is restricted to people mentioned only because I have to, sorted by irrelevance.

I wouldn't be writing these lines hadn't I met Lorenzo Puya and Pantelis: not only am I not accustomed to acknowledge unknown names, but I can't imagine how I could have possibly made it till the end of the PhD without the constant support of their strong mathematics and exceptional code skills. In full truth, one gets also credit for the British humor and someone else only gets the math acknowledgments; better known as Felafel (for apparent reasons) and often accompanied by adjectives to remind people how enlighted he is, noble-hearted Puya would always nurture the confidence in a prosperous PhD by welcoming my Belgian mornings with a reassuring *"You suck!"*, let alone his altruistic commitment to assist cycling impaired colleagues. But if my PhD found a purpose at all I owe it to Ilkay, who would constantly remind me how relieved he was when looking at me, for his problems seemed nothing against how miserable my life was. And yet it is a mystery how our profound late night talks, always starting from heated social and political debates, would inevitably end up with praises to some M.lle Privat (whom I also acknowledge bien sûr)! Of course all my gratitude to Eylül for her charity services.

First of a long and prestigious Kumar dynasty and by far the most natural Italian news reader, my enthusiastic pingpong mate Ajay introduced me to the thrills of waking up at 5am to explore new dimensions of entertainment. I will never forget how inspirational our erudite musical taste was for the entire IMT DYSCO group!

हिंदी भारत की राष्ट्रीय भाषा है *(Ajay Kumar Sampathirao)*

Barely acquainted with the single honorific, the coming of my soon-to-be Leuven flatmate has been one of the most shocking experiences I recall. Double Kumar — so he demanded people to address him — in parallel to his Bachelor-honored PhD studies would run important economical transactions in the shade of the Gandhi Boulevard headquarters of Italian banks. Thankful to those people around him who would compensate the limits of his non-scissorproof credit card, he would attribute them the authorship of romantic text messages composed in inspired Belgian nights. Domagoj would patiently tolerate and rather save his efforts for sentimental liaisons targeting daughters of wealthy bakers or automatic laundry owners, although sometimes bursting in rage with "Zitto Felafel!" echoing in Leuven. And while Vihang prides himself of being the Kumarest, no human will ever possibly outdilip Manas Mejari. After days of intense (and tight!) Danish cohabitation with huggable Valentina (true name omitted for decency concerns), our group introduced and tested in the field

new efficient paradigms of IEEE-formatted postcards. Although dolphin lover Marina Andrić would strongly disagree — for true art, she believes, is the kind of postcards *she* receives instead — this is arguably the biggest joint contribution achieved on a night train.

During these years I had the fortunate chance to get to know people from all around the world and from very different backgrounds. I still can't hold the tears when I think of all those students that despite their disadvantaged SES still manage to accomplish a successful and honored PhD. It only consoles me that the Flemish government is generous enough to support the struggle of KW (full name omitted to avoid retaliations from the family) against the prearranged farmwork envisioned by the parents, and that phone companies would know how best to reward her extensive work-related surfings with premium memberships to leading companies of the field of interest. On the opposite extreme, never could I imagine a wealth in the likes of Swiss-Chinese Qinghua's even existed! Arguably the sexiest 最胖小宝贝儿 (relative to her sense of beauty and elegance), additionally to having been once the first in China (as she would confess on first meetings) thanks to her undeniable smartness and maturity Miss University was unanimously elected chair shortly after joining the KU group, a position which she *comfortably* held till the very last day. More in favor of a gender-quota-based career, at the cost of betraying her solid sobriety pledge, envious colleague Lynn 𓀠𓁹𓂝𓃰𓄿𓅱𓆑 (pron. /'hɒthaʊs/) would always belittle the rich lady by naming her after a beer from the green island of Shandong. Lynn would also know when and to whom best to tell jokes; honoring the memory of Ricardo, she would start with his favorite *"There is a black guy..."*. Yet none would deny Lynn is the bestest ever, and this opinion has clearly nothing to do with how convenient her station wagon is when it comes to changing apartment. And among all these immigrants it is a relief to have a Leuven compatriot in the next office: Zahra, a typical Flemish lady, wouldn't join lunch or coffee break unless properly summoned by knocking on the wall. Admittedly spoilt by the charitable neighbors who would often help her typing on the keyboard, at last she convinced her promotor to overlook the capacity of her office and be assigned a deskmate. And speaking of kind neighbors I can't refrain from praising Francis Hilda; cleverly disguised as a sheer Schuurman at first, only after the arrival of some rich Indian did he gradually disclose his generous instinct of offering Duvels. Being not a 3-to-26.5 year old pregnant lady, whether or not the cunning-humored wealthy Prince of Punjabi is worth such favors is not for me to debate. Arguably not a proper guy and definitely not lsc, we however all agree that he cooks well. The distinct contrast with humble stableman Mathijs hurts our sensitive eyes; tumbled between industry and academy, he is a blatant example of the distinct separation between theory and practice, for no truly risk averse car owner would ever lend his vehicle for Swedish housing purposes. And then there are Yinxue (who made us totally forget about her predecessor

King Hua, so we are told his or her name sounded like), Arun and Marcin, who however haven't had enough time to compromise themselves in the STADIUS history (yet). And congratulations Michael for the arrival of Raphaël!

At the center of the universe to make it spin — thus right in the middle of the section are They mentioned — life itself wouldn't be if not for the allmighty Presijdent of all that is presijdentiable Parijsi, sided by Their vice cattle obstetrician and First Gentleman Lars the Great. By far the most generous emperor ever existed, Daniele would leave to his Schrijnmakersstraat subject(s) the pleasures of impersonating buckets puring on the ground, and would make sure they received the so longed celebrations on birthday and other scomode occasions. Never shall They run out of Grana Padano thanks to Hello Kaity Valentijna, and no credit shall Martijn be praised with for pronouncing Gasthuisberg so long as she lives. Among other faithful servants, Giulia and her risotto would make the authour gain 2.5Kg in one night under the psychological supervision of food disease expert Giorgia, tonari no Agnese (known to some as Panny) would know how best to entertain guests with weapons cleverly disguised as cooking tools, Alessandra (Petty) by sitting in her ladily positions would offer innovative solutions to the cold Belgian weather to those in front of her, and at the same time from the basement to the rooftops people could track the movements of Mario (Brambi) by following the echo of his mild tones and imperceptible Milan accent.

The turning point of my PhD can arguably be identified with the advent of Masoud. Proud of his outstanding islamic impact factor (that I always craved for) and after leveraging on asymmetric divergences to establish himself in the research group, Aghaye Ghaderi with clever communication skills introduced me to the world of the Rakhbari, filling me and my academic production with spiritual meaning. But it was the arrival of Khanome Ghaderi the event that brought a definite change and was most appreciated by all STADIUS members, for at last in the Boss era we all know exactly how long an after lunch break should last, differently from the open-ended propositions of neutral-accented Bottegal, always starting with a *"Coffeeeee?"* and followed by ritual bullism against the weakest (-minded) of his herd. I feel much obliged to acknowledge Bottegal for acknowledging me first, for my (modestly speaking) fluency in Veneto dialects and mastery of sheep breeding are now worldwide established in the scientific community of Researchgate. Known to the most as Prosciuttoressa (better not to investigate on the embarassing origin of the name), true friend Federica from the false friend Vedelago would never find the courage to press charges against the abusive partner who imposes on others his arguable passion for chugging sparkling vinegar. And for how cute her twin Cipo is, I will never hide my preference for the cuter Ciottoli. The image of sheep herds brings my memories back to

Bertrand and our Championnat de la Bergerie, his praises to the only person worth being complimented not as stupid as he looks and our rescue missions in the department ducts 5 meters from the ground.

Admittedly with a pretentious nuance, I can't hide how honored I am to have assisted the lectures of Prof. Borgioli, who thanks to his scientific merits climbed through all academic degrees in few days. I can also proudly declare having shaked hands with flagship of Julia community and worldwide streamed Ph.D. Antonello.

A special thank to Anita and Davide, his homonyms Boschezza and D'Arenzo, Rita, Emi, Olympia, Vasilis, Rafa and Yeshim, grandpa Carollo and the other soft-worded leaf enumerator Valerio, Yahia, Laura Janφy and Dem, Manuela, Chiara and many others for the best memories in IMT.

I am also very thankful to the Lund group, including Prof. Chakraborrty of course, and more locally Mattias, Richard, Gautham and Chris, and the amazingly interesting seminars of Martinka on defects of semiconductors. And how can't I mention my beloved Ukranian wife? For instance by deliberately choosing to acknowledge my Jordan husband Sara instead, whom I could never thank enough for bravely shielding me from the threats of other men. And how is it possible to summarize in few words a magnificent specimen such as Vig, the only living being allergic to Vietnamese eggs (how lucky Van Tien was conceived the western way!), or the adventures with Guillerme and his favorite car rental companies?

In fact, I can't possibly cover all the people that deserve an acknowledgment, and each of them in full fairness should be dedicated an entire thesis. Let me simply say that if there is any one here tonight whom I have not offended, I apologize!

Contents

Preface vii

Abstract xix

List of symbols xxi

List of abbreviations xxiv

Vita & publications xxvii

1 Introduction 1

1.1 Contributions and structure the thesis . . . 3

1.2 Preliminary material . . . 6

1.2.1 Matrices and vectors . . . 6

1.2.2 Sequences . . . 7

1.2.3 Extended-real-valued functions . . . 8

1.2.4 Self-mappings . . . 9

1.2.5 Set-valued mappings . . . 9

1.2.6 Subdifferential . . . 10

1.2.7 (Hypo)convexity . . . 11

1.2.8 Smoothness . . . 12

1.2.9 Proximal map and Moreau envelope . . . 18

1.2.10 Image function . . . 22

2 A general framework for the analysis of nonconvex splitting algorithms 27

2.1 Analysis of fixed-point iterations . . . 27

2.2 Fixed-point iterations in optimization . . . 31
2.3 Proximal majorization-minimization . . . 33
2.3.1 Proximal majorizing models . . . 34
2.3.2 Properties . . . 35
2.3.3 Partial ordering . . . 37
2.4 Criticality . . . 39
2.5 Generalized proximal majorization-minimization . . . 42
2.6 Representation of proximal algorithms . . . 44
2.6.1 Notational conventions . . . 45
2.6.2 The criticality threshold . . . 46

3 Proximal envelopes **48**
3.1 Majorization-minimization value functions . . . 48
3.2 Properties . . . 50
3.2.1 Inequalities . . . 50
3.2.2 Equivalence . . . 52
3.2.3 Regularity . . . 54
3.2.4 The KL property . . . 55
3.3 Lyapunov functions for proximal algorithms . . . 60
3.3.1 Sufficient decrease: a priori estimates . . . 60
3.4 Convergence of GPMM algorithms . . . 63

4 Acceleration of nonconvex splitting algorithms **68**
4.1 A new backtracking paradigm . . . 68
4.2 The CLyD algorithmic framework . . . 70
4.3 Choice of directions . . . 74
4.3.1 (L-)BFGS . . . 75
4.3.2 A modified Broyden scheme . . . 76

4.3.3 Anderson acceleration 76
4.4 Global and (super)linear convergence 77
4.5 Superlinear convergence 78

5 Forward-backward splitting **83**
5.1 Introduction 83
5.2 FBS as a PMM algorithm 84
5.3 Forward-backward envelope 86
5.3.1 Regularity properties 87
5.3.2 First-order differentiability 88
5.3.3 Second-order differentiability 92
5.4 Convergence results 95
5.5 A quasi-Newton FBS 97
5.5.1 Global and (super)linear convergence 99
5.6 Simulations 101
5.6.1 Dictionary learning 101
5.6.2 Nonconvex sparse approximation 103

6 Douglas-Rachford splitting **106**
6.1 Introduction 106
6.2 DRS as a GPMM algorithm 109
6.3 Douglas-Rachford envelope 111
6.3.1 Regularity properties 112
6.3.2 The DRE as a Lyapunov function 113
6.4 Convergence results 117
6.4.1 Tightness of the ranges 120
6.5 A quasi-Newton DRS 122
6.5.1 Global and (super)linear convergence 123

7 Alternating direction method of multipliers **126**

7.1 Introduction . . . 126

7.1.1 Overview on nonconvex ADMM . . . 127

7.2 A universal equivalence of ADMM and DRS . . . 129

7.2.1 An unconstrained problem reformulation . . . 129

7.2.2 From ADMM to DRS . . . 130

7.3 Convergence results . . . 132

7.4 Sufficient conditions . . . 137

7.4.1 Lower semicontinuity . . . 137

7.4.2 Smoothness . . . 139

7.5 A quasi-Newton ADMM . . . 142

7.6 Simulations . . . 144

7.6.1 Sparse principal component analysis . . . 144

8 SuperMann **148**

8.1 Introduction . . . 148

8.1.1 Contributions . . . 149

8.1.2 Chapter organization . . . 149

8.2 Motivating examples . . . 150

8.3 Notation and known results . . . 153

8.3.1 Hilbert spaces and bounded linear operators . . . 153

8.3.2 Nonexpansive operators and Fejér sequences . . . 154

8.4 General abstract framework . . . 155

8.4.1 Global weak convergence . . . 157

8.4.2 Local linear convergence . . . 159

8.4.3 Main idea . . . 163

8.5 Generalized Mann Iterations . . . 164

8.5.1 The classical Krasnosel'skiĭ-Mann scheme 164
8.5.2 Generalized Mann projections 166
8.5.3 Line search for GKM 167
8.6 The SuperMann scheme 169
8.6.1 Global and linear convergence 170
8.6.2 Superlinear convergence 171
8.6.3 The modified Broyden scheme 175
8.6.4 Parameters selection in *SuperMann* 177
8.6.5 Comparisons with other methods 178
8.7 Simulations 180
8.7.1 Cone programs 180
8.7.2 Lasso 182
8.7.3 Constrained linear optimal control 184

Conclusions **188**

Future directions 189

Bibliography **194**

Abstract

Due to their simplicity and versatility, splitting algorithms are often the methods of choice for many optimization problems arising in engineering. "Splitting" complex problems into simpler subtasks, their complexity scales well with problem size, making them particularly suitable for large-scale applications where other popular methods such as IP or SQP cannot be employed.

There are, however, two major downsides: 1) there is no satisfactory theory in support of their employment for nonconvex problems, and 2) their efficacy is severely affected by ill conditioning. Many attempts have been made to overcome these issues, but only incomplete or case-specific theories have been established, and some enhancements have been proposed which however either fail to preserve the simplicity of the original algorithms, or can only offer local convergence guarantees.

This thesis aims at overcoming these downsides. First, we provide novel tight convergence results for the popular DRS and ADMM schemes for nonconvex problems, through an elegant unified framework reminiscent of Lyapunov stability theory. "Proximal envelopes", whose analysis is here extended to nonconvex problems, prove to be the suitable Lyapunov functions. Furthermore, based on these results we develop enhancements of splitting algorithms, the first that 1) preserve complexity and convergence properties, 2) are suitable for nonconvex problems, and 3) achieve asymptotic superlinear rates.

List of symbols

$\mathbf{1}$	Vector of suitable size with all elements equal to 1	6
$\mathbf{1}_n$	$\mathbb{R}^n$ vector with all elements equal to 1	6
$(a^k)_{k\in K}$	Sequence indexed by elements of the set K	7
$\mathscr{A}_\lambda$	GPMM algorithm $\mathscr{A}$ with relaxation λ	45
$A^\top$	Transpose of matrix A	7
$\mathrm{B}(x;r)$	Open ball centered at x with radius r	6
$\overline{\mathrm{B}}(x;r)$	Closed ball centered at x with radius r	6
$\mathcal{B}(\mathcal{H})$	Bounded linear operators $\mathcal{H}\to\mathcal{H}$	154
$\operatorname{bdry} E$	Boundary of set E	6
$C^{1,1}(\mathbb{R}^n)$	Differentiable functions $\mathbb{R}^n\to\mathbb{R}$ with Lipschitz gradient	12
$C^k(\mathbb{R}^n)$	k times continuously differentiable functions $\mathbb{R}^n\to\mathbb{R}$	12
$C^{k+}(\mathbb{R}^n)$	Functions $C^k(\mathbb{R}^n)$ with locally Lispchitz k-th derivative	12
(Ch)	Image function of C and h	23
$\operatorname{cl} E$	Closure of set E	6
$\operatorname{conv} E$	Convex hull of E	11
δ_S	Indicator function of set S	8
$\delta_{i,j}$	Kronecker symbol	6
$DR(\bar{x})$	Semiderivative of R at $\bar{x}$	12
$\mathrm{d}^2h(\bar{x}\|v)[d]$	Second-order epi-derivative of h at x for v along direction d	92
$\operatorname{diag} v$	Diagonal matrix with elements of vector v on the diagonal	6
$\operatorname{dist}(x,S)$	Distance of x from S	10
$\operatorname{dom} h$	Domain of extended-real- or set-valued mapping h	8, 9
$\operatorname{epi} h$	Epigraph of extended-real-valued function h	8
$\operatorname{fix} F$	Fixed set of (set-valued) mapping F	10
$\mathcal{F}^\lambda$	λ-relaxation of set-valued mapping $\mathcal{F}$	43
$\mathcal{F}_\gamma^{\mathscr{A}_\lambda}$	Fixed-point mapping of GPMM algorithm $\mathscr{A}$ with stepsize γ and relaxation λ	45
$\varphi_\gamma^{\mathscr{A}}$	$\mathcal{F}$-envelope relative to GPMM algorithm $\mathscr{A}$ with stepsize γ	50
$h : A\to B$	Single-valued function	8
$H : A\rightrightarrows B$	Set-valued mapping	9
$\Gamma^{\mathscr{A}}$	Criticality threshold of GPMM algorithm $\mathscr{A}$	46

γ_h	Prox-boundedness threshold of h	18
$\operatorname{gph} H$	Graph of extended-real function or set-valued mapping H	9
$h'(x;d)$	directional derivative of h at x along d	13
h^γ	Moreau envelope of h	18
I	Identity matrix of suitable size	6
I_n	Identity $n \times n$ matrix	6
id	Identity mapping	9
$\operatorname{int} E$	Interior of set E	6
$JR(\bar{x})$	Jacobian matrix of R at $\bar{x}$	12
$\ker A$	Kernel (null space) of matrix A	6
ℓ^1	Set of summable sequences	7
ℓ^2	Set of square-summable sequences	7
L_h	Lipschitz modulus of ∇h, for $h \in C^{1,1}(\mathbb{R}^n)$	12
$L_{h,C}$	Lipschitz modulus of ∇h relative to matrix C	139
L^*	Adjoint of linear operator L	154
$\lambda_{\max}(H)$	Maximum eigenvalue of $H \in \operatorname{Sym}(\mathbb{R}^n)$	7
$\lambda_{\min}(H)$	Minimum eigenvalue of $H \in \operatorname{Sym}(\mathbb{R}^n)$	7
$\operatorname{lev}_{\leq \alpha} h$	α-sublevel set of extended-real-valued function h	8
$\overline{\mathcal{M}}_0$	Maximal majorizing model	33
$\mathfrak{M}_\varphi$	Family of proximal majorizing models	34
$\overline{\mathfrak{M}}_\varphi$	Family of majorizing models	33
$\mathcal{M}_\gamma^{\mathscr{A}}$	PMM model of GPMM algorithm $\mathscr{A}$ with stepsize γ	44
$\varphi^{\mathcal{M}}$	$\mathcal{M}$-envelope relative to PMM model $\mathcal{M}$	49
$\mathbb{N}$	Natural numbers $\{0, 1, 2, \cdots\}$	6
$\|\cdot\|$	Euclidean or matrix norm	7
$\|\cdot\|_p$	ℓ_p (semi)norm, for $p \in [0, \infty]$	7
$\|\cdot\|_Q$	(Semi)norm induced by $Q \in \operatorname{Sym}_+(\mathbb{R}^n)$	7
$O(\,\cdot\,)$	Big-O infinitesimal Bachmann-Landau notation	8
$o(\,\cdot\,)$	Small-o infinitesimal Bachmann-Landau notation	8
$[r]_+$	Positive part of r: $\max\{0, r\}$	6
$[r]_-$	Negative part of r: $\max\{0, -r\}$	6
Π_S	(Set-valued) projection onto S	10
$\operatorname{prox}_{\gamma h}$	(Set-valued) proximal mapping of h	18
$\mathbb{R}$	Real numbers $(-\infty, \infty)$	6
$\mathbb{R}_+$	Positive reals $[0, \infty]$	6
$\mathbb{R}_{++}$	Strictly positive reals $(0, \infty]$	6
$\overline{\mathbb{R}}$	Extended-real numbers $(-\infty, \infty]$	6
$\operatorname{range} A$	Range (column span) of matrix A	6
$\operatorname{rank} A$	Rank of matrix A	7
$\mathcal{R}_\gamma^{\mathscr{A}}$	Residual mapping of GPMM algorithm $\mathscr{A}$ with stepsize γ	45
σ_h	Hypoconvexity modulus of $h \in C^{1,1}(\mathbb{R}^n)$	14
$\sigma_{h,C}$	Hypoconvexity modulus of h relative to matrix C	139

∂h	(Limiting) subdifferential of h	10
$\partial_B h$	Bouligand subdifferential of h	10
∂_C	Clarke generalized Jacobian	152
$\partial^\infty h$	Horizon subdifferential of h	10
$\hat{\partial} h$	Regular subdifferential of h	10
$\nabla h\ \nabla^2 h$	(Classical) gradient and Hessian of h	10
$\prec\ \preceq\ \succeq\ \succ$	Partial order relations in $\mathrm{Sym}(\mathbb{R}^n)$ and $\overline{\mathfrak{M}}_\varphi$	7, 37
$\mathrm{Sym}(\mathbb{R}^n)$	Symmetric $n \times n$ real matrices	7
$\mathrm{Sym}_+(\mathbb{R}^n)$	Symmetric $n \times n$ positive semidefinite real matrices	7
$\mathrm{Sym}_{++}(\mathbb{R}^n)$	Symmetric $n \times n$ positive definite real matrices	7
$\mathcal{T}^{\mathcal{M}}$	PMM mapping of PMM model $\mathcal{M}$	35
$\mathcal{T}^{\mathscr{A}}_\gamma$	Shorthand for $\mathcal{T}^{\mathcal{M}^{\mathscr{A}}_\gamma}$	45
$\mathcal{W}$	Weak sequential cluster points	154
$\mathbb{Z}$	Integer numbers $\{0, \pm 1, \pm 2, \cdots\}$	6
$\operatorname{zer} F$	Zeros of (set-valued) mapping F	10

List of abbreviations

ADMM	Alternating direction method of multipliers
AFBA	Asymmetric forward-backward-adjoint
AMM	Alternating minimization method
CG	Conjugate gradient
CLyD	Continuous-Lyapunov descent framework (Alg. 4.1)
DRE	Douglas-Rachford envelope
DRS	Douglas-Rachford splitting
FBE	Forward-backward envelope
FBS	Forward-backward splitting
FFBS	Fast forward-backward splitting
FNE	Firmly nonexpansive
FP	Fixed point
GKM	Generalized Krasnosel'skiĭ-Mann
GPMM	Generalized proximal majorization minimization
iff	If and only if
KL	Kurdyka-Łojasiewicz
KM	Krasnosel'skiĭ-Mann
lsc	Lower semicontinuous
MM	Majorization-minimization
NE	Nonexpansive
osc	Outer semicontinuous
PMM	Proximal majorization minimization
PPA	Proximal point algorithm
PRS	Peaceman-Rachford splitting
QP	Quadratic program
SCS	Splitting conic solver

SPCA	Sparse principal component analysis
SQP	Sequential quadratic programming

Vita

Mach 8, 1988	Born, Florence, Italy
2006–2010	B.Sc. in Mathematics Final mark: 110/110 cum laude University of Florence, Italy
2010–2013	M.Sc. in Mathematics Final mark: 110/110 cum laude University of Florence, Italy
Since 2013	Ph.D. in Computer, Decision and Systems Science IMT School for Advanced Studies Lucca, Italy
2015–2016	Visiting student KU Leuven, Belgium ESAT — Department of Electrical Engineering
Since 2016	Ph.D. student jointly at KU Leuven, Belgium ESAT — Department of Electrical Engineering

Publications

[117] **A. Themelis**, M. Ahookhosh and P. Patrinos.
On the acceleration of forward-backward splitting via an inexact Newton method.
(to appear as book chapter in *Splitting Algorithms, Modern Operator Theory, and Applications*, Springer)
https://arxiv.org/abs/1811.02935

[107] A. Sathya, P. Sopasakis, R. Van Parys, **A. Themelis**, G. Pipeleers and P. Patrinos,
"Embedded nonlinear model predictive control for obstacle avoidance using PANOC,"
2018 European Control Conference (ECC), Limassol, 2018 (to appear)
https://lirias.kuleuven.be/handle/123456789/617689

[115] L. Stella, **A. Themelis**, P. Sopasakis and P. Patrinos,
"A simple and efficient algorithm for nonlinear model predictive control,"
2017 IEEE 56th Annual Conference on Decision and Control (CDC), Melbourne, VIC, 2017, pp. 1939–1944.
http://ieeexplore.ieee.org/document/8263933/

[110] P. Sopasakis, **A. Themelis**, J. Suykens and P. Patrinos,
"A primal-dual line search method and applications in image processing,"
2017 25th European Signal Processing Conference (EUSIPCO), Kos, 2017, pp. 1065–1069.
http://ieeexplore.ieee.org/document/8081371/

[119] **A. Themelis** and P. Patrinos.
Douglas-Rachford splitting and ADMM for nonconvex optimization: tight convergence results.
(under 2nd review round in the SIAM Journal of Optimization since November 2018)
https://arxiv.org/abs/1709.05747

[118] **A. Themelis** and P. Patrinos.
SuperMann: a superlinearly convergent algorithm for finding fixed points of nonexpansive operators.
(under 2nd review round in the IEEE Transactions on Automatic Control journal since March 2018)
https://arxiv.org/abs/1609.06955

[114] L. Stella, **A. Themelis** and P. Patrinos.
Newton-type alternating minimization algorithm for convex optimization.
IEEE Transactions on Automatic Control **64**(2), February 2019 (to appear)
https://ieeexplore.ieee.org/document/8472357

[120] **A. Themelis**, L. Stella and P. Patrinos.
Forward-backward envelope for the sum of two nonconvex functions: further properties and nonmonotone linesearch algorithms,
SIAM Journal on Optimization 2018 **28**(3):2274-2303, 2018.
https://epubs.siam.org/doi/10.1137/16M1080240

[113] L. Stella, **A. Themelis** and P. Patrinos.
Forward-backward quasi-Newton methods for nonsmooth optimization problems,
Computational Optimization and Applications (2017) **67:**443.
http://link.springer.com/article/10.1007/s10589-017-9912-y

[121] **A. Themelis**, S. Villa, P. Patrinos and A. Bemporad,
"Stochastic gradient methods for stochastic model predictive control,"
2016 European Control Conference (ECC), Aalborg, 2016, pp. 154-159.
http://ieeexplore.ieee.org/document/7810279/

Selection of talks without proceedings

1. *Proximal envelopes.*
   ECC 2018 Workshop on "Advances in Distributed and Large-Scale Optimization," Limassol (Cyprus), Jun. 12-15, 2018.
   http://www.ecc18.eu/index.php/workshop-6/
2. *Newton-type operator splitting algorithms.*
   EUCCO 2016: 4th European Conference on Computational Optimization, Leuven (Belgium), Sep. 12-14, 2016.
   https://kuleuvencongres.be/eucco2016/programme
3. *A variable metric stochastic aggregated gradient algorithm for convex optimization.*
   EURO 2016: 28th European Conference on Operational Research, Poznan (Poland). Jul. 3-6, 2016.
   https://www.euro-online.org/conf/euro28/edit_session?sessid=154
4. *A Globally and Superlinearly Convergent Algorithm for Finding Fixed Points of Nonexpansive Operators.*
   CORE@50: Center for Operations Research and Econometrics Conference, Louvain la Neuve (Belgium). May 23-27, 2016

Chapter 1

Introduction

Operator splitting techniques (also known as proximal algorithms), introduced in the 50's for solving PDEs and optimal control problems, have been successfully used to reduce complex problems into a series of simpler subproblems. The most well-known operator splitting methods are the alternating direction method of multipliers (ADMM), forward-backward splitting (FBS) also known as proximal gradient method in composite convex minimization, Douglas-Rachford splitting (DRS) and the alternating minimization method (AMM) [91]. Operator splitting techniques offer several advantages over traditional optimization methods such as sequential quadratic programming and interior point methods: (1) they can easily handle nonsmooth terms and abstract linear operators, (2) each iteration requires only simple arithmetic operations, (3) the algorithms scale gracefully as the dimension of the problem increases, and (4) they naturally lead to parallel and distributed implementation. Therefore, operator splitting methods cope well with limited amount of hardware resources making them particularly attractive for (embedded) control [111], signal processing [32], and distributed optimization [17, 60].

The key idea behind these techniques when applied to convex optimization is to reformulate the optimality conditions of the problem at hand into a problem of finding a fixed point of a nonexpansive operator and then apply relaxed fixed-point iterations. Although sometimes a fast convergence rate can be observed, the norm of the fixed-point residual decreases, at best, with Q-linear rate, and due to an inherent sensitivity to ill conditioning oftentimes the Q-factor is close to one. Moreover, all operator splitting methods are basically "open loop", since the tuning parameters, such as stepsizes and preconditioning, must be set before their execution. In fact, such methods are very sensitive to the choice of parameters. All these are serious obstacles when it comes to using such types of algorithms when speed and efficiency are imperative, as it is the case of

real-time applications on embedded hardware.

As an attempt to solve the issue, people have considered the employment of variable metrics to reshape the geometry of the problem and enhance convergence rates [34]. However, unless such metrics have a very specific structure, even for simple problems the cost of operating in the new geometry outweights the benefits. Another interesting approach that is gaining more and more popularity tries to exploit possible sparsity patterns by means of chordal decomposition techniques [127]. These methods can improve scalability and reduce memory usage, but unless the problem comes with an inherent sparse structure they yield no tangible benefit.

Alternatively, the task of searching fixed points of an operator T can be translated into that of finding zeros of the corresponding residual $R = \mathrm{id} - T$. Many methods with fast asymptotic convergence rates such as Newton-type exist that can be employed for efficiently solving nonlinear equations, see, *e.g.*, [44, §7] and [61]. However, such methods converge only when close enough to the solution, and in order to globalize the convergence there comes the need of a merit function to perform a line search along candidate directions of descent. The typical choice of the square residual $\|Rx\|^2$ unfortunately is of no use, as in meaningful applications R is nonsmooth. On top of this, even when a suitable merit function is found one still needs to deal with the frequent pathology of linesearch methods in nonsmooth optimization that inhibits the achievement of fast convergence rates, well known for SQP-type algorithms and referred to as the *Maratos effect* [80], see also [61, §6.2].

The already tough challenge of overcoming these issues becomes exceptionally complicated if one further drops the assumption of *convexity*. Indeed, although originally designed and analyzed for convex problems, many splitting algorithms have been observed to perform well when applied to certain classes of structured nonconvex optimization problems. However, yet two more major issues have to be taken into account. First, the elegant link with monotone operator theory onto which the convergence of many splitting algorithms is based no longer holds. Secondly, many regularity properties are lost in the transition, to an extent that well-behaved Lipschitz-continuous mappings give way to operators that are defined only in a *set-valued* sense.

Proximal envelopes proved to be a valuable tool for addressing these issues. First introduced in [92, 93], these functions generalize the well-known Moreau envelope together with its connections with the proximal point algorithm to other splitting schemes. Some splitting algorithms were shown to be equivalent to gradient methods on the corresponding envelopes, leading to the reformulation of nonsmooth and constrained problems as the unconstrained minimization of smooth functions whence classical Newton-type methods can be employed. This

promising approach finds however two main limitations. First, it can only be applied to problems where functions are either smooth or convex. Secondly, it does not fully respect the simplicity of the original splitting algorithms, as it requires additional operations such as Hessian evaluations.

1.1 Contributions and structure the thesis

Inspired by such achievements, yet aware of their limitations, this thesis proposes new envelope-based algorithms that (i) are suitable for fully nonconvex problems, (ii) share operation and iteration complexity with plain splitting algorithms, and (iii) achieve fast asymptotic rates of convergence (under local assumptions) without suffering pathological behaviors such as the Maratos effect. Envelope functions are also shown to be valuable tools for extending the convergence analysis of classical splitting algorithms to the nonconvex setting. In fact, the in-depth analysis of different splitting schemes in a setting as general as possible led to the discovery of many common patterns.

♠ These are discussed in **Chapter 2**, where a new framework for the analysis of nonconvex splitting algorithms is introduced. The common denominator is identified in the presence of a *"proximal" majorization-minimization* component in every step, that is to say, an operation involving the minimization of an (at least quadratic) upper bound of the original problem. Classical proximal algorithms, possibly up to a change of variable, are thus reinterpreted in this context.

♠ In **Chapter 3**, an *envelope function* is defined for each algorithm in the proposed framework, and its regularity properties and basic inequalities are discussed in full generality. Based on these findings, a convergence theory for proximal algorithms is developed.

♠ Building on the investigated convergence framework, **Chapter 4** proposes a new envelope-based globalization strategy that allows to customize splitting algorithms with arbitrary update directions. Without any further assumption, the scheme is shown to accept unit stepsize when the selected directions are superlinear (in the sense of [44, §7.5]), proving its robustness against pathologies such as the Maratos effect. The employment of quasi-Newton directions is also investigated, and a Broyden scheme is shown to yield superlinear convergence under some assumptions at the limit point.

Although the leading ideas have been sketched in an oral exposition,[1] the material of the three chapters summarized above has been exclusively developed in the writing of the thesis. The three chapters outlined next are instead based on published or submitted papers, although suitably amended so as to conform with the proposed general framework for the sake of a more uniform and compact exposition.

♠ **Chapter 5** deals with the forward-backward splitting algorithm (FBS). Thanks to the general convergence analysis developed in the previous chapters, once FBS is shown to fit in the investigated framework, inclusive of a possible *relaxation parameter* λ its convergence is directly inferred. To the best of our knowledge, this is the first result that extends the convergence of FBS for nonconvex problems with $\lambda \neq 1$. Quasi-Newton enhancements are also presented, and the efficacy of the methodology is then verified with numerical simulations.

Based on:

> **A. Themelis**, L. Stella and P. Patrinos. *Forward-backward envelope for the sum of two nonconvex functions: further properties and nonmonotone linesearch algorithms*, SIAM Journal on Optimization 2018 **28**(3):2274-2303, 2018. https://epubs.siam.org/doi/10.1137/16M1080240
>
> L. Stella, **A. Themelis**, P. Sopasakis and P. Patrinos, "A simple and efficient algorithm for nonlinear model predictive control," 2017 IEEE 56th Annual Conference on Decision and Control (CDC), Melbourne, VIC, 2017, pp. 1939–1944. http://ieeexplore.ieee.org/document/8263933/
>
> A. Sathya, P. Sopasakis, R. Van Parys, **A. Themelis**, G. Pipeleers and P. Patrinos, "Embedded nonlinear model predictive control for obstacle avoidance using PANOC,"
> 2018 European Control Conference (ECC), Limassol, 2018 (to appear)

♠ **Chapter 6** deals with the Douglas-Rachford splitting algorithm (DRS). Although some convergence results could directly be derived with the same quick arguments employed for FBS, thanks to a more sophisticated analysis we identified the tightest possible range of parameters enabling convergence. The optimality of the findings is assessed by means of suitable counterexamples. A quasi-Newton DRS algorithm is then presented; this was already discussed in the first submission of the preprint [119], but has been removed from the last version due to space limitations.

Based on:

> **A. Themelis** and P. Patrinos. *Douglas-Rachford splitting and ADMM for nonconvex optimization: tight convergence results*

[1] **A. Themelis,** *Proximal envelopes.* ECC 2018 Workshop on "Advances in Distributed and Large-Scale Optimization," Limassol (Cyprus), Jun. 12-15, 2018. http://www.ecc18.eu/index.php/workshop-6/

(under 2nd review round in the SIAM Journal of Optimization since November 2018)
https://arxiv.org/abs/1709.05747

♠ **Chapter 7** deals with the ADMM algorithm. Expanding on a primal equivalence of the algorithms, the tight convergence results derived in the previous chapter are translated into tight results for ADMM. Also for ADMM the employment of quasi-Newton directions is considered, and the induced speed-up confirmed with numerical simulations.

Based on:

A. Themelis and P. Patrinos. *Douglas-Rachford splitting and ADMM for nonconvex optimization: tight convergence results*
(under 2nd review round in the SIAM Journal of Optimization since November 2018)
https://arxiv.org/abs/1709.05747

♠ Although not directly related to envelope functions, the framework investigated in **Chapter 8** reflects the pursuit of certified fast methods that preserve operation and iteration complexity as plain splitting algorithms. This is indeed the role of the *SuperMann scheme*, an algorithmic framework that applies to *any splitting algorithm*, although only limited to the convex case. The name owes to an intended pun involving the *super*linear rates it achieves and the fact that it generalizes *Mann*-type iterations. As it was the case of the envelope-based algorithms, a Broyden method is shown to yield the desired superlinear rates of convergence under assumptions at the limit point; surprisingly, however, no isolatedness of the solution is required, but merely metric subregularity.

Based on:

A. Themelis and P. Patrinos. *SuperMann: a superlinearly convergent algorithm for finding fixed points of nonexpansive operators*
(under 2nd review round in the IEEE Transactions on Automatic Control journal since March 2018)
https://arxiv.org/abs/1609.06955

P. Sopasakis, **A. Themelis**, J. Suykens and P. Patrinos,
"A primal-dual line search method and applications in image processing,"
2017 25th European Signal Processing Conference (EUSIPCO), Kos, 2017, pp. 1065–1069.
http://ieeexplore.ieee.org/document/8081371/

1.2 Preliminary material

Our notation is standard and follows that of optimization and analysis books [10, 20, 57, 102, 106]. For the sake of clarity we now properly specify the adopted conventions, and briefly recap known definitions and facts. The interested reader is referred to the above-mentioned monographs for the details.

The set of natural numbers is denoted by $\mathbb{N}$, and we adopt the convention that $0 \in \mathbb{N}$. The sets of integer and real numbers are denoted by $\mathbb{Z}$ and $\mathbb{R}$, respectively. The set of extended-real numbers is denoted by $\overline{\mathbb{R}} \coloneqq \mathbb{R} \cup \{\infty\}$. Unless differently specified, we adopt the convention that $1/0 = \infty$.

Given $a, b \in \overline{\mathbb{R}}$ we indicate with $(a, b) \coloneqq \{x \in \mathbb{R} \mid a < x < b\}$ and $[a, b] \coloneqq \{x \in \overline{\mathbb{R}} \cup \{-\infty\} \mid a \leq x \leq b\}$, respectively, the open and closed (possibly extended-real) intervals having a and b as endpoints. Intervals $(a, b]$ and $[a, b)$ are defined accordingly. Occasionally, (a, b) may also indicate a pair or a vector in $\mathbb{R}^2$, however the context will always be explicit enough to avoid confusion. The set of positive real numbers is indicated as $\mathbb{R}_+ \coloneqq [0, \infty)$, and that of strictly positive real numbers as $\mathbb{R}_{++} \coloneqq (0, \infty)$.

The positive and negative parts of $r \in \mathbb{R}$ are defined as $[r]_+ \coloneqq \max\{0, r\}$ and $[r]_- \coloneqq \max\{0, -r\}$, respectively. Notice that $[r]_+$ and $[r]_-$ are positive numbers such that $r = [r]_+ - [r]_-$.

The sum of two sets $A, B \subseteq \mathbb{R}^n$ is meant in the Minkowski sense, namely $A + B = \{a + b \mid a \in A,\ b \in B\}$; the difference is defined accordingly. In case $A = \{a\}$ is a singleton, we will write $a + B$ as shorthand for $\{a\} + B$, and similarly if B is a singleton.

The CLOSURE and INTERIOR of $E \subseteq \mathbb{R}^n$ are denoted as $\operatorname{cl} E$ and $\operatorname{int} E$, respectively. The BOUNDARY of E is $\operatorname{bdry} E \coloneqq \operatorname{cl} E \setminus \operatorname{int} E$. With $\mathrm{B}(x; r)$ and $\overline{\mathrm{B}}(x; r)$ we indicate, respectively, the open and closed balls centered at x with radius r.

1.2.1 Matrices and vectors

The $n \times n$ identity matrix is denoted as I_n, and the $\mathbb{R}^n$ vector with all elements equal to 1 as $\mathbf{1}_n$; whenever n is clear from context we simply write I and $\mathbf{1}$, respectively. We use the Kronecker symbol $\delta_{i,j}$ for the (i, j)-th entry of I. Given $v \in \mathbb{R}^n$, with $\operatorname{diag} v$ we indicate the $n \times n$ diagonal matrix whose i-th diagonal entry is v_i.

The range and nullspace (or kernel) of a matrix $A \in \mathbb{R}^{m \times n}$ are denoted by $\operatorname{range} A \coloneqq \{Ax \mid x \in \mathbb{R}^n\}$ and $\ker A \coloneqq \{v \in \mathbb{R}^n \mid Av = 0\}$, respectively. The

rank of A is denoted by $\operatorname{rank} A$, and its transpose by $A^\top$.

With $\mathrm{Sym}(\mathbb{R}^n)$, $\mathrm{Sym}_+(\mathbb{R}^n)$, and $\mathrm{Sym}_{++}(\mathbb{R}^n)$, we denote respectively the set of symmetric, symmetric positive semidefinite, and symmetric positive definite matrices in $\mathbb{R}^{n\times n}$.

The minimum and maximum eigenvalues of $H \in \mathrm{Sym}(\mathbb{R}^n)$ are denoted as $\lambda_{\min}(H)$ and $\lambda_{\max}(H)$, respectively. For $Q, R \in \mathrm{Sym}(\mathbb{R}^n)$ we write $Q \succeq R$ to indicate that $Q - R \in \mathrm{Sym}_+(\mathbb{R}^n)$, and similarly $Q \succ R$ indicates that $Q - R \in \mathrm{Sym}_{++}(\mathbb{R}^n)$. Any matrix $Q \in \mathrm{Sym}_+(\mathbb{R}^n)$ induces the semi-norm $\|\cdot\|_Q$ on $\mathbb{R}^n$, where $\|x\|_Q^2 \coloneqq \langle x, Qx\rangle$; in case $Q = \mathrm{I}$, that is, for the Euclidean norm, we omit the subscript and simply write $\|\cdot\|$. No ambiguity occurs in adopting the same notation for the induced matrix norm, namely $\|M\| \coloneqq \max\{\|Mx\| \mid x \in \mathbb{R}^n,\ \|x\| = 1\}$ for $M \in \mathbb{R}^{n\times n}$. For $p \in [1,\infty]$, the ℓ^p norm on $\mathbb{R}^n$ is denoted by $\|\cdot\|_p$, where

$$\|x\|_\infty \coloneqq \max\{|x_i| \mid i = 1\dots n\}, \quad \text{and} \quad \|x\|_p \coloneqq \left(\textstyle\sum_{i=1}^n |x_i|^p\right)^{1/p},$$

for $p \in [1,\infty)$. The definition extends to $p \in (0,1)$ as well, although in this case $\|\cdot\|_p$ is not subadditive and thus is only a quasi-norm. The ℓ^0 quasi-norm, namely

$$\|x\|_0 \coloneqq \text{number of nonzero entries of } x,$$

additionally fails to be homogeneous.

1.2.2 Sequences

The notation $(a^k)_{k\in K}$ represents a sequence indexed by elements of the set K, and given a set E we write $(a^k)_{k\in K} \subset E$ to indicate that $a^k \in E$ for all indices $k \in K$. We say that $(a^k)_{k\in K} \subset \mathbb{R}^n$ is SUMMABLE if $\sum_{k\in K} \|a^k\|$ is finite, and SQUARE-SUMMABLE if $(\|a^k\|^2)_{k\in K}$ is summable. As a shorthand notation we may write $(x^k)_{k\in\mathbb{N}} \in \ell^1$ and $(x^k)_{k\in\mathbb{N}} \in \ell^2$ to indicate that $(x^k)_{k\in\mathbb{N}}$ is summable and square summable, respectively.

We say that the sequence converges to a point $a \in \mathbb{R}^n$

- Q-LINEARLY if there exists $\rho \in [0,1)$ such that $\|a^{k+1} - a\| \le \rho\|a^k - a\|$ for all k's;
- R-LINEARLY if there exists a sequence $(\varepsilon_k)_{k\in\mathbb{N}}$ Q-linearly convergent to 0 such that $\|a^k - a\| \le \varepsilon_k$;
- SUPERLINEARLY if either $a^k = a$ for some $k \in \mathbb{N}$, or $\|a^{k+1}-a\|/\|a^k-a\| \to 0$ as $k \to \infty$.

We will often adopt the *big-O* and *small-o* notation: given a sequence $(x^k)_{k\in\mathbb{N}} \subset \mathbb{R}$ and $(\varepsilon_k)_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$, we write $x^k \in O(\varepsilon_k)$ and $x^k \in o(\varepsilon_k)$ to indicate that

$$\limsup_{k\to\infty} \frac{|x^k|}{\varepsilon_k} < \infty \quad \text{and} \quad \lim_{k\to\infty} \frac{|x^k|}{\varepsilon_k} = 0,$$

respectively.

1.2.3 Extended-real-valued functions

Given a function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$, its EPIGRAPH is the set

$$\operatorname{epi} h := \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid h(x) \le \alpha\},$$

while its DOMAIN is

$$\operatorname{dom} h := \{x \in \mathbb{R}^n \mid h(x) < \infty\},$$

and for $\alpha \in \mathbb{R}$ its α-LEVEL SET is

$$\operatorname{lev}_{\le\alpha} h := \{x \in \mathbb{R}^n \mid h(x) \le \alpha\}.$$

Function h is said to be LOWER SEMICONTINUOUS (LSC) if $\operatorname{epi} h$ is a closed set in $\mathbb{R}^{n+1}$ (h is also said to be CLOSED); equivalently, h is lsc iff for all $\bar{x} \in \mathbb{R}^n$ it holds that

$$h(\bar{x}) \le \liminf_{x\to\bar{x}} h(x).$$

All level sets of an lsc function are closed. We say that h is PROPER if $\operatorname{dom} h \neq \emptyset$, and that it is LEVEL BOUNDED if for all $\alpha \in \mathbb{R}$ the level set $\operatorname{lev}_{\le\alpha} h$ is a bounded subset of $\mathbb{R}^n$.

The INDICATOR FUNCTION of a set $S \subseteq \mathbb{R}^n$ is the function $\delta_S : \mathbb{R}^n \to \overline{\mathbb{R}}$ defined as

$$\delta_S(x) = \begin{cases} 0 & \text{if } x \in S, \\ \infty & \text{otherwise.} \end{cases}$$

If S is nonempty and closed, then δ_S is proper and lsc.

$h : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be STRICTLY CONTINUOUS at $\bar{x} \in \operatorname{dom} h$ if

$$\limsup_{\substack{x,y\to\bar{x} \\ x\neq y}} \frac{\|h(x) - h(y)\|}{\|x - y\|} < \infty.$$

Having h strictly continuous at every point of a set $D \subseteq \operatorname{dom} h$ is equivalent to h being locally Lipschitz continuous on D [106, §9].

1.2.4 Self-mappings

In this subsection we analyze single-valued mappings from $\mathbb{R}^n$ to itself. Given $\mu > 0$, a function $G : \mathbb{R}^n \to \mathbb{R}^n$ is said to be μ-COCOERCIVE if

$$\langle G(x) - G(y), x - y\rangle \geq \mu\|G(x) - G(y)\|^2 \qquad \forall x, y \in \mathbb{R}^n, \tag{1.1}$$

and μ-STRONGLY MONOTONE if

$$\langle G(x) - G(y), x - y\rangle \geq \mu\|x - y\|^2 \qquad \forall x, y \in \mathbb{R}^n. \tag{1.2}$$

We say that G is MONOTONE if (either of) the inequalities above hold with $\mu = 0$. Notice that the IDENTITY mapping $\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$ is an example of cocoercive and strongly monotone mapping, and that, more generally, μ-cocoercivity implies μ^{-1}-Lipschitz continuity.

Lemma 1.1. *Any L-Lipschitz continuous and μ-strongly monotone mapping $G : \mathbb{R}^n \to \mathbb{R}^n$ is a Lipschitz homeomorphism; that is, other than being Lipschitz continuous, it is also invertible and its inverse is Lipschitz continuous as well (with modulus μ^{-1}).*

Proof. By upper bounding the inner product of (1.2) with the Cauchy-Schwartz inequality we obtain

$$\mu\|x - y\|^2 \leq \|x - y\|\|G(x) - G(y)\| \quad \forall x, y \in \mathbb{R}^n.$$

In particular, G is injective, and if it has an inverse that must be μ^{-1}-Lipschitz continuous. Moreover, since $\psi(x) \coloneqq G(x) - \mu x$ is monotone and continuous, [106, Ex. 12.7 and Thm. 12.12] ensures that $G(x) = \psi(x) + \mu x$ is also surjective, hence the claim. □

1.2.5 Set-valued mappings

We use the notation $H : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ to indicate a point-to-set function $H : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^m)$, where $\mathcal{P}(\mathbb{R}^m)$ is the power set of $\mathbb{R}^m$ (the set of all subsets of $\mathbb{R}^m$). The GRAPH of H is the set

$$\operatorname{gph} H \coloneqq \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m \mid y \in H(x)\},$$

while its DOMAIN is

$$\operatorname{dom} h \coloneqq \{x \in \mathbb{R}^n \mid H(x) \neq \emptyset\}.$$

We say that H is OUTER SEMICONTINUOUS (OSC) at $\bar{x} \in \operatorname{dom} H$ if for any $\varepsilon > 0$ there exists $\delta > 0$ such that $H(x) \subseteq H(\bar{x}) + \mathrm{B}(0; \varepsilon)$ for all $x \in \mathrm{B}(0; \delta)$.

In particular, this implies that whenever $(x^k)_{k\in\mathbb{N}} \subseteq \operatorname{dom} H$ converges to x and $(y^k)_{k\in\mathbb{N}}$ converges to y with $y^k \in H(x^k)$ for all k, it holds that $y \in H(x)$. We say that H is osc (without mention of a point) if H is osc at every point of its domain or, equivalently, if $\operatorname{gph} H$ is a closed subset of $\mathbb{R}^n \times \mathbb{R}^m$.

For notational simplicity, in case $H(x)$ is a singleton we may sometimes treat it as a point rather than a set, allowing notational abuses such as $H(x) = y$ as opposed to $H(x) = \{y\}$.

The PROJECTION onto a nonempty and closed set $S \subseteq \mathbb{R}^n$ will be meant in the set-valued sense; namely, $\Pi_S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is defined by $\Pi_S(x) = \operatorname{arg\,min}_{z\in S} \|z - x\|$. With $\operatorname{dist}(x, S) := \inf_{z\in\mathbb{R}^n} \|z - x\|$ we indicate the DISTANCE of x from S.

Given $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, we say that a point x is FIXED (for F) if $x \in F(x)$, while x is a ZERO (of F) if $0 \in F(x)$. The FIXED SET (*i.e.*, the set of fixed points) and the ZERO SET (*i.e.*, the set of zeros) of F are respectively denoted by

$$\operatorname{fix} F := \{x \in \mathbb{R}^n \mid x \in F(x)\},$$

and

$$\operatorname{zer} F := \{x \in \mathbb{R}^n \mid 0 \in F(x)\}.$$

1.2.6 Subdifferential

Given a proper and lsc function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$, we denote by $\hat{\partial} h : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ the REGULAR SUBDIFFERENTIAL of h, where

$$v \in \hat{\partial} h(\bar{x}) \quad \Leftrightarrow \quad \liminf_{\substack{x\to\bar{x}\\ x\neq\bar{x}}} \frac{h(x) - h(\bar{x}) - \langle v, x - \bar{x}\rangle}{\|x - \bar{x}\|} \geq 0. \tag{1.3}$$

The (limiting) SUBDIFFERENTIAL of h is $\partial h : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, where $v \in \partial h(\bar{x})$ iff there exists a sequence $(x^k, v^k)_{k\in\mathbb{N}} \subseteq \operatorname{gph} \hat{\partial} h$ such that

$$\lim_{k\to\infty} (x^k, h(x^k), v^k) = (x, h(x), v).$$

The set of HORIZON SUBGRADIENTS of h at x is $\partial^\infty h$, defined as $\partial h(x)$ except that $v^k \to v$ is meant in the "cosmic" sense, namely $\lambda_k v^k \to v$ for some $\lambda_k \searrow 0$.

Finally, the BOULIGAND SUBDIFFERENTIAL of h at x is $\partial_B h : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, where $v \in \partial_B h(\bar{x})$ iff there exists a sequence $(x^k)_{k\in\mathbb{N}} \to x$ such that h is differentiable at x^k for all k's and $\nabla h(x^k) \to v$ as $k \to \infty$.

Lemma 1.2 ([106, Thm. 10.1])**.** *Let $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lsc. If $\bar{x}$ is a local minimizer for h, then $0 \in \hat{\partial} h(\bar{x})$.*

Lemma 1.3 (Basic subdifferential rules)**.** *Let $g, h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lsc functions. For all $\bar{x} \in \mathbb{R}^n$ the following hold:*

(i) For any $t > 0$ one has $\partial(th)(\bar{x}) = t\partial h(\bar{x})$ and $\hat{\partial}(th)(\bar{x}) = t\hat{\partial}h(\bar{x})$.

(ii) h is strictly continuous at $\bar{x}$ iff $\bar{x} \in \operatorname{dom} h$ and $\partial^\infty h(\bar{x}) = \{0\}$.

(iii) If h is strictly continuous at $\bar{x}$, then $\partial(g + h)(\bar{x}) \subseteq \partial g(\bar{x}) + \partial h(\bar{x})$.

(iv) If h is strictly continuous at $\bar{x}$ and $\partial h(\bar{x})$ has at most one element, then h is strictly differentiable at $\bar{x}$.

(v) If h is differentiable at $\bar{x}$, then $\hat{\partial}h(x) = \{\nabla h(\bar{x})\}$.

(vi) If h is continuously differentiable around $\bar{x}$, then

- *$\partial h(\bar{x}) = \hat{\partial}h(\bar{x}) = \{\nabla h(\bar{x})\}$,*
- *$\partial(g + h)(\bar{x}) = \partial g(\bar{x}) + \nabla h(\bar{x})$, and*
- *$\hat{\partial}(g + h)(\bar{x}) = \hat{\partial}g(\bar{x}) + \nabla h(\bar{x})$.*

Proof.

♠ 1.3*(i)*. See [106, Eq. (10.6)].

♠ 1.3*(ii)*. See [106, Thm. 9.13].

♠ 1.3*(iii)*. See [106, Ex. 10.10].

♠ 1.3*(iv)*. See [106, Thm. 9.18].

♠ 1.3*(v)* & 1.3*(vi)*. See [106, Ex. 8.8]. □

1.2.7 (Hypo)convexity

A convex combination of two points $x, y \in \mathbb{R}^n$ is any point $(1 - t)x + ty$ with $t \in [0, 1]$. A set $D \subseteq \mathbb{R}^n$ is convex if whenever $x, y \in D$ also any of their convex combinations belongs to D. The CONVEX HULL of a set $E \subseteq \mathbb{R}^n$, denoted as $\operatorname{conv} E$, is the smallest convex set that contains E (the intersection of convex sets is still convex). Specifically,

$$\operatorname{conv} E := \Big\{ \textstyle\sum_{i=1}^k \alpha_i x_i \mid k \in \mathbb{N},\ x_i \in E,\ \alpha_i \geq 0,\ \sum_{i=1}^k \alpha_i = 1 \Big\}.$$

A function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex if $\operatorname{epi} f$ is a convex set; equivalently, h is convex if for any $x, y \in \mathbb{R}^n$ and $t \in [0, 1]$ it holds that $h((1-t)x+ty) \leq (1-t)h(x)+th(y)$. In particular, the domain of a convex function is a convex set.

Given $\sigma \in \mathbb{R}$, we say that a function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ is σ-HYPOCONVEX if $h - \frac{\sigma}{2}\|\cdot\|^2$ is a convex function. Thus, convexity is equivalent to 0-hypoconvexity; if $\sigma > 0$, then not only is h convex, but it is said to be strongly convex with modulus $\sigma > 0$ (or σ-strongly convex). Any strongly convex function is level bounded and has a unique minimizer.

Lemma 1.4. *Let a function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $\sigma \in \mathbb{R}$ be fixed. The following are equivalent:*

(a) h is σ-hypoconvex.

(b) $h(y) \geq h(x) + \langle v_x, y - x\rangle + \frac{\sigma}{2}\|x - y\|^2$ for all $x, y \in \mathbb{R}^n$ and $v_x \in \partial h(x)$.

(c) $\langle v_x - v_y, x - y\rangle \geq \sigma\|x - y\|^2$ for all $x, y \in \mathbb{R}^n$, $v_x \in \partial h(x)$ and $v_y \in \partial h(y)$.

Proof. These are well-known facts when $\sigma = 0$, that is, for convex functions, see *e.g.*, [10, Thm. 20.25]. The other claims readily follow by applying the equivalence to the convex function $\psi(x) = h(x) - \frac{\sigma}{2}\|x\|^2$, in light of the fact that $\partial\psi(x) = \partial h(x) - \sigma x$, as it follows from Lem. 1.3*(vi)*. □

1.2.8 Smoothness

The class of functions $h : \mathbb{R}^n \to \mathbb{R}$ that are k times continuously differentiable is denoted as $C^k(\mathbb{R}^n)$; the subset of those with locally Lispchitz k-th derivative is denoted as $C^{k+}(\mathbb{R}^n)$. We write $h \in C^{1,1}(\mathbb{R}^n)$ to indicate that $h \in C^1(\mathbb{R}^n)$ and that ∇h is (globally) Lipschitz continuous with modulus L_h. To simplify the terminology, we will say that such an h is L_h-SMOOTH.

Definition 1.5. *We say that $R : \mathbb{R}^n \to \mathbb{R}^n$ is*

(i) STRICTLY DIFFERENTIABLE *at $\bar{x}$ if the Jacobian matrix $JR(\bar{x}) := \left(\frac{\partial R_i}{\partial x_j}(\bar{x})\right)_{i,j}$ exists and*

$$\lim_{\substack{y,z \to \bar{x} \\ y \neq z}} \frac{\|Ry - Rx - JR(\bar{x})(y - x)\|}{\|y - x\|} = 0; \tag{1.4}$$

(ii) SEMIDIFFERENTIABLE *at $\bar{x}$ if there exists a continuous and positively homogeneous function $DR(\bar{x}) : \mathbb{R}^n \to \mathbb{R}^n$,[2] called the* SEMIDERIVATIVE *of R at $\bar{x}$, such that*

$$Rx = R\bar{x} + DR(\bar{x})[x - \bar{x}] + o(\|x - \bar{x}\|);$$

[2] That is, such that $R(tx) = tR(x)$ for all $x \in \mathbb{R}^n$ and $t > 0$.

(iii) CALMLY SEMIDIFFERENTIABLE *at $\bar{x}$ if there exists a neighborhood $U_{\bar{x}}$ of $\bar{x}$ in which R is semidifferentiable and such that for all $w \in \mathbb{R}^n$ with $\|w\| = 1$ the function $U_{\bar{x}} \ni x \mapsto DR(x)[w]$ is Lipschitz continuous at $\bar{x}$.*

Due to an ambiguity in the literature, *strict* differentiability is sometimes referred to as *strong* differentiability [59, 90]. We choose to stick the proposed terminology, following [106]. Semidifferentiability is clearly a milder property than differentiability in that the mapping $DR(\bar{x})$ needs not be linear. More precisely, as long as R is strictly continuous, then semidifferentiability is equivalent to directional differentiability [44, Prop. 3.1.3] and the semiderivative is sometimes called *B-derivative* [59, 44]. The three concepts in Definition 1.5 are related as *(iii)* $\Rightarrow$ *(i)* $\Rightarrow$ *(ii)* [90, Thm. 2] and neither requires the existence of the (classical) Jacobian around $\bar{x}$. Recall that a function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ is DIRECTIONALLY DIFFERENTIABLE at $x \in \operatorname{dom} h$ if for every $d \in \mathbb{R}^n$ the (possibly infinite) limit

$$h'(x; d) \coloneqq \lim_{\tau \to 0^+} \frac{h(x + \tau d) - h(x)}{\tau}$$

exists. The quantity $h'(x; d)$ is the DIRECTIONAL DERIVATIVE of h at x along direction d. The following result provides characterization of smoothness under convexity.

Theorem 1.6. *Let $\psi \in C^1(\mathbb{R}^n)$ be a convex function. The following are equivalent:*

(a) *ψ is L_ψ-smooth.*

(b) *$\frac{1}{L_\psi}\|\nabla\psi(x) - \nabla\psi(y)\|^2 \leq \langle \nabla\psi(x) - \nabla\psi(y), x - y\rangle$ for all $x, y \in \mathbb{R}^n$.*

(c) *$0 \leq \langle \nabla\psi(x) - \nabla\psi(y), x - y\rangle \leq L_\psi\|x - y\|^2$ for all $x, y \in \mathbb{R}^n$.*

(d) *$\psi(y) \geq \psi(x) + \langle \nabla\psi(x), y - x\rangle + \frac{1}{2L_\psi}\|\nabla\psi(y) - \nabla\psi(x)\|^2$ for all $x, y \in \mathbb{R}^n$.*

Proof. See [84, Thm. 2.1.5]. □

Lemma 1.7. *Let $h \in C^1(\mathbb{R}^n)$ and $\sigma \in \mathbb{R}$ be fixed. The following are equivalent:*

(a) *h is σ-hypoconvex.*

(b) *$h(y) \geq h(x) + \langle \nabla h(x), y - x\rangle + \frac{\sigma}{2}\|x - y\|^2$ for all $x, y \in \mathbb{R}^n$.*

(c) *$\langle \nabla h(x) - \nabla h(y), y - x\rangle \geq \sigma\|x - y\|^2$ for all $x, y \in \mathbb{R}^n$.*

Proof. Direct consequence of Lem. 1.4, in light of the fact that $\partial h = \nabla h$, cf. Lem. 1.3*(vi)*. □

Hypoconvexity of smooth functions

If $h \in C^{1,1}(\mathbb{R}^n)$ is L_h-smooth, then so is $-h$, and from Lemma 1.7 we then infer that h is $(-L_h)$-hypoconvex. In fact, while hypoconvexity of h amounts to the existence of a quadratic lower bound for h at any point, similarly, smoothness entails the existence of a quadratic upper bound. In general, however, a smooth function could be σ-hypoconvex for some σ not necessarily equal to, but at least larger or equal than $-L_f$. Of course, the upper bound in (1.5) forces $\sigma \leq L_f$. This leads to the following result.

Theorem 1.8. *Any function $h \in C^{1,1}(\mathbb{R}^n)$ is σ_h-hypoconvex for some $\sigma_h \in [-L_h, L_h]$. In fact, for any $h \in C^1(\mathbb{R}^n)$ the following properties are equivalent:*

(a) h is L_h-smooth and σ_h-hypoconvex.

(b) $\sigma_h \geq -L_h$ and for all $x, y \in \mathbb{R}^n$

$$\tfrac{\sigma_h}{2}\|x-y\|^2 \leq h(y) - \big[h(x) + \langle \nabla h(x), y-x\rangle\big] \leq \tfrac{L_h}{2}\|y-x\|^2. \tag{1.5}$$

(c) $\sigma_h \geq -L_h$ and for all $x, y \in \mathbb{R}^n$

$$(L_h+\sigma_h)\langle \nabla h(x) - \nabla h(y), x-y\rangle \geq \sigma_h L_h \|x-y\|^2 + \|\nabla h(x) - \nabla h(y)\|^2.$$

(d) $\sigma_h \geq -L_h$ and for all $x, y \in \mathbb{R}^n$

$$\sigma_h\|x-y\|^2 \leq \langle \nabla h(x) - \nabla h(y), x-y\rangle \leq L_h\|x-y\|^2. \tag{1.6}$$

Clearly, all the claims remain valid if σ_h is replaced by any $\sigma \in [-L_h, \sigma_h]$; in particular, one can always consider $\sigma_h = -L_h$.[3]

Proof. That h is $(-L_h)$-hypoconvex has already been discussed.

♠ 1.8*(a)* ⇒ 1.8*(b)*. Follows from Lem. 1.7 and [21, Prop. A.24].

♠ 1.8*(b)* ⇒ 1.8*(c)*. The claim is trivial if $\sigma_h = L_h$, for this corresponds to having $h = \frac{L_h}{2}\|\cdot\|^2$. Otherwise, the lower bound in (1.5) implies σ_h-hypoconvexity of h, as it follows from Lem. 1.7. The upper bound, instead, ensures that the function $\psi(x) = \frac{L_h}{2}\|x\|^2 - h(x)$ satisfies

$$\psi(y) \geq \psi(x) + \langle \nabla\psi(x), y-x\rangle \quad \forall x, y \in \mathbb{R}^n.$$

[3] If $\sigma_h \geq -L_h$ and $L_h \geq 0$ are not imposed, then the smoothness modulus L_h in Thm. 1.8*(a)* has to be replaced by $\max\{|L_h|, |\sigma_h|\}$.

Therefore, ψ is convex, as it follows from Lem. 1.7*(b)*. We have

$$\begin{aligned} 0 &\leq \langle \nabla\psi(x) - \nabla\psi(y), x - y\rangle \\ &= L_h\|x-y\|^2 - \langle \nabla h(x) - \nabla h(y), x - y\rangle \\ &\leq (L_h - \sigma_h)\|x-y\|^2, \end{aligned} \tag{1.7}$$

where the first inequality follows from Thm. 1.6*(c)*. From Thm. 1.6 we then conclude that ψ is (convex and) L_ψ-smooth, with $L_\psi = L_h - \sigma_h$, hence that we may replace 0 in the first term of the chain of inequalities with $\frac{1}{L_\psi}\|\nabla\psi(x) - \nabla\psi(y)\|^2$. Inequality (1.7) then becomes

$$\tfrac{1}{L_h-\sigma_h}\|\nabla\psi(x) - \nabla\psi(y)\|^2 \leq L_h\|x-y\|^2 - \langle \nabla h(x) - \nabla h(y), x - y\rangle.$$

Multiplying by the strictly positive constant $L_h - \sigma_h$ yields

$$\begin{aligned} &L_h(L_h-\sigma_h)\|x-y\|^2 - (L_h-\sigma_h)\langle \nabla h(x) - \nabla h(y), x - y\rangle \\ &\geq \|\nabla\psi(x) - \nabla\psi(y)\|^2 \\ &= L_h^2\|x-y\|^2 + \|\nabla h(x) - \nabla h(y)\|^2 - 2L_h\langle \nabla h(x) - \nabla h(y), x - y\rangle. \end{aligned}$$

By suitably rearranging, the sought inequality follows.

♠ 1.8*(c)* ⇒ 1.8*(d)*. Expressing the inequality in terms of $\psi := h - \frac{\sigma}{2}\|\cdot\|^2$, we have

$$\begin{aligned} &\|\nabla\psi(x) - \nabla\psi(y)\|^2 + \sigma_h^2\|x-y\|^2 + 2\sigma_h\langle \nabla\psi(x) - \nabla\psi(y), x - y\rangle \\ &\leq (L_h+\sigma_h)\langle \nabla\psi(x) - \nabla\psi(y), x - y\rangle + (L_h+\sigma_h)\sigma_h\|x-y\|^2 - \sigma_h L_h\|x-y\|^2 \\ &= (L_h+\sigma_h)\langle \nabla\psi(x) - \nabla\psi(y), x - y\rangle + \sigma_h^2\|x-y\|^2, \end{aligned}$$

hence

$$(L_h-\sigma_h)\langle \nabla\psi(x) - \nabla\psi(y), x - y\rangle \geq \|\nabla\psi(x) - \nabla\psi(y)\|^2.$$

This shows that $\nabla\psi$ is $\frac{1}{L_h-\sigma_h}$-cocoercive, hence that ψ is convex and $(L_h - \sigma_h)$-smooth in light of Thm. 1.6*(b)*. We then have

$$\sigma_h\|x-y\|^2 \leq \sigma_h\|x-y\|^2 + \langle \nabla\psi(x) - \nabla\psi(y), x - y\rangle \leq L_h\|x-y\|^2,$$

where the inequalities is due to Thm. 1.6*(c)*. The claim then follows from the fact that $\sigma_h\|x-y\|^2 + \langle \nabla\psi(x) - \nabla\psi(y), x - y\rangle = \langle \nabla h(x) - \nabla h(y), x - y\rangle$.

♠ 1.8*(d)* ⇒ 1.8*(a)*. σ_h-hypoconvexity follows from Lem. 1.7. The upper bound

in (1.6) implies that the function $\psi(x) = \frac{L_h}{2}\|x\|^2 - h(x)$ is convex. We may now trace the proof of the implication '1.8*(b)* $\Rightarrow$ 1.8*(c)*' to infer that ψ is $(L_h - \sigma_h)$-smooth, hence that h is L_h-smooth. □

Lemma 1.9 (Subdifferential characterization of smoothness)**.** *Let $h : \mathbb{R}^n \to \mathbb{R}$ be such that $\partial h(x) \neq \emptyset$ for all $x \in \mathbb{R}^n$, and suppose that there exist $L \geq 0$ and $\sigma \in [-L, L]$ such that*

$$\sigma\|x_1 - x_2\|^2 \leq \langle v_1 - v_2, x_1 - x_2\rangle \leq L\|x_1 - x_2\|^2 \tag{1.8}$$

holds for all $x_i \in \mathbb{R}^n$, $v_i \in \partial h(x_i)$, $i = 1, 2$. Then, $h \in C^{1,1}(\mathbb{R}^n)$ is L-smooth and σ-hypoconvex.

Proof. The claimed hypoconvexity follows from [106, Ex. 12.28]. It suffices to show that h is continuously differentiable, so that $\partial h = \nabla h$ and the claim then follows from Thm. 1.8*(d)*. To this end, without loss of generality we may assume that $\sigma \geq 0$, since h is continuously differentiable iff so is $h - \frac{\sigma}{2}\|\cdot\|^2$. Thus, for all $x_i \in \mathbb{R}^n$, $v_i \in \partial h(x_i)$, $i = 1, 2$, one has

$$\begin{aligned} h(x_1) &\geq h(x_2) + \langle v_2, x_1 - x_2\rangle \\ &= h(x_2) + \langle v_2 - v_1, x_1 - x_2\rangle + \langle v_1, x_1 - x_2\rangle \\ &\geq h(x_2) - L\|x_1 - x_2\|^2 + \langle v_1, x_1 - x_2\rangle, \end{aligned}$$

where the first inequality follows from convexity of h (it is 0-hypoconvex by assumption). Rearranging,

$$h(x_2) \leq h(x_1) + \langle v_1, x_2 - x_1\rangle + L\|x_1 - x_2\|^2 \quad \forall x_i \in \mathbb{R}^n,\ v_1 \in \partial h(x_1),\ i = 1, 2.$$

Let $\tilde{h} := h - \langle v_1, \cdot\,\rangle$, so that $0 \in \partial h(x_1)$. Due to convexity, $x_1 \in \operatorname{argmin} \tilde{h}$, hence for all $w \in \mathbb{R}^n$ and $v_1' \in \partial h(x_1)$ one has

$$\begin{aligned} \tilde{h}(x_1) \leq \tilde{h}(w) &\leq h(x_1) + \langle v_1', w - x_1\rangle + L\|w - x_1\|^2 - \langle v, w\rangle \\ &= \tilde{h}(x_1) + \langle v_1' - v_1, w - x_1\rangle + L\|w - x_1\|^2. \end{aligned}$$

By selecting $w = x_1 - \frac{1}{2L}(v_1' - v_1)$, one obtains $\|v_1 - v_1'\|^2 \leq 0$, hence necessarily $v_1 = v_1'$. From the arbitrarity of $x_1 \in \mathbb{R}^n$ and $v_1, v_1' \in \partial h(x_1)$ it follows that ∂h is everywhere single valued, and the sought continuous differentiability of h then follows from [106, Cor. 9.19]. □

Theorem 1.10 (Lower bounds for smooth functions)**.** *Let $h \in C^{1,1}(\mathbb{R}^n)$ be L_h-smooth and σ_h-hypoconvex. Then, for all $x, y \in \mathbb{R}^n$ it holds that*

$$h(y) \geq h(x) + \langle \nabla h(x), y - x \rangle + \rho(y, x),$$

where

(i) $\rho(y, x) = \frac{\sigma_h}{2}\|y - x\|^2$.

If $-L_h < \sigma_h \leq 0$, then one also has

(ii) $\rho(y, x) = \frac{\sigma_h L_h}{2(L_h + \sigma_h)}\|y - x\|^2 + \frac{1}{2(L_h + \sigma_h)}\|\nabla h(y) - \nabla h(x)\|^2$.

Clearly, all inequalities remain valid if one replaces σ_h and L_h with σ and L, respectively, as long as $L \geq L_h$ and $-L \leq \sigma \leq \sigma_h$.

Proof.

♠ 1.10*(i)*. Already shown in Thm. 1.8*(b)*.

♠ 1.10*(ii)*. Function $\psi := h - \frac{\sigma}{2}\|\cdot\|^2$ is L_ψ-smooth and convex, with $L_\psi = L_h - \sigma_h$. By expressing the inequality in Thm. 1.6*(d)* with respect to h, one obtains

$$\begin{aligned} h(y) \geq h(x) &+ \langle \nabla h(x), y - x \rangle + \tfrac{\sigma_h L_h}{2(L_h - \sigma_h)}\|y - x\|^2 \\ &+ \tfrac{1}{2(L_h - \sigma_h)}\|\nabla h(y) - \nabla h(x)\|^2 - \tfrac{\sigma_h}{L_h - \sigma_h}\langle \nabla h(y) - \nabla h(x), y - x \rangle. \end{aligned}$$

Since $\sigma_h \leq 0$, then the coefficient of the scalar product in the second line is positive, and we can further lower bound it by means of the inequality in Thm. 1.8*(c)*; the claimed inequality follows after easy algebraic manipulations. □

We remark that in Theorem 1.10*(ii)* one could also use the same reasoning to obtain bounds in the strongly convex case, *e.g.*, by exploiting the inequality in [84, Thm. 2.1.10], namely

$$\langle \nabla h(y) - \nabla h(x), y - x \rangle \leq \tfrac{1}{\mu_h}\|\nabla h(y) - \nabla h(x)\|^2$$

and holding for any μ_h-strongly convex and smooth function h. However, one can easily verify that this choice results in a bound looser than the ones already provided in Theorems 1.10*(i)* and 1.6*(d)*, which is why only the nonstrongly convex case was investigated.

1.2.9 Proximal map and Moreau envelope

Definition 1.11 (Proximal mapping)**.** *The* PROXIMAL MAPPING *of* $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ *with parameter* $\gamma > 0$ *is the set-valued map* $\operatorname{prox}_{\gamma h} : \mathbb{R}^n \rightrightarrows \operatorname{dom} h$ *defined as*

$$\operatorname{prox}_{\gamma h}(x) := \operatorname*{arg\,min}_{w \in \mathbb{R}^n} \left\{ h(w) + \tfrac{1}{2\gamma}\|w - x\|^2 \right\}. \tag{1.9}$$

We say that a function h is PROX-BOUNDED if $h + \frac{1}{2\gamma}\|\cdot\|^2$ is lower bounded for some $\gamma > 0$. The supremum of all such γ — which is possibly infinite, as it is the case when h is lower bounded or convex — is the *threshold of prox-boundedness of* h, denoted as γ_h. The value function of the minimization problem defining the proximal mapping is the MOREAU ENVELOPE with parameter γ, denoted $h^\gamma : \mathbb{R}^n \to \mathbb{R}$, namely

$$h^\gamma(x) := \inf_{w \in \mathbb{R}^n} \left\{ h(w) + \tfrac{1}{2\gamma}\|w - x\|^2 \right\}. \tag{1.10}$$

Some basic properties of $\operatorname{prox}_{\gamma h}$ and h^γ are collected in the following result.

Proposition 1.12. *Let* $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be proper, lsc and prox-bounded. Then, for every* $\gamma \in (0, \gamma_h)$ *the following hold:*

(i) $\operatorname{prox}_{\gamma h}$ *is osc, locally bounded, and nonempty- and compact-valued.*

(ii) $h^\gamma : \mathbb{R} \to \mathbb{R}$ *is real valued and strictly continuous.*

(iii) For all $x \in \mathbb{R}^n$, $\hat{\partial} h^\gamma(x) = \{\nabla h^\gamma(x)\}$ *if* h^γ *is differentiable at* x*, and is empty otherwise.*

(iv) $\partial h^\gamma(x) = \partial_B h^\gamma(x) \subseteq \frac{1}{\gamma}(x - \operatorname{prox}_{\gamma h}(x))$ *for all* $x \in \mathbb{R}^n$.

(v) h^γ *is differentiable at* x *iff* $\operatorname{prox}_{\gamma h}(x) = \{\bar{x}\}$ *is a singleton, in which case* $\nabla h^\gamma(x) = \frac{1}{\gamma}(x - \bar{x})$ *and, in fact,* h^γ *is strictly differentiable at* x.

(vi) $\frac{1}{\gamma}(x - \bar{x}) \in \hat{\partial} h(\bar{x})$ *for all* $x \in \mathbb{R}^n$ *and* $\bar{x} \in \operatorname{prox}_{\gamma h}(x)$.

Proof.

♠ 1.12*(i)*. See [106, Thm. 1.25].

♠ 1.12*(ii)*. See [106, Thm. 10.32].

♠ 1.12*(iii)*. Follows from [106, Cor. 9.21 and Thm. 10.32].

♠ 1.12*(iv)*. Follows from 1.12*(iii)* and the definition of Bouligand subdifferential.

♠ 1.12*(v)*. h^γ is differentiable at x iff so is $-h^\gamma$, in which case

$$\{-\nabla h^\gamma(x)\} = \partial(-h^\gamma(x)) = -\tfrac{1}{\gamma}\big(x - \operatorname{conv}\operatorname{prox}_{\gamma h}(x)\big),$$

where the equalities follow from [106, Thm. 9.18 and Ex. 10.32]; the same references ensure also that h^γ is in fact strictly differentiable at x in this case. The equations above holds iff $\operatorname{prox}_{\gamma h}$ is a singleton, and the claimed formula for $\nabla h^\gamma(x)$ then also follows.

♠ 1.12*(vi)*. Since $\bar{x}$ minimizes $\psi(w) = h(w) + \frac{1}{2\gamma}\|w - x\|^2$, the necessary optimality conditions (cf. Lem. 1.2) read $0 \in \hat{\partial}\psi(\bar{x}) = \hat{\partial}h(\bar{x}) + \frac{1}{\gamma}(\bar{x} - x)$, where the equality follows from Lem. 1.3*(vi)*. □

Lemma 1.13. *Let $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be γ_h-prox-bounded. Then, for every $\sigma \in \mathbb{R}$ the function $\tilde{h} := h + \frac{\sigma}{2}\|\cdot\|^2$ is prox-bounded with threshold $\gamma_{\tilde{h}} \geq \frac{1}{\gamma_h^{-1} + [\sigma]_-}$. Moreover, for all $\gamma \in \big(0, \min\{\gamma_{\tilde{h}}, 1/[\sigma]_-\}\big)$*

$$\operatorname{prox}_{\gamma\tilde{h}}(x) = \operatorname{prox}_{\frac{\gamma}{1+\gamma\sigma}h}\big(\tfrac{1}{1+\gamma\sigma}x\big)$$

and

$$\tilde{h}^\gamma(x) = h^{\frac{\gamma}{1+\gamma\sigma}}\big(\tfrac{1}{1+\gamma\sigma}x\big) + \tfrac{\sigma}{2(1+\gamma\sigma)}\|x\|^2.$$

Proof. For $\gamma > 0$ and $x, w \in \mathbb{R}^n$, we have

$$\begin{aligned}
\tilde{h}(w) + \tfrac{1}{2\gamma}\|w - x\|^2 &= h(w) + \tfrac{\sigma}{2}\|w\|^2 + \tfrac{1}{2\gamma}\|w - x\|^2 \\
&= h(w) + \tfrac{1}{2}\big(\sigma + \tfrac{1}{\gamma}\big)\|w\|^2 - \tfrac{1}{\gamma}\langle w, x\rangle + \tfrac{1}{2\gamma}\|x\|^2 \\
&= h(w) + \tfrac{1}{2\frac{\gamma}{1+\gamma\sigma}}\big\|w - \tfrac{1}{1+\gamma\sigma}x\big\|^2 + \tfrac{\sigma}{2(1+\gamma\sigma)}\|x\|^2.
\end{aligned}$$

If γ is in bounded as in the statement, then the coefficient of the quadratic term in w is strictly positive and strictly larger than $\frac{1}{2\gamma_h}$. By taking the minimizers and minimum with respect to w we obtain the claimed expressions of $\operatorname{prox}_{\gamma\tilde{h}}$ and $\tilde{h}^\gamma$. □

Regularity properties

Theorem 1.14 (Proximal properties of hypoconvex functions)**.** *Suppose that $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ is σ-hypoconvex. Then, h is prox-bounded with $\gamma_h \geq 1/[\sigma]_-$. Moreover, for all $\gamma \in (0, 1/[\sigma]_-)$ the following hold:*

(i) $\operatorname{prox}_{\gamma h}$ *is single valued and satisfies* $\operatorname{prox}_{\gamma h} = (\mathrm{id} + \gamma\partial h)^{-1}$*; that is, for any* $x \in \mathbb{R}^n$, $\operatorname{prox}_{\gamma h}(x)$ *is the only point* $u \in \mathbb{R}^n$ *such that* $x \in u + \gamma\partial h(u)$.

(ii) $\operatorname{prox}_{\gamma h}$ *is* $\frac{1}{1+\gamma\sigma}$*-Lipschitz continuous and* $(1+\gamma\sigma)$*-cocoercive, and for* $x_i \in \mathbb{R}^n$ *with* $u_i \coloneqq \operatorname{prox}_{\gamma h}(x_i)$, $i = 1, 2$, *one has*

$$(1+\gamma\sigma)\|u_1 - u_2\|^2 \leq \langle u_1 - u_2, x_1 - x_2\rangle \leq \tfrac{1}{1+\gamma\sigma}\|x_1 - x_2\|^2. \tag{1.11}$$

(iii) The Moreau envelope h^γ *is differentiable with* $\nabla h^\gamma(x) = \frac{1}{\gamma}(x - \operatorname{prox}_{\gamma h}(x))$*; in fact, it is* L_{h^γ}*-smooth and* σ_{h^γ}*-hypoconvex, with* $L_{h^\gamma} = \max\left\{\frac{1}{\gamma}, \frac{|\sigma|}{1+\gamma\sigma}\right\}$ *and* $\sigma_{h^\gamma} = \frac{\sigma}{1+\gamma\sigma}$.

(iv) If h *is twice differentiable at* $u \coloneqq \operatorname{prox}_{\gamma h}(x)$*, then*

- $\operatorname{prox}_{\gamma h}$ *is differentiable at* x *with* $J\operatorname{prox}_{\gamma h}(x) = \left[\mathrm{I} + \gamma\nabla^2 h(u)\right]^{-1}$*, and*
- h^γ *is twice differentiable at* x *with* $\nabla^2 h^\gamma(x) = \frac{1}{\gamma}\left[\mathrm{I} - J\operatorname{prox}_{\gamma h}(x)\right]$.

Proof. If $\sigma \geq 0$ then h is convex and thus prox-bounded with threshold $\gamma_h = \infty = 1/[\sigma]_-$. Otherwise, for any $r > -\sigma = -[\sigma]_-$ we have that

$$h(x) + \tfrac{r}{2}\|x\|^2 = \underbrace{h(x) - \tfrac{\sigma}{2}\|x\|^2}_{\text{convex}} + \underbrace{\tfrac{r+\sigma}{2}}_{>0}\|x\|^2$$

is strongly convex, hence lower bounded. By considering $\gamma = 1/r$ it readily follows that $\gamma_h \geq 1/[\sigma]_-$.

Let now $\gamma \in (0, 1/[\sigma]_-)$ be fixed.

♠ 1.14*(i)*. Fix $x \in \mathbb{R}^n$ and consider the function $\psi(w) = \frac{1}{2}\|w - x\|^2 + \gamma h(w)$. Observe that ψ is $(1+\gamma\sigma)$-strongly convex (with $1 + \gamma\sigma > 0$); by definition of $\operatorname{prox}_{\gamma h}$ we have

$$u \in \operatorname{prox}_{\gamma h}(x) \quad\Leftrightarrow\quad u \in \arg\min\psi \quad\Leftrightarrow\quad 0 \in \hat\partial\psi(u) = \gamma\hat\partial h(u) + u - x,$$

where the second implication follows from Lem. 1.2 and the last one from Lem.s 1.3*(i)* and 1.3*(vi)*.

♠ 1.14*(ii)*. Let $x_1, x_2 \in \mathbb{R}^n$ be fixed, and consider $u_i \coloneqq \operatorname{prox}_{\gamma h}(x_i)$, $i = 1, 2$. If $u_1 = u_2$ there is nothing to show, thus let us suppose that $u_1 \neq u_2$. Then, due to 1.14*(i)*, $v_i \coloneqq \frac{1}{\gamma}(x_i - u_i) \in \hat\partial h(u_i)$, $i = 1, 2$. We have

$$\|u_1 - u_2\|\|x_1 - x_2\| \geq \langle u_1 - u_2, x_1 - x_2\rangle$$

$$= \|u_1 - u_2\|^2 + \gamma\langle v_1 - v_2, u_1 - u_2\rangle$$

$$\geq (1 + \gamma\sigma)\|u_1 - u_2\|^2,$$

where the last inequality follows from Lem. 1.4*(c)*. This shows the first inequality in (1.11), as well as the claimed Lipschitz continuity by simply dividing by $\|u_1 - u_2\|$. In turn, the second inequality in (1.11) follows by using Lipschitz continuity on the term $\|u_1 - u_2\|$.

♠ 1.14*(iii)*. From Prop. 1.12 it follows that h^γ is a strictly continuous function on $\mathbb{R}^n$ with $\partial h^\gamma(x) \subseteq \{{}^{(x-u)}/_{\gamma}\}$, where $u \coloneqq \operatorname{prox}_{\gamma h}(x)$. By invoking Lem. 1.3*(iv)* we conclude that h^γ is everywhere differentiable with $\nabla h^\gamma(x) = \frac{1}{\gamma}(x - \operatorname{prox}_{\gamma h}(x))$. Let $x_1, x_2 \in \mathbb{R}^n$ be fixed, and consider $u_i \coloneqq \operatorname{prox}_{\gamma h}(x_i)$, $i = 1, 2$. Then,

$$\langle \nabla h^\gamma(x_1) - \nabla h^\gamma(x_2), x_1 - x_2\rangle = \tfrac{1}{\gamma}\|x_1 - x_2\|^2 - \tfrac{1}{\gamma}\langle u_1 - u_2, x_1 - x_2\rangle$$

$$\overset{(1.11)}{\geq} \tfrac{\sigma}{1+\gamma\sigma}\|x_1 - x_2\|^2,$$

proving the claimed hypoconvexity, as it follows from Lem. 1.7. From (1.11) it also follows that the scalar product in the first line of the inequality above is positive, hence

$$\langle \nabla h^\gamma(x_1) - \nabla h^\gamma(x_2), x_1 - x_2\rangle \leq \tfrac{1}{\gamma}\|x_1 - x_2\|^2.$$

From Thm. 1.8 we infer that ∇h^γ is Lipschitz continuous with modulus $L_{h_\gamma} = \max\left\{\frac{1}{\gamma}, \frac{|\sigma|}{1+\gamma\sigma}\right\}$ as claimed.

♠ 1.14*(iv)*. Since $\nabla h^\gamma = \frac{1}{\gamma}(\mathrm{id} - \operatorname{prox}_{\gamma h})$, it suffices to prove the claim for h^γ. For convex h, the assert is shown in [67, Thm. 3.1]. If h is σ-hypoconvex with $\sigma < 0$, then $\tilde{h} = f - \frac{\sigma}{2}\|\cdot\|^2$ is convex and satisfies

$$h^\gamma(x) = \tilde{h}^{\frac{\gamma}{1+\gamma\sigma}}\left(\tfrac{1}{1+\gamma\sigma}x\right) + \tfrac{\sigma}{2(1+\gamma\sigma)}\|x\|^2,$$

see Lem. 1.13. By using the chain rule of differentiation and rearranging with simple algebra the claimed expression follows. □

Theorem 1.15 (Proximal properties of smooth functions)**.** *Suppose that $h \in C^{1,1}(\mathbb{R}^n)$ is L_h-smooth and σ_h-hypoconvex. Then, additionally to all the claims of Thm. 1.14, for all $\gamma \in (0, {}^1/_{[\sigma_h]_-})$ the following also hold:*

(i) The point $u = \operatorname{prox}_{\gamma h}(s)$ is the only one such that $u + \gamma\nabla h(u) = s$.

(ii) $\mathrm{prox}_{\gamma h}$ *is* $\frac{1}{1+\gamma L_h}$*-strongly monotone: for* $x_i \in \mathbb{R}^n$ *with* $u_i \coloneqq \mathrm{prox}_{\gamma h}(x_i)$, $i = 1, 2$, *one has*

$$\tfrac{1}{1+\gamma L_h}\|x_1 - x_2\|^2 \leq \langle u_1 - u_2, x_1 - x_2\rangle. \tag{1.12}$$

(iii) The Moreau envelope h^γ *is* L_{h^γ}*-smooth with* $L_{h^\gamma} = \max\left\{\frac{L_h}{1+\gamma L_h}, \frac{|\sigma_h|}{1+\gamma\sigma_h}\right\}$.

Proof. 1.15*(i)* directly follows from Thm. 1.14*(i)*, since $\partial h = \nabla h$. Now, let x_i, u_i, $i = 1, 2$, as in the statement be fixed.

♠ 1.15*(ii)*. Define $\psi(x) \coloneqq \gamma h(x) + \frac{1}{2}\|x\|^2$ and observe that ψ is L_ψ-smooth and σ_ψ-strongly convex, with $L_\psi = 1 + \gamma L_h$ and $\sigma_\psi = 1 + \gamma\sigma_h$. It follows from Thm. 1.14*(i)* that $x_i = \nabla\psi(u_i)$, $i = 1, 2$. Cocoercivity of $\nabla\psi$, see Thm. 1.6*(b)*, then implies

$$\begin{aligned}\langle u_1 - u_2, x_1 - x_2\rangle &= \langle u_1 - u_2, \nabla\psi(u_1) - \nabla\psi(u_2)\rangle \\ &\geq \tfrac{1}{1+\gamma L_h}\|\nabla\psi(u_1) - \nabla\psi(u_2)\|^2, \\ &= \tfrac{1}{1+\gamma L_h}\|x_1 - x_2\|^2,\end{aligned}$$

hence the claimed strong monotonicity.

♠ 1.15*(iii)*. From Thm. 1.14*(iii)* we have that $\nabla h^\gamma(x_i) = \frac{1}{\gamma}(x_i - u_i)$, hence

$$\begin{aligned}\tfrac{\sigma_h}{1+\gamma\sigma_h}\|x_1 - x_2\|^2 &\leq \langle \nabla h^\gamma(x_1) - \nabla h^\gamma(x_2), x_1 - x_2\rangle \\ &= \tfrac{1}{\gamma}\|x_1 - x_2\|^2 - \tfrac{1}{\gamma}\langle u_1 - u_2, x_1 - x_2\rangle \\ &= \tfrac{1}{\gamma}\Big(1 - \tfrac{1}{1+\gamma L_h}\Big)\|x_1 - x_2\|^2 = \tfrac{L_h}{1+\gamma L_h}\|x_1 - x_2\|^2,\end{aligned}$$

where the first inequality follows from hypoconvexity of h^γ, see Thm. 1.14*(iii)*, and the second one from the proven strong monotonicity of $\mathrm{prox}_{\gamma h}$. □

1.2.10 Image function

The notion of *image function*, also known as *infimal post-composition* or *epi-composition* [8, 10, 106] will play an important role in Chapter 7.

Definition 1.16 (Image function)**.** *Given $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a linear operator $C \in \mathbb{R}^{m\times n}$, the* IMAGE FUNCTION *$(Ch) : \mathbb{R}^m \to [-\infty, +\infty]$ is defined as*

$$(Ch)(s) := \inf_{w\in\mathbb{R}^n} \{h(w) \mid Cw = s\}.$$

Proposition 1.17. *Let $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lsc, and $C \in \mathbb{R}^{p\times n}$. Suppose that for some $\beta > 0$ the set-valued mapping $X_\beta : \mathbb{R}^p \rightrightarrows \mathbb{R}^n$, defined by*

$$X_\beta(s) := \operatorname*{arg\,min}_{x\in\mathbb{R}^n} \left\{h(x) + \tfrac{\beta}{2}\|Cx - s\|^2\right\}, \tag{1.13}$$

is nonempty for all $s \in \mathbb{R}^p$. Then,

(i) The image function (Ch) is proper.

(ii) $(Ch)(Cx_\beta) = h(x_\beta)$ for all $s \in \mathbb{R}^p$ and $x_\beta \in X_\beta(s)$.

(iii) $\operatorname{prox}_{(Ch)/\beta} \supseteq CX_\beta$.

Proof.

♠ 1.17*(i)*. If $\bar{s} \notin C \operatorname{dom} h$, then $(Ch)(\bar{s}) = \infty$. Otherwise, suppose $\bar{s} = C\bar{x}$ for some $\bar{x} \in \operatorname{dom} h$. Then,

$$-\infty < \min_x \left\{h(x) + \tfrac{\beta}{2}\|Cx - \bar{s}\|^2\right\} \le \inf_{x:\, Cx=\bar{s}} \left\{h(x) + \tfrac{\beta}{2}\|Cx - \bar{s}\|^2\right\} \overset{(def)}{=} (Ch)(\bar{s}),$$

which is upper bounded by the finite quantity $h(\bar{x})$.

♠ 1.17*(ii)*. Since $C(x_\beta + v) = Cx_\beta$ iff $v \in \ker C$, for all $s \in \mathbb{R}^p$ and $x_\beta \in X_\beta(s)$ necessarily $h(x_\beta) \le h(x_\beta + v)$. Consequently,

$$(Ch)(Cx_\beta) \le h(x_\beta) \le \inf_{v\in\ker C} h(x_\beta + v) = \inf_{x:\, Cx=Cx_\beta} h(x) = (Ch)(Cx_\beta).$$

♠ 1.17*(iii)*. Fix $\bar{s} \in \mathbb{R}^p$, and let $x_\beta \in X_\beta(\bar{s})$. Then, from 1.17*(ii)* and the optimality of x_β we have

$$(Ch)(Cx_\beta) + \tfrac{\beta}{2}\|Cx_\beta - \bar{s}\|^2 = h(x_\beta) + \tfrac{\beta}{2}\|Cx_\beta - \bar{s}\|^2 \le h(x) + \tfrac{\beta}{2}\|Cx - \bar{s}\|^2$$

for all $x \in \mathbb{R}^n$. In particular, this holds for all $s \in \mathbb{R}^p$ and x such that $Cx = s$, hence

$$(Ch)(Cx_\beta) + \tfrac{\beta}{2}\|Cx_\beta - \bar{s}\|^2 \le \inf_{x:\, Cx=s} \left\{h(x) + \tfrac{\beta}{2}\|Cx - \bar{s}\|^2\right\} = (Ch)(s) + \tfrac{\beta}{2}\|s - \bar{s}\|^2$$

proving $Cx_\beta \in \operatorname{prox}_{(Ch)/\beta}(\bar{s})$. □

Proposition 1.18. *For an lsc function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $C \in \mathbb{R}^{p\times n}$, let $X : \mathbb{R}^p \rightrightarrows \mathbb{R}^n$ be defined as*

$$X(s) \coloneqq \operatorname*{arg\,min}_{x\in\mathbb{R}^n} \{h(x) \mid Cx = s\}. \tag{1.14}$$

Then, for all $\bar{s} \in C \operatorname{dom} h$ and $\bar{x} \in X(\bar{s})$ it holds that

$$C^\top \hat{\partial}(Ch)(\bar{s}) \subseteq \hat{\partial} h(\bar{x}).$$

Proof. Let $\bar{v} \in \hat{\partial}(Ch)(C\bar{x})$. Then,

$$\begin{aligned}
&\liminf_{\substack{x\to\bar{x}\\ x\neq\bar{x}}} \frac{h(x) - h(\bar{x}) - \langle C^\top\bar{v}, x-\bar{x}\rangle}{\|x-\bar{x}\|}\\
&= \liminf_{\substack{x\to\bar{x}\\ x\neq\bar{x}}} \frac{h(x) - (Ch)(C\bar{x}) - \langle \bar{v}, C(x-\bar{x})\rangle}{\|x-\bar{x}\|}\\
&\geq \liminf_{\substack{x\to\bar{x}\\ x\neq\bar{x}}} \frac{(Ch)(Cx) - (Ch)(C\bar{x}) - \langle \bar{v}, C(x-\bar{x})\rangle}{\|x-\bar{x}\|}\\
&= \liminf_{\substack{x\to\bar{x}\\ x\neq\bar{x}}} \frac{(Ch)(Cx) - (Ch)(C\bar{x}) - \langle \bar{v}, C(x-\bar{x})\rangle}{\|C(x-\bar{x})\|} \frac{\|C(x-\bar{x})\|}{\|x-\bar{x}\|}\\
&\geq 0,
\end{aligned}$$

where in the last inequality we used the fact that $\bar{v} \in \hat{\partial}(Ch)(C\bar{x})$. □

Proposition 1.19 (Strong convexity of the image function)**.** *Suppose that $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper, lsc, and σ_h-strongly convex. Then, for every $C \in \mathbb{R}^{p\times n}$ the image function (Ch) is $\sigma_{(Ch)}$-strongly convex with $\sigma_{(Ch)} = \sigma_h/\|C\|^2$.*

Proof. Plain convexity of (Ch) is shown in [10, Prop. 12.36(ii)]. Since h is strongly convex, for every $s \in C \operatorname{dom} h = \operatorname{dom}(Ch)$ there exists a unique $x_s \in \mathbb{R}^n$ such that $Cx_s = s$ and $(Ch)(s) = h(x_s)$. Let $v_s \in \partial(Ch)(s)$. Then, it follows from Prop. 1.18 that $C^\top v_s \in \partial h(x_s)$, hence, for all $s' \in \operatorname{dom}(Ch)$

$$\begin{aligned}
h(x_{s'}) &\geq h(x_s) + \langle C^\top v_s, x_{s'} - x_s\rangle + \tfrac{\sigma_h}{2}\|x_{s'} - x_s\|^2\\
&\geq h(x_s) + \langle v_s, s' - s\rangle + \tfrac{\sigma_h}{2\|C\|^2}\|s' - s\|^2,
\end{aligned}$$

where in the second inequality the identities $Cx_{s'} = s'$ and $Cx_s = s$ were used. Since $h(x_s) = (Ch)(s)$ and $h(x_{s'}) = (Ch)(s')$, from the arbitrarity of s, s', and v_s, the claimed strong convexity follows. □

In order to proceed to the next result, we first need to introduce the following important notion for parametric minimization.

Definition 1.20 (Locally uniform level boundedness [106, Def. 1.16])**.** *We say that a function $M : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ with values $M(w, x)$ is* LEVEL BOUNDED IN w LOCALLY UNIFORMLY IN x *if for all $\alpha \in \mathbb{R}$ and $\bar{x} \in \mathbb{R}^n$ there exists $\varepsilon > 0$ such that the set*

$$\{(w, x) \in \mathbb{R}^m \times \mathbb{R}^n \mid M(w, x) \leq \alpha,\ \|x - \bar{x}\| \leq \varepsilon\}$$

is bounded.

Theorem 1.21. *Let $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be lsc and $C \in \mathbb{R}^{p\times n}$. Suppose that for some $\beta > 0$ the function $h + \frac{\beta}{2}\|C \cdot - s\|^2$ is level bounded for all $s \in \mathbb{R}^p$. Then, the following hold:*

(i) *(Ch) is proper and lsc.*

(ii) *For all $s \in C \operatorname{dom} h$ the set of minimizers $X(s)$ as in (1.14) is nonempty; moreover, X is locally bounded, and it is osc with respect to (Ch)-attentive convergence: for all $\bar{s} \in C \operatorname{dom} h$*

$$\limsup_{k\to\infty} X(s_k) \subseteq X(\bar{s})$$

whenever $\big(s^k, (Ch)(s^k)\big) \to \big(\bar{s}, (Ch)(\bar{s})\big)$ as $k \to \infty$.

(iii) *For all $\bar{s} \in C \operatorname{dom} h$ and $\bar{x} \in X(\bar{s})$ one has*

$$C^\top \partial(Ch)(\bar{s}) \subseteq \bigcup_{\bar{x}\in X(\bar{s})} \partial h(\bar{x}).$$

Proof. The level boundedness condition ensures that $H(x, s) := h(x)+\delta_{\{0\}}(Cx-s)$ is level bounded in x, locally uniformly in s, cf. Def. 1.20. The first two claims then follow from [106, Thm. 1.32].

Let $\bar{v} \in \partial(Ch)(\bar{s})$ be fixed. Then, there exits a sequence $(s^k, v^k)_{k\in\mathbb{N}} \subseteq \operatorname{gph} \hat{\partial} h$ such that $\big(s^k, (Ch)(s^k), v^k\big) \to \big(\bar{s}, (Ch)(\bar{s}), \bar{v}\big)$ as $k \to \infty$. For each $k \in \mathbb{N}$ let $x^k \in X(s^k)$; then, $(x^k)_{k\in\mathbb{N}}$ is bounded and all its accumulation points belong to $X(\bar{s})$; thus, up to possibly extracting, $x^k \to \bar{x}$ for some $\bar{x} \in X(\bar{s})$ as $k \to \infty$. Then,

$$C^\top \bar{v} = \lim_{k\to\infty} C^\top v^k \overset{1.18}{\in} \limsup_{k\to\infty} \hat{\partial} h(x^k) \subseteq \partial h(\bar{x}),$$

where the last inclusion follows from the definition of ∂h and the fact that

$$h(x^k) = (Ch)(s^k) \to (Ch)(\bar{s}) = h(\bar{x}).$$

The claimed inclusion then follows from the arbitrarity of $\bar{v} \in \partial(Ch)(\bar{s})$. □

Lemma 1.22. *Let $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex and $C \in \mathbb{R}^{p\times n}$ be surjective. Then, (Ch) is convex, and as long as the set of minimizers $X(\bar{s})$ is nonempty (see (1.14)), it holds that*

$$\partial(Ch)(\bar{s}) = \{y \mid A^\top y \in \partial h(\bar{x})\},$$

where $\bar{x}$ is any element of $X(\bar{s})$. In particular, if h is differentiable at some point in $X(\bar{s})$, then (Ch) is differentiable at $\bar{s}$.

Proof. See [57, Thm. D.4.5.1 and Cor. D.4.5.2]. □

Chapter 2

A general framework for the analysis of nonconvex splitting algorithms

2.1 Analysis of fixed-point iterations

A Lyapunov stability approach

One of the most appealing properties of splitting algorithms is their twofold simplicity. First, in most applications their main building blocks amount to relatively cheap algebraic operations; secondly, they are inherently modular, as the same operations are repeated in a recursive fashion through a time-invariant black box. As a result, a splitting algorithm can be efficiently implemented with few lines of code. There is also an important theoretical advantage enabled by such a recursive nature, as the analysis of different algorithms, to a certain extent, can be reduced to that of the common underlying framework of *fixed-point iterations*.

As a prelude to the thesis, in this section we bring the investigated algorithms down to the essential, in the attempt to detect the minimal requirements needed for the development of a sensible theory. To this end, we start by considering fixed-point iterations of the form

$$s^0 \in \mathbb{R}^n, \quad s^{k+1} \in \mathcal{F}(s^k), \quad k = 0, 1, \dots \tag{FP}$$

where the set-valued fixed-point mapping $\mathcal{F}$ complies with the following requirement.

Assumption 2.I. $\mathcal{F} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is osc and nonempty valued (*i.e.*, with full domain).

In the next sections we will then specialize this framework to cases in which the fixed-point mapping has a particular structure, yet is still general enough to cover a vast range of splitting algorithms for nonconvex optimization. We begin with an asymptotic analysis of (FP).

Proposition 2.1. *Suppose that the sequence $(s^k)_{k\in\mathbb{N}}$ generated by (FP) satisfies $\|s^k - s^{k+1}\| \to 0$ as $k \to \infty$. Then, all its accumulation points are fixed points of $\mathcal{F}$.*

Proof. For an arbitrary accumulation point $\bar{s}$, consider a strictly increasing sequence $(k_j)_{j\in\mathbb{N}} \subseteq \mathbb{N}$ such that $s^{k_j} \to \bar{s}$ as $j \to \infty$. Since $\|s^{k_j+1} - s^{k_j}\| \to 0$, the shifted sequence $(s^{k_j+1})_{j\in\mathbb{N}}$ also converges to $\bar{s}$. Thus,

$$\bar{s} = \lim_{j\to\infty} s^{k_j+1} \in \limsup_{j\to\infty} \mathcal{F}(s^{k_j}) \subseteq \mathcal{F}(\bar{s}),$$

where the last inclusion is due to outer semicontinuity of $\mathcal{F}$. □

Having the fixed-point residual $s^k - s^{k+1}$ vanishing is a necessary requirement for ensuring that all accumulation points are fixed. To see this, consider the mapping

$$\mathcal{F}(s) = \begin{cases} \{-1\} & \text{if } s > 0, \\ \{0, \pm 1\} & \text{if } s = 0, \\ \{1\} & \text{otherwise,} \end{cases} \tag{2.1}$$

defined on $\mathbb{R}$. It can be easily verified that $\mathcal{F}$ is osc and that 0 is the unique fixed point. However, starting from $s_0 \neq 0$, the fixed-point iteration sequence will be $s^k = \operatorname{sgn}(s_0)(-1)^k$, hence with ± 1 as accumulation points, none of which belongs to $\operatorname{fix}\mathcal{F}$.

Having $\|s^k - s^{k+1}\| \to 0$ as $k \to \infty$ plays a fundamental role from an algorithmic perspective; for instance, termination criteria based on (the norm of) the fixed-point residual can be imposed, which will be satisfied in a finite number of iterations. However, Proposition 2.1 only investigates the consequences of having the residual vanishing, but it gives no hint as to how such condition can be guaranteed. As the example (2.1) demonstrates, Assumption 2.I alone is not enough for this purpose, and the challenge then turns to providing sufficient properties broad enough to cover the widest possible range of fixed-point iterations. As it will be better detailed in Chapter 8, one such condition involves Lipschitzian properties of $\mathcal{F}$, such as contractiveness or averagedness, which turn out to be general enough to cover algorithms stemming from monotone operator theory. A key property of such iterations is the so-called Fejér monotonicity,

which entails the existence of a constant $c > 0$ such that

$$\operatorname{dist}(\mathcal{F}(s), \operatorname{fix}\mathcal{F})^2 \leq \operatorname{dist}(s, \operatorname{fix}\mathcal{F})^2 - \tfrac{c}{2}\|s - \mathcal{F}(s)\|^2.^{1} \tag{2.2}$$

From a dynamical system perspective, the square distance (from the fixed set) acts as a "potential function" that stabilizes the dynamical system $s^{k+1} = \mathcal{F}(s^k)$. Due to the lower boundedness of the distance function, one can telescope the inequality above for the fixed-point iterations $s^{k+1} = \mathcal{F}(s^k)$ to infer that $(\|s^k - s^{k+1}\|^2)_{k\in\mathbb{N}}$ is summable, hence in particular that $\|s^k - s^{k+1}\| \to 0$ as $k \to \infty$.

Although such a requirement is still too restrictive to encompass the set-valued operators of nonconvex splitting algorithms, the dynamical system interpretation of (2.2) is surely inspirational. Indeed, as long as it is only the vanishing of the fixed-point residual that is concerned, it really makes no difference whether it is the square distance rather than an arbitrary lower bounded function to act as a potential. This is the key point of our analysis, which indeed boils down to the existence of a potential function that behaves the same way the square distance does for Fejér-monotonic sequences (with the due modifications to account for the possible set-valued nature of $\mathcal{F}$). This leads to the following definition.

Definition 2.2 (Lyapunov function)**.** *We say that $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}$ is a* Lyapunov function *for* (FP) *if it satisfies the following properties:*

P1 Lower Boundedness: $\inf \mathcal{L} > -\infty$.

P2 Sufficient Decrease: *there exists a* "sufficient decrease constant" *$c > 0$ such that*

$$\mathcal{L}(s^+) \leq \mathcal{L}(s) - \tfrac{c}{2}\|s - s^+\|^2 \quad \text{for all } s \in \mathbb{R}^n \text{ and } s^+ \in \mathcal{F}(s).$$

We now proceed to formalize the intuition that any Lyapunov function as in Definition 2.2 is a suitable replacement for the square distance in (2.2). The following result will be useful to this end.

Lemma 2.3. *Suppose that $(s^k)_{k\in\mathbb{N}} \subseteq \mathbb{R}^n$ is bounded and satisfies $\|s^k - s^{k+1}\| \to 0$ as $k \to \infty$. Then, the set of accumulation points ω of $(s^k)_{k\in\mathbb{N}}$ is nonempty and compact, and such that $\operatorname{dist}(s^k, \omega) \to 0$ as $k \to \infty$.*

Proof. This is shown in [25, Rem. 5] (the claim therein is slightly misstated, as the needed boundedness requirement is not explicitly mentioned). □

[1] Fejér monotonicity is, in fact, a stronger property which will be introduced in Definition 8.2. This inequality is more akin to the *linear monotonicity* described in [76], although linearity is not required here.

Theorem 2.4 (Subsequential convergence)**.** *Suppose that a fixed-point mapping $\mathcal{F}$ as in Assumption 2.I admits a Lyapunov function $\mathcal{L}$. Then, the following hold for the fixed-point iterations (FP):*

(i) The fixed-point residual $(\|s^k - s^{k+1}\|)_{k\in\mathbb{N}}$ is square-summable; in particular, $\min_{j\le k}\|s^j - s^{j+1}\| \in O(1/\sqrt{k})$.

(ii) Every accumulation point of $(s^k)_{k\in\mathbb{N}}$ satisfies $\bar{s} \in \operatorname{fix}\mathcal{F}$.

(iii) If $(s^k)_{k\in\mathbb{N}}$ is bounded (as it is the case when $\mathcal{L}$ is level bounded), then the set of accumulation points ω is nonempty, compact and connected, and $\operatorname{dist}(s^k,\omega)\to 0$ as $k\to\infty$.

Proof.

♠ 2.4*(i)*. The sufficient decrease property of $\mathcal{L}$ ensures the existence of a constant $c>0$ such that

$$\mathcal{L}(s^{k+1}) \le \mathcal{L}(s^k) - \tfrac{c}{2}\|s^k - s^{k+1}\|^2 \quad \forall k\in\mathbb{N}.$$

Since $\mathcal{L}$ is real valued, for any $K\ge 1$ we may telescope the inequality for $k=1,2,\dots,K$ to arrive to

$$\sum_{k=1}^{K}\|s^k - s^{k+1}\|^2 \le \frac{2}{c}\sum_{k=1}^{K}\big(\mathcal{L}(s^k)-\mathcal{L}(s^{k+1})\big) = \frac{2}{c}\big(\mathcal{L}(s^1)-\mathcal{L}(s^{K+1})\big)$$
$$\le \frac{2}{c}\big(\mathcal{L}(s^1)-\inf\mathcal{L}\big).$$

Since $\mathcal{L}$ is lower bounded, the partial sums $\sum_{k=1}^{K}\|s^k - s^{k+1}\|^2$ are upper bounded by a same finite constant for all $K\ge 1$. By letting $K\to\infty$, square summability follows. As to the claimed $O(1/\sqrt{k})$ rate, notice that, for all $K\in\mathbb{N}$,

$$\infty > S := \sum_{k=0}^{\infty}\|s^k - s^{k+1}\|^2 \ge \sum_{k=0}^{K}\|s^k - s^{k+1}\|^2 \ge \sum_{k=0}^{K}\min_{j\le k}\|s^j - s^{j+1}\|^2$$
$$\ge (K+1)\min_{j\le K}\|s^j - s^{j+1}\|^2,$$

where in the last inequality we used the decreasing behavior of the sequence $(\min_{j\le k}\|s^j - s^{j+1}\|^2)_{k\in\mathbb{N}}$. Thus, $\min_{j\le K}\|s^j - s^{j+1}\| \le \sqrt{S/K+1}$ for all $K\in\mathbb{N}$.

♠ 2.4*(ii)* & 2.4*(iii)*. Since $\|s^k - s^{k+1}\|\to 0$, the first assert follows from Prop. 2.1, and the second from Lem. 2.3. □

We conclude the section with a remark on some properties relating fixed-point mappings and their Lyapunov functions.

Lemma 2.5. *Suppose that the fixed-point mapping $\mathcal{F}$ as in Assumption 2.I admits a Lyapunov function $\mathcal{L}$. Then,*

(i) $\mathcal{F}(s)$ *is compact for all* $s \in \mathbb{R}^n$.

(ii) $\mathcal{L}(s) - \inf \mathcal{L} \geq \frac{c}{2}\|s - \bar{s}\|^2$ *for all* $s \in \mathbb{R}^n$ *and* $\bar{s} \in \mathcal{F}(s)$.

In particular,

(iii) $\mathcal{F}(s_\star) = \{s_\star\}$ *for all* $s_\star \in \operatorname{arg\,min} \mathcal{L}$ *(hence* $\operatorname{arg\,min} \mathcal{L} \subseteq \operatorname{fix} \mathcal{F}$*).*

Proof. The set $\mathcal{F}(s)$ must be bounded for all $s \in \mathbb{R}^n$, for otherwise either the lower boundedness 2.2.P1 or the sufficient decrease property 2.2.P2 would be violated. That $\mathcal{F}(s)$ is closed holds regardless of whether $\mathcal{F}$ admits a Lyapunov function or not, owing to the fact that any sequence contained in $\mathcal{F}(s)$ has, by definition of osc, all accumulation points in $\mathcal{F}(s)$. Moreover, for any $\bar{s} \in \mathcal{F}(s)$ we have

$$\mathcal{L}(s) - \inf \mathcal{L} \geq \mathcal{L}(s) - \mathcal{L}(\bar{s}) \geq \tfrac{c}{2}\|s - \bar{s}\|^2,$$

where the second inequality follows from the sufficient decrease property. The last claim follows straightforwardly. □

We terminate here the abstract fixed-point framework and begin to specialize the study to the solution of nonconvex optimization problems. The rest of the chapter is dedicated to establishing the class of investigated fixed-point mappings. In Chapter 3 we will then analyze their convergence by introducing *proximal envelopes*, which will prove to be particularly suitable Lyapunov functions.

2.2 Fixed-point iterations in optimization

The challenges of nonconvexity

We now begin to specialize the fixed-point framework (FP) for solving optimization problems

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \;\; \varphi(x), \tag{P}$$

where $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a proper, lsc, extended-real valued function with nonempty set of minimizers. Unless differently specified, these minimal requirements on the cost function φ will be assumed in the sequel.

There are two main issues that need be addressed. First, the mapping $\mathcal{F}$ must be consistent with the investigated problem, as the limit points of its fixed-point iterations must somehow be related to the solutions of (P). This, in turn, raises the issue of properly defining what a "solution" is, for "solving" (P) may acquire a meaning broader than that of finding minimizers of φ. This second point arises because of the possible nonconvex nature of φ. Under convexity, instead, first-order optimality is necessary and sufficient for global optimality, and indeed splitting algorithms address the equivalent problem of finding first-order optimal points; the sought (global) minimizers are then obtained, up to possibly operating a change of variable (as it is the case of the Douglas-Rachford splitting where a proximal mapping relates fixed points of the operator with solutions to the optimization problem). Although this is not the case in the more general framework investigated here, once again we shall gain some insight from the convex realm.

In the previous section we characterized the limit point(s) of fixed-point iterations of $\mathcal{F}$ as those belonging to $\operatorname{fix}\mathcal{F}$. The idea is then to seek fixed points of $\mathcal{F}$, all of which, up to possibly operating a change of variable, shall satisfy some necessary condition for optimality for (P). This leads to the following criterion of *compatibility* between the fixed-point mapping $\mathcal{F}$ and problem (P).

Definition 2.6 (Compatibility)**.** *We say that a fixed-point mapping $\mathcal{F} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is* COMPATIBLE *with problem* (P) *if the following properties are satisfied:*

P1 *$\mathcal{F}$ complies with Assumption 2.I.*

P2 *There exists an L_G-Lipschitz continuous and μ_G-strongly monotone function $G : \mathbb{R}^n \to \mathbb{R}^n$ such that*

$$\arg\min \varphi \subseteq G(\operatorname{fix}\mathcal{F}) \subseteq \operatorname{zer}\hat{\partial}\varphi.$$

The strong monotonicity of G implies, in particular, that G is invertible with μ_G^{-1}-Lipschitz inverse. As expectable, the lack of convexity may result in the need of additional assumptions on the problem, and these will happen to ensure this apparently overly restrictive property.

The majorization-minimization (MM) principle will be the core of our approach. As the name suggests, an MM step amounts to the minimization of a surrogate function that is pointwise greater than the real objective φ. Our algorithmic framework will consist of "generalized" MM schemes, in which "pure" MM steps may be composed with the transformation mapping G appearing in Definition 2.6. The rest of the chapter is devoted to developing the needed theory. We begin with the analysis of "pure" MM schemes, and later complete the picture by including the claimed generalization.

2.3 Proximal majorization-minimization

PMM schemes

As the name suggests, *majorization-minimization* algorithms address the minimization of a function φ by iteratively minimizing a larger surrogate function, or model. The approach is convenient whenever the minimization of the model is easier than that of the original function φ. A classical definition of *majorizing model* can be formulated as follows [64].

Definition 2.7 (Majorizing model)**.** *The function $\mathcal{M} : \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$ is a* MAJORIZING MODEL *for the proper, lsc, and lower bounded function φ if*

P1 $\mathcal{M}(x;x) = \varphi(x)$ *for all* $x \in \mathbb{R}^n$*, and*

P2 $\mathcal{M}(w;x) \geq \varphi(w)$ *for all* $x, w \in \mathbb{R}^n$*.*

The collection of all majorizing models for φ is denoted by $\overline{\mathfrak{M}}_\varphi$.

Given a model $\mathcal{M}$, a majorization minimization (MM) step at x consists of selecting $x^+ \in \operatorname{argmin}_w \mathcal{M}(w;x)$; due to the tangency condition 2.7.P1, one can easily infer that any such x^+ satisfies $\varphi(x^+) \leq \varphi(x)$. Let us consider two (de)motivating examples.

Example 2.8 (Maximal model)**.** Let

$$\overline{\mathcal{M}}_0(w;x) \coloneqq \begin{cases} \varphi(x) & \text{if } w = x, \\ \infty & \text{otherwise.} \end{cases}$$

Then, $\overline{\mathcal{M}}_0$ is a majorizing model for φ, in fact it is the pointwise largest such function. However, this model is quite useless for the sake of minimizing φ, for one MM step at any point $x \in \mathbb{R}^n$ necessarily yields $x^+ = x$ if $x \in \operatorname{dom} \varphi$, and no x^+ even exists otherwise. □

Example 2.9 (Minimal model)**.** Let

$$\mathcal{M}(w;x) \coloneqq \varphi(w).$$

This time, $\mathcal{M}$ is the pointwise smallest majorizing model for φ. Once again this model turns out to be of no help in the minimization of φ, being one MM step as hard as directly minimizing φ. In fact, $\operatorname{argmin}_w \mathcal{M}(w;x) = \operatorname{argmin} \varphi$ for any $x \in \mathbb{R}^n$. □

These two examples are a clear indication that for the sake of a sensible theory some additional requirements on the employed models are in order. The survey

[116] offers a nice overview on MM algorithms where these aspects are discussed. A common requirement is directional differentiability of both the cost function and the model together with a compatibility requirement on the directional derivatives [15, 108]. Alternatively, [78] requires the model to differ from the cost function by a smooth term. The additional assumptions are usually imposed both to relax classical MM iterations, for instance with randomized or distributed variants or by using approximate majorizing models, and to ensure that limit points $x_\star$ of MM iterations satisfy *directional stationarity*, namely the inequality $\varphi'(x_\star; d) \geq 0$ for all $d \in \mathbb{R}^n$, which is implied by the stationarity condition $0 \in \hat{\partial}\varphi(x_\star)$.

The MM framework proposed in this thesis is based on the notion of *proximal* majorizing models, defined in the next subsection, and makes no regularity assumptions on the problem or the model, other than mere lower semicontinuity.

2.3.1 Proximal majorizing models

In accordance with the quadratic majorization in Definition 1.11 of proximal mapping, we will restrict the analysis to "proximal" majorizing models, defined as follows.

Definition 2.10 (Proximal majorizing model)**.** *The function $\mathcal{M}(w; x) : \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$ is a* PROXIMAL MAJORIZING MODEL *for φ if*

P0 *$\mathcal{M}$ is lsc,*

P1 *$\mathcal{M}(x; x) = \varphi(x)$ for all $x \in \mathbb{R}^n$, and*

P2 *there exist $m_1, m_2 > 0$ such that*

$$\varphi(w) + \tfrac{m_1}{2}\|w - x\|^2 \leq \mathcal{M}(w; x) \leq \varphi(w) + \tfrac{m_2}{2}\|w - x\|^2$$

for all $x, w \in \mathbb{R}^n$.

The collection of all proximal majorizing models for φ is denoted by $\mathfrak{M}_\varphi$.

To streamline the notation, we may omit the 'majorizing' part and refer to elements of $\mathfrak{M}_\varphi$ simply as PROXIMAL MODELS.

Definition 2.11 (Continuity of proximal models)**.** *We say that a proximal model $\mathcal{M}$ is continuous if the sections $x \mapsto \mathcal{M}(w; x)$ are continuous for all $w \in \mathbb{R}^n$.*

An MM step relative to a proximal majorizing model $\mathcal{M}$ will be referred to as a PROXIMAL MM (PMM) STEP. The set-valued mapping that associates a point x to all its proximal MM steps through $\mathcal{M}$ will be called PMM MAPPING, and denoted as $\mathcal{T}^{\mathcal{M}}$. Namely, $\mathcal{T}^{\mathcal{M}} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is defined as

$$\mathcal{T}^{\mathcal{M}}(x) \coloneqq \operatorname*{arg\,min}_{w \in \mathbb{R}^n} \mathcal{M}(w; x). \tag{2.3}$$

Example 2.12 (Proximal point)**.** For $\gamma > 0$, let

$$\mathcal{M}^{\mathrm{PP}}_{\gamma}(w; x) \coloneqq \varphi(x) + \tfrac{1}{2\gamma}\|w - x\|^2.$$

Clearly, $\mathcal{M}^{\mathrm{PP}}_{\gamma}$ is a proximal majorizing model, specifically with $m_1 = m_2 = 1/\gamma$ as parameters in property 2.10.P2. The associated PMM mapping is $\mathcal{T}^{\mathcal{M}^{\mathrm{PP}}_{\gamma}} = \operatorname{prox}_{\gamma\varphi}$, the proximal mapping of φ with parameter γ. □

In the next subsection we list some of the advantages that the properties of proximal majorizing models pose over the classical Definition 2.7.

2.3.2 Properties

Theorem 2.13 (Regularity of the PMM mapping)**.** *For any proximal majorizing model $\mathcal{M} \in \mathfrak{M}_{\varphi}$, the PMM mapping $\mathcal{T}^{\mathcal{M}}$ is osc, nonempty- and compact-valued. In particular, $\mathcal{T}^{\mathcal{M}}$ satisfies Assumption 2.I.*

Proof. Due to the lower bound in 2.10.P2, we have

$$\mathcal{M}(w; x) \geq \varphi(w) + \tfrac{m_1}{2}\|w - x\|^2 \geq \inf \varphi + \tfrac{m_1}{2}\|w - x\|^2.$$

Since φ is proper and lower bounded, it holds that $\inf \varphi \in \mathbb{R}$ and therefore $\mathcal{M}$ is level bounded in w locally uniformly in x, cf. Def. 1.20. Moreover, for any $z \in \operatorname{dom} \varphi$ (such z exists due to properness of φ) we have

$$\inf_{w \in \mathbb{R}^n} \mathcal{M}(w; x) \leq \mathcal{M}(z; x) \leq \varphi(z) + \tfrac{m_2}{2}\|z - x\|^2 < \infty,$$

hence the parametric infimum with respect to w is everywhere finite. The claimed properties of $\mathcal{T}^{\mathcal{M}}$ then follow from [106, Thm. 1.17]. □

Lemma 2.14 (Basic inequality)**.** *Let $\mathcal{M} \in \mathfrak{M}_{\varphi}$ and $x \in \mathbb{R}^n$ be fixed. Then,*

$$\mathcal{M}(\bar{x}; x) \leq \varphi(x) \quad \textit{for all } \bar{x} \in \mathcal{T}^{\mathcal{M}}(x),$$

and equality holds iff $x \in \operatorname{fix} \mathcal{T}^{\mathcal{M}}$.

Proof. For any $\bar{x} \in \mathcal{T}^{\mathcal{M}}(x)$ we have

$$\varphi(x) \overset{2.10.\text{P1}}{=} \mathcal{M}(x;x) \geq \inf_{w\in\mathbb{R}^n} \mathcal{M}(w;x) \overset{(def)}{=} \mathcal{M}(\bar{x};x).$$

Thus, equality holds iff $\mathcal{M}(x;x) = \inf_{w\in\mathbb{R}^n} \mathcal{M}(w;x)$, which is equivalent to having $x \in \operatorname{arg\,min}_{w\in\mathbb{R}^n} \mathcal{M}(w;x) \overset{(def)}{=} \mathcal{T}^{\mathcal{M}}(x)$. □

We conclude the subsection with a useful inequality that relates the norm of the subgradients after one MM step with the fixed-point residual.

Lemma 2.15 (An error-bound-like inequality)**.** *Let $\mathcal{M}$ be a proximal model for φ, and suppose that the difference $\delta(w) := \mathcal{M}(w;x) - \varphi(w)$ is differentiable with L_δ-Lipschitz gradient. Then,*

$$\operatorname{dist}(0, \hat{\partial}\varphi(\bar{x})) \leq L_\delta \|x - \bar{x}\| \quad \textit{for all } x \in \mathbb{R}^n \textit{ and } \bar{x} \in \mathcal{T}^{\mathcal{M}}(x).$$

Proof. Due to property 2.10.P2, it holds that

$$\tfrac{m_1}{2}\|w - x\|^2 \leq \delta(w) \leq \tfrac{m_2}{2}\|w - x\|^2,$$

and in particular $\nabla\delta(x) = 0$. Combined with the L_δ-Lipschitz continuity of $\nabla\delta$, we obtain $\|\nabla\delta(\bar{x})\| \leq L_\delta\|x - \bar{x}\|$. Since $\bar{x}$ minimizes $w \mapsto \mathcal{M}(w;x)$, we have

$$0 \in \hat{\partial}_w \mathcal{M}(\bar{x};x) = \hat{\partial}\varphi(\bar{x}) + \nabla\delta(\bar{x}),$$

hence $-\nabla\delta(\bar{x}) \in \hat{\partial}\varphi(\bar{x})$. Thus, $\operatorname{dist}(0, \hat{\partial}\varphi(\bar{x})) \leq \| - \nabla\delta(\bar{x})\| \leq L_\delta\|x - \bar{x}\|$. □

The requirements for δ in Lemma 2.15 are actually restrictive, but they suffice to our purposes. Indeed, if δ is strictly continuous, then the inclusion (as opposed to equality) $0 \in \hat{\partial}_w \mathcal{M}(\bar{x};x) \subseteq \hat{\partial}\varphi(\bar{x}) + \hat{\partial}\delta(\bar{x})$ still holds (cf. Lem. 1.3*(iii)*), and we then infer that $-v \in \hat{\partial}\varphi(\bar{x})$ for some $v \in \hat{\partial}\delta(\bar{x})$. In order to ensure a bound of the form $\|v\| \leq L\|x - \bar{x}\|$, it then suffices to require the following *calmness* condition: there exists $L \geq 0$ such that

$$\max_{v\in\hat{\partial}\delta(w)} \|v\| \leq L\|w - x\| \quad \text{for all } w \in \mathbb{R}^n.$$

Thus, we infer from [106, Thm. 9.13(a)-(f)] that *a sufficient condition for Lemma 2.15 to hold is having δ locally Lipschitz-continuous with modulus growing at most linearly with respect to $\|w - x\|$* (which is indeed the case when δ is Lipschitz-differentiable).

The inequality in Lemma 2.15 is closely related to the error bound condition under which linear convergence of some proximal algorithms can be established, see [77, 41]. The key difference is that an error bound would require the subdifferential at x, as opposed to the one at $\bar{x} \in \mathcal{T}^{\mathcal{M}}(x)$. Although less powerful, the given inequality has still some useful consequences.

2.3.3 Partial ordering

As the extreme Examples 2.8 and 2.9 confirm, it is intuitive that the more a model $\mathcal{M}$ penalizes points $(w;x)$ far from $(x;x)$, the more the minimizers defining $\mathcal{T}^{\mathcal{M}}$ will be close to x, hence the likelier for x to be a fixed point. In this sense, being a fixed point for a "high" model may be regarded as a loose property; on the contrary, if x is fixed with respect to a "low" model, the extreme scenario depicted in Example 2.9 may suggest that x is closer to be a (local) optimum. In the next pages we will confirm this intuition, and we will indeed quantify, to some extent, local optimality of a point in terms of how much a model can be "pushed down" without affecting the status of being a fixed point. "Pushing down" is to be meant in the sense of considering a lower model; it thus becomes necessary to formalize how models can be compared one another. An intuitive such option is to define a pointwise ordering, agreeing that $\mathcal{M}$ is *lower* than $\mathcal{M}'$ if $\mathcal{M}(w;x) \leq \mathcal{M}'(w;x)$ holds for all points $x, w \in \mathbb{R}^n$. Nevertheless, to enforce some degree of uniformity when comparing models it is convenient to work with a coarser partial ordering, which we define next.

Definition 2.16. *Given two majorizing models $\mathcal{M}, \mathcal{M}' \in \overline{\mathfrak{M}}_\varphi$, we write $\mathcal{M} \prec \mathcal{M}'$ to indicate that $\mathcal{M} \leq \mathcal{M}'$ pointwise, and*

$$\mathcal{M}(w;x) < \mathcal{M}'(w;x) \quad \text{for all } (w;x) \in \operatorname{dom} \mathcal{M} \text{ with } w \neq x.$$

The relation $\mathcal{M} \preceq \mathcal{M}'$ then indicates that either $\mathcal{M} = \mathcal{M}'$ or $\mathcal{M} \prec \mathcal{M}'$; the relations $\succ$ and $\succeq$ are defined accordingly.

The restriction $w \neq x$ in Definition 2.16 rules out points on the diagonal $\{(x;x) \mid x \in \mathbb{R}^n\}$ where all elements of $\overline{\mathfrak{M}}_\varphi$ must agree, as prescribed by property 2.7.P1. Further restricting to points $(w;x) \in \operatorname{dom} \mathcal{M}$ ensures that infinite values are not compared. In the case of proximal models $\mathcal{M}, \mathcal{M}' \in \mathfrak{M}_\varphi$, in light of property 2.10.P2 having $\mathcal{M}(w;x) = \infty$ is equivalent to having $x \notin \operatorname{dom} \varphi$, and in particular all proximal models have same domain. Thus, $\mathcal{M} \prec \mathcal{M}'$ for proximal models indicates that the strict inequality $<$ holds pointwise wherever the $\mathcal{M}$ and $\mathcal{M}'$ can differ.

Lemma 2.17 (Existence of extrema)**.** *Relative to the partial ordering $\succeq$, any family $\mathscr{A} \subseteq \mathfrak{M}_\varphi$ of proximal models admits an (lsc) supremum and an infimum*

in $\overline{\mathfrak{M}}_\varphi$*. In particular, the family of all proximal models has* $(w;x) \mapsto \varphi(w)$ *as infimum and* $\overline{\mathcal{M}}_0$ *as in Example 2.8 as supremum.*

Proof. That $(w;x) \mapsto \varphi(w)$ and $\overline{\mathcal{M}}_0$ are the extrema of $\mathfrak{M}_\varphi$ is trivial. Suppose that $\mathscr{A} \neq \emptyset$, and consider the pointwise supremum $\bar{M} \coloneqq \sup(\mathscr{A}, \geq)$. Clearly, $\bar{M} \in \overline{\mathfrak{M}}_\varphi$, and since proximal models are lsc by definition, it follows from [106, Prop. 1.26(a)] that $\bar{M}$ is lsc as well. Notice that $\mathcal{M} \preceq \mathcal{M}'$ implies the pointwise inequality $\mathcal{M} \leq \mathcal{M}'$, hence one can easily verify that $\bar{M}$ is the sought supremum. The claim for the infimum is similar, except that lower semicontinuity cannot be deduced. □

Although they always exist, extrema of families of majorizing models should be treated with care, lest intuitive properties of total order relations are mistakenly attributed to a partial ordering such as $\succeq$. For instance, having $\sup(\mathscr{A}, \succeq) = \overline{\mathcal{M}}_0$ for a family $\mathscr{A}$ of majorizing models does not imply, given a model $\bar{\mathcal{M}}$, the existence of a model $\mathcal{M} \in \mathscr{A}$ such that $\mathcal{M} \succeq \bar{\mathcal{M}}$. To see this, consider

$$\bar{\mathcal{M}}(w;x) \coloneqq \varphi(w) + \|w - x\|^2$$

and $\mathscr{A} \coloneqq \{\mathcal{M}_k \mid k \in \mathbb{N}\}$, where

$$\mathcal{M}_k(w;x) \coloneqq \varphi(w) + \tfrac{1}{2}\|w - x\|^2 + k\|w - x\|.$$

Then, $(\mathscr{A}, \succeq)$ is even totally ordered and its supremum is $\overline{\mathcal{M}}_0$, yet no model $\mathcal{M}_k$ satisfies $\mathcal{M}_k \succeq \bar{\mathcal{M}}$ (in fact, not even $\mathcal{M}_k \geq \bar{\mathcal{M}}$).

The following result shows how by raising proximal models towards the supremum $\overline{\mathcal{M}}_0$ the pathological behavior depicted in Example 2.8 is approached.

Lemma 2.18. *Let* $(\mathcal{M}_k)_{k\in\mathbb{N}} \subset (\mathfrak{M}_\varphi, \succeq)$ *be an increasing sequence of proximal models such that* $\sup(\mathcal{M}_k)_{k\in\mathbb{N}} = \overline{\mathcal{M}}_0$*. Then,* $\limsup_{k\to\infty} \mathcal{T}^{\mathcal{M}_k}(x) = \{x\}$ *for any* $x \in \operatorname{dom}\varphi$*.*

Proof. Due to property 2.10.P2 and since $(\mathcal{M}_k)_{k\in\mathbb{N}}$ is increasing, there exists $m_1 > 0$ (independent of k) such that

$$\mathcal{M}_k(w;z) \geq \varphi(w) + \tfrac{m_1}{2}\|w - z\|^2 \quad \text{for all } w, z \in \mathbb{R}^n \text{ and } k \in \mathbb{N}. \tag{2.4}$$

Let $x \in \operatorname{dom}\varphi$ be fixed and let $\bar{x}_k \in \mathcal{T}^{\mathcal{M}_k}(x)$ be arbitrary. We have

$$\varphi(x) \overset{2.14}{\geq} \mathcal{M}_k(\bar{x}_k;x) \overset{(2.4)}{\geq} \varphi(\bar{x}_k) + \tfrac{m_1}{2}\|x - \bar{x}_k\|^2 \geq \inf\varphi + \tfrac{m_1}{2}\|x - \bar{x}_k\|^2.$$

Since $\varphi(x) < \infty$, it follows that $(\bar{x}^k)_{k\in\mathbb{N}}$ is bounded. To arrive to a contradiction, suppose that a subsequence $(\bar{x}^{k_j})_{j\in\mathbb{N}}$ converges to a point $\bar{x} \neq x$, and let $B \coloneqq \mathrm{B}(\bar{x}; \|x - \bar{x}\|)$. Then, with similar arguments we obtain

$$\varphi(x) \overset{2.14}{\geq} \limsup_{j\in\mathbb{N}} \mathcal{M}_{k_j}(\bar{x}^{k_j}; x) \geq \limsup_{j\in\mathbb{N}} \inf_{w\in B} \mathcal{M}_{k_j}(w; x) = \infty,$$

where the second inequality follows from the fact that $\bar{x}^{k_j} \in B$ for large enough j's, and the last equality from the fact that $\overline{\mathcal{M}}_0(w; x) = \infty$ for all $w \in B$. This contradicts the fact that $x \in \operatorname{dom}\varphi$. □

2.4 Criticality

In this section we finally establish the intuitive connections brought forth in the previous section linking fixed points with local optima. As a byproduct, the compatibility (in the sense of Definition 2.6) of a PMM mapping $\mathcal{T}^{\mathcal{M}}$ and the minimization of φ will be established. Theorem 2.22 will then provide a first hint in support of the chosen partial ordering between models. Further evidence will be given later on with the introduction in Definition 2.31 of a "threshold" function, that measures the "optimality" of a point by analyzing the fixed sets of sufficiently large models.

We begin by establishing a terminology to replace the vague term "solution" (to problem (P)) with a more specific dedicated expression. A constructive way to assess whether a point complies with this definition will be given in Corollary 2.24.

Definition 2.19 (Criticality)**.** *We say that $\bar{x} \in \mathbb{R}^n$ is a* CRITICAL POINT *for φ if there exists a proximal model $\mathcal{M} \in \mathfrak{M}_\varphi$ such that $\bar{x} \in \operatorname{fix}\mathcal{T}^{\mathcal{M}}$.*

In the MM setting addressed in [15], this property is referred to as *strong stationarity* and is shown to be stronger than directional stationarity and necessary for global optimality. These facts still hold in the proximal MM framework addressed in this thesis, as the next result shows; see also Theorem 2.25 and Prop. 2.33.

Proposition 2.20 (Higher-order stationarity of critical points)**.** *Suppose that $\bar{x}$ is critical for φ; then,*

$$\varphi(x) - \varphi(\bar{x}) \geq O(\|x - \bar{x}\|^2).$$

In particular, not only $0 \in \hat{\partial}\varphi(\bar{x})$*, but for all* $\vartheta \in [0,1)$ *the following stronger stationarity property holds:*

$$\liminf_{\substack{x \to \bar{x} \\ x \neq \bar{x}}} \frac{\varphi(x) - \varphi(\bar{x})}{\|x - \bar{x}\|^{1+\vartheta}} \geq 0. \tag{2.5}$$

Proof. Let $\mathcal{M} \in \mathfrak{M}_\varphi$ be such that $\bar{x} \in \operatorname{fix} \mathcal{T}^{\mathcal{M}}$. We have,

$$\begin{aligned} \varphi(x) - \varphi(\bar{x}) &\overset{2.10.\text{P1}}{=} \varphi(x) - \mathcal{M}(\bar{x};\bar{x}) \overset{2.10.\text{P2}}{\geq} \mathcal{M}(x;\bar{x}) - \tfrac{m_2}{2}\|x-\bar{x}\|^2 - \mathcal{M}(\bar{x};\bar{x}) \\ &\geq \inf_{w \in \mathbb{R}^n} \mathcal{M}(w;\bar{x}) - \mathcal{M}(\bar{x};\bar{x}) - \tfrac{m_2}{2}\|x-\bar{x}\|^2 \\ &= -\tfrac{m_2}{2}\|x-\bar{x}\|^2, \end{aligned}$$

where the last equality follows from the fact that $\bar{x} \in \mathcal{T}^{\mathcal{M}}(\bar{x})$. □

The bound $\vartheta < 1$ in the higher-order stationarity property (2.5) is tight, and we need to accept the fact that saddle points or maxima cannot be avoided, as shown in the next example.

Example 2.21. Consider $\varphi : \mathbb{R}^2 \to \mathbb{R}$ given by

$$\varphi(x) = \max\left\{-1, (x_1 - x_2)^2 - \tfrac{3}{2}x_2^2\right\},$$

which has a saddle point at $\bar{x} = (0,0)$. However, $\bar{x}$ also happens to be critical: consider the proximal point model $\mathcal{M}_\gamma^{\text{PP}}$ of Example 2.12 with $\gamma < 1/2$. Then

$$\mathcal{T}^{\mathcal{M}_\gamma^{\text{PP}}}(\bar{x}) = \operatorname{prox}_{\gamma\varphi}(0,0) = \{(0,0)\} = \{\bar{x}\}.$$

□

The next result shows that the partial ordering $\succeq$ among proximal models is paralleled by an inclusion of the fixed sets, in the sense that the lower the model, the stronger the property of being a fixed point. It also shows an interesting fact when the relation is strict which will be important later on in the thesis when regularity properties will be discussed.

Theorem 2.22. *For any pair of proximal models* $\mathcal{M}, \mathcal{M}' \in \mathfrak{M}_\varphi$*, the following hold:*

(i) If $\mathcal{M} \preceq \mathcal{M}'$*, then* $\operatorname{fix} \mathcal{T}^{\mathcal{M}} \subseteq \operatorname{fix} \mathcal{T}^{\mathcal{M}'}$.

(ii) If $\mathcal{M} \prec \mathcal{M}'$*, then* $\mathcal{T}^{\mathcal{M}'}(x) = \{x\}$ *for all* $x \in \operatorname{fix} \mathcal{T}^{\mathcal{M}}$.

Proof. To ease the notation, let us denote $\mathcal{T} \coloneqq \mathcal{T}^{\mathcal{M}}$ and $\mathcal{T}' \coloneqq \mathcal{T}^{\mathcal{M}'}$.

♠ 2.22*(i)*. Let $x \in \operatorname{fix} \mathcal{T}$. Then, for all $\bar{x}' \in \mathcal{T}'(x)$ we have

$$\varphi(x) \overset{2.10.\text{P1}}{=} \mathcal{M}(x;x) \leq \mathcal{M}(\bar{x}';x) \leq \mathcal{M}'(\bar{x}';x) \overset{(def)}{=} \inf_{w\in\mathbb{R}^n} \mathcal{M}'(w;x)$$

$$\leq \mathcal{M}'(x;x) \overset{2.10.\text{P1}}{=} \varphi(x),$$

where the first inequality follows from the fact that $x \in \mathcal{T}(x)$. Therefore, $\mathcal{M}'(x;x) = \inf_{w\in\mathbb{R}^n} \mathcal{M}'(w;x)$, hence $x \in \mathcal{T}'(x)$.

♠ 2.22*(ii)*. Let $\bar{x}' \in \mathcal{T}'(x)$; to arrive to a contradiction, suppose that $\bar{x}' \neq x$. Then, $\mathcal{M}(\bar{x}';x) \leq \mathcal{M}'(\bar{x}';x) < \infty$, hence $\mathcal{M}(\bar{x}';x) < \mathcal{M}'(\bar{x}';x)$, cf. Def. 2.16. We have

$$\varphi(x) \overset{2.14}{=} \mathcal{M}'(\bar{x}';x) > \mathcal{M}(\bar{x}';x) \geq \inf_{w\in\mathbb{R}^n} \mathcal{M}(w;x) = \mathcal{M}(x;x) \overset{2.10.\text{P1}}{=} \varphi(x),$$

where the second equality follows from the fact that $x \in \mathcal{T}(x)$. Thus, we obtained the contradiction $\varphi(x) > \varphi(x)$. □

Corollary 2.23. *Let $\bar{x}$ be a critical point for φ, and let $\mathscr{A} \subset (\mathfrak{M}_\varphi, \succeq)$ be a totally ordered family of proximal models such that $\sup \mathscr{A} = \overline{\mathcal{M}}_0$. Then, there exists $\mathcal{M} \in \mathscr{A}$ such that $\bar{x} \in \operatorname{fix} \mathcal{T}^{\mathcal{M}}$.*

Proof. Suppose that $\bar{x} \in \operatorname{fix} \bar{\mathcal{M}}$ for some proximal model $\bar{\mathcal{M}}$ (not necessarily in $\mathscr{A}$), and let $(\mathcal{M}_k)_{k\in\mathbb{N}} \subseteq \mathscr{A}$ be an increasing sequence pointwise converging to $\overline{\mathcal{M}}_0$. For $k \in \mathbb{N}$, let $\bar{x}^k \in \mathcal{T}^{\mathcal{M}_k}(\bar{x})$. From Lem. 2.18 it follows that $(\bar{x}^k)_{k\in\mathbb{N}}$ is contained in a bounded set B ($\bar{x}$ is critical, and in particular $\bar{x} \in \operatorname{dom} \varphi$). Then, there exists $k \in \mathbb{N}$ such that $\mathcal{M}_k(w;\bar{x}) \geq \bar{\mathcal{M}}(w;\bar{x})$ for all $w \in B$. Thus,

$$\varphi(\bar{x}) \overset{2.10.\text{P1}}{=} \bar{\mathcal{M}}(\bar{x};\bar{x}) = \inf_{w\in\mathbb{R}^n} \bar{\mathcal{M}}(w;\bar{x}) \leq \bar{\mathcal{M}}(\bar{x}^k;\bar{x}) \leq \mathcal{M}_k(\bar{x}^k;\bar{x}) \overset{2.14}{\leq} \varphi(\bar{x}),$$

from which we infer that $\mathcal{M}_k(\bar{x}^k;\bar{x}) = \varphi(\bar{x})$. Lem. 2.14 then ensures that $\bar{x} \in \mathcal{T}^{\mathcal{M}_k}(\bar{x})$. □

Notice that, although closely related, the result cannot be shown by directly invoking Thm. 2.22*(i)*. In fact, the existence of a model in $\mathscr{A}$ (globally) greater than $\mathcal{M}$ cannot be guaranteed, as discussed in Section 2.3.3.

As seen in Example 2.12, $\operatorname{prox}_{\gamma\varphi}$ is a PMM mapping, specifically the one relative to the proximal model $\mathcal{M}_\gamma^{\text{PP}}(w;x) = \varphi(w) + \frac{1}{2\gamma}\|w - x\|^2$. The family

$\mathrm{PP} = (\mathcal{M}_\gamma^{\mathrm{PP}})_{\gamma>0}$ is totally ordered and satisfies $\mathcal{M}_\gamma^{\mathrm{PP}} \nearrow \overline{\mathcal{M}}_0$ as $\gamma \searrow 0$. Corollary 2.23 can thus be invoked, resulting in the following constructive criterion for checking whether a point is critical or not.

Corollary 2.24. *A point $\bar{x}$ is critical for φ iff there exists $\gamma > 0$ such that $\bar{x} \in \operatorname{prox}_{\gamma\varphi}(\bar{x})$.*

The next result completes the assert of Proposition 2.20 by providing a converse implication: local minimizers are critical points, and critical points are *almost* second-order optimal.

Theorem 2.25 (Criticality of local minima)**.** *Let x be a local minimum of φ; then x is critical. In fact, for any totally ordered family $\mathscr{A} \subset (\mathfrak{M}_\varphi, \succeq)$ of proximal models satisfying $\sup \mathscr{A} = \overline{\mathcal{M}}_0$ there exists $\mathcal{M} \in \mathscr{A}$ such that $x \in \operatorname{fix} \mathcal{T}^{\mathcal{M}}$.*

Proof. Since $\mathscr{A}$ is totally ordered, there exists a sequence $(\mathcal{M}_k)_{k\in\mathbb{N}} \subseteq \mathscr{A}$ such that $\mathcal{M}_k \nearrow \overline{\mathcal{M}}_0$ as $k \to \infty$. To ease the notation, let $\mathcal{T}^k := \mathcal{T}^{\mathcal{M}_k}$. Due to property 2.10.P2 and the monotonicity of $(\mathcal{M}_k)_{k\in\mathbb{N}}$, there exists $m_1 > 0$ such that $\mathcal{M}_k(w; x) \geq \varphi(w) + \frac{m_1}{2}\|w - x\|^2$ for all $x, w \in \mathbb{R}^n$. Let x be a local minimum for φ, and for all $k \in \mathbb{N}$ let $\bar{x}^k \in \mathcal{T}^k(x)$. It follows from Lem. 2.18 that $\bar{x}^k \to x$ as $k \to \infty$. In particular, since x is a local minimum, there exists $\bar{k}' \in \mathbb{N}$ such that $\varphi(\bar{x}^k) \geq \varphi(x)$ for all $k \geq \bar{k}'$. Therefore,

$$\varphi(x) \overset{2.14}{\geq} \mathcal{M}_{\bar{k}'}(\bar{x}^{\bar{k}'}; x) \geq \varphi(\bar{x}^{\bar{k}'}) + \tfrac{m_1}{2}\|x - \bar{x}^{\bar{k}'}\|^2 \geq \varphi(x) + \tfrac{m_1}{2}\|x - \bar{x}^{\bar{k}'}\|^2,$$

where the second inequality follows from the fact that $\bar{x}^{\bar{k}'} \in \mathcal{T}^{\bar{k}'}(x)$. We then conclude that $x = \bar{x}^{\bar{k}'} \in \mathcal{T}^{\bar{k}'}(x)$. □

2.5 Generalized proximal majorization-minimization

GPMM schemes

Given a proximal model $\mathcal{M} \in \mathfrak{M}_\varphi$, the regularity properties assessed in Theorem 2.13 ensure that $\mathcal{F} = \mathcal{T}^{\mathcal{M}}$ fits into the general fixed-point framework (FP). Apart from having all accumulation points critical for φ, hence being *compatible* with problem (P) in the sense of Definition 2.6, such a "pure" majorization-minimization scheme has also the advantage of being a descent algorithm on the cost function, in the sense that $(\varphi(x^k))_{k\in\mathbb{N}}$ is monotonically (strictly) decreasing. However, limiting the analysis to these iterative schemes only would rule out many splitting algorithms that do not have a "pure" MM nature, such as the Douglas-Rachford splitting or the sibling ADMM. For the sake of developing a

universal theory, we sacrifice the simplicity of a pure MM scheme by including a possible change of variable G as introduced in Definition 2.6.

Definition 2.26 (Generalized proximal MM schemes)**.** *Given*

P1 *a proximal model* $\mathcal{M} : \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$ *for* φ*, and*

P2 *a* TRANSIENT *mapping, that is, an* L_G*-Lipschitz continuous and* μ_G*-strongly monotone mapping* $G : \mathbb{R}^n \to \mathbb{R}^n$*,*

with $\mathcal{F} \sim (\mathcal{M}, G)$ *we indicate the collection of fixed-point mappings* $\mathcal{F}^\lambda : \mathbb{R}^n \to \mathbb{R}^n$*, indexed over a* RELAXATION *parameter* $\lambda \neq 0$*, defined as*

$$\mathcal{F}^\lambda := \mathrm{id} - \lambda(\mathrm{id} - \mathcal{T}^{\mathcal{M}}) \circ G. \tag{2.6}$$

Fixed-point iterations of $\mathcal{F}^\lambda$ *constitute a* GENERALIZED PROXIMAL MM (GPMM) SCHEME*, or simply* (PURE) PROXIMAL MM (PMM) SCHEME *in case* $G \equiv \mathrm{id}$*.*

It follows from Lemma 1.1 that transients are invertible and their inverse is Lipschitz continuous as well. Although these requirements of Lipschitz continuity and strong monotonicity could actually be dropped, as plain or strict continuity would suffice to our purposes in most cases, for the sake of a simpler exposition we prefer to stick to these assumptions, which, in any case, will be satisfied in all the investigated splitting algorithms. The next result assesses the fundamental *compatibility* of the scheme (2.6) and the optimization problem (P); the role of G as *transition* from the fixed-point variable s of the mappings $\mathcal{F}^\lambda$ to the optimization variable x will then be clear. In particular, we will see that from fixed points of G we can recover stationary points of φ, due to a one-to-one correspondence with fixed points of the pure MM scheme $\mathcal{T}^{\mathcal{M}}$.

Theorem 2.27 (Compatibility of GPMM schemes)**.** *Let* $\mathcal{M} \in \mathfrak{M}_\varphi$ *be a proximal model for* φ*. Then, for every continuous bijection* $G : \mathbb{R}^n \to \mathbb{R}^n$ *and* $\lambda \in \mathbb{R} \setminus \{0\}$*, the mapping* $\mathcal{F}^\lambda : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *as in* (2.6) *is osc, with* $\mathcal{F}^\lambda(s)$ *nonempty and compact for all* $s \in \mathbb{R}^n$*. Moreover, the following inclusions hold:*

$$\arg\min \varphi \;\subseteq\; G(\operatorname{fix} \mathcal{F}^\lambda) \;=\; \operatorname{fix} \mathcal{T}^{\mathcal{M}} \;\subseteq\; \operatorname{zer} \hat{\partial}\varphi. \tag{2.7}$$

In particular, the fixed-point mapping $\mathcal{F}^\lambda$ *is compatible with* φ*.*

Proof. The properties of $\mathcal{F}^\lambda$ follow from the similar ones of $\mathcal{T}^\mathcal{M}$ shown in Thm. 2.13 and the continuity of G. We have

$$\bar{s} \in \mathcal{F}^\lambda(\bar{s}) \quad \Leftrightarrow \quad \bar{s} \in \bar{s} - \lambda\big[G(\bar{s}) - \mathcal{T}^\mathcal{M}\big(G(\bar{s})\big)\big] \quad \Leftrightarrow \quad G(\bar{s}) \in \mathcal{T}^\mathcal{M}\big(G(\bar{s})\big),$$

where the first implication follows from the definition of $\mathcal{F}^\lambda$, and the second one from the invertibility of G and the fact that $\lambda \neq 0$. Thus, $G(\operatorname{fix}\mathcal{F}^\lambda) = \operatorname{fix}\mathcal{T}^\mathcal{M}$. Suppose now that $x \in \arg\min\varphi$ and let $\bar{x} \in \mathcal{T}^\mathcal{M}(x)$. Then,

$$\min\varphi = \mathcal{M}(x,x) \geq \mathcal{M}(\bar{x},x) \overset{2.10.\text{P2}}{\geq} \varphi(\bar{x}) + \tfrac{m_1}{2}\|x-\bar{x}\|^2 \geq \min\varphi + \tfrac{m_1}{2}\|x-\bar{x}\|^2, \overset{2.10.\text{P1}}{}$$

from which it follows that $\bar{x} = x$, hence the inclusion $\arg\min\varphi \subseteq \operatorname{fix} T$. Finally, the inclusion $\operatorname{fix}\mathcal{T}^\mathcal{M} \subseteq \operatorname{zer}\hat{\partial}\varphi$ follows from Prop. 2.20. □

2.6 Representation of proximal algorithms

To finalize the general framework investigated in this chapter, let us formally define how to represent a proximal algorithm in terms of majorizing models and transient functions. To this end, for the sake of an example let us consider the most elementary MM scheme, namely, the proximal point algorithm (PPA). As briefly discussed in Example 2.12, for $\gamma > 0$ we have that $\operatorname{prox}_{\gamma\varphi}$ is the (pure) PMM mapping $\mathcal{T}^\mathcal{M}(x)$ relative to the proximal model $\mathcal{M}(w;x) = \varphi(x) + \frac{1}{2\gamma}\|w-x\|^2$. Thus, the model $\mathcal{M}$ (together with the identity transient mapping $G = \mathrm{id}$) in the generalized MM framework captures PPA with all possible relaxation parameters, yet is bound to a unique stepsize γ, having

$$\mathcal{F}^\lambda(s) = (1-\lambda)s + \lambda\operatorname{prox}_{\gamma\varphi}(s)$$

for all $\lambda \neq 0$. This is somehow an unavoidable consequence of the different nature of γ and λ, the former being intrisic in the fixed-point black box, and the latter simply amounting to an a posteriori averaging.

This is readily solved by identifying PPA with a *family* of models $\mathrm{PP} = (\mathcal{M}_\gamma^{\mathrm{PP}})_{\gamma>0}$ indexed over a parameter γ. Minimal requirements on the parametrization $\gamma \mapsto \mathcal{M}_\gamma^\mathscr{A}$ facilitate operating with such a collection, and the inclusion of possible transients allows to recover all generalized PMM schemes. This leads to the following definition.

Definition 2.28. *A* GENERALIZED PROXIMAL MM (GPMM) ALGORITHM *for problem* (P) *is a collection*

$$\mathscr{A} = (\mathcal{M}_\gamma, G_\gamma)_{\gamma\in(0,\bar{\gamma})}$$

indexed over a STEPSIZE *parameter* γ *ranging between* 0 *and* $\bar{\gamma} \in (0, \infty]$*, where*

P1 $(\mathcal{M}_\gamma, G_\gamma)$ *is a GPMM scheme as in* (2.6) *for all* $\gamma \in (0, \bar{\gamma}]$*,*

P2 $\mathcal{M}_\gamma \succ \mathcal{M}_{\gamma'}$ *whenever* $0 < \gamma < \gamma' < \bar{\gamma}$ *(hence in particular the models form a totally ordered family), and*

P3 $\sup_{\gamma \in (0,\bar{\gamma})} \mathcal{M}_\gamma = \overline{\mathcal{M}}_0$ *(equivalently,* $\mathcal{M}_\gamma \nearrow \overline{\mathcal{M}}_0$ *as* $\gamma \searrow 0$ *pointwise).*

2.6.1 Notational conventions

To give more emphasis to the GPMM family and to the parameters γ and λ, we will adopt the following conventions:

- $\mathscr{A}_\lambda$ will indicate the GPMM algorithm with relaxation parameter λ. The GPMM fixed-point mapping with stepsize γ and relaxation λ will thus be indicated by $\mathcal{F}_\gamma^{\mathscr{A}_\lambda}$.
- We will write $\mathcal{T}_\gamma^{\mathscr{A}}$ in place of $\mathcal{T}^{\mathcal{M}_\gamma^{\mathscr{A}}}$ (this definition is independent of λ); in particular,
$$\mathcal{F}_\gamma^{\mathscr{A}_\lambda} := \mathrm{id} - \lambda(\mathrm{id} - \mathcal{T}_\gamma^{\mathscr{A}}) \circ G_\gamma. \tag{2.8}$$
- We shall adopt the superscript notation $\mathcal{M}_\gamma^{\mathscr{A}}$ to emphasize that the model belongs to the GPMM collection $\mathscr{A}$.
- When $G_\gamma \equiv \mathrm{id}$ for all $\gamma \in (0, \bar{\gamma})$, that is, when representing a *pure* PMM algorithm, the transients G_γ may be omitted from the notation.
- The symbols $m_1(\gamma)$ and $m_2(\gamma)$ will denote the constants m_1 and m_2 in property 2.10.P2, respectively, relative to the proximal model $\mathcal{M}_\gamma$. Similarly, μ_{G_γ} and L_{G_γ} will denote the strong convexity and Lipschitz moduli, respectively, of the transients G_γ as in property 2.26.P2. Whenever clear from context, γ may be removed to ease the notation.

It is also convenient to introduce the RESIDUAL MAPPING $\mathcal{R}_\gamma^{\mathscr{A}} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, given by

$$\mathcal{R}_\gamma^{\mathscr{A}}(s) := \tfrac{1}{\gamma}\big(\mathrm{id} - \mathcal{T}_\gamma^{\mathscr{A}}\big) \circ G_\gamma(s), \tag{2.9}$$

so that the GPMM fixed-point mapping $\mathcal{F}_\gamma^{\mathscr{A}_\lambda}$ can be expressed as

$$\mathcal{F}_\gamma^{\mathscr{A}_\lambda}(s) = s - \lambda\gamma\mathcal{R}_\gamma^{\mathscr{A}}(s).$$

To ease the notation and avoid repetitions, unless differently specified, it will be assumed that $\mathscr{A} = (\mathcal{M}_\gamma, G_\gamma)_{\gamma\in(0,\bar\gamma)}$ is a GPMM algorithm for (the proper and lsc) function $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ and that $\gamma \in (0, \bar\gamma)$ is a stepsize.

Example 2.29 (Proximal point as a GPMM algorithm)**.** The proximal point algorithm fits into the GPMM framework through the following representation:

$$\mathrm{PP} = (\mathcal{M}_\gamma)_{\gamma>0} \quad \text{where} \quad \mathcal{M}_\gamma^{\mathrm{PP}}(w; x) = \varphi(w) + \tfrac{1}{2\gamma}\|w - x\|^2.$$

In particular, $m_1(\gamma) = m_2(\gamma) = 1/\gamma$ and $L_{G_\gamma} = \mu_{G_\gamma} = 1$ are the constants in properties 2.10.P2 and 2.26.P2. □

2.6.2 The criticality threshold

The following result is an adaptation of Theorem 2.22 to GPMM algorithms, owing to property 2.28.P2; simply put, the higher the value of γ, hence the lower the model, the stronger the property of being a fixed point of $\mathcal{T}_\gamma^{\mathscr{A}}$.

Proposition 2.30. *Whenever* $0 < \gamma < \gamma' < \bar\gamma$ *it holds that* $\operatorname{fix}\mathcal{T}_\gamma^{\mathscr{A}} \supseteq \operatorname{fix}\mathcal{T}_{\gamma'}^{\mathscr{A}}$.

Recall from Corollary 2.23 that in every totally ordered family of models with $\overline{\mathcal{M}}_0$ as supremum, for all critical points x there exists a *frontier* model above which x is fixed, and below which x is not. In light of properties 2.28.P2 and 2.28.P3, we can thus represent the *criticality threshold* in terms of the parameter γ, as in the following definition.

Definition 2.31 (Criticality threshold)**.** *The* CRITICALITY THRESHOLD *of a GPMM algorithm* $\mathscr{A} = (\mathcal{M}_\gamma, G_\gamma)_{\gamma\in(0,\bar\gamma)}$ *is the function* $\Gamma^{\mathscr{A}} : \mathbb{R}^n \to [0, \bar\gamma]$ *defined as*

$$\Gamma^{\mathscr{A}}(x) := \sup\left\{\gamma \in (0,\bar\gamma) \mid x \in \operatorname{fix}\mathcal{T}_\gamma^{\mathscr{A}}\right\},$$

with the convention $\sup\emptyset = 0$*. In other words,* $\Gamma^{\mathscr{A}}(x)$ *is the unique index in* $[0, \bar\gamma]$ *such that*

P1 $\mathcal{T}_\gamma^{\mathscr{A}}(x) = \{x\}$ *for all* $\gamma < \Gamma^{\mathscr{A}}(x)$*, and*

P2 $x \notin \mathcal{T}_\gamma^{\mathscr{A}}(x)$ *for all* $\gamma > \Gamma^{\mathscr{A}}(x)$*.*

In particular, a point x *is critical iff* $\Gamma^{\mathscr{A}}(x) > 0$*.*

It is important to observe that the characterization of criticality of a point x highlighted in Definition 2.31, namely the fact that $\Gamma^{\mathscr{A}}(x) > 0$, does not depend

on the GPMM algorithm $\mathscr{A}$. This confirms the criterion proposed in Corollary 2.24 where $\mathscr{A} = \mathrm{PP}$, the proximal point algorithm, was considered. In other words, *if x is a critical point, then* ***for every GPMM algorithm*** $\mathscr{A}$ *it is a fixed point of the PMM mapping $\mathcal{T}_\gamma^{\mathscr{A}}$ for all stepsizes small enough.* In the next chapters we will see how the criticality threshold plays a fundamental role in establishing regularity properties of *envelope functions* at critical points.

Example 2.32. Relative to the proper, lsc, and lower bounded function $\varphi : \mathbb{R} \to \overline{\mathbb{R}}$ defined by

$$\varphi(x) = \tfrac{1}{2}x^2 + \delta_{\mathbb{Z}}(x),$$

let us consider the proximal point algorithm as in Example 2.29. Notice that every point $x \in \operatorname{dom}\varphi = \mathbb{Z}$ is a local minimum, hence it must be critical as ensured by Theorem 2.25. In fact, it can be easily verified that

$$\operatorname{prox}_{\gamma\varphi}(x) = \Pi_{\mathbb{Z}}\left(\tfrac{x}{1+\gamma}\right) \quad \text{for all } x \in \mathbb{R},$$

and that the inclusion $\Pi_{\mathbb{Z}}\left(\frac{x}{1+\gamma}\right) \ni x$ holds iff $x \in \mathbb{Z}$ and $\gamma < 1/2|n|$ (with '$\frac{1}{0} = \infty$'). Thus,

$$\Gamma^{\mathrm{PP}}(x) = \begin{cases} \infty & \text{if } x = 0, \\ \frac{1}{2|n|} & \text{if } x \in \mathbb{Z} \setminus \{0\}, \\ 0 & \text{if } x \notin \mathbb{Z} \end{cases}$$

is the criticality threshold of the proximal point algorithm for φ. □

The fact that the global minimum $x = 0$ has the highest threshold is not a coincidence. This is a straightforward consequence of properties 2.10.P1 and 2.10.P2 of proximal models.

Proposition 2.33 (*Total* criticality of global minima)**.** *Let $x_\star \in \arg\min\varphi$. Then, $\mathcal{T}_\gamma^{\mathscr{A}}(x_\star) = \{x_\star\}$ for every $\gamma \in (0, \bar{\gamma})$. In particular, $\Gamma^{\mathscr{A}}(x_\star) = \bar{\gamma}$.*

Notice that strong local minimality is not enough to ensure total criticality, as Example 2.32 clearly demonstrates.

Chapter 3

Proximal envelopes

MM Lyapunov functions

3.1 Majorization-minimization value functions

The smoothness properties of the Moreau envelope of a proper, convex, and lsc function φ (cf. Thm. 1.14*(iii)*) make it possible to address the constrained and nonsmooth minimization of φ by means of gradient descent on the smooth envelope function φ^γ with stepsize $0 < \tau < 2/L_{\varphi^\gamma} = 2\gamma$. As first noticed by Rockafellar [103], this simply amounts to (relaxed) fixed-point iterations of the proximal point operator, namely

$$x^+ = (1-\lambda)x + \lambda \operatorname{prox}_{\gamma\varphi}(x), \tag{3.1}$$

where $\lambda = \tau/\gamma \in (0,2)$ is a relaxation parameter. The scheme, known as proximal point algorithm and first introduced by Martinet [81], is well covered by the broad theory of monotone operators, where convergence properties can be easily derived with simple tools of Fejérian monotonicity, see *e.g.*, [10, Thm.s 23.41 and 27.1]. Nevertheless, not only does the interpretation as gradient method provide a beautiful theoretical link, but it also enables the employment of acceleration techniques exclusively stemming from smooth unconstrained optimization, such as Nesterov's extrapolation [54] or quasi-Newton schemes [30], see also [19] for extensions to the dual formulation.

Even if φ is nonconvex, although not anymore differentiable the Moreau envelope still exhibits more regularity over the original function φ, being it real valued (as opposed to extended-real valued) and, in fact, strictly continuous (cf. Prop. 1.12*(ii)*). The quadratic penalty appearing in the subproblem that defines the proximal mapping, cf. Def. 1.11, has a regularization effect on function φ when

considering the value function φ^γ, namely, the Moreau envelope.

The appeal of the Moreau envelope goes beyond its regularity. Proximal point iterations $x^+ \in \operatorname{prox}_{\gamma\varphi}(x)$ are easily seen to generate a sequence such that

$$\varphi(x^+) \leq \varphi(x) - \tfrac{1}{2\gamma}\|x - x^+\|^2,$$

and one can expand the arguments in the proof of Theorem 2.4 to infer that every accumulation point of the sequence is stationary; in terms of the fixed-point framework of Section 2.1, the cost function φ itself, although extended-real valued, serves as Lyapunov function for proximal point iterations. This, however, is no longer the case if one considers relaxed iterations as in (3.1) with $\lambda \neq 1$; as a matter of fact, the nonconvexity of $\operatorname{dom}\varphi$ may result in having $\varphi(x^+) = \infty$. This limitation is readily solved if one considers the Moreau envelope instead, for it can be easily verified that relaxed proximal point iterations (3.1) satisfy

$$\varphi^\gamma(x^+) \leq \varphi^\gamma(x) - \tfrac{2-\lambda}{2\lambda\gamma}\|x - x^+\|^2; \tag{3.2}$$

see Example 3.21 for the details. Real valuedness (and lower boundedness) of φ^γ make the Moreau envelope a suitable Lyapunov function for the fixed-point iterations (3.1) for any $\lambda \in (0, 2)$, and one can again infer (subsequential) convergence of the proximal point algorithm for any relaxation $\lambda \in (0, 2)$, as opposed to $\lambda = 1$ only.

These observations suggest to extend the definition of *envelope function* to the more general, yet closely related, proximal majorizing models investigated in the previous chapter. While all the argumentations can quite easily be extended for all *pure* proximal MM algorithms, the presence of transient functions G_γ makes the analysis of *generalized* proximal MM algorithms more complicated. Once again, it is important to clearly distinguish the variable s of the fixed-point mapping $\mathcal{F}_\gamma^{\mathscr{A}_\lambda}$ from the optimization variable x. This distinction will require the introduction of *two* envelope functions, a *model envelope* ($\mathcal{M}$-envelope) that operates on the optimization variable x, and an *algorithmic (fixed-point) envelope* ($\mathcal{F}$-envelope) that operates on the fixed-point variable s. In case of *pure* MM schemes, the two will coincide; more generally, they are related by the transient mapping G_γ.

Definition 3.1 ($\mathcal{M}$- and $\mathcal{F}$-envelope functions)**.** *Let $\mathscr{A} = (\mathcal{M}_\gamma, G_\gamma)_{\gamma\in(0,\bar\gamma)}$ be a GPMM algorithm for φ. The* $\mathcal{M}$-ENVELOPE (MODEL ENVELOPE) *of φ with parameter $\gamma \in (0, \bar\gamma)$ is the function $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}} : \mathbb{R}^n \to \mathbb{R}$ given by*

$$\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x) := \min_{w\in\mathbb{R}^n} \mathcal{M}_\gamma^{\mathscr{A}}(w; x)$$

(independent of the transient G_γ), while its $\mathcal{F}$-ENVELOPE *(*FIXED-POINT *or* ALGORITHMIC ENVELOPE*) is* $\varphi_\gamma^{\mathscr{A}} : \mathbb{R}^n \to \mathbb{R}$ *defined by*

$$\varphi_\gamma^{\mathscr{A}}(s) \coloneqq \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}\big(G_\gamma(s)\big) = \min_{w\in\mathbb{R}^n} \mathcal{M}_\gamma^{\mathscr{A}}\big(w; G_\gamma(s)\big).$$

Notice that for all $x \in \mathbb{R}^n$ and $\bar{x} \in \mathcal{T}_\gamma^{\mathscr{A}}(x)$ one has

$$\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x) = \mathcal{M}_\gamma^{\mathscr{A}}(\bar{x}; x),$$

and similarly, for all $s \in \mathbb{R}^n$

$$\varphi_\gamma^{\mathscr{A}}(s) = \mathcal{M}_\gamma^{\mathscr{A}}(\bar{x}; x), \quad \text{where } x = G_\gamma(s) \text{ and } \bar{x} \in \mathcal{T}_\gamma^{\mathscr{A}}(x).$$

In particular, evaluating the $\mathcal{M}$- and $\mathcal{F}$-envelopes requires exactly the same operations needed for one iteration of the GPMM algorithm with stepsize γ. Therefore, once a GPMM step has been performed, *the evaluation of the envelopes comes at the sole cost of evaluating the proximal model at known points.*

3.2 Properties

3.2.1 Inequalities

The following result extends some known inequalities relating a function to its Moreau envelope.

Theorem 3.2 ($\mathcal{M}$-envelope: sandwich property)**.** *For all $x \in \mathbb{R}^n$ the following hold for the $\mathcal{M}$-envelope:*

(i) $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x) \le \varphi(x)$*, with equality holding iff* $x \in \mathcal{T}_\gamma^{\mathscr{A}}(x)$.

(ii) $-\frac{m_2}{2}\|x-\bar{x}\|^2 \le \varphi(\bar{x}) - \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x) \le -\frac{m_1}{2}\|x-\bar{x}\|^2$ *for all* $\bar{x} \in \mathcal{T}_\gamma^{\mathscr{A}}(x)$.

(iii) $\inf \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}} = \inf \varphi$ *and* $\operatorname{arg\,min} \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}} = \operatorname{arg\,min} \varphi$.

Proof.

♠ 3.2*(i)*. We have $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x) = \inf_{w\in\mathbb{R}^n} \mathcal{M}_\gamma^{\mathscr{A}}(w; x) \le \mathcal{M}_\gamma^{\mathscr{A}}(x; x) = \varphi(x)$, where the last equality is due to 2.10.P1. From this we also easily infer the claimed necessary and sufficient condition for equality.

♠ 3.2*(ii)*. By definition of $\mathcal{T}_\gamma^{\mathscr{A}}$, for all points $\bar{x} \in \mathcal{T}_\gamma^{\mathscr{A}}(x)$ we have $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x) = \mathcal{M}_\gamma^{\mathscr{A}}(\bar{x}; x)$ and the claimed inequalities follow from the bounds in 2.10.P2.

♠ 3.2*(iii)*. Consider a sequence $(x^k)_{k\in\mathbb{N}}$ such that $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x^k) \to \inf \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}$, and for all $k \in \mathbb{N}$ let $\bar{x}^k \in \mathcal{T}_\gamma^{\mathscr{A}}(x^k)$. We have

$$\inf \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}} \overset{3.2(i)}{\leq} \inf \varphi \leq \varphi(\bar{x}^k) \overset{3.2(ii)}{\leq} \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x^k) \to \inf \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}} \quad \text{as } k \to \infty,$$

proving that $\inf \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}} = \inf \varphi$. Combined with the inequality $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}} \leq \varphi$ we infer that $\operatorname{argmin} \varphi \subseteq \operatorname{argmin} \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}$.

Suppose now that $x \in \operatorname{argmin} \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}$ and let $\bar{x} \in \mathcal{T}_\gamma^{\mathscr{A}}(x)$; from 3.2*(ii)* we have that

$$\inf \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}} \overset{3.2(ii)}{\leq} \varphi(\bar{x}) \leq \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x) - \tfrac{m_1}{2}\|x - \bar{x}\|^2 = \inf \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}} - \tfrac{m_1}{2}\|x - \bar{x}\|^2,$$

hence $x = \bar{x} \in \operatorname{argmin} \varphi$. □

A trivial change of variable yields the following equivalent result involving the $\mathcal{F}$-envelope.

Corollary 3.3 ($\mathcal{F}$-envelope: sandwich property)**.** *For all $s \in \mathbb{R}^n$, denoting $x := G_\gamma(s)$,*

(i) $\varphi_\gamma^{\mathscr{A}}(s) \leq \varphi(x)$*, with equality holding iff* $x \in \mathcal{T}_\gamma^{\mathscr{A}}(x)$.

(ii) $-\frac{m_2}{2}\|x - \bar{x}\|^2 \leq \varphi(\bar{x}) - \varphi_\gamma^{\mathscr{A}}(s) \leq -\frac{m_1}{2}\|x - \bar{x}\|^2$ *for all* $\bar{x} \in \mathcal{T}_\gamma^{\mathscr{A}}(x)$.

(iii) $\inf \varphi_\gamma^{\mathscr{A}} = \inf \varphi$ *and* $G_\gamma(\operatorname{argmin} \varphi_\gamma^{\mathscr{A}}) = \operatorname{argmin} \varphi$.

Proposition 3.4 (Connection with the Moreau envelope)**.** *We have*

$$\varphi^{1/m_1} \leq \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}} = \varphi_\gamma^{\mathscr{A}} \circ G_\gamma^{-1} \leq \varphi^{1/m_2}.$$

Proof. Let $x \in \mathbb{R}^n$ and $\bar{x} \in \mathcal{T}_\gamma^{\mathscr{A}}(x)$ be fixed. We have

$$\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x) = \mathcal{M}_\gamma^{\mathscr{A}}(\bar{x}; x) \overset{2.10.P2}{\geq} \varphi(\bar{x}) + \tfrac{m_1}{2}\|x - \bar{x}\|^2 \geq \varphi^{1/m_1}(x).$$

Moreover, for all $p \in \operatorname{prox}_{\varphi/m_2}(x)$ we have

$$\varphi^{1/m_2}(x) = \varphi(p) + \tfrac{m_2}{2}\|x - p\|^2 \overset{2.10.P2}{\geq} \mathcal{M}_\gamma^{\mathscr{A}}(p; x) \geq \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x).$$

□

The following result shows that local minimizers of φ are stable with respect to fixed-point iterations of $\mathcal{T}_\gamma^{\mathscr{A}}$.

Proposition 3.5 (Stability of minimizers)**.** *Let $x_\star$ be a local minimum for φ. Then, for all $\gamma < \Gamma^{\mathscr{A}}(x_\star)$ there exists $\varepsilon = \varepsilon(\gamma) > 0$ such that*

$$\varphi(\bar{x}) \geq \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(\bar{x}) \geq \varphi(x_\star) \quad \text{for all } x \in \mathrm{B}(x_\star;\varepsilon) \text{ and } \bar{x} \in \mathcal{T}_\gamma^{\mathscr{A}}(x).$$

Proof. The inequality $\varphi(\bar{x}) \geq \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(\bar{x})$ is due to Thm. 3.2*(i)* and holds globally. As to the other inequality, since $\mathcal{T}_\gamma^{\mathscr{A}}(x_\star) = \{x_\star\}$, cf. property 2.31.P1, we may invoke Thm. 3.6 to infer that $x_\star$ is a local minimum for $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}$. The claim now follows from the outer semicontinuity of $\mathcal{T}_\gamma^{\mathscr{A}}$, cf. Thm. 2.13. In fact, for an arbitrary sequence $(x^k, \bar{x}^k)_{k\in\mathbb{N}} \subset \operatorname{gph} \mathcal{T}_\gamma^{\mathscr{A}}$ with $x^k \to x_\star$ as $k \to \infty$, necessarily $\bar{x}^k \to x_\star$, hence eventually $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(\bar{x}^k) \geq \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x_\star) = \varphi(x_\star)$, where the equality follows from Thm. 3.2*(i)*. □

3.2.2 Equivalence

Theorem 3.6 (Equivalence of local minimality)**.** *For any $\bar{s} \in \operatorname{fix} \mathcal{F}_\gamma^{\mathscr{A}_\lambda}$, denoting $\bar{x} := G_\gamma(\bar{s})$ (hence $\bar{x} \in \operatorname{fix} \mathcal{T}_\gamma^{\mathscr{A}}$), the following statements are equivalent:*

(a) $\bar{x}$ is a (strong) local minimum for $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}$.

(b) $\bar{s}$ is a (strong) local minimum for $\varphi_\gamma^{\mathscr{A}}$.

When any of the property above holds, then $\bar{x}$ is a (strong) local minimum for φ; the converse implication holds provided that $\mathcal{T}_\gamma^{\mathscr{A}}(\bar{x}) = \{\bar{x}\}$ (or, equivalently, that $\mathcal{F}_\gamma^{\mathscr{A}_\lambda}(\bar{s}) = \{\bar{s}\}$).

Proof. The equivalence of (strong) local minimality between the envelopes is a direct consequence of the Lipschitz continuity and Lipschitz invertibility of G_γ. That (strong) local minimality for $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}$ implies that for φ follows from the fact that $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}$ "supports" φ at $\bar{x}$, namely that $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}} \leq \varphi$ and $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(\bar{x}) = \varphi(\bar{x})$, cf. Thm. 3.2*(i)*.

It remains to show the converse. Suppose that $\mathcal{T}_\gamma^{\mathscr{A}}(\bar{x}) = \{\bar{x}\}$ and that there exists $\mu \geq 0$ such that $\varphi(x) \geq \varphi(\bar{x}) + \frac{\mu}{2}\|x - \bar{x}\|^2$ for all x sufficiently close to $\bar{x}$. Let $\delta := \frac{1}{2}\min\{\mu, m_1\} \geq 0$, and note that $\delta = 0$ iff $\mu = 0$. Thus, contrary to the claim suppose that for all $k \in \mathbb{N}$ there exists $x^k \in \mathrm{B}(\bar{x}; 1/k)$ such that

$\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x^k) < \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(\bar{x}) + \frac{\delta}{2}\|x^k - \bar{x}\|^2$. Let $\bar{x}^k \in \mathcal{T}_\gamma^{\mathscr{A}}(x^k)$; since $\mathcal{T}_\gamma^{\mathscr{A}}$ is osc and $\mathcal{T}_\gamma^{\mathscr{A}}(\bar{x}) = \{\bar{x}\}$, necessarily $\bar{x}_k \to \bar{x}$ as $k \to \infty$. We have

$$\begin{aligned}\varphi(\bar{x}^k) &\overset{3.2(i)}{\leq} \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x^k) - \tfrac{m_1}{2}\|x^k - \bar{x}^k\|^2 \\ &< \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(\bar{x}) + \tfrac{\delta}{2}\|x^k - \bar{x}\|^2 - \tfrac{m_1}{2}\|x^k - \bar{x}^k\|^2 \\ &= \varphi(\bar{x}) + \tfrac{\delta}{2}\|x^k - \bar{x}\|^2 - \tfrac{m_1}{2}\|x^k - \bar{x}^k\|^2.\end{aligned}$$

By using the inequality $\frac{1}{2}\|a - c\|^2 \leq \|a - b\|^2 + \|b - c\|^2$ holding for all vectors $a, b, c \in \mathbb{R}^n$, we have

$$\begin{aligned}\varphi(\bar{x}^k) &< \varphi(\bar{x}) + \delta\|\bar{x}^k - \bar{x}\|^2 + \big(\delta - \tfrac{m_1}{2}\big)\|x^k - \bar{x}^k\|^2 \\ &\leq \varphi(\bar{x}) + \tfrac{\mu}{2}\|\bar{x}^k - \bar{x}\|^2,\end{aligned}$$

where the last inequality follows from the definition of δ. Thus, we obtain $\varphi(\bar{x}^k) < \varphi(\bar{x}) + \frac{\mu}{2}\|\bar{x}^k - \bar{x}\|^2$ for all $k \in \mathbb{N}$, hence the contradiction since $\bar{x}^k$ is arbitrarily close to $\bar{x}^k$. □

The necessity of single valuedness of $\mathcal{T}_\gamma^{\mathscr{A}}(\bar{x})$ for inferring the converse implication can be demonstrated with a simple example. Consider the proximal point algorithm as in Example 2.29 applied to the minimization of $\varphi(x) := \frac{1}{2}x^2 + \delta_{\{0,1\}}(x)$ on $\mathbb{R}$. Clearly, $\bar{x} = 1$ is a strong local minimum, and it can be easily verified that

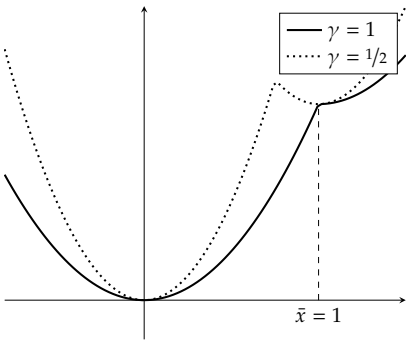


Figure 3.1: *Where the PMM mapping $\mathcal{T}_\gamma^{\mathscr{A}}$ is not single valued, local minimality may fail to be preserved. The pathology occurs at the criticality threshold; for smaller stepsizes, (strong) local minimality is instead preserved.*

$$\operatorname{prox}_{\gamma\varphi}(1) = \begin{cases} 1 & \text{if } \gamma < 1, \\ \{0, 1\} & \text{if } \gamma = 1, \\ 0 & \text{otherwise.} \end{cases}$$

For every $\gamma < 1 = \Gamma^{\mathrm{PP}}(\bar{x})$, the single valuedness of $\operatorname{prox}_{\gamma\varphi}(\bar{x})$ results in the strong local minimality of $\bar{x}$ for the envelope function φ^γ, inherited by that

on the original function φ. On the contrary, for $\gamma = 1$ the point $\bar{x}$ is not even stationary for φ^γ, in the sense that $0 \notin \hat{\partial}\varphi^\gamma(\bar{x})$. Figure 3.1 provides a graphical representation of the pathology occurring at the threshold value $\gamma = 1$.

Theorem 3.7 (Equivalence of level boundedness)**.** *The following are equivalent:*

(a) φ *is level bounded;*

(b) $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}$ *is level bounded;*

(c) $\varphi_\gamma^{\mathscr{A}}$ *is level bounded.*

Proof.

♠ 3.7*(b)* $\Rightarrow$ 3.7*(a)*. From Thm. 3.2*(i)* we know that $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}} \leq \varphi$, hence if $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}$ is level bounded then so is φ.

♠ 3.7*(b)* $\Leftarrow$ 3.7*(a)*. Suppose that $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}$ is not level bounded. Then, there exists $\alpha \in \mathbb{R}$ and $(x^k)_{k\in\mathbb{N}} \subseteq \operatorname{lev}_{\leq\alpha} \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}$ such that $\|x^k\| \to \infty$ as $k \to \infty$. For all $k \in \mathbb{N}$, let $\bar{x}^k \in \mathcal{T}_\gamma^{\mathscr{A}}(x^k)$; then, it follows from Thm. 3.2*(ii)* that

$$\varphi(\bar{x}^k) \leq \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x^k) - \tfrac{m_1}{2}\|x^k - \bar{x}^k\|^2 \leq \alpha - \tfrac{m_1}{2}\|x^k - \bar{x}^k\|^2 \quad \text{for all } k \in \mathbb{N}.$$

If $(\bar{x}^k)_{k\in\mathbb{N}}$ is bounded, then φ is lower unbounded; otherwise, $\|\bar{x}^k\| \to \infty$ as $k \to \infty$. Either way, φ cannot be level bounded.

♠ 3.7*(b)* $\Leftrightarrow$ 3.7*(c)*. This follows from the continuity of G_γ and G_γ^{-1}, as any continuous function maps bounded sets to bounded sets. □

3.2.3 Regularity

Proposition 3.8 (Continuity)**.** *Both* $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}$ *and* $\varphi_\gamma^{\mathscr{A}}$ *are lsc and with full domain. If, additionally, the model* $\mathcal{M}_\gamma^{\mathscr{A}}$ *is continuous (cf. Def. 2.11), then* $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}$ *and* $\varphi_\gamma^{\mathscr{A}}$ *are continuous.*

Proof. The claim on the $\mathcal{M}$-envelope follows from [106, Thm. 1.17(c)] (which applies, as shown in the proof of Thm. 2.13). In particular, the full domain property is a consequence of the fact that $\operatorname{argmin}_w \mathcal{M}_\gamma^{\mathscr{A}}(w; x) \neq \emptyset$ for all $x \in \mathbb{R}^n$. Since $\varphi_\gamma^{\mathscr{A}} = \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}} \circ G_\gamma$ and G_γ is continuous, all the claims are equally valid for the $\mathcal{F}$-envelope. □

Proposition 3.9 (Quadratic upper bound)**.** *For all $x, x_\star \in \mathbb{R}^n$ it holds that*

$$\varphi_\gamma^{\mathscr{A}}(G_\gamma^{-1}(x)) \overset{(def)}{=} \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x) \leq \varphi(x_\star) + \tfrac{m_2}{2}\|x - x_\star\|^2.$$

In particular, if $x_\star \in \mathcal{T}_\gamma^{\mathscr{A}}(x_\star)$ is a fixed point, then

$$\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x) \leq \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x_\star) + \tfrac{m_2}{2}\|x - x_\star\|^2 \quad \forall x \in \mathbb{R}^n.$$

Proof. This is a direct consequence of the quadratic upper bound property 2.10.P2 of $\mathcal{M}_\gamma^{\mathscr{A}}$, namely

$$\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x) \overset{(def)}{=} \min_{w \in \mathbb{R}^n} \mathcal{M}_\gamma^{\mathscr{A}}(w; x) \leq \mathcal{M}_\gamma^{\mathscr{A}}(x_\star; x) \overset{2.10.\text{P2}}{\leq} \varphi(x_\star) + \tfrac{m_2}{2}\|x - x_\star\|^2,$$

and the fact that $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x_\star) = \varphi(x_\star)$ whenever $x_\star$ is a fixed point of $\mathcal{T}_\gamma^{\mathscr{A}}$, cf. Thm. 3.2*(i)*. □

As a consequence of the quadratic upper bound, even though the envelopes may fail to be continuous the discontinuity "jumps" are bounded by quantities that depend on the residual.

Proposition 3.10 (Continuity at fixed points)**.** *For every $x \in \mathbb{R}^n$ the $\mathcal{M}$-envelope satisfies*

$$0 \leq \limsup_{x' \to x} \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x') - \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x) \leq \tfrac{m_2 - m_1}{2} \operatorname{dist}(x, \mathcal{T}_\gamma^{\mathscr{A}}(x))^2.$$

In particular, envelope functions are continuous at fixed points.

Proof. The first inequality is due to lsc, cf. Prop. 3.8. Let $x, x' \in \mathbb{R}^n$ be fixed, and let $\bar{x} \in \mathcal{T}_\gamma^{\mathscr{A}}(x)$. Then,

$$\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x') \leq \varphi(\bar{x}) + \tfrac{m_2}{2}\|x' - \bar{x}\|^2 \leq \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x) - \tfrac{m_1}{2}\|x - \bar{x}\|^2 + \tfrac{m_2}{2}\|x' - \bar{x}\|^2,$$

where the first inequality follows from Prop. 3.9 and the second one from the sandwiching property, cf. Thm. 3.2*(ii)*. By letting $x' \to x$, from the arbitrarity of $\bar{x} \in \mathcal{T}_\gamma^{\mathscr{A}}(x)$ the sought upper bound follows. □

3.2.4 The KL property

As disclosed at the beginning of the chapter, proximal envelopes will be shown to be suitable Lyapunov functions for fixed-point iterations of GPMM algorithms.

Once this is proven, subsequential convergence of the algorithms can be deduced from the more general analysis detailed in Theorem 2.4. Nevertheless, there are favorable cases in which global convergence to a unique limit point can be established, even with asymptotic linear rates. The key ingredient is the so-called *Kurdyka-Łojasiewicz property*, which we define next.

Definition 3.11 (KL property)**.** *A proper and lsc function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ has the* KURDYKA-ŁOJASIEWICZ (KL) PROPERTY *at $x_\star \in \operatorname{dom} \partial h$ if there exist a concave* DESINGULARIZING FUNCTION *(or* KL FUNCTION*) $\psi : [0, \eta] \to [0, +\infty)$ for some $\eta > 0$ and a neighborhood $U_{x_\star}$ of $x_\star$, such that*

P1 *$\psi(0) = 0$;*

P2 *ψ is C^1 with $\psi' > 0$ on $(0, \eta)$;*

P3 *for all $x \in U_{x_\star}$ s.t. $h(x_\star) < h(x) < h(x_\star) + \eta$ it holds that*

$$\psi'\big(h(x) - h(x_\star)\big) \operatorname{dist}\big(0, \partial h(x)\big) \geq 1. \tag{3.3}$$

Lemma 3.12 (Uniformized KL function [25, Lem. 6])**.** *Suppose that a function h is constant on a compact nonempty and connected set ω, with value, say, $h_\star$. If h has the KL property at all points $x_\star \in \omega$, then there exist $\eta, \varepsilon > 0$ and a function $\psi : [0, \eta] \to [0, \infty)$ such that*

P1 *$\psi(0) = 0$;*

P2 *ψ is C^1 with $\psi' > 0$ on $(0, \eta)$;*

P3 *for all points x such that $\operatorname{dist}(x, \omega) < \varepsilon$ and $h_\star < h(x) < h_\star + \eta$ it holds that*

$$\psi'\big(h(x) - h_\star\big) \operatorname{dist}\big(0, \partial h(x)\big) \geq 1. \tag{3.4}$$

The KL property is a mild requirement enjoyed by semialgebraic functions and by subanalytic functions which are continuous on their domain [24, 23] see also [74, 75, 63]. We remind that a set $A \subseteq \mathbb{R}^n$ is SEMIALGEBRAIC if it can be expressed as

$$A = \bigcup_{i=1}^{p} \bigcap_{j=1}^{q} \{x \in \mathbb{R}^m \mid P_{ij}(x) = 0,\ Q_{ij}(x) < 0\}$$

for some polynomial functions $P_{ij}, Q_{ij} : \mathbb{R}^n \to \mathbb{R}$, and that a function $h : \mathbb{R}^n \to \overline{\mathbb{R}}^m$ (in fact, even set valued $h : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$) is SEMIALGEBRAIC if $\operatorname{gph} h$ is a semialgebraic subset of $\mathbb{R}^{n+m}$. Since semialgebraic functions are closed under parametric minimization, semialgebraic models yield semialgebraic evelopes.

More precisely, in all such cases the desingularizing function can be taken of the form $\psi(s) = \varrho s^{\vartheta}$ for some $\varrho > 0$ and $\vartheta \in (0, 1]$, in which case it is usually referred to as a ŁOJASIEWICZ FUNCTION. The following result states this formally.

Theorem 3.13 (Łojasiewicz property for semialgebraic models)**.** *Suppose that the proximal model $\mathcal{M}_\gamma^{\mathscr{A}}$ is semialgebraic. Then, $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}$ is semialgebraic, and in particular has the Łojasiewicz property. The same holds for $\varphi_\gamma^{\mathscr{A}}$ if, additionally, G_γ is semialgebraic.*

Proof. As detailed in [6, §2], parametric minimization of a semialgebraic function is still semialgebraic, hence $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}$ is semialgebraic and thus has the Łojasiewicz property [24]. The claim on $\varphi_\gamma^{\mathscr{A}}$ follows from the fact that semialgebraicity is preserved under composition [23, Prop. 2.2.6(i)]. □

Inequality (3.3) (or (3.4)) indicates that, up to a reparametrization with ψ, the function h grows linearly around the point $x_\star$ (or the region ω). Intuitively, a p-th power growth of h can be *desingularized* with a p-th root, as illustrated in the next example.

Lemma 3.14 (p-th power growth and Łojasiewicz property)**.** *Let $x_\star$ be a stationary point for h, and let $h_\star \coloneqq h(x_\star)$. Suppose that there exists a (convex) neighborhood $\mathcal{U}_\star$ of $x_\star$ in which the following properties hold:*

P1 (local convexity) *function $\tilde{h} \coloneqq h + \delta_{\mathcal{U}_\star}$ is convex;*

P2 (p-th power growth) *$h \geq h_\star + \frac{c}{2}\operatorname{dist}(\,\cdot\,, \omega)^p$ for some $c > 0$ and $p \geq 1$, where $\omega \coloneqq h^{-1}(\{h_\star\})$.*

Then, h has the Łojasiewicz property at $x_\star$ with exponent $\vartheta = 1/p$.

Proof. Because of convexity of $\tilde{h}$ and the fact that $\tilde{h} = h \geq h_\star$ on $\mathcal{U}_\star$, the set

$$\operatorname{lev}_{\leq h_\star} \tilde{h} = h^{-1}(\{h_\star\}) \cap \mathcal{U}_\star$$

is convex. In particular, up to possibly restricting $\mathcal{U}_\star$, we may assume that $\mathcal{U}_\star$ is closed and that $\emptyset \neq \Pi_\omega(x) \subseteq \mathcal{U}_\star$ for any $x \in \mathcal{U}_\star$. Let $\eta \coloneqq \sup_{\mathcal{U}_\star} h - h_\star$ and let $\psi(s) \coloneqq \varrho s^{\vartheta}$ for some $\varrho > 0$ and $\vartheta \in (0, 1]$ to be determined. Fix any $x \in \mathcal{U}_\star$ such that $h_\star < h(x) < h_\star + \eta$ (equivalently, $x \in \mathcal{U}_\star$ such that $h(x) \neq h_\star$), and let $\bar{x} \in \Pi_\omega(x)$. Then, $\bar{x} \in \mathcal{U}_\star$ and $h(\bar{x}) = h_\star$. Due to the locality of the definition of the subdifferential ∂, notice that $\partial h(x) \subseteq \partial \tilde{h}(x)$. Then, for any $v_x \in \partial h(x)$ one has

$$\tfrac{c}{2}\|x - \bar{x}\|^p = \tfrac{c}{2}\operatorname{dist}(x, \omega)^p \leq h(x) - h_\star = \tilde{h}(x) - \tilde{h}(\bar{x}) \leq \langle v_x, x - \bar{x} \rangle$$

$$\leq \|v_x\|\|x - \bar{x}\|,$$

where the second inequality holds without an $o(\|x - x_\star\|)$ term due to convexity of $\tilde{h}$. From the arbitrarity of $v_x \in \partial h(x)$, it follows that

$$\tfrac{c}{2}\|x - \bar{x}\|^p \leq h(x) - h_\star \leq \operatorname{dist}(0, \partial h(x))\|x - \bar{x}\|.$$

In particular,

$$\begin{aligned}\psi'\big(h(x) - h_\star\big)\operatorname{dist}\big(0, \partial h(x)\big) &= \varrho\vartheta(h(x) - h_\star)^{\vartheta-1}\operatorname{dist}\big(0, \partial h(x)\big)\\ &\geq \varrho\vartheta\frac{(h(x) - h_\star)^\vartheta}{\|x - \bar{x}\|}\\ &\geq \tfrac{sc^\vartheta}{2^\vartheta p}\|x - \bar{x}\|^{p\vartheta-1}\varrho.\end{aligned}$$

By selecting $\vartheta = 1/p \in (0, 1]$ and $\varrho = \frac{p}{(c/2)^{1/p}} > 0$ the sought KL inequality (3.3) is obtained. □

The requirement of convexity cannot be removed from Lemma 3.14, unless additional assumptions are made. To see this, consider function $h(x) \coloneqq x^2(2 + \sin\frac{1}{x})$, defined on $\mathbb{R}$. Then, h is continuously differentiable and has the quadratic growth at $x_\star = 0$, but it does not admit a Łojasiewicz function (in fact, not even a KL function). This is because $h(x) - h(x_\star) > 0$ for any $x \neq x_\star$, and for every neighborhood $\mathcal{U}_\star$ of $x_\star$ there exists $x \in U_\star \setminus \{x_\star\}$ such that $h'(x_\star) = 0$, as it can be easily verified, hence the KL inequality (3.3) can never be satisfied.

The next result establishes the equivalence of the KL property on the $\mathcal{M}$- and $\mathcal{F}$-envelopes.

Theorem 3.15. *Suppose that $\varphi^{\mathcal{M}^{\mathscr{A}}_\gamma}$ is strictly continuous. Then, for every nonempty set $\omega \subseteq \mathbb{R}^n$ the following are equivalent:*

(a) $\varphi^{\mathcal{M}^{\mathscr{A}}_\gamma}$ has the KL property on ω.

(b) $\varphi^{\mathscr{A}}_\gamma$ has the KL property on $G_\gamma^{-1}(\omega)$.

In fact, up to strictly positive scalings, there is a one-to-one correspondence between the KL functions for $\varphi^{\mathscr{A}}_\gamma$ and those for $\varphi^{\mathcal{M}^{\mathscr{A}}_\gamma}$.

Proof. We will prove the result for arbitrary strictly continuous functions $h : \mathbb{R}^n \to \mathbb{R}$ and Lipschitz homeomorphisms $G : \mathbb{R}^n \to \mathbb{R}^n$.

Let $\bar{x} \in \omega$ be fixed and let L be a Lipschitz modulus for G^{-1}. By combining Lem. 1.3*(ii)* with [106, Thm.s 10.49 and 9.62] we have that

$$\partial(h \circ G)(x) \subseteq \left\{ M^\top v \mid M \in \partial_C G(x),\ v \in \partial h(G(x)) \right\} \quad \text{for all } x \in \omega_\varepsilon.$$

Due to L-Lipschitz continuity of G^{-1}, it holds that $\|M^\top v\| \geq \frac{1}{L}\|v\|$ for any $M \in \partial_C G(\bar{x})$ and $v \in \mathbb{R}^n$. Therefore,

$$\begin{aligned} \operatorname{dist}\big(0, \partial(h \circ G)(\bar{x})\big) &\geq \min\left\{ \|M^\top v\| \mid M \in \partial_C G(\bar{x}),\ v \in \partial h(G(\bar{x})) \right\} \\ &\geq \tfrac{1}{L} \min\left\{ \|v\| \mid v \in \partial h(G(\bar{x})) \right\} \\ &= \tfrac{1}{L} \operatorname{dist}\big(0, \partial h(G(\bar{x}))\big). \end{aligned}$$

Suppose now that ψ is a KL function for h at $G(\bar{x})$. Then, for x close enough to $\bar{x}$ and with $h(G(x)) > h(G(\bar{x}))$ we have that $h(G(x)) - h(G(\bar{x}))$ is in the domain of ψ, hence

$$\operatorname{dist}\big(0,\ \partial(h \circ G)(x)\big) \geq \tfrac{1}{L} \operatorname{dist}\big(0,\ \partial h(G(x))\big) \geq \tfrac{1}{L\psi(h(G(x)) - h(G(\bar{x})))},$$

proving that $L\psi$ is a KL function for $h \circ G$ at $\bar{x}$. By interchanging G with G^{-1}, the converse implication is similarly derived. □

In conclusion of this section we establish sufficient conditions ensuring the Łojasiewicz property with exponent $\vartheta = 1/2$ when the residual $\mathcal{R}^{\mathscr{A}}_\gamma$, as defined in (2.9), is differentiable.

Proposition 3.16 (Residual nonsingularity and Łojasiewicz property)**.** *Let $s_\star$ be a fixed point for $\mathcal{F}^{\mathscr{A}}_\gamma$. Suppose that the following hold:*

- A1 *$\mathcal{R}^{\mathscr{A}}_\gamma$ is differentiable at $s_\star$ with nonsingular Jacobian $J\mathcal{R}^{\mathscr{A}}_\gamma(s_\star)$.*
- A2 *there exists $m > 0$ such that $\operatorname{dist}(0, \partial\varphi^{\mathscr{A}}_\gamma(s)) \leq m \operatorname{dist}(0, \mathcal{R}^{\mathscr{A}}_\gamma(s))$ holds for all s close to $s_\star$.*

Then, $\varphi^{\mathscr{A}}_\gamma$ has the Łojasiewicz property with exponent $\vartheta = 1/2$ at $s_\star$.

Proof. Let $\psi(t) = ct^{1/2}$ with $c > 0$ to be determined. For s close to $s_\star$, necessarily $\mathcal{R}^{\mathscr{A}}_\gamma$ is single valued. Moreover, the nonsingularity assumption entails the existence of $\alpha > 0$ such that $\|\mathcal{R}^{\mathscr{A}}_\gamma(s)\| \geq \alpha\|s - s_\star\|$. Here, we used the fact that

$s_\star$ is critical and hence $\mathcal{R}_\gamma^{\mathscr{A}}(s_\star) = 0$. Whenever $\varphi_\gamma^{\mathscr{A}}(s) > \varphi_\star \coloneqq \varphi_\gamma^{\mathscr{A}}(s_\star)$, we have

$$\begin{aligned}\psi'\big(\varphi_\gamma^{\mathscr{A}}(s) - \varphi_\star\big)\operatorname{dist}\big(0, \partial\varphi_\gamma^{\mathscr{A}}(s)\big) &= \frac{c}{2\sqrt{\varphi_\gamma^{\mathscr{A}}(s) - \varphi_\star}}\operatorname{dist}\big(0, \partial\varphi_\gamma^{\mathscr{A}}(s)\big)\\ &\geq \frac{cm}{2\sqrt{\varphi_\gamma^{\mathscr{A}}(s) - \varphi_\star}}\|\mathcal{R}_\gamma^{\mathscr{A}}(s)\|\\ &\geq \frac{cm\alpha}{\sqrt{2m_2}},\end{aligned}$$

where in the last inequality we used the quadratic bound of Prop. 3.9. By taking $c = \frac{\sqrt{2m_2}}{m\alpha}$ we obtain the sought Łojasiewicz function ψ. □

3.3 Lyapunov functions for proximal algorithms

We are now ready to show how proximal envelopes fit into the fixed-point Lyapunov framework for ensuring convergence of GPMM algorithms. Specifically, we will show that the $\mathcal{F}$-envelope $\varphi_\gamma^{\mathscr{A}}$ serves as Lyapunov function for the GPMM iterations of $\mathcal{F}_\gamma^{\mathscr{A}_\lambda}$, provided that the stepsize γ is small enough and that the relaxation λ is sufficiently close to 1. To this end, it will suffice to show that $\varphi_\gamma^{\mathscr{A}}$ satisfies the sufficient decrease property 2.2.P1, as formalized in the next result.

Lemma 3.17 (Necessity and sufficiency of the sufficient decrease)**.** *The $\mathcal{F}$-envelope $\varphi_\gamma^{\mathscr{A}}$ is a Lyapunov function for $\mathcal{F}_\gamma^{\mathscr{A}_\lambda}$ iff there exists $c > 0$ such that*

$$\varphi_\gamma^{\mathscr{A}}(\bar{s}) \leq \varphi_\gamma^{\mathscr{A}}(s) - \tfrac{c}{2}\|s - \bar{s}\|^2$$

holds for all $s \in \mathbb{R}^n$ and $\bar{s} \in \mathcal{F}_\gamma^{\mathscr{A}_\lambda}(s)$.

Proof. Since $\inf \varphi_\gamma^{\mathscr{A}} = \inf \varphi$ and since $\varphi_\gamma^{\mathscr{A}}$ has full domain, cf. Cor. 3.3*(iii)* and Prop. 3.8, both the lower boundedness prescribed by property 2.2.P1 and the real valuedness are covered. The only missing ingredient is the sufficient decrease property 2.2.P2, which is exactly what required in the statement. □

3.3.1 Sufficient decrease: a priori estimates

In this subsection we provide sufficient conditions involving the parameters m_1, m_2, L_{G_γ} and μ_{G_γ} of properties 2.10.P2 and 2.26.P2 (all depending on the stepsize γ), that ensure that the $\mathcal{F}$-envelope is a Lyapunov function for the GPMM

fixed-point iterations of $\mathcal{F}_\gamma^{\mathscr{A}_\lambda}$. Although possibly very loose estimates, they can be computed with elementary algebra without the need of any in-depth analysis, resulting in easy criteria for determining ranges of stepsizes γ and relaxation parameters λ with which $\mathcal{F}_\gamma^{\mathscr{A}_\lambda}$ fixed-point iterations are (subsequentially) convergent. In the dedicated Chapter 6, a more sophisticated analysis will tighten such ranges for the Douglas-Rachford splitting; in fact, (some of) the given ranges will prove to be optimal.

Theorem 3.18. *For all $\lambda \neq 0$, $s \in \mathbb{R}^n$, and $s^+ \in \mathcal{F}_\gamma^{\mathscr{A}_\lambda}(s)$ it holds that*

$$\varphi_\gamma^{\mathscr{A}}(s^+) \leq \varphi_\gamma^{\mathscr{A}}(s) - \frac{m_2}{2}\left(\frac{2\mu_{G_\gamma}}{\lambda} - \frac{1-\rho_{\mathcal{M}}}{\lambda^2} - L_{G_\gamma}^2\right)\|s - s^+\|^2, \tag{3.5}$$

where $\rho_{\mathcal{M}} \coloneqq {}^{m_1}\!/_{m_2}$.

Proof. Let $x \coloneqq G_\gamma(s)$ and $x^+ \coloneqq G_\gamma(s^+)$; then, there exists $\bar{x} \in \mathcal{T}_\gamma^{\mathscr{A}}(x)$ such that $s^+ = s - \lambda(x - \bar{x})$. We have

$$\begin{aligned}
\varphi_\gamma^{\mathscr{A}}(s^+) = \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x^+) &\leq \mathcal{M}_\gamma^{\mathscr{A}}(\bar{x}; x^+) \\
&\overset{2.10.\text{P2}}{\leq} \varphi(\bar{x}) + \tfrac{m_2}{2}\|x^+ - \bar{x}\|^2 \\
&\overset{3.2(ii)}{\leq} \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x) - \tfrac{m_1}{2}\|x - \bar{x}\|^2 + \tfrac{m_2}{2}\|x^+ - \bar{x}\|^2 \\
&= \varphi_\gamma^{\mathscr{A}}(s) - \tfrac{m_1}{2}\|x - \bar{x}\|^2 \\
&\quad + \tfrac{m_2}{2}\|x^+ - x\|^2 + \tfrac{m_2}{2}\|x - \bar{x}\|^2 - m_2\langle x - x^+, x - \bar{x}\rangle.
\end{aligned}$$

Since $x - \bar{x} = \frac{1}{\lambda}(s - s^+)$, the inequality becomes

$$\begin{aligned}
\varphi_\gamma^{\mathscr{A}}(s^+) &\leq \varphi_\gamma^{\mathscr{A}}(s) + \tfrac{m_2 - m_1}{2\lambda^2}\|s - s^+\|^2 + \tfrac{m_2}{2}\|x^+ - x\|^2 - \tfrac{m_2}{\lambda}\langle x - x^+, s - s^+\rangle \\
&\leq \varphi_\gamma^{\mathscr{A}}(s) - \left(\tfrac{m_2\mu_G}{\lambda} - \tfrac{m_2 - m_1}{2\lambda^2} - \tfrac{m_2 L_G^2}{2}\right)\|s - s^+\|^2,
\end{aligned}$$

hence the claimed expression. □

Corollary 3.19 (Sufficient decrease of GPMM schemes)**.** *Let $\rho_{\mathcal{M}_\gamma} \coloneqq {}^{m_1}\!/_{m_2}$, $\rho_{G_\gamma} \coloneqq {}^{\mu_{G_\gamma}}\!/_{L_{G_\gamma}}$, and $\Delta \coloneqq 1 - \rho_{G_\gamma}^2(1 - \rho_{\mathcal{M}_\gamma})$. Then, $\varphi_\gamma^{\mathscr{A}}$ is a Lyapunov function for $\mathcal{F}_\gamma^{\mathscr{A}_\lambda}$ provided that*

$$\frac{1 - \sqrt{\Delta}}{\rho_{G_\gamma} L_{G_\gamma}} < \lambda < \frac{1 + \sqrt{\Delta}}{\rho_{G_\gamma} L_{G_\gamma}}.$$

In particular, for any such stepsize γ and relaxation λ the convergence results of Theorem 3.22 apply to the fixed-point iterations of $\mathcal{F}_\gamma^{\mathscr{A}\lambda}$.

Proof. Denoting $\xi := (\mu_{G_\gamma}\lambda)^{-1}$, it follows from Thm. 3.18 that $\varphi_\gamma^{\mathscr{A}}(s^+) \leq \varphi_\gamma^{\mathscr{A}}(s) - \frac{m_2\mu_{G_\gamma}^2}{2}c\|s-s^+\|^2$, where $c := 2\xi - (1-\rho_{\mathcal{M}_\gamma})\xi^2 - \rho_{G_\gamma}^2$. By imposing $c > 0$ and solving with respect to ξ one obtains

$$\frac{1-\sqrt{\Delta}}{1-\rho_{\mathcal{M}_\gamma}} < \xi < \frac{1+\sqrt{\Delta}}{1-\rho_{\mathcal{M}_\gamma}},$$

where $\Delta := 1-\rho_{G_\gamma}^2(1-\rho_{\mathcal{M}_\gamma}$ is a strictly positive constant (since $\rho_{G_\gamma}, \rho_{\mathcal{M}_\gamma} \in (0,1]$). After easy algebraic manipulations one obtains the range of λ as in the statement, hence that $\varphi_\gamma^{\mathscr{A}}$ satisfies the sufficient decrease property 2.2.P2. The claim then follows from Lem. 3.17. □

For pure PMM schemes, the analysis is much simpler. This fact, a direct consequence of Corollary 3.19 and stated next, extends and details the analysis of the proximal point algorithm discussed at the beginning of the chapter, see Example 3.21.

Corollary 3.20 (Sufficient decrease of PMM schemes)**.** *Suppose that $G_\gamma \equiv \mathrm{id}$ for all γ's (hence $\mathcal{F}_\gamma^{\mathscr{A}\lambda} \equiv \mathcal{T}_\gamma^{\mathscr{A}}$ and $\varphi_\gamma^{\mathscr{A}} \equiv \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}$). Then, for all $\lambda > 0$, $x \in \mathbb{R}^n$ and $x^+ \in \mathcal{F}_\gamma^{\mathscr{A}\lambda}(x)$ it holds that*

$$\varphi_\gamma^{\mathscr{A}}(x^+) \leq \varphi_\gamma^{\mathscr{A}}(x) - \tfrac{m_2}{2\lambda^2}\big(\rho_{\mathcal{M}} - (1-\lambda)^2\big)\|x-x^+\|^2,$$

where $\rho_{\mathcal{M}} := {}^{m_1}\!/_{m_2}$. In particular, if $1-\sqrt{\rho_{\mathcal{M}}} < \lambda < 1+\sqrt{\rho_{\mathcal{M}}}$, then $\varphi_\gamma^{\mathscr{A}} = \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}$ is a Lypunov function for $\mathcal{F}_\gamma^{\mathscr{A}\lambda}$, hence the convergence results of Theorem 3.22 apply to its fixed-point iterations.

Proof. The inequality follows from Thm. 3.18 and the feasible range from Cor. 3.19; in fact, $G_\gamma = \mathrm{id}$ is μ_{G_γ}-strongly monotone and L_{G_γ}-Lipschitz continuous with $\mu_{G_\gamma} = L_{G_\gamma} = 1$. □

Example 3.21 (Proximal point algorithm)**.** As detailed in Example 2.29, the proximal point is a pure PMM algorithm with $m_1(\gamma) = m_2(\gamma) = {}^1\!/_\gamma$. Its model and algorithmic envelopes coincide, and equal the Moreau envelope φ^γ. Corollary 3.20 then readily applies, resulting in

$$\varphi^\gamma(x^+) \leq \varphi^\gamma(x) - \tfrac{2-\lambda}{2\gamma\lambda}\|x-x^+\|^2$$

for all $x^+ \in \mathcal{F}_\gamma^{\mathrm{PP}_\lambda}(x) = (1-\lambda)x + \lambda \operatorname{prox}_{\gamma\varphi}(x)$, thus confirming what revealed in (3.2). □

3.4 Convergence of GPMM algorithms

We conclude the chapter by furnishing the claimed convergence results of GPMM algorithms. As expected, subsequential convergence will easily follow from the similar result in the more general fixed-point framework that opened Chapter 2. The KL property, discussed and analyzed in Section 3.2.4, will allow the development of stronger convergence results, which will pattern the similar ones shown in particular cases, see *e.g.*, [5].

Theorem 3.22. *Consider the fixed-point iterations $(s^k)_{k\in\mathbb{N}}$ generated by $\mathcal{F}_\gamma^{\mathscr{A}_\lambda}$, and let $x^k := G_\gamma(s^k)$. If $\varphi_\gamma^{\mathscr{A}}$ satisfies the sufficient decrease property 2.2.P2 for $\mathcal{F}_\gamma^{\mathscr{A}_\lambda}$, then the following hold:*

(i) The fixed-point residual $(\operatorname{dist}(0, \mathcal{R}_\gamma^{\mathscr{A}}(s^k)))_{k\in\mathbb{N}}$ is square summable; in particular, $\min_{j\le k} \operatorname{dist}(0, \mathcal{R}_\gamma^{\mathscr{A}}(s^j)) \in O(1/\sqrt{k})$.

(ii) The set ω of accumulation points of the sequence $(x^k)_{k\in\mathbb{N}}$ satisfies $\omega \subseteq \operatorname{fix} \mathcal{T}_\gamma^{\mathscr{A}} \subseteq \operatorname{zer} \hat{\partial}\varphi$.

(iii) If φ is level bounded, then $(s^k)_{k\in\mathbb{N}}$ and $(x^k)_{k\in\mathbb{N}}$ are bounded, and ω is a nonempty, compact, and connected set satisfying $\operatorname{dist}(x^k, \omega) \to 0$ as $k \to \infty$.

(iv) $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}} \equiv \varphi$ on ω, the value being the limit of the (decreasing) sequence $(\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x^k))_{k\in\mathbb{N}}$.

Proof. We know from Lem. 3.17 that $\varphi_\gamma^{\mathscr{A}}$ is a Lyapunov function for $\mathcal{F}_\gamma^{\mathscr{A}_\lambda}$. Moreover, by definition of $\mathcal{F}_\gamma^{\mathscr{A}_\lambda}$, for all $k \in \mathbb{N}$ there exists $\bar{x}^k \in \mathcal{T}_\gamma^{\mathscr{A}}(x^k)$ such that $s^{k+1} = s^k - \lambda(x^k - \bar{x}^k)$.

♠ 3.22*(i)*. Follows from Thm. 2.4*(i)* together with the fact that

$$\operatorname{dist}(x^k, \mathcal{T}_\gamma^{\mathscr{A}}(x^k)) \le \|x^k - \bar{x}^k\| = \tfrac{1}{\lambda}\|s^k - s^{k+1}\|.$$

♠ 3.22*(ii)*. Follows from Thm. 2.4*(ii)* and (2.7).

♠ 3.22*(iii)*. Follows from Thm. 2.4*(iii)* together with the fact that the continuous function G_γ maps bounded sets to bounded sets.

♠ 3.22*(iv)*. That $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}} \equiv \varphi$ on ω follows from Thm. 3.2*(i)* in light of the inclusion $\omega \subseteq \operatorname{fix} \mathcal{T}_\gamma^{\mathscr{A}}$. Due to properties 2.2.P1 and 2.2.P2 of Lyapunov functions, the sequence $(\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x^k))_{k\in\mathbb{N}} = (\varphi_\gamma^{\mathscr{A}}(s^k))_{k\in\mathbb{N}}$ is decreasing and admits a finite limit, be it $\varphi_\star$. From Thm. 3.2*(ii)* we have that

$$\left|\varphi(\bar{x}^k) - \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x^k)\right| \leq \tfrac{m_2}{2}\|x^k - \bar{x}^k\|^2,$$

and since $\|x^k - \bar{x}^k\| \to 0$ we infer that $\varphi(\bar{x}^k) \to \varphi_\star$. Let $x_\star \in \omega$ be fixed, and consider a subsequence $(x^{j_k})_{k\in\mathbb{N}}$ that converges to $x_\star$. Then, also $(\bar{x}^{j_k})_{k\in\mathbb{N}} \to x_\star$. We have

$$\varphi(x_\star) \leq \varphi_\star \leq \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x^{j_k}) \leq \mathcal{M}_\gamma^{\mathscr{A}}(x_\star; x^{j_k}) \overset{2.10.\text{P2}}{\leq} \varphi(x_\star) + \tfrac{m_2}{2}\|x^{j_k} - x_\star\|^2$$

where the first inequality is due to lower semicontinuity, the second one to the fact that $(\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x^k))_{k\in\mathbb{N}}$ is decreasing, and the third one from the definition of $\mathcal{M}_\gamma^{\mathscr{A}}$-envelopes. Since $(x^{j_k})_{k\in\mathbb{N}} \to x_\star$, we conclude that $\varphi(x_\star) = \varphi_\star$, and the claim follows from the arbitrarity of $x_\star \in \omega$. □

Theorem 3.23 (Global convergence)**.** *Suppose that the following hold:*

A1 *φ is level bounded;*

A2 *$\varphi_\gamma^{\mathscr{A}}$ satisfies the sufficient decrease property 2.2.P2 for $\mathcal{F}_\gamma^{\mathscr{A}_\lambda}$;*

A3 *$\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}$ has the KL property;*

A4 *there exists $m > 0$ such that $\operatorname{dist}(0, \partial\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x)) \leq m \operatorname{dist}(x, \mathcal{T}_\gamma^{\mathscr{A}}(x))$ holds for all x close to $\operatorname{fix} \mathcal{T}_\gamma^{\mathscr{A}}$.*

Let $(s^k)_{k\in\mathbb{N}}$ be a sequence generated by fixed-point iterations of $\mathcal{F}_\gamma^{\mathscr{A}_\lambda}$, and let $x^k := G_\gamma(s^k)$. Then, the following hold:

(i) $(x^k)_{k\in\mathbb{N}}$ converges to a point $x_\star \in \operatorname{fix} \mathcal{T}_\gamma^{\mathscr{A}} \subseteq \operatorname{zer} \hat{\partial}\varphi$ (hence $(s^k)_{k\in\mathbb{N}}$ converges to $G_\gamma^{-1}(x_\star)$).

(ii) The fixed-point residual $(\operatorname{dist}(0, \mathcal{R}_\gamma^{\mathscr{A}}(s^k)))_{k\in\mathbb{N}}$ is summable, and in particular $\min_{j\leq k} \operatorname{dist}(0, \mathcal{R}_\gamma^{\mathscr{A}}(s^k)) \in O(1/k)$.

Proof. From Thm.s 3.22*(iii)* and 3.22*(iv)* we have that the sequence $(x^k)_{k\in\mathbb{N}}$ remains bounded, that the set of accumulation points ω is a nonempty, compact and connected set such that $\operatorname{dist}(x^k, \omega) \to 0$ as $k \to \infty$, and that $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}$ is constant on ω. Let $\varphi_\star$ be the value of $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}$ on ω. Let the constants $\delta, \varepsilon > 0$ and

the uniformized KL function ψ be as in Lem. 3.12. Up to possibly discarding the first iterates, without loss of generality we may assume that $\operatorname{dist}(x^k, \omega) < \varepsilon$ and $\varphi^{\mathcal{M}^{\mathscr{A}}_\gamma}(x^k) < \varphi_\star + \eta$ for all $k \in \mathbb{N}$, so that

$$\Delta_k \coloneqq \psi\big(\varphi^{\mathcal{M}^{\mathscr{A}}_\gamma}(x^k) - \varphi_\star\big)$$

are well-defined quantities for all $k \in \mathbb{N}$. Similarly, we may also assume that all points x^k are close enough to $\omega \subseteq \operatorname{fix} \mathcal{T}^{\mathscr{A}}_\gamma$ so that the bound 3.23A4 holds. Let $c > 0$ be the sufficient decrease constant of the Lyapunov function $\varphi^{\mathscr{A}}_\gamma$ (cf. 2.2.P2), and let $\bar{x}^k \in \mathcal{T}^{\mathscr{A}}_\gamma(x^k)$ be such that $s^{k+1} = s^k - \lambda(x^k - \bar{x}^k)$. Then,

$$\begin{aligned}\varphi^{\mathcal{M}^{\mathscr{A}}_\gamma}(x^k) - \varphi^{\mathcal{M}^{\mathscr{A}}_\gamma}(x^{k+1}) &= \varphi^{\mathscr{A}}_\gamma(s^k) - \varphi^{\mathscr{A}}_\gamma(s^{k+1}) \\ &\geq \tfrac{c}{2}\|s^k - s^{k+1}\|^2 = \tfrac{c\lambda^2}{2}\|x^k - \bar{x}^k\|^2. \end{aligned} \tag{3.6}$$

We have

$$\begin{aligned}\Delta_k - \Delta_{k+1} &\geq \psi'\big(\varphi^{\mathcal{M}^{\mathscr{A}}_\gamma}(x^k) - \varphi_\star\big)\big(\varphi^{\mathcal{M}^{\mathscr{A}}_\gamma}(x^k) - \varphi^{\mathcal{M}^{\mathscr{A}}_\gamma}(x^{k+1})\big) \\ &\overset{3.12.\text{P3}}{\geq} \frac{1}{\operatorname{dist}\big(0, \partial\varphi^{\mathcal{M}^{\mathscr{A}}_\gamma}(x^k)\big)}\big(\varphi^{\mathcal{M}^{\mathscr{A}}_\gamma}(x^k) - \varphi^{\mathcal{M}^{\mathscr{A}}_\gamma}(x^{k+1})\big) \\ &\overset{(3.6)}{\geq} \frac{c\lambda^2\|x^k - \bar{x}^k\|^2}{2\operatorname{dist}\big(0, \partial\varphi^{\mathcal{M}^{\mathscr{A}}_\gamma}(x^k)\big)} \\ &\overset{3.23\text{A4}}{\geq} \frac{c\lambda^2\|x^k - \bar{x}^k\|^2}{2m\operatorname{dist}\big(x^k, \mathcal{T}^{\mathscr{A}}_\gamma(x^k)\big)} \\ &\geq \frac{c\lambda^2}{2m}\|x^k - \bar{x}^k\|. \end{aligned} \tag{3.7}$$

We may then telescope the inequality to obtain

$$\Delta_0 \geq \sum_{k\in\mathbb{N}}(\Delta_k - \Delta_{k+1}) \geq \frac{c\lambda^2}{2m}\sum_{k\in\mathbb{N}}\|x^k - \bar{x}^k\| = \frac{c\lambda^2}{2m}\sum_{k\in\mathbb{N}}\|s^k - s^{k+1}\|,$$

where the first inequality follows from the fact that $\Delta_k \geq 0$. This shows that the sequence $(s^k)_{k\in\mathbb{N}}$ has finite length, and therefore is convergent, hence so is $(x^k)_{k\in\mathbb{N}}$ due to continuity of G_γ. That the limit of $(x^k)_{k\in\mathbb{N}}$ belongs to $\operatorname{fix}\mathcal{T}^{\mathscr{A}}_\gamma \subseteq \operatorname{zer}\hat{\partial}\varphi$ is a consequence of Thm. 3.22*(ii)*, and the claim on the rate of convergence can be shown by arguing as in the proof of Thm. 2.4*(i)*. □

Theorem 3.24 (Linear convergence)**.** *Suppose that the assumptions of Theorem 3.23 are satisfied, and that the KL function can be taken of the form $\psi(t) = ct^{\vartheta}$ for some $c > 0$ and $\vartheta \geq 1/2$. Then, the sequences $(s^k)_{k\in\mathbb{N}}$, $(x^k)_{k\in\mathbb{N}}$, and $\operatorname{dist}(0, \mathcal{R}_\gamma^{\mathscr{A}}(s^k))$ are R-linearly convergent.*

Proof. Let $x^k \coloneqq G_\gamma(s^k)$ and $r^k \coloneqq x^k - \bar{x}^k$. We know from Thm. 3.23 that the sequence $(x^k)_{k\in\mathbb{N}}$ converges to a point $x_\star$, and that $r^k \to 0$ as $k \to \infty$. For all k's large enough such that x^k is sufficiently close to $x^\star$, we have

$$\|r^k\| \geq \operatorname{dist}(x^k, \mathcal{T}_\gamma^{\mathscr{A}}(x^k)) \overset{3.23\text{A}4}{\geq} \tfrac{1}{m} \operatorname{dist}\big(0, \partial\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x^k)\big)$$

$$\text{(due to 3.23A3)} \geq \frac{1}{m\psi'\big(\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x^k) - \varphi(x_\star)\big)}$$

$$= \frac{1}{mc\vartheta}\big(\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x^k) - \varphi(x_\star)\big)^{1-\vartheta}. \tag{3.8}$$

Therefore,

$$\Delta_k \coloneqq \psi\big(\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x^k) - \varphi(x_\star)\big) = c\big(\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x^k) - \varphi(x_\star)\big)^{\vartheta} \leq c\big(mc\vartheta\|r^k\|\big)^{\frac{\vartheta}{1-\vartheta}},$$

where the last inequality follows from (3.8). Since $r^k \to 0$ as $k \to \infty$, for all k's large enough it holds that $mc\vartheta\|r^k\| \leq 1$. Moreover, having $\vartheta \geq 1/2$ implies that $\frac{\vartheta}{1-\vartheta} \geq 1$, hence $\big(mc\vartheta\|r^k\|\big)^{\frac{\vartheta}{1-\vartheta}} \leq mc\vartheta\|r^k\|$ for all k's large enough. We may thus continue the inequality as

$$\Delta_k \leq mc^2\vartheta\|r^k\|. \tag{3.9}$$

Let $B_k \coloneqq \sum_{j\geq k}\|r^j\|$. Then,

$$\|s^k - s_\star\| \leq \sum_{j\geq k}\|s^j - s^{j+1}\| = \lambda\sum_{j\geq k}\|x^j - \bar{x}^j\| = \lambda B_k. \tag{3.10}$$

We have

$$B_k = \sum_{j\geq k}\|r^j\| \overset{(3.7)}{\leq} \frac{2m}{c\lambda^2}\sum_{j\geq k}(\Delta_j - \Delta_{j+1}) = \frac{2m}{c\lambda^2}\Delta_k$$

$$\text{(due to (3.9))} \leq \frac{2\vartheta m^2 c}{\lambda^2}\|r^k\| = \frac{2\vartheta m^2 c}{\lambda^2}(B_k - B_{k+1}).$$

By suitably rearranging, we obtain that for all k's large enough it holds that

$$B_{k+1} \leq \big(1 - \tfrac{\lambda^2}{2\vartheta m^2 c}\big) B_k,$$

hence, that $(B_k)_{k\in\mathbb{N}}$ is asymptotically Q-linearly convergent. The claimed R-linear convergence rates then follow from (3.10) and from the fact that $\|r^k\| \leq B_k$. □

Chapter 4

Acceleration of nonconvex splitting algorithms

4.1 A new backtracking paradigm
for continuous Lyapunov functions

In the next chapters we will see that, under due assumptions, many splitting algorithms fit into the proposed GPMM framework. In this perspective, the theory developed so far serves a twofold purpose. First, it establishes a novel (and unified) convergence analysis of *known* splitting algorithms applied to nonconvex problems. Secondly, it sets the ground for building *new* methods on top of the known ones, which at negligible additional cost per iteration may result in an outstanding performance improvement. This chapter deals with this second objective. Interestingly, most (if not all) known splitting algorithms that fit into the generalized proximal MM framework are based on *continuous* models, in the sense of Definition 2.11. As a result, not only are the corresponding $\mathcal{M}$- and $\mathcal{F}$-envelope functions real valued, but they are also continuous, as ensured by Proposition 3.8.

Continuity of the Lyapunov function is the key property over which this chapter depends. In fact, although the same arguments could even be applied to the abstract Lyapunov framework briefly investigated in Section 2.1 by simply restricting the analysis to continuous functions $\mathcal{L}$, now that we are well acquainted with proximal envelopes and GPMM algorithms this degree of generality is no longer needed. Nevertheless, for the sake of understanding how plain *continuity* can be of any use, let us suppose that $\mathcal{L} \in C^0(\mathbb{R}^n)$ is a Lyapunov function for some osc and nonempty-valued fixed-point mapping $\mathcal{F} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, and let $c > 0$ be the sufficient decrease constant as in property 2.2.P2. Suppose that the current iterate is $s \in \mathbb{R}^n$, and let $d \in \mathbb{R}^n$ be an arbitrary candidate *update direction* at s. What d is, and how it is retrieved is irrelevant at the mo-

ment; suffice it to say that the choice of an update direction d represents our degree of freedom for extending a *known* (splitting) algorithm while maintaining its (subsequential) convergence properties, and that "ideally" we would like to replace the fixed-point update $s \mapsto s^+ \in \mathcal{F}(s)$ with the chosen $s^+ = s + d$, for we have reason to believe this choice will lead us closer to a fixed point of $\mathcal{F}$. In order to distinguish between the proposed modification and the original fixed-point iteration, let us establish the convention of denoting $\bar{s} \in \mathcal{F}(s)$ (as opposed to s^+ which we shall reserve to the modified update), and refer to it as a *nominal* update. Let us also suppose that the current iterate s is not a fixed point of $\mathcal{F}$, for otherwise a solution would be found and there would be no reason to investigate any further. Due to the sufficient decrease property of the Lyapunov function $\mathcal{L}$, we know that

$$\mathcal{L}(\bar{s}) \leq \mathcal{L}(s) - \tfrac{c}{2}\|s - \bar{s}\|^2 \quad \text{for any } \textit{nominal} \text{ update } \bar{s} \in \mathcal{F}(s).$$

However, nothing can be said as to whether $\mathcal{L}(s+d)$ is also (sufficiently) smaller than $\mathcal{L}(s)$ or not, nor can we hope to enforce the condition with a classical backtracking $s + \tau d$ for small $\tau > 0$, as no notion of descent is known to $\mathcal{L}$ (which is continuous but not necessarily differentiable); moreover the direction d is even arbitrary. Is there a way to design a linesearch ensuring that the wanted update, or something close to it, also satisfies a sufficient decrease? Here is where continuity comes into the picture.

Let us replace the sufficient decrease constant c with a smaller value, say, αc for some $\alpha \in (0, 1)$. Then, not only does $\bar{s}$ satisfy the sufficent decrease with constant αc, but due to continuity of $\mathcal{L}$ so do all the points around: loosely speaking,

$$\mathcal{L}(s') \leq \mathcal{L}(s) - \tfrac{\alpha c}{2}\|s - \bar{s}\|^2 \quad \text{for all } s' \textit{ close} \text{ to } \bar{s}. \tag{4.1}$$

The idea is then to "push" the candidate update $s + d$ towards the "safe" update $\bar{s}$ until the *relaxed* decrease condition (4.1) holds. One way to do so is through a linesearch along the segment connecting the "ideal" update $s + d$ and the "safe" nominal update $\bar{s}$, as follows:

$\tau \leftarrow 1$
repeat $s^+ = (1 - \tau)\bar{s} + \tau(s + d)$
until $\mathcal{L}(s^+) \leq \mathcal{L}(s) - \frac{\alpha c}{2}\|s - \bar{s}\|^2$.

The caveat of this approach is that it assumes that $\mathcal{L}$ is an *explicit* function, in the sense that its value can be computed at any point. On the contrary, this was not an issue in the previous chapters where only nominal algorithms were investigated. To underline the significance of the claim, recall that for convex splitting algorithms $\mathcal{L} = \operatorname{dist}(\,\cdot\,, \operatorname{fix}\mathcal{F})^2$ is a suitable Lyapunov function, cf. (2.2), which is also continuous. Unfortunately, however, such a Lyapunov function is of no use, as its value, in meaningful problems, cannot be evaluated.

A solution to this problem will be proposed in Chapter 8, where by means of properties of projections and *averaged* mappings $\mathcal{L}$ will be approximated by using available information.

This solution, however, does not apply in the nonconvex setting investigated here, which is where envelope functions yet again prove their worth. Indeed, when $\mathcal{L} = \varphi_\gamma^{\mathscr{A}}$ is an envelope function, one evaluation requires only one nominal step: any $\bar{s} \in \mathcal{F}_\gamma^{\mathscr{A}_\lambda}(s)$ is expressed as

$$\bar{s} = s - \lambda(x - \bar{x}) \quad \text{for some } \bar{x} \in \mathcal{T}_\gamma^{\mathscr{A}}(x),$$

hence

$$\varphi_\gamma^{\mathscr{A}}(s) = \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x) = \mathcal{M}_\gamma^{\mathscr{A}}(\bar{x}; x).$$

4.2 The CLyD algorithmic framework

Algorithm 4.1. CONTINUOUS-LYAPUNOV DESCENT FRAMEWORK

REQUIRE
- γ, λ s.t. the continuous function $\varphi_\gamma^{\mathscr{A}}$ satisfies the sufficient decrease 2.2.P2 with $c > 0$ for $\mathcal{F}_\gamma^{\mathscr{A}_\lambda}$
- scaling factor $\alpha \in (0, 1)$ for sufficient decrease constant
- initial iterate $s^0 \in \mathbb{R}^n$
- tolerance $\varepsilon > 0$

PROVIDE x_* with $\operatorname{dist}\big(x_*, \mathcal{T}_\gamma^{\mathscr{A}}(x_*)\big) \le \varepsilon$

1: **for** $k = 0, 1, 2, \ldots$ **do**

2: Do one nominal $\mathcal{F}_\gamma^{\mathscr{A}_\lambda}$-step: $x^k = G_\gamma(s^k)$, $\bar{x}^k \in \mathcal{T}_\gamma^{\mathscr{A}}(s^k)$, $\bar{s}^k = s^k - \lambda(x^k - \bar{x}^k)$

3: **if** $\|x^k - \bar{x}^k\| \le \varepsilon$ **then**

4: **return** $x_* = \bar{x}^k$

5: Select an update direction $d^k \in \mathbb{R}^n$ at s^k

6: Let $\tau_k \in \{2^{-i} \mid i \in \mathbb{N}\}$ be the largest such that

$$\varphi_\gamma^{\mathscr{A}}(s^{k+1}) \le \varphi_\gamma^{\mathscr{A}}(s^k) - \alpha\tfrac{c}{2}\|s^k - \bar{s}^k\|^2, \tag{4.2}$$

where $s^{k+1} := (1 - \tau_k)\bar{s}^k + \tau_k(s^k + d^k)$

The ideas discussed in the previous section lead to the *Continuous-Lyapunov Descent algorithm (CLyD)*, detailed in Algorithm 4.1. We begin by showing that the proposed algorithm maintains the subsequential convergence properties of the underlying nominal scheme.

Theorem 4.1 (Subsequential convergence of (nonmonotone) CLyD)**.** *The following hold for the iterates generated by CLyD (Alg. 4.1) with tolerance* $\varepsilon = 0$*:*

(i) The residual $(\|x^k - \bar{x}^k\|)_{k\in\mathbb{N}}$ *is square-summable; in particular, it vanishes with rate* $\min_{j\le k} \operatorname{dist}(x^j, \mathcal{T}_\gamma^{\mathscr{A}}(x^j)) \in O(1/\sqrt{k})$.

(ii) The set ω *of accumulation points of* $(x^k)_{k\in\mathbb{N}}$ *satisfies* $\omega \subseteq \operatorname{fix} \mathcal{T}_\gamma^{\mathscr{A}} \subseteq \operatorname{zer} \hat{\partial}\varphi$.

If, additionally, $\|d^k\| \to 0$ *as* $k \to \infty$*, then the following also hold:*

(iii) If φ *is level bounded, then* $(s^k)_{k\in\mathbb{N}}$, $(x^k)_{k\in\mathbb{N}}$, $(\bar{s}^k)_{k\in\mathbb{N}}$, *and* $(\bar{x}^k)_{k\in\mathbb{N}}$ *are bounded, and* ω *is a nonempty, compact and connected set satisfying* $\operatorname{dist}(x^k, \omega) \to 0$ *as* $k \to \infty$.

(iv) $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}} \equiv \varphi$ *on* ω*, the value being the limit of the (decreasing) sequence* $(\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x^k))_{k\in\mathbb{N}}$.

All the claims remain valid if the linesearch condition (4.2) is replaced by the following nonmonotone version:

$$\varphi_\gamma^{\mathscr{A}}(s^{k+1}) \le \bar{\mathcal{L}}_k - \alpha\tfrac{c}{2}\|s^k - \bar{s}^k\|^2, \tag{4.3}$$

where, for any sequence $(t_k)_{k\in\mathbb{N}} \subseteq [0,1]$ *bounded away from* 0, $\bar{\mathcal{L}}_k$ *are recursively defined as follows:*

$$\bar{\mathcal{L}}_k := \begin{cases} \varphi_\gamma^{\mathscr{A}}(s^0) & \text{if } k = 0, \\ (1-t_k)\bar{\mathcal{L}}_{k-1} + t_k \varphi_\gamma^{\mathscr{A}}(s^k) & \text{otherwise.} \end{cases}$$

Proof. The feasibility of the linesearch for arbitrary directions d^k has been extensively discussed in the previous section. Moreover, the first part of the proof is similar to that of Thm. 3.22, so we simply outline the details.

♠ 4.1*(i)*. We have

$$\sum_{k\in\mathbb{N}} \|s^k - \bar{s}^k\|^2 \overset{(4.2)}{\le} \frac{2}{\alpha c} \sum_{k\in\mathbb{N}} \big(\varphi_\gamma^{\mathscr{A}}(s^k) - \varphi_\gamma^{\mathscr{A}}(s^{k+1})\big) \le \frac{2}{\alpha c}\big(\varphi_\gamma^{\mathscr{A}}(s^0) - \inf\varphi\big) \overset{2.2.\text{P1}}{<} \infty,$$

where in the third inequality we used the fact that $\inf \varphi_\gamma^{\mathscr{A}} = \inf \varphi$, cf. Cor. 3.3*(iii)*.

♠ 4.1*(ii)*. Suppose that a subsequence $(s^{k_j})_{j\in\mathbb{N}}$ converges to a point $s_\star$. Since $(\|s^k - \bar{s}^k\|)_{k\in\mathbb{N}} \to 0$, it also holds that $\bar{s}^{k_j} \to s_\star$ as $j \to \infty$. Thus,

$$s_\star = \lim_{j\to\infty} \bar{s}^{k_j} \in \limsup_{j\to\infty} \mathcal{F}_\gamma^{\mathscr{A}_\lambda}(s^{k_j}) \subseteq \mathcal{F}_\gamma^{\mathscr{A}_\lambda}(s_\star),$$

where the last inclusion follows from the fact that $\mathcal{F}_\gamma^{\mathscr{A}_\lambda}$ is osc.

♠ 4.1*(iii)*. Since $(s^k)_{k\in\mathbb{N}} \subset \operatorname{lev}_{\leq \varphi_\gamma^{\mathscr{A}}(s^0)}$, the sequence is bounded provided $\varphi_\gamma^{\mathscr{A}}$ is level bounded, which in turn is equivalent to having φ level bounded, cf. Thm. 3.7. Then, due to continuity of G_γ, necessarily also $(x^k)_{k\in\mathbb{N}}$ is bounded.

If, additionally, $\|d^k\| \to 0$ as $k \to \infty$, then since $\tau_k \in [0,1]$ for all k's we have

$$\|s^{k+1} - s^k\| = \big\|(1-\tau_k)\bar{s}^k + \tau_k(s^k + d^k) - s^k\big\| \leq (1-\tau_k)\|s^k - \bar{s}^k\| + \tau_k\|d^k\| \to 0$$

as $k \to \infty$. The proof now follows from Lem. 2.3.

Let us now prove the claim for the nonmonotone variant. We start by showing that the linesearch is indeed feasible, and that for all $k \in \mathbb{N}$ the following hold:

(v) $\bar{\mathcal{L}}_k \geq \varphi_\gamma^{\mathscr{A}}(s^k)$.

(vi) $\bar{\mathcal{L}}_{k+1} \leq \bar{\mathcal{L}}_k - t_{k+1}\alpha c\|s^k - \bar{s}^k\|^2$.

For $k = 0$, inequality 4.1*(v)* holds as equality, and in particular the nonmonotone linesearch condition (4.3) is satisfied by small enough stepsizes. Suppose now that up to iteration $k \geq 0$ the inequality holds and in particular the nonmonotone linesearch is feasible; then,

$$\begin{aligned}\bar{\mathcal{L}}_{k+1} &= (1-t_{k+1})\bar{\mathcal{L}}_k + t_{k+1}\varphi_\gamma^{\mathscr{A}}(s^{k+1}) \\ &\geq (1-t_{k+1})\varphi_\gamma^{\mathscr{A}}(s^{k+1}) + t_{k+1}\varphi_\gamma^{\mathscr{A}}(s^{k+1}) \\ &= \varphi_\gamma^{\mathscr{A}}(s^{k+1}),\end{aligned}$$

where in the inequality the nonmonotone linesearch (4.3) was used. Hence, 4.1*(v)* holds for all k's, and the linesearch is always feasible. The inequality in 4.1*(vi)* then readily follows from the fact that $\bar{\mathcal{L}}_{k+1} \geq \varphi_\gamma^{\mathscr{A}}(s^{k+1})$. In particular, $\bar{\mathcal{L}}_k \geq \inf \varphi_\gamma^{\mathscr{A}} > -\infty$ for all k; we may then telescope the linesearch inequality

(4.3) to arrive to

$$\infty > \bar{\mathcal{L}}_0 - \inf \varphi_\gamma^{\mathscr{A}} \geq \alpha c \sum_{k\in\mathbb{N}} t_k \|s^k - \bar{s}^k\|^2 \geq \alpha c t_{\min} \sum_{k\in\mathbb{N}} \|s^k - \bar{s}^k\|^2,$$

where $t_{\min} := \inf_{k\in\mathbb{N}} t_k$, which is strictly positive by assumption. We may now trace the proof of the monotone variant to arrive to the same conclusions. □

An interesting observation is that *for pure PMM schemes with no relaxation*, that is, if $G = \mathrm{id}$ and $\lambda = 1$, *continuity of the envelope function is not required for the well definedness of the algorithm.* In fact, in this case upper bounds on the number of backtracking of τ_k can be established without any continuity requirement. To see this, let $x, d \in \mathbb{R}^n$ be fixed, and let $\bar{x} \in \mathcal{T}_\gamma^{\mathscr{A}}(x)$ be the result of a nominal PMM-step. For $\tau \in [0,1]$, consider $x_\tau^+ := (1-\tau)\bar{x} + \tau(x+d)$. Then,

$$\begin{aligned}
\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x_\tau^+) &\overset{3.9}{\leq} \varphi(\bar{x}) + \tfrac{m_2}{2}\|\bar{x} - x_\tau^+\|^2 \\
&\leq \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x) - \tfrac{c}{2}\|\bar{x} - x\|^2 + \tfrac{m_2}{2}\|\bar{x} - x_\tau^+\|^2 \\
&= \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x) - \tfrac{c}{2}\|\bar{x} - x\|^2 + \tfrac{m_2\tau^2}{2}\|\bar{x} - x - d\|^2 \\
&\leq \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x) - \tfrac{c}{2}\|\bar{x} - x\|^2 + m_2\tau^2\big(\|\bar{x} - x\|^2 + \|d\|^2\big).
\end{aligned}$$

Thus, $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x_\tau^+) \leq \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x) - \alpha\frac{c}{2}\|x - \bar{x}\|^2$ when $\tau \leq \sqrt{\frac{(1-\alpha)c}{2m_2}\frac{1}{1+\|d\|^2/\|x-\bar{x}\|^2}}$. Since τ is halven every time the linesearch condition (4.2) does not hold, it follows that

$$G = \mathrm{id},\ \lambda = 1 \quad \Rightarrow \quad \tau_k \geq \tfrac{1}{2}\sqrt{\tfrac{(1-\alpha)c}{2m_2}\tfrac{1}{1+\|d^k\|^2/\|x^k-\bar{x}^k\|^2}} \quad \forall k$$

in CLyD (Alg. 4.1). In order to drop the assumption of continuity as well as to upper bound the number of τ-bactrackings for any GPMM scheme with relaxation, one could consider replacing the s-update rule with

$$s^+ = (1-\tau)G^{-1}(\bar{x}_0^+) + \tau(s+d),$$

where $s_0^+ \in \mathcal{F}_\gamma^{\mathscr{A}_\lambda}(s)$ is the result of a nominal $\mathcal{F}_\gamma^{\mathscr{A}_\lambda}$-update, $x_0^+ = G(s_0^+)$, and $\bar{x}_0^+ \in \mathcal{T}_\gamma^{\mathscr{A}}(x_0^+)$. In fact, similarly to the chain of inequalities above,

$$\begin{aligned}
\varphi_\gamma^{\mathscr{A}}(s^+) &\overset{(def)}{=} \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(G(s^+)) \\
&\overset{\text{(Prop. 3.9)}}{\leq} \varphi(\bar{x}_0^+) + \tfrac{m_2}{2}\|G(s^+) - \bar{x}_0^+\|^2
\end{aligned}$$

$$
\begin{aligned}
\text{(Thm. 3.2}\textit{(ii)}\text{)} &\leq \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x_0^+) + \tfrac{m_2}{2}\|G(s^+) - \bar{x}_0^+\|^2 \\
&= \varphi_\gamma^{\mathscr{A}}(s_0^+) + \tfrac{m_2}{2}\|G(s^+) - \bar{x}_0^+\|^2 \\
&\leq \varphi_\gamma^{\mathscr{A}}(s) - \tfrac{c}{2}\|s - s_0^+\|^2 + \tfrac{m_2}{2}\|G(s^+) - \bar{x}_0^+\|^2.
\end{aligned}
$$

Then, we may use L_G-Lipschitz continuity of the transient mapping G to infer that

$$
\|G(s^+) - \bar{x}_0^+\| \leq L_G\|s^+ - G^{-1}(\bar{x}_0^+)\| = \tau L_G\|G^{-1}(\bar{x}_0^+) - s - d\|,
$$

and a lower bound on τ ensuring the linesearch condition (4.2) can then be easily established, without the need of continuity of $\varphi_\gamma^{\mathscr{A}}$.

The problem of this strategy is that it requires inverting the transient mapping G (for computing the point $G^{-1}(x_0^+)$, an operation that is not prescribed by the nominal $\mathcal{F}_\gamma^{\mathscr{A}_\lambda}$-iterations. As detailed in the next section, by using the backtracking proposed in CLyD (Alg. 4.1) we can instead improve the underlying GPMM scheme without the need to complicate the oracle of its iterations. For this reason and for simplicity of the exposition, although most convergence results could easily be extended, the variant here introduced will not be further discussed in the thesis.

4.3 Choice of directions

Although the proposed algorithmic framework is robust to any choice of directions d^k, on the contrary its efficacy is greatly affected by the specific selection. This section provides an overview on some update directions d^k that can be conveniently considered.

The termination criterion for CLyD (Alg. 4.1) is based on (the norm of) the fixed-point residual of the underlying splitting schemes

$$
\mathcal{R}_\gamma^{\mathscr{A}}(s) = \tfrac{1}{\gamma}(x - \bar{x}) = \tfrac{1}{\gamma\lambda}(s^+ - s).
$$

Under some assumptions, which will be investigated case by case in the remaining chapters, close to critical points the residual mapping $\mathcal{R}_\gamma^{\mathscr{A}}$ becomes a well-behaved single-valued function, possibly enjoying Lipschitzian or differentiability properties. As a result, one ends up solving a system of nonlinear equations, namely finding $s_\star$ such that $\mathcal{R}_\gamma^{\mathscr{A}}(s_\star) = 0$.

As a way to speed up convergence, one possibility is to employ directions stemming from fast methods for nonlinear equations, namely

$$d^k = -H_k \mathcal{R}_\gamma^{\mathscr{A}}(s^k),$$

where the linear operator H_k mimicks $J\mathcal{R}_\gamma^{\mathscr{A}}(s^k)^{-1}$. When the residual is differentiable or admits some Jacobian approximations, one can indeed consider an exact Newton step as update direction. However, the combination of nonconvexity and nonsmoothness in the investigated problems makes this property quite uncommon. Moreover, even when this is the case, the computation of (generalized) Jacobians and the consequent solution of linear system to retrieve the Newton direction fails to preserve the simple oracle of the nominal splitting algorithms.

For these reasons, we limit the analysis to *quasi-Newton* schemes which, starting from any invertible matrix H_0 (typically a positve multiple of the identity) perform low-rank updates based on available quantities. Such quantities are pairs of vectors (p_k, q_k), where p_k is the difference between iterates and q_k the difference of the respective fixed-point residuals: denoting $s_0^{k+1} := s^k + d^k$ the update tried first in the backtracking (that is, with $\tau = 1$), these vectors are given by

$$\begin{cases} p_k = s_0^{k+1} - s^k \\ q_k = r_0^{k+1} - r^k, \end{cases} \quad \text{with} \quad r^k := \tfrac{1}{\lambda\gamma}(s^k - \bar{s}^k) \in \mathcal{R}_\gamma^{\mathscr{A}}(s^k) \tag{4.4}$$

and similarly $r_0^k \in \mathcal{R}_\gamma^{\mathscr{A}}(s_0^k)$. As it will be clear in the proof of Theorem 4.7, this particular choice of p_k and q_k rather than the conventional $p_k = s^{k+1} - s^k$ and $q_k = r^{k+1} - r^k$, is suited for the proposed innovative linesearch. We will now list a few update rules for H_k based on the indicated pairs.

4.3.1 (L-)BFGS

Start with $H_0 \succ 0$ and update as follows:

$$H_{k+1} = H_k + \frac{\langle p_k, q_k\rangle + \langle H_k q_k, q_k\rangle}{(\langle p_k, q_k\rangle)^2} p_k p_k^\top - \frac{H_k q_k s_k^\top + s_k q_k^\top H_k}{\langle p_k, q_k\rangle}.$$

Whenever $\langle p_k, q_k\rangle \leq 0$, one can either set $H_{k+1} = H_k$ or use a different vector p_k as proposed in [100]. The limited-memory variant L-BFGS [88, Alg. 7.4], which does not require storage of full matrices H_k or matrix-vector products but only storage of the last few pairs and scalar products, can be conveniently considered.

Although very well performing in practice, to the best of our knowledge fast convergence of BFGS can only be shown when the Jacobian at the limit point is symmetric, which hardly ever holds in our framework. We suspect, however, that the good performance of BFGS derives from the observation that, when it exists, the Jacobian of $\mathcal{R}_\gamma^{\mathscr{A}}$ is similar (in the sense of conjugacy) to a symmetric matrix.

4.3.2 A modified Broyden scheme

Fix $\bar\vartheta \in (0,1)$, *e.g.*, $\bar\vartheta = 0.2$, an invertible matrix H_0, and update as follows:

$$H_{k+1} = H_k + \frac{p_k - H_k q_k}{\langle p_k, (1/\vartheta_k - 1)p_k + H_k q_k\rangle} p_k^\top H_k, \tag{4.5a}$$

where

$$\vartheta_k := \begin{cases} 1 & \text{if } |\gamma_k| \geq \bar\vartheta \\ \frac{1-\operatorname{sgn}(\gamma_k)\bar\vartheta}{1-\gamma_k} & \text{if } |\gamma_k| < \bar\vartheta \end{cases} \quad \text{and} \quad \gamma_k := \frac{\langle H_k q^k, p^k\rangle}{\|p^k\|^2}, \tag{4.5b}$$

with the convention that $\operatorname{sgn} 0 = 1$. The original Broyden formula [28] corresponds to $\vartheta_k \equiv 1$, whereas this specific selection ensures that all matrices H_k are invertible [99]. Under some regularity assumptions at the limit point, in Theorem 4.7 this modified Broyden method will be shown to trigger superlinear convergence.

We should remark, however, that extensive numerical evidence seems to agree that BFGS directions as in Section 4.3.1 yield the best performance. In fact, with as little as a five-to-ten-vector buffer, the limited memory L-BFGS is extremely beneficial and requires negligible algebraic operations (few scalar products only per iteration). For the time being, the Broyden scheme investigated here serves only for theoretical purposes. Nevertheless, it will find a practical utility in Chapter 8, where a limited-memory variant will be also proposed.

4.3.3 Anderson acceleration

Fix a buffer size $m \geq 1$ and start with $H_0 = \mathrm{I}$. For $k \geq 1$, let

$$H_k = \mathrm{I} + (\mathcal{P}_k - \mathcal{Q}_k)(\mathcal{Q}_k^\top \mathcal{Q}_k)^{-1} \mathcal{Q}_k^\top,$$

where the columns of matrix $\mathcal{P}_k$ are the last vectors $p_{k-M}, \cdots, p_{k-1}$ and those of $\mathcal{Q}_k$ are the last vectors $q_{k-M}, \cdots, q_{k-1}$, with $M = \min\{k, m\}$. If $\mathcal{Q}_k$ is not

full-column rank, for $x \in \mathbb{R}^M$ the product $(\mathcal{Q}_k^\top \mathcal{Q}_k)^{-1}x$ is meant in a least-square sense. This is a limited-memory scheme, which requires only the storage of few vectors and the solution of a small $M \times M$ linear system. Anderson acceleration originated in [2]; here we use the interpretation well explained in [45] of *(inverse) multi-secant* update: H_k is the matrix closest to the identity (with respect to the Frobenius norm) among those satisfying $H_k \mathcal{Q}_k = \mathcal{P}_k$.

4.4 Global and (super)linear convergence

Theorem 4.2 (Global convergence)**.** *Consider the iterates generated by CLyD (Alg. 4.1) with tolerance $\varepsilon = 0$. Suppose that the following hold:*

- A1 *φ is level bounded;*
- A2 *The $\mathcal{M}$-envelope $\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}$ has the KL property;*
- A3 *there exists $m > 0$ such that $\operatorname{dist}(0, \partial\varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(x)) \leq m \operatorname{dist}(x, \mathcal{T}_\gamma^{\mathscr{A}}(x))$ holds for all x close to $\operatorname{fix}\mathcal{T}_\gamma^{\mathscr{A}}$;*
- A4 *there exists $D > 0$ such that $\|d^k\| \leq D\|s^k - \bar{s}^k\|$ for all k's.*

Then, the following hold:

(i) $(x^k)_{k\in\mathbb{N}}$ converges to a point $x_\star \in \operatorname{fix}\mathcal{T}_\gamma^{\mathscr{A}} \subseteq \operatorname{zer}\hat{\partial}\varphi$ (hence $(s^k)_{k\in\mathbb{N}}$ converges to $G_\gamma^{-1}(x_\star)$).

(ii) The residual is summable and in particular $\min_{j\leq k}\operatorname{dist}(x^k, \mathcal{T}_\gamma^{\mathscr{A}}(x^k)) \in O(1/k)$.

Proof. Let L be a Lipschitz modulus for G_γ and G_γ^{-1}. We have

$$\varphi_\gamma^{\mathscr{A}}(s^k) - \varphi_\gamma^{\mathscr{A}}(s^{k+1}) \geq \alpha\tfrac{c}{2}\|s^k - \bar{s}^k\|^2 \geq \alpha\tfrac{c}{2L^2}\|x^k - \bar{x}^k\|^2,$$

which is exactly the inequality (3.6) in Thm. 3.23 (with a different constant). We may thus trace the proof therein up to equation (3.7) to infer that $(\|x^k - \bar{x}^k\|)_{k\in\mathbb{N}}$ is summable. Moreover,

$$\begin{aligned}\|s^{k+1} - s^k\| &\leq (1-\tau_k)\|s^k - \bar{s}^k\| + \tau_k\|d^k\| \overset{4.2\text{A4}}{\leq} (1+D)\|s^k - \bar{s}^k\| \\ &\leq (1+D)L\|x^k - \bar{x}^k\|,\end{aligned}$$

proving that $(\|s^{k+1}-s^k\|)_{k\in\mathbb{N}}$ is summable too, and thus converges to a point $s_\star$. Because of Lipschitz continuity, also $(x^k)_{k\in\mathbb{N}}$ converges to some point $x_\star$, in fact, to $x_\star = G_\gamma^{-1}(s_\star)$. That $x_\star \in \operatorname{fix}\mathcal{T}_\gamma^{\mathscr{A}} \subseteq \operatorname{zer}\hat{\partial}\varphi$ follows from Thm. 4.1*(ii)*. □

Theorem 4.3 (Linear convergence)**.** *Suppose that the assumptions of Theorem 4.2 are satisfied, and that the KL function can be taken of the form* $\psi(s) = cs^\vartheta$ *for some* $c > 0$ *and* $\vartheta \geq 1/2$*. Then, the sequences* $(s^k)_{k\in\mathbb{N}}$*,* $(x^k)_{k\in\mathbb{N}}$*, and* $\operatorname{dist}(x^k, \mathcal{T}_\gamma^{\mathscr{A}}(x^k))$ *are R-linearly convergent.*

Proof. The proof is exactly the same as that of Thm. 3.24, with the only exception that in (3.10) the inequality $\|s^j - s^{j+1}\| \leq L\|x^j - \bar{x}^j\|$ is to be used, as opposed to the equality $\|s^j - s^{j+1}\| = \lambda\|x^j - \bar{x}^j\|$ therein. □

4.5 Superlinear convergence

In the sequel, we will make use of the notion of SUPERLINEAR DIRECTIONS that we define next.

Definition 4.4 (Superlinear directions [44, §7.5])**.** *We say that* $(d^k)_{k\in\mathbb{N}}$ *are* SUPERLINEARLY CONVERGENT DIRECTIONS *for a sequence* $(s^k)_{k\in\mathbb{N}}$ *converging to a point* $s_\star$ *if*

$$\lim_{k\to\infty} \frac{\|s^k + d^k - s_\star\|}{\|s^k - s_\star\|} = 0.$$

The next result consititutes a key component of the methodology, as it shows that the proposed algorithm does not suffer from the *Maratos effect* [80], a well-known obstacle for fast local methods that inhibits the acceptance of the unit stepsize. On the contrary, we will show that whenever the directions $(d^k)_{k\in\mathbb{N}}$ of CLyD (Alg. 4.1) are superlinear, then indeed the unit stepsize is eventually always accepted, and the algorithm reduces to the (undamped) local method $s^{k+1} = s^k + d^k$, and $(x^k)_{k\in\mathbb{N}}$ then converges superlinearly.

Theorem 4.5 (Acceptance of the unit stepsize and superlinear convergence)**.** *Consider the iterates generated by CLyD (Alg. 4.1). Suppose that the following hold:*

A1 $(x^k)_{k\in\mathbb{N}}$ *converges to a strong local minimum of* φ*;*

A2 $(d^k)_{k\in\mathbb{N}}$ *are superlinearly convergent directions with respect to* $(s^k)_{k\in\mathbb{N}}$*;*

A3 $\gamma \neq \Gamma^{\mathscr{A}}(x_\star)$*;*

Then, there exists $\bar{k} \in \mathbb{N}$ *such that*

$$\varphi_\gamma^{\mathscr{A}}(s^k + d^k) \leq \varphi_\gamma^{\mathscr{A}}(s^k) - \alpha\tfrac{c}{2}\|s^k - \bar{s}^k\|^2 \quad \textit{for all } k \geq \bar{k}.$$

In particular, eventually the iterates reduce to $s^{k+1} = s^k + d^k$ *and converge superlinearly.*

Proof. Let L be a Lipschitz modulus for the transient mapping G_γ and its inverse. We know from Thm. 3.6 and the property 2.31.P1 of the criticality threshold that $s_\star = G_\gamma^{-1}(s_\star)$ is a strong local minimum for $\varphi_\gamma^{\mathscr{A}}$: there exist $\varepsilon, \mu > 0$ such that

$$\varphi_\gamma^{\mathscr{A}}(s) - \varphi_\star \geq \tfrac{\mu}{2}\|s - s_\star\|^2 \quad \forall s \in \mathrm{B}(s_\star; \varepsilon),$$

where $\varphi_\star \coloneqq \varphi_\gamma^{\mathscr{A}}(s_\star) = \varphi(x_\star)$ (the second equality is due to Cor. 3.3*(i)*). Combined with the quadratic upper bound of the $\mathcal{F}$-envelope, see Prop. 3.9, we obtain

$$\frac{\varphi_\gamma^{\mathscr{A}}(s^k + d^k) - \varphi_\star}{\varphi_\gamma^{\mathscr{A}}(s^k) - \varphi_\star} \leq \frac{m_2}{\mu}\frac{\|G_\gamma(s^k + d^k) - x_\star\|^2}{\|x^k - x_\star\|^2} \leq \frac{m_2 L^2}{\mu}\frac{\|s^k + d^k - s_\star\|^2}{\|s^k - s_\star\|^2},$$

holding for all k's sufficiently large. In particular,

$$\varepsilon_k \coloneqq \frac{\varphi_\gamma^{\mathscr{A}}(s^k + d^k) - \varphi_\star}{\varphi_\gamma^{\mathscr{A}}(s^k) - \varphi_\star} \to 0 \quad \text{as } k \to \infty.$$

Thus, for all k's large enough we have that $\varepsilon_k \leq 1 - \alpha$, and also $\varphi_\gamma^{\mathscr{A}}(\bar{s}^k) = \varphi^{\mathcal{M}_\gamma^{\mathscr{A}}}(\bar{x}^k) \geq \varphi_\star$ as ensured by Prop. 3.5. Hence, eventually,

$$\begin{aligned}\varphi_\gamma^{\mathscr{A}}(s^k + d^k) - \varphi_\gamma^{\mathscr{A}}(s^k) &= \big(\varphi_\gamma^{\mathscr{A}}(s^k + d^k) - \varphi_\star\big) - \big(\varphi_\gamma^{\mathscr{A}}(s^k) - \varphi_\star\big) \\ &= (\varepsilon_k - 1)\big(\varphi_\gamma^{\mathscr{A}}(s^k) - \varphi_\gamma^{\mathscr{A}}(\bar{s}^k)\big) \\ &\leq -\alpha\tfrac{c}{2}\|s^k - \bar{s}^k\|^2,\end{aligned}$$

which proves the claim. □

Theorem 4.6 (Dennis-Moré condition)**.** *Consider the iterates generated by CLyD (Alg. 4.1). Suppose that the following hold:*

A1 $(s^k)_{k\in\mathbb{N}}$ *converges to a point* $s_\star$ *at which* $\mathcal{R}_\gamma^{\mathscr{A}}$ *is strictly differentiable and with nonsingular Jacobian* $J\mathcal{R}_\gamma^{\mathscr{A}}(s_\star)$.

A2 *The Dennis-Moré condition holds:*

$$\lim_{k\to\infty} \frac{\|\mathcal{R}_\gamma^{\mathscr{A}}(s^k) + J\mathcal{R}_\gamma^{\mathscr{A}}(s_\star)d^k\|}{\|d^k\|} = 0. \tag{4.6}$$

Then, $(d^k)_{k\in\mathbb{N}}$ *are superlinearly convergent directions with respect to* $(s^k)_{k\in\mathbb{N}}$.

Proof. The Dennis-Moré condition (4.6) implies that

$$0 \leftarrow \frac{\mathcal{R}_\gamma^{\mathscr{A}}(s^k) + J\mathcal{R}_\gamma^{\mathscr{A}}(s_\star)d^k - \mathcal{R}_\gamma^{\mathscr{A}}(s^k + d^k)}{\|d^k\|} + \frac{\mathcal{R}_\gamma^{\mathscr{A}}(s^k + d^k)}{\|d^k\|}$$

as $k \to \infty$. Due to strict differentiability, the first term on the right-hand side vanishes, hence so does the second. Moreover, nonsingularity of $\mathcal{R}_\gamma^{\mathscr{A}}(s_\star)$ implies that there exists $\alpha > 0$ such that

$$\|\mathcal{R}_\gamma^{\mathscr{A}}(s)\| = \|\mathcal{R}_\gamma^{\mathscr{A}}(s) - \mathcal{R}_\gamma^{\mathscr{A}}(s_\star)\| \geq \alpha\|s - s_\star\|$$

holds for all s close to $s_\star$. Here, the first equality is due to the fact that $G_\gamma(s_\star)$ is critical, hence $0 = \mathcal{R}_\gamma^{\mathscr{A}}(s_\star)$ (equality, as opposed to inclusion, holds due to the assumption of differentiability). We thus have

$$\begin{aligned} 0 \leftarrow \frac{\|\mathcal{R}_\gamma^{\mathscr{A}}(s^k + d^k)\|}{\|d^k\|} &\geq \alpha \frac{\|s^k + d^k - s_\star\|}{\|d^k\|} \\ &\geq \alpha \frac{\|s^k + d^k - s_\star\|}{\|s^k + d^k - s_\star\| + \|s^k - s_\star\|} \\ &= \alpha \frac{\frac{\|s^k + d^k - s_\star\|}{\|s^k - s_\star\|}}{1 + \frac{\|s^k + d^k - s_\star\|}{\|s^k - s_\star\|}}, \end{aligned}$$

as $k \to \infty$, and in particular $\frac{\|s^k + d^k - s_\star\|}{\|s^k - s_\star\|} \to 0$. □

We conclude the section showing that employing Broyden directions (4.5) enables superlinear convergence rates, provided that $\mathcal{R}_\gamma^{\mathscr{A}}$ is Lipschitz continuously *semidifferentiable* at the limit point, see [59].

Theorem 4.7 (Superlinear convergence with Broyden directions)**.** *Consider the iterates generated by CLyD (Alg. 4.1) with directions* d^k *selected with the modified Broyden method of Section 4.3.2. Suppose that the following hold:*

A1 $(s^k)_{k\in\mathbb{N}}$ *converges to a point* $s_\star$ *at which* $\mathcal{R}_\gamma^{\mathscr{A}}$ *is Lipschitz-continuously semidifferentiable and with nonsingular Jacobian* $J\mathcal{R}_\gamma^{\mathscr{A}}(s_\star)$ *(in particular,* $\mathcal{R}_\gamma^{\mathscr{A}}$ *is strictly differentiable there).*

A2 *there exists* $m > 0$ *such that* $\operatorname{dist}(0, \partial\varphi_\gamma^{\mathscr{A}}(s)) \leq m \operatorname{dist}(0, \mathcal{R}_\gamma^{\mathscr{A}}(s))$ *holds for all* s *close to* $s_\star$.

Then, the Dennis-Moré condition (4.6) *is satisfied, and in particular all the claims of Theorem 4.6 hold.*

Proof. Denoting $G_\star \coloneqq J\mathcal{R}_\gamma^{\mathscr{A}}(s_\star)$, we have

$$\frac{\|q^k - G_\star p^k\|}{\|p^k\|} = \frac{\|\mathcal{R}_\gamma^{\mathscr{A}}(s^{k+1}) - \mathcal{R}_\gamma^{\mathscr{A}}(s^k) - G_\star(s^{k+1} - s^k)\|}{\|s^{k+1} - s^k\|},$$

and since $s^k \to x^\star$, due to [59, Lem. 2.2] there exists $L > 0$ such that $\frac{\|q^k - G_\star p^k\|}{\|p^k\|} \leq L \max\left\{\|s^{k+1} - x^\star\|, \|s^k - x^\star\|\right\}$ for k large enough. Consequently, due to Thm. 3.24 and Prop. 3.16, $\frac{\|q_k - G_\star p_k\|}{\|p_k\|}$ is summable. Let $B_k \coloneqq H_k^{-1}$ and $E_k \coloneqq B_k - G_\star$, and let $\|\cdot\|_F$ denote the Frobenius norm. With a simple modification of the proofs of [59, Thm. 4.1] and [4, Lem. 4.4] that takes into account the scalar $\vartheta_k \in [\bar{\vartheta}, 2 - \bar{\vartheta}]$, we obtain

$$\begin{aligned}\|E_{k+1}\|_F &\leq \left\|E_k\left(\mathrm{I} - \vartheta_k \tfrac{p_k (p_k)^\top}{\|p_k\|^2}\right)\right\|_F + \vartheta_k \frac{\|q_k - G_\star p_k\|}{\|p_k\|} \\ &\leq \|E_k\|_F - \frac{\bar{\vartheta}(2-\bar{\vartheta})}{2\|E_k\|_F} \frac{\|E_k p_k\|^2}{\|p_k\|^2}.\end{aligned}$$

Consequently, $(\|E_k\|_F)_{k\in\mathbb{N}}$ is decreasing, and in particular its supremum $\bar{E} \coloneqq \sup(\|E_k\|_F)_{k\in\mathbb{N}}$ is finite. By rearranging the inequality above, we obtain

$$\begin{aligned}\frac{\bar{\vartheta}(2-\bar{\vartheta})}{2\bar{E}} \sum_{k\in\mathbb{N}} \frac{\|E_k p_k\|^2}{\|p_k\|^2} &\leq \sum_{k\in\mathbb{N}} \frac{\bar{\vartheta}(2-\bar{\vartheta})}{2\|E_k\|_F} \frac{\|E_k p_k\|^2}{\|p_k\|^2} \\ &\leq \sum_{k\in\mathbb{N}} (\|E_k\|_F - \|E_{k+1}\|_F) \\ &\leq \|E_0\|_F.\end{aligned}$$

Therefore, since $p_k = s_0^{k+1} - s^k = d^k = -H_k \mathcal{R}_\gamma^{\mathscr{A}}(s^k)$, we have that

$$\frac{\|E_k p_k\|}{\|p_k\|} = \frac{\|(B_k - G_\star)p_k\|}{\|p_k\|} = \frac{\|\mathcal{R}_\gamma^{\mathscr{A}}(s^k) + G_\star d_k\|}{\|d_k\|}$$

is square summable, hence the Dennis-Moré condition (4.6). □

Chapter 5

Forward-backward splitting

5.1 Introduction

We now consider composite minimization problems

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad \varphi(x) \equiv f(x) + g(x) \tag{5.1}$$

under the following requirements.

Assumption 5.I (FBS: basic assumption)**.** In problem (5.1)

A1 $f \in C^{1,1}(\mathbb{R}^n)$ is L_f-smooth, hence σ_f-hypoconvex with $|\sigma_f| \le L_f$.

A2 $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper and lsc.

A3 A solution exists, that is, $\arg\min \varphi \neq \emptyset$.

Both f and g are allowed to be nonconvex, making (5.1) prototypic for a plethora of applications spanning signal and image processing, machine learning, statistics, control and system identification. A well-known algorithm addressing (5.1) is *forward-backward splitting* (FBS), also known as *proximal gradient method*, amounting to fixed-point iterations

$$x^+ \in \operatorname{prox}_{\gamma g}(x - \gamma\nabla f(x)), \tag{5.2}$$

where $\gamma > 0$ is a stepsize parameter.

The name *forward-backward splitting* is a loan from monotone operator theory, where given a maximally monotone (set-valued) operator A and a cocoercive (single-valued) operator B, the problem of finding $x \in \operatorname{zer}(A + B)$ is addressed

by interleaving *forward steps* $\mathrm{id} - \gamma B$ and *backward steps* $(\mathrm{id} + \gamma A)^{-1}$ for some stepsize parameter $\gamma > 0$. In fact, when both f and g are convex, problem (5.4) is equivalent to finding $x \in \operatorname{zer} \partial\varphi = \operatorname{zer}(A + B)$, where $A \coloneqq \partial g$ is maximally monotone, and $B \coloneqq \nabla f$ is L_f^{-1}-cocoercive [10, Cor. 18.17 and Thm. 20.25]. In this case, the forward step becomes a gradient descent $\mathrm{id} - \gamma \nabla f$ and the backward step the proximal mapping $\operatorname{prox}_{\gamma g}$ [10, Prop. 16.44], hence the name *proximal gradient method* in optimization. Thanks to this theoretical link, when both f and g are convex, relaxed fixed-point iterations of the (single-valued) forward-backward operator, namely

$$x^{k+1} = (1 - \lambda_k)x^k + \lambda_k \operatorname{prox}_{\gamma g}(x^k - \gamma \nabla f(x^k)),$$

are known to converge to a minimizer of φ for any stepsize $\gamma \in (0, 2/L_f)$ and any choice of relaxation parameters $(\lambda_k)_{k\in\mathbb{N}} \subset \big(0,\, 2 - \gamma/2L_f\big)$ as long as $\sum_{k\in\mathbb{N}} \lambda_k\big(2 - \gamma/2L_f - \lambda_k\big) = \infty$ [10, Cor. 28.9]. Moreover, FBS enjoys a global rate $O(1/k)$ in terms of objective value, and accelerated variants, also known as fast forward-backward splitting (FFBS) or *accelerated* proximal gradient method, can be derived thanks to the work of Nesterov [83, 122, 16, 86], that only require minimal additional computations per iteration but achieve the optimal global convergence rate of order $o(1/k^2)$ [7].

When f and/or g are nonconvex, convergence results can be established by viewing FBS as a (pure) PMM algorithm. As we will see in the next section, this requires reducing the range of the stepsize to $\gamma \in (0, 1/L_f)$. Before that, let us first observe that proximal gradient iterations (5.2) are well defined for any stepsize in such range.

Remark 5.1 (Feasible stepsizes for FBS)**.** Under Assumption 5.I, for all $x \in \mathbb{R}^n$ it holds that

$$\inf \varphi \le f(x) + g(x) \le f(0) + \langle \nabla f(0), x\rangle + \tfrac{L_f}{2}\|x\|^2 + g(x),$$

hence, for all $r > L_f$ the function $x \mapsto g(x) + \frac{r}{2}\|x\|^2$ is lower bounded. It then follows from the definition of prox-boundedness that g is prox-bounded with threshold $\gamma_g \ge 1/L_f$. Proposition 1.12 then ensures that $\operatorname{prox}_{\gamma g}$ is nonempty-valued for any stepsize $\gamma < 1/L_f$, thus forward-backward iterations (5.2) are well defined. □

5.2 FBS as a PMM algorithm

We now show that, when Assumption 5.I is satisfied, FBS is a pure PMM scheme, that is to say, FBS fits into the fixed-point iteration framework (2.6)

Forward-backward splitting FB $\sim (\mathcal{M}_\gamma^{\mathrm{FB}})_{\gamma\in(0,1/L_f)}$		
$\bar{\gamma}$:	$1/L_f$	(ensures $m_1 > 0$)
$\mathcal{M}_\gamma^{\mathrm{FB}}(w;x)$:	$g(w) + f(x) + \langle \nabla f(x), w - x\rangle + \frac{1}{2\gamma}\|w-x\|^2$	(5.4)
$\mathcal{T}_\gamma^{\mathrm{FB}}(x)$:	$\operatorname{prox}_{\gamma g}(x - \gamma\nabla f(x))$	
$\mathcal{R}_\gamma^{\mathrm{FB}}(x)$:	$\frac{1}{\gamma}(x - \mathcal{T}_\gamma^{\mathrm{FB}}(x))$	(2.9)
$\varphi_\gamma^{\mathrm{FB}}(x)$:	$f(x) - \frac{\gamma}{2}\|\nabla f(x)\|^2 + g^\gamma(x - \gamma\nabla f(x))$	(5.7)
$m_1(\gamma)$:	$\frac{1-\gamma L_f}{\gamma}$	(5.5)
$m_2(\gamma)$:	$\frac{1-\gamma\sigma_f}{\gamma}$	(5.5)
L_δ :	$\frac{1-\gamma\sigma_f}{\gamma}$	Lem. 2.15

Table 5.1: *FBS with stepsize $\gamma \in (0, 1/L_f)$: parameters of the proximal model. The $\mathcal{F}$-envelope, denoted as $\varphi_\gamma^{\mathrm{FB}}$, is the forward-backward envelope function.*

with $G_\gamma \equiv \mathrm{id}$. To facilitate the reading, all the elements and the parameters are summarized in Table 5.1.

Let us consider one iteration $x \mapsto x^+$ with stepsize $\gamma < 1/L_f$. We have

$$\begin{aligned} x^+ &\in \operatorname{prox}_{\gamma g}(x - \gamma\nabla f(x)) \\ &= \operatorname*{arg\,min}_{w\in\mathbb{R}^n} \left\{ g(w) + \tfrac{1}{2\gamma}\|w - x + \gamma\nabla f(x)\|^2 \right\} \qquad (5.3) \\ &= \operatorname*{arg\,min}_{w\in\mathbb{R}^n} \left\{ g(w) + \langle \nabla f(x), w - x\rangle + \tfrac{1}{2\gamma}\|w-x\|^2 + \tfrac{\gamma}{2}\|\nabla f(x)\|^2 \right\}. \end{aligned}$$

By adding the constant quantity $f(x) - \frac{\gamma}{2}\|\nabla f(x)\|^2$ into the function being minimized, we obtain that $x^+ \in \operatorname{arg\,min}_{w\in\mathbb{R}^n} \mathcal{M}_\gamma^{\mathrm{FB}}(w;x)$, where

$$\mathcal{M}_\gamma^{\mathrm{FB}}(w;x) := g(w) + f(x) + \langle \nabla f(x), w - x\rangle + \tfrac{1}{2\gamma}\|w-x\|^2. \qquad (5.4)$$

It follows from the quadratic bound (1.5) that $\mathcal{M}_\gamma^{\mathrm{FB}}$ is a proximal model for φ with constants as in property 2.10.P2 given by

$$m_1 = \tfrac{1-\gamma L_f}{\gamma} \quad \text{and} \quad m_2 = \tfrac{1-\gamma\sigma_f}{\gamma}. \qquad (5.5)$$

Lemma 5.2. *Suppose that Assumption 5.I is satisfied. Then, the difference*

$$\delta(w) := \mathcal{M}_\gamma^{\mathrm{FB}}(w;x) - \varphi(w) = f(x) + \langle \nabla f(x), w - x\rangle - f(w) + \tfrac{1}{2\gamma}\|w-x\|^2$$

is L_δ-Lipschitz differentiable with $L_\delta = \frac{1-\gamma\sigma_f}{\gamma}$ for any $x \in \mathbb{R}^n$. In particular, for any $\bar{x} \in \mathcal{T}_\gamma^{\mathrm{FB}}(x)$ it holds that

$$\operatorname{dist}(0, \hat{\partial}\varphi(\bar{x})) \le \tfrac{1-\gamma\sigma_f}{\gamma}\|x - \bar{x}\|.$$

Proof. Since $\nabla\delta(w) = \frac{1}{\gamma}(w - \gamma\nabla f(w)) - \frac{1}{\gamma}(x - \gamma\nabla f(x))$, for all $w, w' \in \mathbb{R}^n$ it holds that

$$\langle \nabla\delta(w) - \nabla\delta(w'), w - w'\rangle = \tfrac{1}{\gamma}\|w-w'\|^2 - \langle \nabla f(w) - \nabla f(w'), w - w'\rangle.$$

From (1.5) we then obtain that the scalar product is bounded as

$$\tfrac{1-\gamma L_f}{\gamma}\|w-w'\|^2 \le \langle \nabla\delta(w) - \nabla\delta(w'), w - w'\rangle \le \tfrac{1-\gamma\sigma_f}{\gamma}\|w-w'\|^2,$$

hence the claimed Lipschitz continuity. The rest of the proof then follows from Lem. 2.15. □

5.3 Forward-backward envelope

Consistently with Definition 3.1 and since FBS is a pure PMM scheme (the transient function is $G_\gamma = \mathrm{id}$), we have

$$\begin{aligned}\varphi_\gamma^{\mathrm{FB}}(x) &\overset{(def)}{=} \min_{w\in\mathbb{R}^n} \mathcal{M}_\gamma^{\mathrm{FB}}(w;x) \\ &= \min_{w\in\mathbb{R}^n} \left\{ g(w) + f(x) + \langle \nabla f(x), w - x\rangle + \tfrac{1}{2\gamma}\|w-x\|^2 \right\}. \end{aligned} \tag{5.6}$$

We name this function FORWARD-BACKWARD ENVELOPE (FBE). Since the minimum in (5.3) is, by definition, $g^\gamma(x - \gamma\nabla f(x))$, and in passing to (5.4) the term $f(x) - \frac{\gamma}{2}\|\nabla f(x)\|^2$ was added, we easily infer the following alternative expression of the FBE in terms of the Moreau envelope of g:

$$\varphi_\gamma^{\mathrm{FB}}(x) = f(x) - \tfrac{\gamma}{2}\|\nabla f(x)\|^2 + g^\gamma(x - \gamma\nabla f(x)). \tag{5.7}$$

The FBE was first introduced in [92] for convex problems with f twice continuously differentiable, and later generalized in [113] by discarding the convexity

assumption of f. Under these assumptions the FBE was shown to be continuously differentiable, see [112, 94] for further details and more differentiabilty properties. The more general analysis dealt here was investigated in [120]; all the results are special cases of the unified analysis provided in Chapter 3.

Theorem 5.3 (FBE: sandwich property)**.** *Suppose that Assumption 5.I is satisfied and let $\gamma \in (0, 1/L_f)$ be fixed. For all $x \in \mathbb{R}^n$ the following hold:*

(i) $\varphi_\gamma^{\mathrm{FB}}(x) \leq \varphi(x)$, *with equality holding iff* $x \in \mathcal{T}_\gamma^{\mathrm{FB}}(x)$.

(ii) $\frac{1-\gamma L_f}{2\gamma}\|x-\bar{x}\|^2 \leq \varphi_\gamma^{\mathrm{FB}}(x) - \varphi(\bar{x}) \leq \frac{1-\gamma\sigma_f}{2\gamma}\|x-\bar{x}\|^2$ *for all* $\bar{x} \in \mathcal{T}_\gamma^{\mathrm{FB}}(x)$.

(iii) $\inf \varphi_\gamma^{\mathrm{FB}} = \inf \varphi$ *and* $\operatorname{arg\,min} \varphi_\gamma^{\mathrm{FB}} = \operatorname{arg\,min} \varphi$.

Proof. Follows from Theorem 3.2*(ii)*, since $m_1(\gamma) = \frac{1-\gamma L_f}{2\gamma}$ and $m_2(\gamma) = \frac{1-\gamma\sigma_f}{2\gamma}$, cf. Table 5.1. □

Theorem 5.4 (FBE: equivalence of local minimality)**.** *Suppose that Assumption 5.I is satisfied, and let $\gamma \in (0, 1/L_f)$ and $\bar{x} \in \operatorname{fix} \mathcal{T}_\gamma^{\mathrm{FB}}$ be fixed. The following hold:*

(i) If $\bar{x}$ is a (strong) local minimum for $\varphi_\gamma^{\mathrm{FB}}$, then it is a (strong) local minimum for φ.

(ii) If $\mathcal{T}_\gamma^{\mathrm{FB}}(\bar{x}) = \{\bar{x}\}$, (e.g., if $\gamma < \Gamma^{\mathrm{FB}}(\bar{x})$) then the converse also holds.

Proof. See Theorem 3.6. □

Theorem 5.5 (FBE: Equivalence of level boundedness)**.** *Suppose that Assumption 5.I is satisfied. For any $\gamma \in (0, 1/L_f)$, φ is level bounded iff $\varphi_\gamma^{\mathrm{FB}}$ is level bounded.*

Proof. See Theorem 3.7. □

5.3.1 Regularity properties

Since f, ∇f, and g^γ are strictly continuous, the following regularity property of the FBE is immediately deduced from the expression (5.7).

Proposition 5.6 (Strict continuity of the FBE)**.** *Suppose that Assumption 5.I is satisfied. For any $\gamma \in (0, 1/L_f)$ the FBE $\varphi_\gamma^{\mathrm{FB}}$ is a strictly continuous function.*

Notice that (nonstrict) continuity of $\varphi_\gamma^{\mathrm{FB}}$ could directly be inferred from Proposition 3.8, being the sections $x \mapsto \mathcal{M}_\gamma^{\mathrm{FB}}(w; x)$ continuous for any $w \in \mathbb{R}^n$.

Proposition 5.7. *Suppose that Assumption 5.1 is satisfied. If $\bar{x}$ is critical, then for all $\gamma \in (0, \Gamma^{\mathrm{FB}}(\bar{x}))$ the Moreau envelope g^γ is strictly differentiable at $\bar{x} - \gamma\nabla f(\bar{x})$ with $\nabla g^\gamma(\bar{x} - \gamma\nabla f(\bar{x})) = -\nabla f(\bar{x})$.*

In particular, if f is (strictly) twice differentiable at $\bar{x}$, then $\varphi_\gamma^{\mathrm{FB}}$ is (strictly) differentiable at $\bar{x}$ with $\nabla\varphi_\gamma^{\mathrm{FB}}(\bar{x}) = 0$.

Proof. It follows from Prop. 1.12 that g^γ is strictly continuous with

$$\partial g^\gamma\big(\bar{x} - \gamma\nabla f(\bar{x})\big) \subseteq \tfrac{1}{\gamma}\big[\bar{x} - \gamma\nabla f(\bar{x}) - \mathcal{T}_\gamma^{\mathrm{FB}}(\bar{x})\big] \overset{2.31.\mathrm{P1}}{=} \{-\nabla f(\bar{x})\},$$

and the claim on g^γ then follows by invoking Lem. 1.3*(iv)*. The last part follows from the chain rule of differentiation. □

5.3.2 First-order differentiability

In the favorable case in which g is convex and $f \in C^2(\mathbb{R}^n)$, the FBE enjoys global continuous differentiability [113]. In our setting, PROX-REGULARITY acts as a surrogate of convexity; the interested reader is referred to [106, §13.F] for a detailed discussion.

Definition 5.8 (Prox-regularity)**.** *Function g is said to be* PROX-REGULAR *at x_0 for $v_0 \in \partial g(x_0)$ if there exist $\rho, \varepsilon > 0$ such that for all $x' \in \mathrm{B}(x_0; \varepsilon)$ and*

$$(x, v) \in \operatorname{gph}\partial g \quad \text{s.t.} \quad x \in \mathrm{B}(x_0;\varepsilon),\ v \in \mathrm{B}(v_0;\varepsilon),\ \text{and } g(x) \le g(x_0) + \varepsilon$$

it holds that $g(x') \ge g(x) + \langle v, x' - x\rangle - \frac{\rho}{2}\|x' - x\|^2$.

Lemma 5.9 ([106, Ex. 13.35])**.** *Function g is prox-regular at $x_\star$ for $\bar{v}$ iff $g - \langle \bar{v}, \cdot\,\rangle$ is prox-regular at $\bar{x}$ for 0.*

To help better visualize this definition, let us consider the local geometrical property that it entails on the function's epigraph [98, Cor. 3.4, Thm. 3.5]. If g is prox-regular at x_0 for v_0 for some constants $\varepsilon, \rho > 0$ as in Definition 5.8, then there exists a neighborhood of $(x_0 + {}^{v_0}\!/_\rho, g(x_0) - {}^1\!/_\rho)$ in which the projection on $\operatorname{epi} g \cap \big(\mathrm{B}(x_0;\varepsilon) \times \mathrm{B}(v_0;\varepsilon)\big)$ is single valued.

Prox-regularity is a mild requirement enjoyed globally and for any subgradient by all convex functions, with $\varepsilon = +\infty$ and $\rho = 0$. When g is prox-regular at x_0 for v_0, then for sufficiently small $\gamma > 0$ the Moreau envelope g^γ is continuously differentiable in a neighborhood of $x_0 + \gamma v_0$ [98]. To our purposes, when needed, prox-regularity of g will be required only at critical points $x_\star$, and only for the

subgradient $-\nabla f(x_\star)$. Therefore, with a slight abuse of terminology we define prox-regularity *of* critical points as follows.

Definition 5.10 (Prox-regularity *of* critical points)**.** *We say that a critical point $x_\star$ is* PROX-REGULAR *if g is prox-regular at $x_\star$ for $-\nabla f(x_\star)$.*

Clearly, if g is convex then any critical point is prox-regular. Prox-regularity *of* critical points is a mild requirement, also considering that the fact of being critical itself entails some regularity properties as shown in Proposition 5.7. We now prove an important result that connects prox-regularity with first-order properties of the FBE.

Theorem 5.11 (Continuous differentiability of $\varphi_\gamma^{\mathrm{FB}}$ and error bound)**.** *Suppose that Assumption 5.1 is satisfied, and let $x_\star$ be a prox-regular critical point. Then, for all $\gamma \in (0, \Gamma^{\mathrm{FB}}(x_\star))$ there exists a neighborhood $U_{x_\star}$ of $x_\star$ on which the following properties hold:*

(i) *T_γ^{FB} and $\mathcal{R}_\gamma^{\mathrm{FB}}$ are strictly continuous, and in particular single-valued.*

(ii) *$\operatorname{dist}(0, \partial\varphi_\gamma^{\mathrm{FB}}(x)) \le (1-\gamma\sigma_f)\|\mathcal{R}_\gamma^{\mathrm{FB}}(x)\|$; in fact, it suffices to have $\mathcal{R}_\gamma^{\mathrm{FB}}(x)$ single-valued for this to hold.*

(iii) *If f is of class C^2 (resp. C^{2+}) around $x_\star$, then $\varphi_\gamma^{\mathrm{FB}} \in C^1$ (resp. $\varphi_\gamma^{\mathrm{FB}} \in C^{1+}$) with $\nabla\varphi_\gamma^{\mathrm{FB}} = [\mathrm{I} - \gamma\nabla^2 f]\mathcal{R}_\gamma^{\mathrm{FB}}$.*

Proof. Due to property 2.31.P1 of the criticality threshold, we have that $\mathcal{M}_\gamma^{\mathrm{FB}}(x_\star; x_\star) < \mathcal{M}_\gamma^{\mathrm{FB}}(x; x_\star)$ whenever $x \neq x_\star$. Expanding as in (5.4), the inequality reduces to

$$g(x) > g(x_\star) - \langle \nabla f(x_\star), x - x_\star\rangle - \tfrac{1}{2\gamma}\|x - x_\star\|^2 \quad \forall x \in \mathbb{R}^n \setminus \{x_\star\}. \tag{5.8}$$

From [98, Thm. 4.4] applied to the "tilted" function $x \mapsto g(x + x_\star) - g(x_\star) - \langle \nabla f(x_\star), x\rangle$ and in light of Lem. 5.9, it follows that there is a neighborhood V of $x_\star - \gamma\nabla f(x_\star)$ in which $\operatorname{prox}_{\gamma g}$ is strictly continuous and g^γ is of class C^{1+} with

$$\nabla g^\gamma(x) = \gamma^{-1}\big(x - \operatorname{prox}_{\gamma g}(x)\big) \quad \forall x \in V.$$

♠ 5.11*(i)*. Follows from the fact that strict continuity is preserved under composition.

♠ 5.11*(ii)*. Since ∇f is strictly continuous, it is differentiable on a set D with negligible complement in $\mathbb{R}^n$. Due to the chain rule of differentiation and the fact that $g^\gamma \in C^1$, $\varphi_\gamma^{\mathrm{FB}}$ is differentiable on D, with

$$\nabla\varphi_\gamma^{\mathrm{FB}}(w) = \big(\mathrm{I} - \gamma\nabla^2 f(w)\big)\big(\nabla f(w) + \nabla g^\gamma(w - \gamma\nabla f(w))\big)$$

$$= \big(\mathrm{I} - \gamma\nabla^2 f(w)\big)\Big(\nabla f(w) + \tfrac{1}{\gamma}\big(w - \gamma\nabla f(w) - \mathcal{T}_\gamma^{\mathrm{FB}}(w)\big)\Big)$$

$$= \big(\mathrm{I} - \gamma\nabla^2 f(w)\big)\mathcal{R}_\gamma^{\mathrm{FB}}(w) \tag{5.9}$$

for all $w \in D$, where the second equality follows from Prop. 1.12*(iv)* and the chain rule of differentiation. We may then invoke [106, Thm. 9.61] to infer that

$$\begin{aligned}\partial\varphi_\gamma^{\mathrm{FB}}(x) &\supseteq \partial_B\varphi_\gamma^{\mathrm{FB}}(x)\\ &= \big\{v \in \mathbb{R}^n \mid \exists (x^k)_{k\in\mathbb{N}} \subset D:\ x^k \to x,\ \nabla\varphi_\gamma^{\mathrm{FB}}(x^k) \to v\big\}\\ &= \limsup_{D\ni w\to x} \big(\mathrm{I} - \gamma\nabla^2 f(w)\big)\mathcal{R}_\gamma^{\mathrm{FB}}(w).\end{aligned}$$

Therefore,

$$\begin{aligned}\operatorname{dist}\big(0,\, \partial\varphi_\gamma^{\mathrm{FB}}(x)\big) &\le \operatorname{dist}\Big(0,\, \limsup_{D\ni w\to x} \big(\mathrm{I} - \gamma\nabla^2 f(w)\big)\mathcal{R}_\gamma^{\mathrm{FB}}(w)\Big)\\ &\le (1-\gamma\sigma_f)\operatorname{dist}\Big(0,\, \limsup_{D\ni w\to x} \mathcal{R}_\gamma^{\mathrm{FB}}(w)\Big)\\ &= (1-\gamma\sigma_f)\operatorname{dist}\big(0,\, \mathcal{R}_\gamma^{\mathrm{FB}}(x)\big),\end{aligned} \tag{5.10}$$

where the last equality follows from osc and single valuedness of $\mathcal{R}_\gamma^{\mathrm{FB}}$ at x.

♠ 5.11*(iii)*. If f is C^2 around $x_\star$ and ∇f is continuous, by possibly narrowing $U_{x_\star}$ we may assume that $f \in C^2(U_{x_\star})$ and $x - \gamma\nabla f(x) \in V$ for all $x \in U_{x_\star}$. The claimed expression for $\nabla\varphi_\gamma^{\mathrm{FB}}$ follows from (5.9). □

Since prox-regularity is enjoyed globally by convex functions, the following special case is a straightforward consequence.

Corollary 5.12 (First-order properties for convex g)**.** *Additionally to Assumption 5.1, suppose that g is convex and that $f \in C^2(\mathbb{R}^n)$ (resp. $f \in C^{2+}(\mathbb{R}^n)$). Then, for all $\gamma > 0$ all the properties in Theorem 5.11 hold globally (i.e., for all $x_\star \in \mathbb{R}^n$ with $U_{x_\star} = \mathbb{R}^n$).*

When $f = 0$, Theorem 5.11 restates the known fact that if g is prox-regular at $x_\star$ for $0 \in \partial g(x_\star)$, then g^γ is continuously differentiable around $x_\star$ with $\nabla g^\gamma(x) = \frac{1}{\gamma}(x - \operatorname{prox}_{\gamma g}(x))$. Notice that the bound $\gamma < \Gamma^{\mathrm{FB}}(x_\star)$ is tight: in general, for $\gamma = \Gamma^{\mathrm{FB}}(x_\star)$ no continuity of $\mathcal{T}_\gamma^{\mathrm{FB}}$ nor continuous differentiability of $\varphi_\gamma^{\mathrm{FB}}$ around $x_\star$ can be guaranteed. In fact, even when $x_\star$ is $\Gamma^{\mathrm{FB}}(x_\star)$-critical, $\mathcal{T}_\gamma^{\mathrm{FB}}$ might even fail to be single-valued and $\varphi_\gamma^{\mathrm{FB}}$ differentiable at $x_\star$. To see this, let us consider once again function φ as in Section 3.2.2. This time, simply for the

sake of generalizing the analysis, let us decompose it as $\varphi = f + g$, where

$$f(x) = \tfrac{1}{2}x^2 \quad \text{and} \quad g(x) = \delta_{\{0,1\}}(x).$$

Clearly, f is L_f-smooth and σ_f hypoconvex with $L_f = \sigma_f = 1$, and the FB operator is

$$\mathcal{T}_\gamma^{\mathrm{FB}}(x) = \Pi_{\{0,1\}}((1-\gamma)x).$$

In particular,

$$\mathcal{T}_\gamma^{\mathrm{FB}}(1) = \Pi_{\{0,1\}}(1-\gamma) = \begin{cases} 1 & \text{if } \gamma < 1/2, \\ \{0,1\} & \text{if } \gamma = 1/2, \\ 0 & \text{otherwise,} \end{cases}$$

which indicates that $\bar{x} = 1$ has FB-criticality threshold $\Gamma^{\mathrm{FB}}(\bar{x}) = 1/2$.

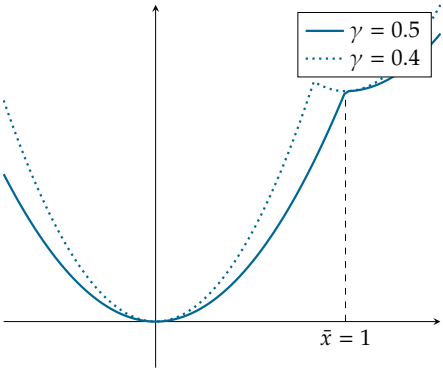


Figure 5.1: *Around prox-regular critical points the FBE $\varphi_\gamma^{\mathrm{FB}}$ is continuously differentiable, provided that the stepsize γ is smaller than the criticality threshold.*

From the expression (5.7) we can write the FBE as

$$\varphi_\gamma^{\mathrm{FB}}(x) = \tfrac{1-\gamma}{2}\|x\|^2 + \tfrac{1}{2\gamma}\operatorname{dist}((1-\gamma)x, S)^2.$$

At the critical point $x = 1$, which satisfies $\Gamma^{\mathrm{FB}}(1) = 1/2$, g is prox-regular for any subgradient. For any $\gamma \in (0, 1/2)$ it is easy to see that $\varphi_\gamma^{\mathrm{FB}}$ is differentiable in a neighborhood of $x = 1$. However, for $\gamma = 1/2$ the distance function has a first-order singularity in $x = 1$, due to the 2-valuedness of $\mathcal{T}_\gamma^{\mathrm{FB}}(1) = \Pi_S(1/2) = \{0,1\}$. As shown in Figure 5.1, the scenario is much similar to what observed in Figure 3.1 for the case of the proximal point algorithm (*i.e.*, with $f = 0$ in the decomposition of φ).

The next example depicts a different kind of pathological situation, namely the lack of prox-regularity at critical points.

Example 5.13 (Prox-nonregularity of critical points)**.** Consider $\varphi = f+g$ where $f(x) = \frac{1}{2}x^2$, $g(x) = \delta_S(x)$ and $S = \{1/n \mid n \in \mathbb{N}_{\geq 1}\} \cup \{0\}$. For $x_0 = 0$ we have

$\Gamma^{\text{FB}}(x_0) = +\infty$, however g fails to be prox-regular at x_0 for $v_0 = 0 = -\nabla f(x_0)$. For any $\rho > 0$ and for any neighborhood V of $(0,0)$ in $\operatorname{gph} g$ it is always possible to find a point arbitrarily close to $(0, -1/\rho)$ with multi-valued projection on V. Specifically, the midpoint $P_n = \big(\frac{1}{2}(\frac{1}{n} + \frac{1}{n+1}),\ -1/\rho\big)$ has a 2-valued projection on $\operatorname{gph} g$ for any $n \in \mathbb{N}_{\geq 1}$, being it $\Pi_{\operatorname{gph} g}(P_n) = \{1/n, 1/n+1\}$. By considering a large n, P_n can be made arbitrarily close to $(0, -1/\rho)$ and at the same time its projection(s) arbitrarily close to $(0,0)$. It follows that g cannot be prox-regular at 0 for 0, for otherwise such projections would be single-valued close enough to $(0,0)$ [98, Cor. 3.4 and Thm. 3.5]. As a result, $g^\gamma(x) = \frac{1}{2\gamma}\operatorname{dist}(x,S)^2$ is not differentiable around $x = 0$, and indeed at each midpoint $\frac{1}{2}(\frac{1}{n} + \frac{1}{n+1})$ for $n \in \mathbb{N}_{\geq 1}$ it has a nonsmooth spike. □

To underline how unfortunate the situation depicted in Example 5.13 is, notice that adding a linear term λx to f for any $\lambda \neq 0$, yet leaving g unchanged, restores the desired prox-regularity of each critical point. Indeed, this is trivially true for any nonzero critical point; besides, g is prox-regular at 0 for any $\lambda > 0$, while for any $\lambda < 0$ the point 0 is not critical. The reason why prox-regularity fails to hold in the above example is due to the density of isolated points close to 0.

5.3.3 Second-order differentiability

In this section we discuss sufficient conditions for twice-differentiability of the FBE at critical points. Additionally to prox-regularity, which is needed for local continuous differentiability, we will also need generalized second-order properties of g. The interested reader is referred to [106, §13] for an extensive discussion on *epi-differentiability*.

Assumption 5.II. With respect to a given critical point $x_\star$

(i) $\nabla^2 f$ exists and is (strictly) continuous around $x_\star$;

(ii) g is prox-regular and (strictly) twice epi-differentiable at $x_\star$ for $-\nabla f(x_\star)$, with its second order epi-derivative being generalized quadratic:

$$\mathrm{d}^2 g(x_\star|-\nabla f(x_\star))[d] = \langle d, Md\rangle + \delta_S(d), \quad \forall d \in \mathbb{R}^n \tag{5.11}$$

where $S \subseteq \mathbb{R}^n$ is a linear subspace and $M \in \mathbb{R}^{n\times n}$. Without loss of generality we take M symmetric, and such that $\operatorname{range}(M) \subseteq S$ and $\ker(M) \supseteq S^\perp$.[1]

[1] This can indeed be done without loss of generality: if M and S satisfy (5.11), then it suffices to replace M with $M' = \frac{1}{2}\Pi_S(M + M^\top)\Pi_S$ to ensure the desired properties.

We say that the assumptions are "strictly" satisfied if the stronger conditions in parenthesis hold.

Twice epi-differentiability of g is a mild requirement, and cases where $\mathrm{d}^2 g$ is generalized quadratic are abundant [104, 105, 95, 96]. Moreover, prox-regular and C^2-partly smooth functions g (see [68, 36]) comprise a wide class of functions that strictly satisfy Assumption 5.II*(ii)* at a critical point $x_\star$ provided that *strict complementarity* holds, namely if $-\nabla f(x_\star) \in \operatorname{relint} \partial g(x_\star)$. In fact, it follows from [36, Thm. 28] applied to the *tilted* function $\tilde{g} = g + \langle \nabla f(x_\star), \cdot \rangle$ (which is still C^2-partly smooth and prox-regular at $x_\star$, cf. [68, Cor. 4.6] and Lem. 5.9) that $\operatorname{prox}_{\gamma\tilde{g}}$ is continuously differentiable around $x_\star$ for γ small enough (in fact, for $\gamma < \Gamma^{\mathrm{FB}}(x_\star)$). From [97, Thm 4.1(g)] we then obtain that $\tilde{g}$ is strictly twice epi-differentiable at $x_\star$ with generalized quadratic second-order epiderivative, and the claim follows by *tilting* back to g.

We now show that the properties required in Assumption 5.II are all that is needed for ensuring first-order properties of the proximal mapping and second-order properties of the FBE at critical points. The result is more general than the one in [113], as here g is allowed to be nonconvex.

Theorem 5.14 (Twice differentiability of $\varphi_\gamma^{\mathrm{FB}}$)**.** *Additionally to Assumption 5.I, suppose that Assumption 5.II is (strictly) satisfied with respect to a critical point $x_\star$. Then, for any $\gamma \in (0, \Gamma^{\mathrm{FB}}(x_\star))$*

(i) $\operatorname{prox}_{\gamma g}$ *is (strictly) differentiable at $x_\star - \gamma\nabla f(x_\star)$ with symmetric and positive semidefinite Jacobian*

$$P_\gamma(x_\star) \coloneqq J \operatorname{prox}_{\gamma g}(x_\star - \gamma\nabla f(x_\star)); \tag{5.12}$$

(ii) $\mathcal{R}_\gamma^{\mathrm{FB}}$ *is (strictly) differentiable at $x_\star$ with Jacobian*

$$J\mathcal{R}_\gamma^{\mathrm{FB}}(x_\star) = \tfrac{1}{\gamma}[\mathrm{I} - P_\gamma(x_\star)Q_\gamma(x_\star)], \tag{5.13}$$

where $Q_\gamma \coloneqq \mathrm{I} - \gamma\nabla^2 f$;

(iii) $\varphi_\gamma^{\mathrm{FB}}$ *is (strictly) twice differentiable at $x_\star$ with symmetric Hessian*

$$\nabla^2\varphi_\gamma^{\mathrm{FB}}(x_\star) = Q_\gamma(x_\star)J\mathcal{R}_\gamma^{\mathrm{FB}}(x_\star). \tag{5.14}$$

Proof.

♠ 5.14*(i)*. It follows from [97, Thm.s 3.8 and 4.1] that $\operatorname{prox}_{\gamma g}$ is (strictly) differentiable at $x^\star - \gamma\nabla f(x^\star)$ iff g (strictly) satisfies assumption 5.II*(ii)*. Consequently, if f is of class C^2 around $x^\star$ (and in particular strictly differentiable at

$x^\star$ [106, Cor. 9.19]), $\mathcal{R}^{\text{FB}}_\gamma(x) = \frac{1}{\gamma}\big(x - \operatorname{prox}_{\gamma g}(x - \gamma\nabla f(x))\big)$ is (strictly) differentiable at $x^\star$ with Jacobian as in (5.13) due to the chain rule of differentiation (and the fact that strict differentiability is preserved by composition). For $\gamma' \in (\gamma, \Gamma^{\text{FB}}(x^\star))$ and $w \in \mathbb{R}^n$ we have

$$\mathrm{d}^2 g(x^\star|-\nabla f(x^\star))[w] = \liminf_{\substack{w'\to w\\ \tau\to 0^+}} \frac{g(x^\star+\tau w') - g(x^\star) + \tau\langle \nabla f(x^\star), w'\rangle}{\tau^2/2}$$

$$\text{(due to (5.8))} \geq -\tfrac{1}{\gamma'}\|w\|^2.$$

The expression (5.11) of the second-order epi-derivative then implies $\langle Mw, w\rangle \geq -\frac{1}{\gamma'}\|w\|^2$ for all $w \in \mathbb{R}^n$ (since $Mw = 0$ for $w \in S^\perp$). Therefore, $\lambda_{\min}(M) \geq -1/\gamma' > -1/\gamma$, proving $I + \gamma M$ to be positive definite, and in particular invertible. To obtain an expression for $P_\gamma(x^\star) = J\operatorname{prox}_{\gamma g}(x^\star - \gamma\nabla f(x^\star))$ we can apply [106, Ex. 13.45] to the function $g + \langle \nabla f(x^\star), \cdot\,\rangle$ so that, letting $\mathrm{d}^2 g = \mathrm{d}^2 g(x^\star|-\nabla f(x^\star))[\,\cdot\,]$ and Π_S the idempotent and symmetric projection matrix on S,

$$\begin{aligned} P_\gamma(x^\star)d &= \operatorname{prox}_{(\gamma/2)\mathrm{d}^2 g}(d) = \operatorname*{arg\,min}_{d'\in S}\left\{\tfrac{1}{2}\langle d', Md'\rangle + \tfrac{1}{2\gamma}\|d'-d\|^2\right\} \\ &= \Pi_S \operatorname*{arg\,min}_{d'\in\mathbb{R}^n}\left\{\tfrac{1}{2}\langle \Pi_S d', M\Pi_S d'\rangle + \tfrac{1}{2\gamma}\|\Pi_S d' - d\|^2\right\} \\ &= \Pi_S\left(\Pi_S[I+\gamma M]\Pi_S\right)^\dagger \Pi_S d \\ &= \Pi_S[I+\gamma M]^{-1}\Pi_S d, \end{aligned} \tag{5.15}$$

where † indicates the pseudo-inverse, and last equality is due to [18, Facts 6.4.12(i)-(ii) and 6.1.6(xxxii)]. Apparently, $JP_\gamma(x^\star)$ is symmetric and positive semidefinite.

♠ 5.14*(ii)*. With basic calculus rules it can be easily verified that, since $\mathcal{R}^{\text{FB}}_\gamma(x^\star) = 0$, $\nabla\varphi^{\text{FB}}_\gamma = Q_\gamma\mathcal{R}^{\text{FB}}_\gamma$ is (strictly) differentiable at $x^\star$ provided that Q_γ is (strictly) continuous at $x^\star$ and $\mathcal{R}^{\text{FB}}_\gamma$ is (strictly) differentiable at $x^\star$.

♠ 5.14*(iii)*. A simple application of the chain rule proves (5.14); moreover, combined with (5.13) we obtain

$$\nabla^2\varphi^{\text{FB}}_\gamma(x^\star) = \tfrac{1}{\gamma}[Q_\gamma(x^\star) - Q_\gamma(x^\star)P_\gamma(x^\star)Q_\gamma(x^\star)],$$

and since both $Q_\gamma(x^\star)$ and $P_\gamma(x^\star)$ are symmetric, so is $\nabla^2\varphi(x^\star)$. □

Again, when $f \equiv 0$ Theorem 5.14 covers the differentiability properties of the

proximal mapping (and consequently the second-order properties of the Moreau envelope, due to the identity $\nabla g^\gamma(x) = \frac{1}{\gamma}(x - \operatorname{prox}_{\gamma g}(x))$) as discussed in [97].

We now provide a key result that links nonsingularity of the Jacobian of the forward-backward residual $\mathcal{R}_\gamma^{\mathrm{FB}}$ to strong (local) minimality for the original cost φ and for the FBE $\varphi_\gamma^{\mathrm{FB}}$, under the generalized second-order properties of Assumption 5.II.

Theorem 5.15 (Conditions for strong local minimality)**.** *Additionally to Assumption 5.I, suppose that Assumption 5.II is satisfied with respect to a critical point $x_\star$, and let $\gamma \in (0, \min\{\Gamma^{\mathrm{FB}}(x_\star), 1/L_f\})$. The following are equivalent:*

(a) $x_\star$ is a strong local minimum for φ.

(b) $x_\star$ is a local minimum for φ and $J\mathcal{R}_\gamma^{\mathrm{FB}}(x_\star)$ is nonsingular.

(c) the (symmetric) matrix $\nabla^2\varphi_\gamma^{\mathrm{FB}}(x_\star)$ is positive definite.

(d) $x_\star$ is a strong local minimum for $\varphi_\gamma^{\mathrm{FB}}$.

(e) $x_\star$ is a local minimum for $\varphi_\gamma^{\mathrm{FB}}$ and $J\mathcal{R}_\gamma^{\mathrm{FB}}(x_\star)$ is nonsingular.

Proof. It follows from Thm. 5.14 that both $J\mathcal{R}_\gamma^{\mathrm{FB}}(x_\star)$ and $\nabla^2\varphi_\gamma^{\mathrm{FB}}(x_\star)$ exist, and that the latter is symmetric.

♠ 5.15*(a)* $\Leftrightarrow$ 5.15*(d)*. Follows from Thm. 5.4.

♠ 5.15*(d)* $\Leftrightarrow$ 5.15*(c)*. Trivial, since $\nabla^2\varphi_\gamma^{\mathrm{FB}}(x_\star)$ exists.

♠ 5.15*(c)* $\Leftrightarrow$ 5.15*(b)*. Apparent from (5.14), since $Q_\gamma(x_\star) \succ 0$. □

5.4 Convergence results

Theorem 5.16 (Finite termination of relaxed FBS)**.** *Under Assumption 5.I, the iterates generated by FBS (Alg. 5.1) satisfy*

$$\varphi_\gamma^{\mathrm{FB}}(x^{k+1}) \leq \varphi_\gamma^{\mathrm{FB}}(x^k) - \tfrac{1-\gamma\sigma_f}{2\gamma\lambda^2}\left(\tfrac{1-\gamma L_f}{1-\gamma\sigma_f} - (1-\lambda)^2\right)\|x^k - x^{k+1}\|^2.$$

In particular, the algorithm terminates in a finite number of iterations and yields a point x_* satisfying $\operatorname{dist}(0, \hat{\partial}\varphi(x_*)) \leq \varepsilon$.

Proof. It follows from Thm. 3.22 and Cor. 3.20 that for any $\varepsilon > 0$ the algorithm terminates in a finite number of iterations. That $\operatorname{dist}(0, \hat{\partial}\varphi(x_*)) \leq \varepsilon$ follows from Lem. 5.2, by observing that $1 - \gamma\sigma_f \leq 1 + \gamma\sigma_f \leq 2$. □

Algorithm 5.1. FORWARD-BACKWARD SPLITTING WITH RELAXATION

REQUIRE
- initial iterate $x_0 \in \mathbb{R}^n$
- stepsize $\gamma \in (0, 1/L_f)$
- tolerance $\varepsilon > 0$
- relaxation $\lambda \in \left(1 - \sqrt{\kappa},\, 1 + \sqrt{\kappa}\right)$, where $\kappa := \frac{1-\gamma L_f}{1-\gamma \sigma_f}$.

PROVIDE x_* with $\varphi(x_*) \leq \varphi(x^0)$ and $\operatorname{dist}(0, \hat{\partial}\varphi(x_*)) \leq \varepsilon$.

for $k = 0, 1, \ldots$ **do**
  $\bar{x}^k \in \operatorname{prox}_{\gamma g}\left(x^k - \gamma \nabla f(x^k)\right)$
  if $\frac{1}{2\gamma}\|x^k - \bar{x}^k\| \leq \varepsilon$ **then**
    return $x_* = \bar{x}^k$
  $x^{k+1} = (1-\lambda)x^k + \lambda \bar{x}^k$

Let us now analyze the asymptotic behavior of relaxed FBS without termination criterion, that is, when setting the tolerance as $\varepsilon = 0$ in FBS (Alg. 5.1).

Theorem 5.17 (Asymptotic convergence of relaxed FBS)**.** *Suppose that Assumption 5.1 is satisfied, and donsider the iterates generated by FBS (Alg. 5.1) with tolerance $\varepsilon = 0$. The following hold:*

(i) The forward-backward residual $(\|x^k - \bar{x}^k\|)_{k\in\mathbb{N}}$ is square-summable; in particular, $\min_{j\leq k} \|x^j - \bar{x}^j\| \in O(1/\sqrt{k})$.

(ii) $(x^k)_{k\in\mathbb{N}}$ and $(\bar{x}^k)_{k\in\mathbb{N}}$ have the same accumulation points, on which φ has the same value (this being the limit of the sequence $(\varphi_\gamma^{\mathrm{FB}}(x^k))_{k\in\mathbb{N}}$ or, equivalently, of $(\varphi(\bar{x}^k))_{k\in\mathbb{N}}$). Moreover, the set ω of such accumulation points satisfies $\omega \subseteq \operatorname{fix} \mathcal{T}_\gamma^{\mathrm{FB}} \subseteq \operatorname{zer} \hat{\partial}\varphi$.

(iii) If φ is level bounded, then $(x^k)_{k\in\mathbb{N}}$ is bounded, and ω is a nonempty, compact and connected set satisfying $\operatorname{dist}(x^k, \omega) \to 0$ as $k \to \infty$.

(iv) $\varphi_\gamma^{\mathrm{FB}} \equiv \varphi$ on ω, the value being the limit of the (decreasing) sequence $(\varphi_\gamma^{\mathrm{FB}}(x^k))_{k\in\mathbb{N}}$ (or, equivalently, of $(\varphi(\bar{x}^k))_{k\in\mathbb{N}}$.

Proof. Follows from Thm. 3.22 in light of Cor. 3.20. (That $(x^k)_{k\in\mathbb{N}}$ and $(\bar{x}^k)_{k\in\mathbb{N}}$ have the same accumulation points follows from the fact that $\|x^k - \bar{x}^k\| \to 0$ as $k \to \infty$). □

Theorem 5.18 (Global convergence of relaxed FBS)**.** *Suppose that Assumption 5.1 is satisfied, and consider the iterates generated by FBS (Alg. 5.1) with tolerance $\varepsilon = 0$. Suppose further that the following hold:*

A1 φ *is level bounded.*

A2 *All accumulation points of the sequence are prox-regular, in the sense of Definition 5.10.*

A3 $\varphi_\gamma^{\mathrm{FB}}$ *has the KL property.*

Then, the following hold:

(i) $(x^k)_{k\in\mathbb{N}}$ *converges to a point* $x_\star \in \operatorname{fix} \mathcal{T}_\gamma^{\mathrm{FB}} \subseteq \operatorname{zer} \hat{\partial}\varphi$.

(ii) The forward-backward residual $(\|x^k - \bar{x}^k\|)_{k\in\mathbb{N}}$ *is summable, and in particular* $\min_{j\le k} \|x^j - \bar{x}^j\| \in O(1/k)$.

Proof. Follows from Thm. 3.23, in light of Cor. 3.20 and Thm. 5.11*(ii)*. □

Theorem 5.19 (Linear convergence of relaxed FBS)**.** *Suppose that Assumption 5.1 is satisfied, and consider the iterates generated by FBS (Alg. 5.1) with tolerance* $\varepsilon = 0$. *Suppose further that the following hold:*

A1 φ *is level bounded.*

A2 *All accumulation points of the sequence are prox-regular, in the sense of Definition 5.10.*

A3 $\varphi_\gamma^{\mathrm{FB}}$ *has the KL property and the KL function is of the form* $\psi(s) = cs^\vartheta$ *for some* $c > 0$ *and* $\vartheta \ge 1/2$.

Then, the sequences $(x^k)_{k\in\mathbb{N}}$ *and* $(\operatorname{dist}(0, \mathcal{R}_\gamma^{\mathrm{FB}}(x^k)))_{k\in\mathbb{N}}$ *are R-linearly convergent.*

Proof. Follows from Thm. 3.24, in light of Cor. 3.20 and Thm. 5.11*(ii)*. □

5.5 A quasi-Newton FBS

Theorem 5.20 (Subsequential convergence of (nonmonotone) CLyD-FBS)**.** *Suppose that Assumption 5.1 is satisfied. Then, the following hold for the iterates generated by CLyD-FBS (Alg. 5.2) with tolerance* $\varepsilon = 0$*:*

(i) The residual $(\|x^k - \bar{x}^k\|)_{k\in\mathbb{N}}$ *is square-summable; in particular, it vanishes with rate* $\min_{j\le k} \operatorname{dist}(x^j, \mathcal{T}_\gamma^{\mathrm{FB}}(x^j)) \in O(1/\sqrt{k})$.

Algorithm 5.2. CLyD-FBS

REQUIRE
- stepsize $\gamma \in (0, 1/L_f)$
- scaling factor $\alpha \in (0, 1)$ for sufficient decrease constant
- initial iterate $x^0 \in \mathbb{R}^n$
- tolerance $\varepsilon > 0$

PROVIDE x_* with $\operatorname{dist}\big(0, \partial\varphi(x_*)\big) \le \varepsilon$

1: **for** $k = 0, 1, 2, \dots$ **do**
2: Do one nominal FB-step: $\bar{x}^k \in \operatorname{prox}_{\gamma g}\left(x^k - \gamma\nabla f(x^k)\right)$
3: **if** $\frac{1}{2\gamma}\|x^k - \bar{x}^k\| \le \varepsilon$ **then**
4: **return** $x_* = \bar{x}^k$
5: Select an update direction $d^k \in \mathbb{R}^n$ at x^k
6: Let $\tau_k \in \left\{2^{-i} \mid i \in \mathbb{N}\right\}$ be the largest such that

$$\varphi_\gamma^{\mathrm{FB}}(x^{k+1}) \le \varphi_\gamma^{\mathrm{FB}}(x^k) - \alpha\tfrac{1-\gamma L_f}{2\gamma}\|x^k - \bar{x}^k\|^2, \tag{5.16}$$

where $x^{k+1} := (1-\tau_k)\bar{x}^k + \tau_k(x^k + d^k)$

(ii) The set ω of accumulation points of $(x^k)_{k\in\mathbb{N}}$ satisfies $\omega \subseteq \operatorname{fix}\mathcal{T}_\gamma^{\mathrm{FB}} \subseteq \operatorname{zer}\hat{\partial}\varphi$.

If, additionally, $\|d^k\| \to 0$ as $k \to \infty$, then the following also hold:

(iii) If φ is level bounded, then $(x^k)_{k\in\mathbb{N}}$ and $(\bar{x}^k)_{k\in\mathbb{N}}$ are bounded, and ω is a nonempty, compact and connected set satisfying $\operatorname{dist}(x^k, \omega) \to 0$ as $k \to \infty$.

(iv) $\varphi_\gamma^{\mathrm{FB}} \equiv \varphi$ on ω, the value being the limit of the (decreasing) sequence $(\varphi_\gamma^{\mathrm{FB}}(x^k))_{k\in\mathbb{N}}$.

All the claims remain valid if the linesearch condition (5.16) is replaced by the following nonmonotone version:

$$\varphi_\gamma^{\mathrm{FB}}(x^{k+1}) \le \bar{\mathcal{L}}_k - \alpha\tfrac{c}{2}\|x^k - \bar{x}^k\|^2, \tag{5.17}$$

where, for any sequence $(t_k)_{k\in\mathbb{N}} \subseteq [0, 1]$ bounded away from 0, $\bar{\mathcal{L}}_k$ are recursively defined as follows:

$$\bar{\mathcal{L}}_k := \begin{cases} \varphi_\gamma^{\mathrm{FB}}(x^0) & \text{if } k = 0, \\ (1-t_k)\bar{\mathcal{L}}_{k-1} + t_k\varphi_\gamma^{\mathrm{FB}}(x^k) & \text{otherwise.} \end{cases}$$

Proof. Follows from Theorem 4.1. □

5.5.1 Global and (super)linear convergence

Theorem 5.21 (Global convergence)**.** *Additionally to Assumption 5.I, suppose that the following hold for the iterates generated by CLyD-FBS (Alg. 5.2) with tolerance $\varepsilon = 0$.*

A1 *φ is level bounded.*

A2 *All accumulation points of the sequence are prox-regular, in the sense of Definition 5.10.*

A3 *The FBE $\varphi_\gamma^{\mathrm{FB}}$ has the KL property.*

A4 *there exists $D > 0$ such that $\|d^k\| \leq D\|x^k - \bar{x}^k\|$ for all k.*

Then, the following hold:

(i) *$(x^k)_{k\in\mathbb{N}}$ converges to a point $x_\star \in \operatorname{fix} \mathcal{T}_\gamma^{\mathrm{FB}} \subseteq \operatorname{zer} \hat{\partial}\varphi$.*

(ii) *The residual is summable and in particular $\min_{j\leq k} \operatorname{dist}(x^j, \mathcal{T}_\gamma^{\mathrm{FB}}(x^j)) \in O(1/k)$.*

Proof. Follows from Theorems 5.11*(ii)* and 4.2. □

Theorem 5.22 (Linear convergence)**.** *Suppose that the assumptions of Theorem 5.21 are satisfied, and that the KL function can be taken of the form $\psi(s) = cs^\vartheta$ for some $c > 0$ and $\vartheta \geq 1/2$. Then, the sequences $(s^k)_{k\in\mathbb{N}}$, $(x^k)_{k\in\mathbb{N}}$, and $\operatorname{dist}(x^k, \mathcal{T}_\gamma^{\mathrm{FB}}(x^k))$ are R-linearly convergent.*

Proof. Follows from Theorem 4.3. □

Theorem 5.23 (Acceptance of the unit stepsize and superlinear convergence)**.** *Suppose that Assumption 5.I is satisfied, and consider the iterates generated by CLyD-FBS (Alg. 5.2). Suppose further that the following hold:*

A1 *$(x^k)_{k\in\mathbb{N}}$ converges to a strong local minimum of φ;*

A2 *$(d^k)_{k\in\mathbb{N}}$ are superlinearly convergent directions with respect to $(x^k)_{k\in\mathbb{N}}$;*

A3 *$\gamma \neq \Gamma^{\mathrm{FB}}(x_\star)$.*

Then, there exists $\bar{k} \in \mathbb{N}$ *such that*

$$\varphi_\gamma^{\mathrm{FB}}(s^k + d^k) \leq \varphi_\gamma^{\mathrm{FB}}(s^k) - \alpha \tfrac{1-\gamma L_f}{2\gamma} \|x^k - \bar{x}^k\|^2 \quad \text{for all } k \geq \bar{k}.$$

In particular, eventually the iterates reduce to $x^{k+1} = x^k + d^k$ *and converge superlinearly.*

Proof. Follows from Theorem 4.5. □

Theorem 5.24 (Dennis-Moré condition)**.** *Suppose that Assumption 5.I is satisfied, and consider the iterates generated by CLyD-FBS (Alg. 5.2). Suppose further that the following hold:*

A1 $(x^k)_{k\in\mathbb{N}}$ *converges to a strong local minimum* $x_\star$ *at which Assumption 5.II is (strictly) satisfied.*

A2 *The Dennis-Moré condition holds:*

$$\lim_{k\to\infty} \frac{\|\mathcal{R}_\gamma^{\mathrm{FB}}(x^k) + J\mathcal{R}_\gamma^{\mathrm{FB}}(x_\star)d^k\|}{\|d^k\|} = 0. \tag{5.18}$$

Then, $(d^k)_{k\in\mathbb{N}}$ *are superlinearly convergent directions with respect to* $(x^k)_{k\in\mathbb{N}}$.

Proof. If follows from Theorems 5.14*(ii)* and 5.15 that $\mathcal{R}_\gamma^{\mathrm{FB}}$ is strictly differentiable at $x_\star$ and has nonsingular Jacobian there. The proof then follows from Theorem 4.6. □

Theorem 5.25 (Superlinear convergence with Broyden directions)**.** *Suppose that Assumption 5.I is satisfied, and consider the iterates generated by CLyD-FBS (Alg. 5.2) with directions* d^k *selected with Broyden method* (4.5)*. Suppose further that the following hold:*

A1 $(x^k)_{k\in\mathbb{N}}$ *converges to a point* $x_\star$ *at which* $\mathcal{R}_\gamma^{\mathrm{FB}}$ *is Lipschitz-continuously semidifferentiable and with nonsingular Jacobian* $J\mathcal{R}_\gamma^{\mathrm{FB}}(x_\star)$ *(in particular,* $\mathcal{R}_\gamma^{\mathrm{FB}}$ *is strictly differentiable there).*

Then, the Dennis-Moré condition (5.18) *is satisfied, and in particular all the claims of Theorem 5.24 hold.*

Proof. Follows from Theorem 4.7 together with the observation that single valuedness of $\mathcal{R}_\gamma^{\mathrm{FB}}$ around $x_\star$ (due to semidifferentiability) ensures the required error bound assumption 4.7A2, cf. Thm. 5.11*(ii)*. □

5.6 Simulations

We now present numerical results with the proposed method. In CLyD-FBS (Alg. 5.2) we used the nonmonotone variant described in Theorem 5.20 with $(t_k)_{k\in\mathbb{N}}$ selected as in [126, 72], namely: $t_k = (\eta c_k + 1)^{-1}$, $c_0 = 1$, and $c_{k+1} = 0.85c_k + 1$. We performed experiments with the following choices of update directions:

- Broyden (modified) as in Section 4.3.2 with $\bar{\vartheta} = 10^{-4}$;
- BFGS as in Section 4.3.1;
- L-BFGS, namely the limited-memory variant of BFGS as in [88, Alg. 7.4] with memory 10.

We only show the results with full quasi-Newton updates (Broyden, BFGS) for one of the examples: for the other experiments we focus on L-BFGS, which is better suited for large-scale problems. Although $J\mathcal{R}_\gamma^{\mathrm{FB}}$ is nonsymmetric at the critical points in general, we observed that the symmetric updates of BFGS and L-BFGS perform very well in practice and outperform the Broyden method.

We compared CLyD-FBS (Alg. 5.2) with the nominal FBS, the inertial FBS (denoted IFBS) proposed in [26, Eq. (7)] with parameter $\beta = 0.2$, and the nonmonotone accelerated FBS (denoted AFBS) proposed in [72, Alg. 2] for fully nonconvex problems. All experiments were performed in MATLAB. The implementation of the methods used in the tests is available online.[2]

5.6.1 Dictionary learning

Expressing large data by means of only few elements from a collection of vectors is an important problem in machine learning and signal processing. The challenge is finding such a collection of vectors, known as *dictionary*, that can accurately represent data signals in the sparsest way. In mathematical terms, given m signals $y_1, \dots, y_m \in \mathbb{R}^n$ we wish to find k *dictionary atoms* $d_1, \dots, d_k \in \mathbb{R}^n$ in such a way that each y_j can be represented, or accurately approximated, as a sparse linear combination of them. If we stack the data in a matrix $Y \in \mathbb{R}^{n\times m}$, and the dictionary atoms in a matrix $D \in \mathbb{R}^{n\times k}$ (to be found), the problem can be expressed as follows [1]

$$\underset{D,C}{\text{minimize}} \quad \tfrac{1}{2}\|Y - DC\|_F^2 \quad \text{subject to} \quad \begin{array}{ll} \|d_i\|_2 = 1 & i = 1, \dots, k, \\ \|c_j\|_0 \le N & j = 1, \dots, m, \\ \|c_j\|_\infty \le T & j = 1, \dots, m, \end{array} \tag{5.19}$$

[2] http://github.com/kul-forbes/ForBES

where $C = [c_1, \dots, c_m] \in \mathbb{R}^{k\times m}$ is a matrix containing the sought coefficients, and $N \in \mathbb{N}$ and $T > 0$ are parameters. Differently from [1], we bound the set of feasible points by means of the ℓ_∞-norm constraint; this artificial constraint ensures that ∇f is globally Lipschitz continuous over the feasible domain. Moreover, we explicitly constrain the norm of the dictionary atoms: this causes no loss of generality, as the objective value of (5.19) is unchanged if the j-th atom d_j and the j-th row of C are scaled by reciprocal factors.

The problem can be expressed in the canonical form (5.1) by letting $f(D, C) = \frac{1}{2}\|Y - DC\|_F^2$ and $g(D, C) = \delta_S(D, C)$, where

$$S = \left\{D \in \mathbb{R}^{n\times k} \mid \|d_j\|_2 = 1,\ j = 1 \dots k\right\} \times \left\{C \in \mathbb{R}^{k\times m} \mid \begin{smallmatrix}\|c_j\|_0 \le N \\ \|c_j\|_\infty \le T,\end{smallmatrix}\ j = 1 \dots m\right\}$$

is the product of Euclidean spheres and box-constrained ℓ_0 balls. Both f and g are nonconvex. The projection of (D, C) onto S is simple and column-wise separable: the columns d_j of D are scaled by their ℓ_2 norm, while the N largest coefficients (in absolute value) of the columns c_j of C are projected onto the box $[-T, T]$ and the other ones are set to zero, see *e.g.*, [14, Alg. 3 and Ex. 4.6].

We tested our algorithm on 50 problems with $N = 3$, $n = 20$, $m = 500$ and $k = 50$, for a total of 26000 variables each. We chose $T = 10^6$ as a large bound for the ℓ_∞ norm of the columns of C. Problems were generated according to [1, §V.A]: first, a dictionary $D_{\text{gen}} \in \mathbb{R}^{20\times 50}$ was randomly generated with normal entries, and each column was normalized to one. Then, a matrix $C_{\text{gen}} \in \mathbb{R}^{50\times 500}$ was constructed with 3 normally distributed nonzero coefficients per column. Then we set $Y = C_{\text{gen}} D_{\text{gen}} + V$, where $V \in \mathbb{R}^{20\times 500}$ is a matrix with normally distributed entries with variance 10^{-2}.

We compared FBS, AFBS and CLyD-FBS (Alg. 5.2)(L-BFGS), using a backtracking procedure to adaptively adjust the stepsize γ. IFBS could not be applied due to the lack of an adaptive stepsize-selection rule for the algorithm [26]. Moreover, we did not test CLyD-FBS (Alg. 5.2) with Broyden and (full) BFGS directions because of the prohibitive overhead of storing and operating with 26000×26000 matrices.

Figure 5.2 shows the performance profile of the algorithms by comparing the time needed to reach an accuracy of $\|r^k\| \le 10^{-4}$ starting from $(D^0, C^0) = (0, 0)$. In most of the cases, CLyD-FBS (Alg. 5.2)(L-BFGS) exhibited a speedup of a factor 5-to-100 with respect to FBS, and 3-to-60 with respect to AFBS, at reaching a critical point.

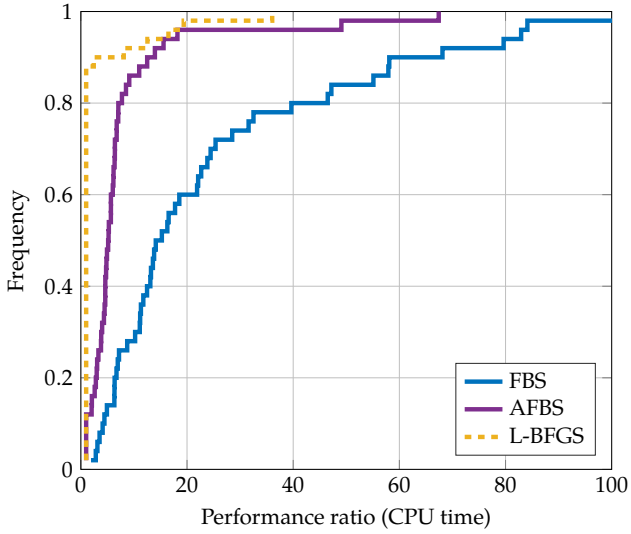


Figure 5.2: *Dictionary learning. Performance profiles of FBS, AFBS and CLyD-FBS (Alg. 5.2) with L-BFGS directions when applied to* 50 *randomly generated problems with* $n = 20$, $m = 500$, $k = 50$, $T = 10^6$ *and* $N = 3$. *The algorithms are executed until tolerance* $\|\mathcal{R}_\gamma^{\text{FB}}(x^k)\| \leq 10^{-4}$ *is reached. In the great majority of cases, the employment of L-BFGS directions with the proposed framework reaches a critical point significantly faster than FBS and its nonconvex accelerated variant AFBS.*

5.6.2 Nonconvex sparse approximation

Here we consider the problem of finding a sparse solution $x \in \mathbb{R}^n$ to a least-squares problem $Ax = b$, where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Sparsity can be induced by constraining or penalizing the ℓ_0 quasi-norm of x, namely the number of nonzero elements of x, but due to the challenges of nonconvexity it is often the case that the ℓ_1 norm is used instead. As well explained and documented in [124], the use of the (square root of the) $\ell_{1/2}$ quasi-norm, namely $\|x\|_{1/2}^{1/2} = \sum_{i=1}^n |x_i|^{1/2}$, is in some sense optimal in trading-off representativeness of the solution and numerical simplicity of the ℓ_0 and ℓ_1 approaches, respectively. The problem then becomes

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \;\; \tfrac{1}{2}\|Ax - b\|_2^2 + \lambda \|x\|_{1/2}^{1/2}, \tag{5.20}$$

n	λ	FBS avg/max (s)	IFBS avg/max (s)	AFBS avg/max (s)	ZeroFPR(L-BFGS) avg/max (s)
500	0.10	0.141/0.405	0.159/0.449	0.135/0.221	0.037/0.088
	0.03	0.498/2.548	0.688/3.962	0.274/0.430	0.084/0.126
	0.01	1.305/5.445	1.721/4.942	0.570/1.157	0.152/0.560
1000	0.10	0.176/0.287	0.231/0.659	0.228/0.483	0.021/0.077
	0.03	0.576/2.756	0.645/4.165	0.382/0.841	0.091/0.275
	0.01	1.864/9.740	2.391/8.311	0.795/1.446	0.222/0.438
2000	0.10	0.291/0.599	0.392/0.719	0.393/0.640	0.025/0.055
	0.03	0.553/1.841	0.602/3.270	0.464/0.702	0.088/0.198
	0.01	2.108/10.934	2.439/8.010	0.979/1.411	0.271/0.464

Table 5.2: *Nonconvex sparse approximation. The table shows average and maximum CPU time required to reach* $\|\mathcal{R}_\gamma^{\text{FB}}(x^k)\| \leq 10^{-6}$ *in* 100 *random experiments. Each algorithm was run on the same set of randomly generated problems, with* $x^0 = 0$.

where $\lambda > 0$ is a regularization parameter. Function $\|x\|_{1/2}^{1/2}$ is separable, and its proximal mapping can be computed in closed form as follows, see [124, Thm. 1]:

$$\left[\operatorname{prox}_{\gamma\|\cdot\|_{1/2}^{1/2}}(x)\right]_i = \tfrac{2}{3}\Big(1 + \cos\tfrac{2}{3}\Big(\pi - \arccos\tfrac{\gamma}{8}(|x_i|/3)^{-3/2}\Big)\Big)x_i, \quad i = 1, \dots, n.$$

We ran numerical experiments consistently with the setting of [37, Sec. 8.2]. We considered different scenarios obtained by changing the regularization term λ and the size of A, keeping a constant column-to-row ratio of $n/m = 5$ for matrix A. Matrix A was generated with random Gaussian entries, with zero mean and variance $1/m$, while vector b was generated as $b = Ax_{\text{orig}} + v$ where $x_{\text{orig}} \in \mathbb{R}^n$ was randomly generated with $k = 5$ nonzero normally distributed entries, and $v \in \mathbb{R}^n$ is a noise vector with zero mean and variance $1/m$.

For each scenario, we solved 100 randomly generated problems and compared the performance of all algorithms in terms of CPU time to reach an accuracy of $\|r^k\| \leq 10^{-6}$. For all algorithms and problems, we used $x^0 = 0$ as the starting iterate. Average and worst-case performance of the algorithms in each of the nine scenarios are illustrated in Table 5.2; apparently, CLyD-FBS (Alg. 5.2) is significantly faster than FBS, IFBS and AFBS, even in a worst-case-to-average comparison.

Figure 5.3 shows the convergence rates of the algorithms in one of the generated problems. Since CLyD-FBS (Alg. 5.2) employs a linesearch, and therefore the complexity of each iteration is unknown a priori, we recorded the number of matrix-vector products by A and $A^\top$ performed during the iterations, and displayed it on the horizontal axis. Apparently, CLyD-FBS (Alg. 5.2) with

Broyden directions achieves superlinear convergence, beating the linear of FBS, IFBS and AFBS. This comparison also confirms the claimed great performance of (L-)BFGS directions.

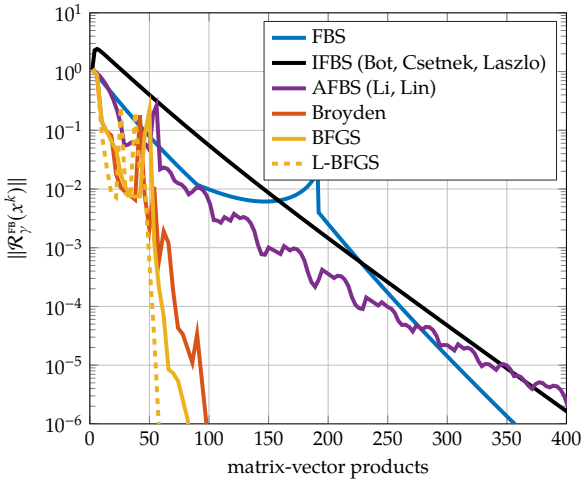


Figure 5.3: *Nonconvex sparse approximation. Convergence of fixed-point residual and cost in FBS, IFBS, AFBS and CLyD-FBS (Alg. 5.2), for different choices of the search directions and for* $n = 1500$, $\lambda = 0.03$.

Chapter 6

Douglas-Rachford splitting

6.1 Introduction

First introduced in [40] for finding numerical solutions of heat differential equations, the Douglas-Rachford splitting (DRS) is now considered a textbook algorithm in convex optimization or, more generally, in monotone inclusion problems. Similarly to FBS, the objective to be minimized is *split* as the sum of two functions, resulting in the following canonical framework addressed by DRS:

$$\underset{s\in\mathbb{R}^p}{\text{minimize}} \;\; \varphi(s) \equiv \varphi_1(s) + \varphi_2(s). \tag{6.1}$$

Here, $\varphi_1, \varphi_2 : \mathbb{R}^p \to \overline{\mathbb{R}}$ are proper, lower semicontinuous (lsc), extended-real-valued functions. Starting from some $s \in \mathbb{R}^p$, one iteration of DRS applied to (6.1) with *stepsize* $\gamma > 0$ and *relaxation* parameter $\lambda > 0$ amounts to

$$\begin{cases} u & \in \operatorname{prox}_{\gamma\varphi_1}(s) \\ v & \in \operatorname{prox}_{\gamma\varphi_2}(2u - s) \\ s^+ & = s + \lambda(v - u). \end{cases} \tag{DRS}$$

The case $\lambda = 1$ corresponds to the classical DRS, whereas for $\lambda = 2$ the scheme is also known as Peaceman-Rachford splitting (PRS). When both φ_1 and φ_2 are convex functions, DRS iterations are known to converge for any $\gamma > 0$ and $\lambda \in (0, 2)$, yielding a minimizer of φ [10, Cor. 28.3].

Although the scheme does not involve any gradients and is thus applicable even when both functions φ_1 and φ_2 are nonsmooth, in order to frame it as a generalized PMM scheme we will need to work under the same assumptions as

FBS.

Assumption 6.I (DRS: basic assumption)**.** In problem (6.1)

A1 $\varphi_1 \in C^{1,1}(\mathbb{R}^n)$ is L_{φ_1}-smooth, hence σ_{φ_1}-hypoconvex with $|\sigma_{\varphi_1}| \leq L_{\varphi_1}$.

A2 $\varphi_2 : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper and lsc.

A3 A solution exists, that is, $\operatorname{arg\,min} \varphi \neq \emptyset$.

Clearly, one can repropose the same remark about prox-boundedness of φ_2 discussed in Remark 5.1.

Remark 6.1 (Feasible stepsizes for DRS)**.** Under Assumption 6.I, both functions φ_1 and φ_2 are prox-bounded with threshold at least $1/L_{\varphi_1}$ (the claim on φ_1 follows from Thm. 1.15). In particular, for any stepsize $\gamma < 1/L_{\varphi_1}$ DRS iterations are well defined. □

Those in Assumption 6.I are indeed the same requirements under which global convergence of nonconvex DRS applied to general problems has been established [71, 69], specifically for the cases $\lambda = 1$ and $\lambda = 2$; other results are problem-specific and mostly concerned with local convergence. These mainly focus on feasibility problems, where the goal is to find points in the intersection of nonempty closed sets A and B subjected to some regularity conditions. This is done by applying DRS to the minimization of the sum of the indicator functions $\varphi_1 = \delta_A$ and $\varphi_2 = \delta_B$. The minimization subproblems in DRS then reduce to (set-valued) projections onto either set, regardless of the stepsize parameter $\gamma > 0$. This is the case of [9], for instance, where A and B are finite unions of convex sets. Local linear convergence when A is affine, under some conditions on the (nonconvex) set B, are shown in [55, 56].

Although this particular application of DRS does not comply with our requirements, as φ_1 fails to be Lipschitz-differentiable, however replacing δ_A with $\varphi_1 = \frac{1}{2}\operatorname{dist}_A^2$ yields an equivalent problem which fits into our framework when A is a convex set. In terms of DRS iterations, this simply amounts to replacing Π_A, the projection onto set A, with a "relaxed" version $\Pi_{A,t} \coloneqq (1-t)\mathrm{id} + t\,\Pi_A$ for some $t \in (0,1)$. Then, it can be easily verified that for any $\alpha, \beta \in (0, +\infty]$ one DRS-step applied to

$$\underset{s \in \mathbb{R}^p}{\text{minimize}} \;\; \tfrac{\alpha}{2}\operatorname{dist}_A^2(s) + \tfrac{\beta}{2}\operatorname{dist}_B^2(s) \tag{6.2}$$

results in

$$s^+ \in (1 - \lambda/2)s + \lambda/2\,\Pi_{B,q}\,\Pi_{A,p}\,s \tag{6.3}$$

for $p = \frac{2\alpha\gamma}{1+\alpha\gamma}$ and $q = \frac{2\beta\gamma}{1+\beta\gamma}$. Notice that (6.3) is the $\lambda/2$-relaxation of the "method of alternating (p,q)-relaxed projections" ((p,q)-MARP) [13]. The (non-relaxed)

(p, q)-MARP is recovered by setting $\lambda = 2$, that is, by applying PRS to (6.2). Local linear convergence of MARP was shown when A and B, both possibly nonconvex, satisfy some constraint qualifications, and also global convergence when some other requirements are met. When set A is convex, then $\frac{\alpha}{2}\operatorname{dist}_A^2$ is convex and α-Lipschitz differentiable; our theory then ensures convergence of the fixed-point residual and subsequential convergence of the iterations (6.3) for any $\lambda \in (0, 2)$, $p \in (0, 1)$ and $q \in (0, 1]$, without any requirements on the (nonempty closed) set B. Here, $q = 1$ is obtained by replacing $\frac{\beta}{2}\operatorname{dist}_B^2$ with δ_B, which can be interpreted as the hard penalization obtained by letting $\beta = \infty$. The non-relaxed MARP is not covered due to the non-strong convexity of dist_A^2, however λ can be set arbitrarily close to 2.

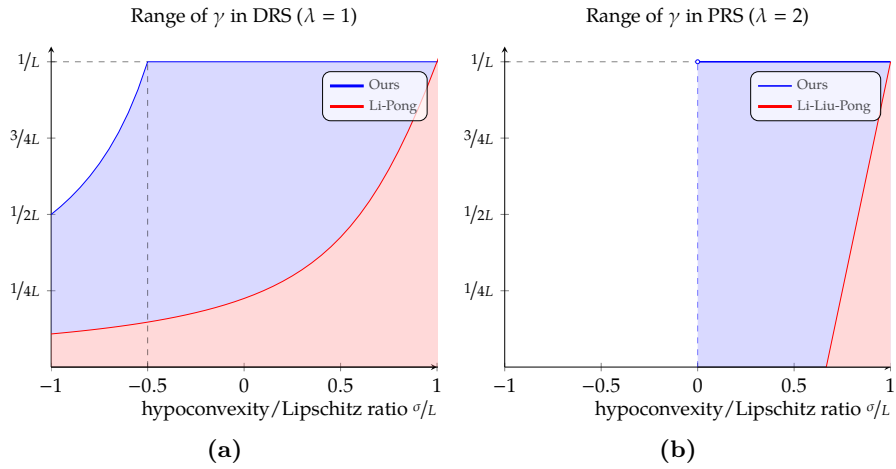


Figure 6.1: *Maximum stepsize γ ensuring convergence of DRS (Fig. 6.1a) and PRS (Fig. 6.1b); comparison of our bounds (blue plot) with [71] for DRS and [69] for PRS. On the x-axis the ratio between hypoconvexity parameter σ and the Lipschitz modulus L of the gradient of the smooth function. On the y-axis, the supremum of stepsize γ such that the algorithms converge.*

The work [71] presents the first general analysis of global convergence of (non-relaxed) DRS for fully nonconvex problems where one function is Lipschitz differentiable. In [69] PRS is also considered under the additional requirement that the smooth function is strongly convex with strong-convexity/Lipschitz moduli ratio of at least $2/3$. Both papers show that for sufficiently small stepsizes one iteration of DRS or PRS yields a sufficient decrease on an augmented Lagrangian. However, due to the lower unboundedness of the augmented Lagrangian the vanishing of the fixed-point residual could not be shown, unless φ has bounded level sets.

Other than completing the analysis to all relaxation parameters $\lambda \in (0,4)$, as opposed to $\lambda \in \{1,2\}$, we improve their results by showing convergence for a considerably larger range of stepsizes and, in the case of PRS, with no restriction on the strong convexity modulus of the smooth function. We also show that our bounds are optimal whenever $\lambda \in (0,2]$. The extent of the improvement is evident in the comparisons outlined in Figure 6.1.

6.2 DRS as a GPMM algorithm

We now show that under Assumption 6.I DRS and all its relaxations fit into the generalized PMM scheme (2.6). As done for FBS, to facilitate the reading all the elements and the parameters are summarized in Table 6.1.

Let us consider one DRS iteration $s \mapsto (u, v, s^+)$ with stepsize $\gamma < 1/L_{\varphi_1}$. Since φ_1 is differentiable, it follows from Proposition 1.12*(vi)* that variable $u = \operatorname{prox}_{\gamma\varphi_1}(s)$ satisfies $\nabla\varphi_1(u) = \frac{1}{\gamma}(s-u)$. As first noted in [93], the v-update thus boils down to

$$v \in \operatorname{prox}_{\gamma\varphi_2}(2u-s) = \operatorname{prox}_{\gamma\varphi_2}(u - \gamma\nabla\varphi_1(u)), \tag{6.4}$$

which is a forward-backward step at u. As seen in the previous chapter, FBS is the pure PMM scheme $(\mathcal{M}_\gamma^{\mathrm{FB}})_{\gamma\in(0,1/L_{\varphi_1})}$; we thus can write DRS iterations as

$$\begin{aligned} s^+ &= s + \lambda(v-u) \\ &\in s - \lambda\big(\mathrm{id} - \mathcal{T}_\gamma^{\mathrm{FB}}\big)(u) \\ &= s - \lambda\big(\mathrm{id} - \mathcal{T}_\gamma^{\mathrm{FB}}\big) \circ \operatorname{prox}_{\gamma\varphi_1}(s). \end{aligned} \tag{6.5}$$

It follows that DRS fits into the generalized framework, identified as

$$\mathrm{DR} \sim (\mathcal{M}_\gamma^{\mathrm{FB}}, \operatorname{prox}_{\gamma\varphi_1})_{\gamma\in(0,1/L_{\varphi_1})}. \tag{6.6}$$

With this identification, we can condensate (6.6) into

$$s^+ \in \mathcal{F}_\gamma^{\mathrm{DR}_\lambda}(s) \coloneqq s + \lambda(v-u),$$

where u, v are as in (DRS).

Remark 6.2. Under Assumption 6.I, in light of Theorem 1.15 the transient mapping $G_\gamma = \operatorname{prox}_{\gamma\varphi_1}$ is μ_{G_γ}-strongly monotone and L_{G_γ}-Lipschitz continuous, with

$$\mu_{G_\gamma} = \tfrac{1}{1+\gamma L_{\varphi_1}} \quad \text{and} \quad L_{G_\gamma} = \tfrac{1}{1+\gamma\sigma_{\varphi_1}}.$$

Douglas-Rachford splitting DR $\sim (\mathcal{M}_\gamma^{\mathrm{DR}}, \mathrm{prox}_{\gamma\varphi_1})_{\gamma\in(0,1/L_{\varphi_1})}$		
$\bar{\gamma}$:	$1/L_{\varphi_1}$	
$\mathcal{M}_\gamma^{\mathrm{DR}}$:	$\mathcal{M}_\gamma^{\mathrm{FB}}$	same MM model
$\mathcal{T}_\gamma^{\mathrm{DR}}$:	$\mathcal{T}_\gamma^{\mathrm{FB}}$	as FBS with
$m_1(\gamma)$:	$\frac{1-\gamma L_{\varphi_1}}{\gamma}$	$f=\varphi_1$ and $g=\varphi_2$
$m_2(\gamma)$:	$\frac{1-\gamma\sigma_{\varphi_1}}{\gamma}$	(see Table 5.1)
L_δ :	$\frac{1-\gamma\sigma_{\varphi_1}}{\gamma}$	
$\mathcal{R}_\gamma^{\mathrm{DR}}(s)$:	$\frac{u-v}{\gamma}$ (with u,v as in (DRS))	(6.5)
$\varphi_\gamma^{\mathrm{DR}}$:	$\varphi_\gamma^{\mathrm{FB}} \circ \mathrm{prox}_{\gamma\varphi_1}$	(5.7)
G_γ :	$\mathrm{prox}_{\gamma\varphi_1}$	(6.5)
G_γ^{-1} :	$\mathrm{id}+\gamma\nabla\varphi_1$	Rem. 6.2
μ_{G_γ} :	$\frac{1}{1+\gamma L_{\varphi_1}}$	Rem. 6.2
L_{G_γ} :	$\frac{1}{1+\gamma\sigma_{\varphi_1}}$	Rem. 6.2

Table 6.1: *Douglas-Rachford splitting with stepsize $\gamma \in (0, 1/L_{\varphi_1})$: parameters of the proximal model. Being based on the same proximal MM model as FBS, the $\mathcal{M}$-envelope $\varphi^{\mathcal{M}_\gamma^{\mathrm{DR}}} = \varphi^{\mathcal{M}_\gamma^{\mathrm{FB}}} = \varphi_\gamma^{\mathrm{FB}}$ is the FBE. The $\mathcal{F}$-envelope, denoted as $\varphi_\gamma^{\mathrm{DR}}$, is the Douglas-Rachford envelope function $\varphi_\gamma^{\mathrm{DR}} = \varphi_\gamma^{\mathrm{FB}} \circ \mathrm{prox}_{\gamma\varphi_1}$.*

Consequently, the inverse of G_γ, namely

$$G_\gamma^{-1} = \mathrm{id} + \gamma\nabla\varphi_1,$$

is $(1+\gamma L_{\varphi_1})$-Lipschitz continuous and $(1+\gamma\sigma_{\varphi_1})$-strongly monotone. □

Since the proximal model of DRS is the same as that of FBS, the error bound in Lemma 5.2 can be imported verbatim.

Lemma 6.3. *Suppose that Assumption 6.I is satisfied, and let $s \mapsto (u, v, s^+)$ be a DRS update with stepsize $\gamma \in (0, 1/L_{\varphi_1})$. Then,*

$$\operatorname{dist}(0, \hat{\partial}\varphi(v)) \leq \tfrac{1-\gamma\sigma_f}{\gamma}\|u - v\|.$$

6.3 Douglas-Rachford envelope

The $\mathcal{F}$-envelope associated to the GPMM scheme DRS is the Douglas-Rachford envelope (DRE), namely

$$\varphi_\gamma^{\mathrm{DR}}(s) := \varphi_\gamma^{\mathrm{FB}} \circ \operatorname{prox}_{\gamma\varphi_1}(s). \tag{6.7}$$

A glance at the expression (5.6) of the FBE shows that

$$\varphi_\gamma^{\mathrm{DR}}(s) = \varphi_1(u) + \varphi_2(v) + \langle \nabla\varphi_1(u), v - u\rangle + \tfrac{1}{2\gamma}\|v - u\|^2 \tag{6.8}$$

$$= \mathscr{L}_{1/\gamma}(u, v, \tfrac{1}{\gamma}(u - s)), \tag{6.9}$$

where (u, v) are the variables appearing in one iteration of DRS starting from s, and $\mathscr{L}_{1/\gamma}$ is the $(1/\gamma)$-augmented Lagrangian associated to the equivalent problem formulation

$$\operatorname*{minimize}_{x,z\in\mathbb{R}^n} \ \varphi_1(x) + \varphi_2(z) \quad \text{subject to } \ x - z = 0,$$

namely,

$$\mathscr{L}_{1/\gamma}(x, z, y) = \varphi_1(x) + \varphi_2(z) + \langle y, x - z\rangle + \tfrac{1}{2\gamma}\|x - z\|^2.$$

In fact, the Lagrange multiplier is

$$y = -\nabla\varphi_1(u) = \tfrac{1}{\gamma}(s - u), \tag{6.10}$$

as it follows from Theorem 1.15*(i)*.

The DRE was first introduced in [93] for convex problems with φ_1 twice continuously differentiable. Under these assumptions the DRE was shown to be continuously differentiable. The more general analysis dealt here was investigated in [119]. Further properties of the DRE can be imported verbatim from the unified analysis provided in Chapter 3.

Theorem 6.4 (DRE: sandwich property)**.** *Suppose that Assumption 6.I is satisfied, and let $\gamma \in (0, 1/L_{\varphi_1})$ be fixed. For all $s \in \mathbb{R}^n$, $u = \operatorname{prox}_{\gamma\varphi_1}(s)$, and $v \in \operatorname{prox}_{\gamma\varphi_2}(2u - s)$, the following hold:*

(i) $\varphi_\gamma^{\mathrm{DR}}(s) \leq \varphi(u)$, *with equality holding iff* $s \in \mathcal{F}_\gamma^{\mathrm{DR}\lambda}(s)$.

(ii) $\frac{1-\gamma L_{\varphi_1}}{2}\|u-v\|^2 \leq \varphi_\gamma^{\mathrm{DR}}(s) - \varphi(v) \leq \frac{1-\gamma\sigma_{\varphi_1}}{2}\|u-v\|^2$.

(iii) $\inf \varphi_\gamma^{\mathrm{FB}} = \inf \varphi$ *and* $\operatorname{prox}_{\gamma\varphi_1}(\arg\min \varphi_\gamma^{\mathrm{FB}}) = \arg\min \varphi$.

Proof. Follows from Cor. 3.3. □

Theorem 6.5 (DRE: equivalence of local minimality)**.** *Suppose that Assumption 6.1 is satisfied, and let* $\gamma \in (0, 1/L_{\varphi_1})$, $\bar{s} \in \operatorname{fix} \mathcal{F}_\gamma^{\mathrm{DR}\lambda}$, *and* $\bar{u} \coloneqq \operatorname{prox}_{\gamma\varphi_1}(\bar{s})$ *be fixed. The following hold:*

(i) If $\bar{s}$ *is a (strong) local minimum for* $\varphi_\gamma^{\mathrm{DR}}$, *then* $\bar{u}$ *is a (strong) local minimum for* φ.

(ii) If $\mathcal{F}_\gamma^{\mathrm{DR}\lambda}(\bar{s}) = \{\bar{s}\}$, *(e.g., if* $\gamma < \Gamma^{\mathrm{DR}}(\bar{u})$*) then the converse also holds.*

Proof. See Thm. 3.6. □

Theorem 6.6 (DRE: Equivalence of level boundedness)**.** *Suppose that Assumption 6.1 is satisfied. Then, for any* $\gamma \in (0, 1/L_{\varphi_1})$, φ *is level bounded iff* $\varphi_\gamma^{\mathrm{FB}}$ *is level bounded.*

Proof. See Thm. 3.7. □

6.3.1 Regularity properties

Proposition 6.7 (Strict continuity of the DRE)**.** *Suppose that Assumption 6.1 is satisfied. Then, for any* $\gamma \in (0, 1/L_{\varphi_1})$ *the DRE* $\varphi_\gamma^{\mathrm{DR}}$ *is a strictly continuous function.*

Proof. Since $\varphi_\gamma^{\mathrm{DR}} = \varphi_\gamma^{\mathrm{FB}} \circ \operatorname{prox}_{\gamma\varphi_1}$, the claim follows from the strict continuity of $\varphi_\gamma^{\mathrm{FB}}$ and the Lipschitz continuity of $\operatorname{prox}_{\gamma\varphi_1}$, cf. Prop. 5.6 and Rem. 6.2. □

Proposition 6.8. *Suppose that Assumption 6.1 is satisfied, and let* $\bar{s}$ *be such that* $\mathcal{F}_\gamma^{\mathrm{DR}\lambda}(\bar{s}) = \{\bar{s}\}$. *If* φ_1 *is (strictly) twice differentiable at* $\bar{s}$, *then* $\varphi_\gamma^{\mathrm{DR}}$ *is (strictly) differentiable at* $\bar{s}$ *with* $\nabla\varphi_\gamma^{\mathrm{DR}}(\bar{s}) = 0$.

Proof. Let $\bar{u} \coloneqq \operatorname{prox}_{\gamma\varphi_1}(\bar{s})$. Then, since $\mathcal{T}_\gamma^{\mathrm{DR}} = \mathcal{T}_\gamma^{\mathrm{FB}}$ we have that $\mathcal{T}_\gamma^{\mathrm{FB}}(\bar{u}) = \{\bar{u}\}$, and Prop. 6.8 ensures that $\varphi_\gamma^{\mathrm{FB}}$ is (strictly) differentiable at $\bar{u}$ with $\nabla\varphi_\gamma^{\mathrm{FB}}(\bar{u}) = 0$.

Then, for all s_i and $u_i \coloneqq \operatorname{prox}_{\gamma\varphi_1}(s_i)$, $i = 1, 2$, we have that

$$\frac{|\varphi_\gamma^{\mathrm{DR}}(s_1) - \varphi_\gamma^{\mathrm{DR}}(s_2)|}{\|s_1 - s_2\|} = \frac{|\varphi_\gamma^{\mathrm{FB}}(u_1) - \varphi_\gamma^{\mathrm{FB}}(u_2)|}{\|s_1 - s_2\|} \leq (1 + \gamma L_{\varphi_1}) \frac{|\varphi_\gamma^{\mathrm{FB}}(u_1) - \varphi_\gamma^{\mathrm{FB}}(u_2)|}{\|u_1 - u_2\|}$$

where the last equality follows from the strong monotonicity of $\operatorname{prox}_{\gamma\varphi_1}$, cf. Rem. 6.2. (Strict) differentiability of $\varphi_\gamma^{\mathrm{DR}}$ at $\bar{s}$ with $\nabla\varphi_\gamma^{\mathrm{DR}}(\bar{s}) = 0$ then easily follows from that of $\varphi_\gamma^{\mathrm{FB}}$, cf. Prop. 5.7. □

6.3.2 The DRE as a Lyapunov function

We now proceed to showing that the DRE can conveniently serve as Lyapunov function for DRS, so that one can directly import all the convergence results developed in the general framework of Section 3.3.

A first result can be derived from Theorem 3.18 with no effort by simply plugging the constants $m_1, m_2, \mu_{G_\gamma}, L_{G_\gamma}$ corresponding to the GPMM scheme of DRS with stepsize γ, cf. Table 6.1. Indeed, after simple algebra one obtains that the $\varphi_\gamma^{\mathrm{DR}}$ is a Lyapunov function for $\mathcal{F}_\gamma^{\mathrm{DR}_\lambda}$ provided that $\xi \coloneqq \gamma L_{\varphi_1}$ solves the following cubic inequality

$$(1 - p + p^3)\xi^3 + (2 - 2p + p^2)\xi^2 + (1 - 2p)\xi - 1 < 0, \tag{6.11}$$

where $p \coloneqq {}^{\sigma_{\varphi_1}}\!/_{L_{\varphi_1}} \in [-1, 1]$, and that λ is bounded in some range contained in $(0, 2)$. Although for any $p \in [-1, 1]$ one can always find small enough stepsizes γ such that $\xi = \gamma L_{\varphi_1}$ satisfies (6.11), it turns out that this estimate is extremely loose. The following result uses more sophisticated inequalities and provides sensibly better ranges. In fact, we will show in Section 6.4.1 that for any relaxation $\lambda \in (0, 2]$ the given bounds are tight, as DRS is not ensured to converge otherwise. For the sake of a comparison, in the worst-case scenario $p = -1$, (6.11) imposes $\gamma L_{\varphi_1} < \frac{\sqrt{5}-2}{2} \approx 0.24$ and additional constraints on λ, whereas the tight bound is $\gamma L_{\varphi_1} < 1 - {}^{\lambda}\!/_2$ for any $\lambda \in (0, 2)$; if φ_1 is convex, hence $p = 0$, (6.11) imposes $\gamma L_{\varphi_1} < 0.47$, while the tight bound is $\gamma L_{\varphi_1} < 1$ for any $\lambda \in (0, 2)$. The result also investigates the employment of stepsizes $\lambda \in [2, 4)$; it will also be shown that no guarantee of convergence of DRS can be established for $\lambda \notin (0, 4)$.

Theorem 6.9 (Sufficient decrease on the DRE)**.** *Suppose that Assumption 6.I is satisfied, and consider one DRS update $s \mapsto (u, v, s^+)$ for some stepsize $\gamma < \min\left\{\frac{2-\lambda}{2[\sigma_{\varphi_1}]_-}, \frac{1}{L_{\varphi_1}}\right\}$ and relaxation $\lambda \in (0, 2)$. Then,*

$$\varphi_\gamma^{\mathrm{DR}}(s) - \varphi_\gamma^{\mathrm{DR}}(s^+) \geq \tfrac{c}{(1+\gamma L_{\varphi_1})^2} \|s - s^+\|^2, \tag{6.12}$$

where, denoting $p_{\varphi_1} := \sigma_{\varphi_1}/L_{\varphi_1} \in [-1,1]$, $c > 0$ *is defined by*[1]

$$c = \frac{2-\lambda}{2\lambda\gamma} - \begin{cases} L_{\varphi_1} \max\left\{ \frac{[p_{\varphi_1}]_-}{2(1-[p_{\varphi_1}]_-)},\ \frac{1}{2} - \frac{\gamma L_{\varphi_1}}{\lambda} \right\} & \text{if } p_{\varphi_1} \geq \frac{\lambda}{2} - 1, \\ \frac{[\sigma_{\varphi_1}]_-}{\lambda} & \text{otherwise.} \end{cases} \tag{6.13}$$

If φ_1 *is strongly convex, then* (6.12) *also holds for*

$$2 \leq \lambda < \frac{4}{1+\sqrt{1-p_{\varphi_1}}} \quad \text{and} \quad \frac{p_{\varphi_1}\lambda - \delta}{4\sigma_{\varphi_1}} < \gamma < \frac{p_{\varphi_1}\lambda + \delta}{4\sigma_{\varphi_1}}, \tag{6.14}$$

where $\delta := \sqrt{(p_{\varphi_1}\lambda)^2 - 8p_{\varphi_1}(\lambda-2)}$, *in which case*

$$c = \tfrac{2-\lambda}{2\lambda\gamma} + \tfrac{\sigma_{\varphi_1}}{\lambda}\left(\tfrac{1}{2} - \tfrac{\gamma L_{\varphi_1}}{\lambda}\right). \tag{6.15}$$

Proof. Let (u^+, v^+) be generated by one DRS iteration starting at s^+. Then,

$$\varphi_\gamma^{\mathrm{DR}}(s^+) = \min_{w\in\mathbb{R}^n} \left\{ \varphi_1(u^+) + \varphi_2(w) + \langle \nabla\varphi_1(u^+), w - u^+\rangle + \tfrac{1}{2\gamma}\|w-u^+\|^2 \right\}$$

and the minimum is attained at $w = v^+$. Therefore, letting ρ be as in Thm. 1.10,

$$\begin{aligned}
\varphi_\gamma^{\mathrm{DR}}(s^+) &\leq \varphi_1(u^+) + \langle \nabla\varphi_1(u^+), v-u^+\rangle + \varphi_2(v) + \tfrac{1}{2\gamma}\|u^+ - v\|^2 \\
&= \overbrace{\varphi_1(u^+) + \langle \nabla\varphi_1(u^+), u-u^+\rangle}^{\geq} + \langle \nabla\varphi_1(u^+), v-u\rangle + \varphi_2(v) \\
&\quad + \tfrac{1}{2\gamma}\|u^+ - v\|^2 \\
&\overset{\text{(Thm. 1.10)}}{\leq} \overbrace{\varphi_1(u) - \rho(u,u^+)} + \langle \nabla\varphi_1(u^+), v-u\rangle + \varphi_2(v) + \tfrac{1}{2\gamma}\|u^+ - v\|^2 \\
&= \varphi_1(u) - \rho(u,u^+) + \langle \nabla\varphi_1(u), v-u\rangle + \varphi_2(v) + \tfrac{1}{2\gamma}\|u^+ - v\|^2 \\
&\quad + \langle \nabla\varphi_1(u^+) - \nabla\varphi_1(u), v-u\rangle \\
&= \varphi_\gamma^{\mathrm{DR}}(s) - \rho(u,u^+) + \langle \nabla\varphi_1(u^+) - \nabla\varphi_1(u), v-u\rangle + \tfrac{1}{2\gamma}\|u-u^+\|^2 \\
&\quad + \tfrac{1}{\gamma}\langle u^+ - u, u - v\rangle.
\end{aligned}$$

[1] A one-line expression is $c = \frac{2-\lambda}{2\lambda\gamma} - \min\left\{ \frac{[p_{\varphi_1}]_-}{\lambda},\ L_{\varphi_1}\max\left\{ \frac{[\sigma_{\varphi_1}]_-}{2(1-[p_{\varphi_1}]_-)},\ \frac{1}{2} - \frac{\gamma L_{\varphi_1}}{\lambda} \right\} \right\}$.

Since $u - v = \frac{1}{\lambda}(s - s^+) = \frac{1}{\lambda}(u - u^+) + \frac{\gamma}{\lambda}(\nabla\varphi_1(u) - \nabla\varphi_1(u^+))$, as it follows from Thm. 1.15*(i)*, it all simplifies to

$$\varphi_\gamma^{\mathrm{DR}}(s) - \varphi_\gamma^{\mathrm{DR}}(s^+) \geq \tfrac{2-\lambda}{2\gamma\lambda}\|u-u^+\|^2 - \tfrac{\gamma}{\lambda}\|\nabla\varphi_1(u^+) - \nabla\varphi_1(u)\|^2 + \rho(u,u^+). \quad (6.16)$$

It will suffice to show that

$$\varphi_\gamma^{\mathrm{DR}}(s) - \varphi_\gamma^{\mathrm{DR}}(s^+) \geq c\|u - u^+\|^2;$$

inequality (6.12) will then follow from the $\frac{1}{1+\gamma L_{\varphi_1}}$-strong monotonicity of $\mathrm{prox}_{\gamma\varphi_1}$, see Thm. 1.15*(ii)*. We now proceed by cases.

♠ **Case 1:** $\lambda \in (0, 2)$**.**

Let $\sigma := -[\sigma_{\varphi_1}]_- = \min\{\sigma_{\varphi_1}, 0\}$ and $L \geq L_{\varphi_1}$ be such that $L + \sigma > 0$; the value of such an L will be fixed later. Then, $\sigma \leq 0$ and φ_1 is L-smooth and σ-hypoconvex. We may thus choose $\rho(u, u^+)$ as in Thm. 1.10*(ii)* with these values of L and σ. Inequality (6.16) then becomes

$$\tfrac{\varphi_\gamma^{\mathrm{DR}}(s) - \varphi_\gamma^{\mathrm{DR}}(s^+)}{L} \geq \left(\tfrac{2-\lambda}{2\lambda\xi} + \tfrac{p}{2(1+p)}\right)\|u^+ - u\|^2 + \tfrac{1}{L^2}\left(\tfrac{1}{2(1+p)} - \tfrac{\xi}{\lambda}\right)\|\nabla\varphi_1(u^+) - \nabla\varphi_1(u)\|^2,$$

where $\xi := \gamma L$ and $p := \sigma/L \in (-1, 0]$. Since $\nabla\varphi_1$ is L_{φ_1}-Lipschitz continuous, the claim holds provided that the constant

$$\frac{c}{L} = \begin{cases} \frac{2-\lambda}{2\lambda\xi} + \frac{p}{2(1+p)} & \text{if } 0 < \frac{1}{2(1+p)} - \frac{\xi}{\lambda}, \\ \frac{2-\lambda}{2\lambda\xi} + \frac{p}{2(1+p)} + \frac{L_{\varphi_1}^2}{L^2}\left(\frac{1}{2(1+p)} - \frac{\xi}{\lambda}\right) & \text{otherwise,} \end{cases} \quad (6.17)$$

is strictly positive. Now, let us consider two subcases:

- **Case 1a:** $0 < \lambda \leq 2(1 + \sigma/L_{\varphi_1})$.
  Then, $\sigma \geq -\frac{2-\lambda}{2}L_{\varphi_1} > -L_{\varphi_1}$ and we can take $L = L_{\varphi_1}$. Consequently, $p = \sigma/L_{\varphi_1}$, $\xi = \gamma L_{\varphi_1}$, and (6.17) becomes

  $$\frac{c}{L_{\varphi_1}} = \frac{2-\lambda}{2\lambda\gamma L_{\varphi_1}} + \begin{cases} \frac{p}{2(1+p)} & \text{if } \gamma < \frac{\lambda}{2(1+p)}, \\ \frac{1}{2} - \frac{\gamma L_{\varphi_1}}{\lambda} & \text{otherwise.} \end{cases} \quad (6.18)$$

  Let us verify that in this case any γ such that $\gamma < 1/L_{\varphi_1}$ yields a strictly positive coefficient c. If $0 < \gamma L_{\varphi_1} < \frac{\lambda}{2(1+p)} \leq 1$, then

  $$\tfrac{c}{L_{\varphi_1}} = \tfrac{2-\lambda}{2\lambda\gamma L_{\varphi_1}} + \tfrac{p}{2(1+p)} > \tfrac{2-\lambda}{2\lambda} + \tfrac{p}{\lambda} = \tfrac{1+p}{\lambda} - \tfrac{1}{2} \geq 0,$$

where in the inequality we used the fact that $\lambda < 2$ and $p \leq 0$. If instead $\frac{\lambda}{2(1+p)} < \gamma L_{\varphi_1} < 1$, then

$$\frac{c}{L_{\varphi_1}} = \frac{2-\lambda}{2\lambda\gamma L_{\varphi_1}} + \frac{1}{2} - \frac{\gamma L_{\varphi_1}}{\lambda} > \frac{2-\lambda}{2\lambda} + \frac{1}{2} - \frac{1}{\lambda} = 0.$$

Either way, the sufficient decrease constant c is strictly positive. Since $\sigma = -[\sigma_{\varphi_1}]_-$ and

$$\frac{2-\lambda}{2\lambda\gamma} + \frac{\sigma}{2(1+p)} \leq \frac{2-\lambda}{2\lambda\gamma} + \frac{L_{\varphi_1}}{2} - \frac{\gamma L_{\varphi_1}^2}{\lambda} \quad \Leftrightarrow \quad \gamma \leq \frac{\lambda}{2(L_{\varphi_1}+\sigma)},$$

from (6.18) we conclude that c is as in (6.12).

- **Case 1b:** $2(1 + \sigma/L_{\varphi_1}) < \lambda < 2$.
  Necessarily $\sigma < 0$, for otherwise the range of λ would be empty. In particular, $\sigma = \sigma_{\varphi_1}$, and the lower bound on λ can be expressed as $\sigma_{\varphi_1} < -\frac{2-\lambda}{2} L_{\varphi_1}$. Consequently, $L := \frac{-2\sigma_{\varphi_1}}{2-\lambda}$ is strictly larger than L_{φ_1}, and in particular $\sigma + L = \sigma_{\varphi_1} + L > 0$. The ratio of σ and L is thus $p = \frac{\lambda}{2} - 1$, and (6.17) becomes

$$c = \frac{2-\lambda}{2\lambda\gamma} + \begin{cases} \frac{\sigma_{\varphi_1}}{\lambda} & \text{if } \gamma < \frac{2-\lambda}{-2\sigma_{\varphi_1}} \\ \frac{\sigma_{\varphi_1}}{\lambda} - \frac{\gamma L_{\varphi_1}^2}{\lambda} + \frac{2-\lambda}{-2\sigma_{\varphi_1}\lambda} L_{\varphi_1}^2 & \text{otherwise.} \end{cases} \tag{6.19}$$

  Let us show that, when $\gamma < \frac{2-\lambda}{-2\sigma_{\varphi_1}} = \frac{1}{L}$, also in this case the sufficient decrease constant c is strictly positive. We have

$$\frac{c}{L} = \frac{2-\lambda}{2\lambda\gamma L} + \frac{\sigma_{\varphi_1}}{\lambda}\frac{1}{L} > \frac{2-\lambda}{2\lambda} + \frac{\sigma_{\varphi_1}}{\lambda}\frac{2-\lambda}{-2\sigma_{\varphi_1}} = 0,$$

  hence the claim. This concludes the proof for the case $\lambda \in (0, 2)$.

♠ **Case 2:** $\lambda \geq 2$.

In this case we need to assume that φ_1 is strongly convex, that is, that $\sigma_{\varphi_1} > 0$. Instead of considering a single expression of ρ, we will rather take a convex combination of those in Thm.s 1.10*(i)* and 1.6*(d)*, namely

$$\rho(u, u^+) = (1-\alpha)\frac{\sigma_{\varphi_1}}{2}\|u - u^+\|^2 + \alpha\frac{1}{2L_{\varphi_1}}\|\nabla\varphi_1(u) - \nabla\varphi_1(u^+)\|^2$$

for some $\alpha \in [0, 1]$ to be determined. (6.16) then becomes

$$\frac{\varphi_\gamma^{\mathrm{DR}}(s) - \varphi_\gamma^{\mathrm{DR}}(s^+)}{L_{\varphi_1}} \geq \left(\frac{2-\lambda}{2\lambda\xi} + \frac{(1-\alpha)p}{2}\right)\|u-u^+\|^2 + \frac{1}{L_{\varphi_1}^2}\left(\frac{\alpha}{2} - \frac{\xi}{\lambda}\right)\|\nabla\varphi_1(u) - \nabla\varphi_1(u^+)\|^2,$$

where $\xi := \gamma L_{\varphi_1}$ and $p := \sigma_{\varphi_1}/L_{\varphi_1} \in (0, 1]$. By restricting $\xi \in (0, 1)$, since $\lambda \geq 2$ one can take $\alpha := 2\xi/\lambda \in (0, 1)$ to make the coefficient multiplying the gradient

norm vanish. We then obtain

$$\tfrac{c}{L_{\varphi_1}} = \tfrac{2-\lambda}{2\lambda\xi} + \tfrac{(\lambda-2\xi)p}{2\lambda}. \tag{6.20}$$

Imposing $c > 0$ results in the following second-order equation in variable ξ,

$$2p\xi^2 - p\lambda\xi + (\lambda - 2) < 0. \tag{6.21}$$

The discriminant is $\Delta := (p\lambda)^2 - 8p(\lambda - 2)$, which, for $\lambda \geq 2$, is strictly positive iff

$$2 \leq \lambda < \tfrac{4}{1+\sqrt{1-p}} \quad \vee \quad \lambda > \tfrac{4}{1-\sqrt{1-p}}.$$

Denoting $\delta := \sqrt{\Delta} = \sqrt{(p\lambda)^2 - 8p(\lambda-2)}$, the solution to (6.21) is $\frac{p\lambda-\delta}{4p} < \xi < \frac{p\lambda+\delta}{4p}$. However, the case $\lambda \geq 4$ has to be discarded, as $\frac{p\lambda-\delta}{4p} > 1$ in this case, contradicting the fact that $p \leq 1$. To see this, suppose $\lambda \geq 4$. Then,

$$\begin{aligned} \tfrac{p\lambda-\delta}{4p} < 1 \quad &\Leftrightarrow \quad p(\lambda-4) < \delta \\ &\Leftrightarrow \quad p^2(\lambda-4)^2 < \Delta = (p\lambda)^2 - 8p(\lambda-2) \\ &\Leftrightarrow \quad p(2-\lambda) < 2 - \lambda, \end{aligned}$$

hence $p > 1$, which contradicts the fact that $\sigma_{\varphi_1} \leq L_{\varphi_1}$. Thus, the only feasible ranges are the ones given in (6.14), hence the claimed sufficient decrease constant c, cf. (6.20). □

Corollary 6.10. *Under Assumption 6.I, for any stepsize γ and relaxation λ as in Theorem 6.9, the DRE $\varphi_\gamma^{\mathrm{DR}}$ is a Lyapunov function for the fixed-point iterations of $\mathcal{F}_\gamma^{\mathrm{DR}_\lambda}$.*

6.4 Convergence results

Theorem 6.11 (Finite termination of relaxed DRS)**.** *The iterates generated by DRS (Alg. 6.1) satisfy*

$$\varphi_\gamma^{\mathrm{DR}}(s^{k+1}) \leq \varphi_\gamma^{\mathrm{DR}}(s^k) - \tfrac{c\lambda^2}{(1+\gamma L_{\varphi_1})^2}\|u^k - v^k\|^2,$$

where $c > 0$ is as in Theorem 6.9. In particular, the algorithm terminates in a finite number of iterations and yields a point x_* satisfying $\operatorname{dist}(0, \hat{\partial}\varphi(x_*)) \leq \varepsilon$.

Proof. Follows from Thm. 3.22 in light of Cor. 3.20 and Lem. 6.3. The fact that $\operatorname{dist}(0, \hat{\partial}\varphi(x_*)) \leq \varepsilon$ follows the same arguments as in the proof of Thm.

Algorithm 6.1. DOUGLAS-RACHFORD SPLITTING WITH RELAXATION

REQUIRE
- initial iterate $s_0 \in \mathbb{R}^n$;
- tolerance $\varepsilon > 0$;
- stepsize and relaxation $\gamma, \lambda > 0$ as follows:

$$\begin{cases} 0 < \gamma < \min\left\{\frac{2-\lambda}{2[\sigma_{\varphi_1}]_-}, \frac{1}{L_{\varphi_1}}\right\} & \text{and} \quad \lambda \in (0,2), \quad \text{or} \\ \frac{p\lambda-\delta}{4\sigma_{\varphi_1}} < \gamma < \frac{p\lambda+\delta}{4\sigma_{\varphi_1}} & \text{and} \quad 2 \leq \lambda < \frac{4}{1+\sqrt{1-p}}, \end{cases}$$

where $p := {}^{\sigma_{\varphi_1}}\!/_{L_{\varphi_1}}$ and $\delta := \sqrt{(p\lambda)^2 - 8p(\lambda-2)}$.

PROVIDE x_* with $\varphi_\gamma^{\mathrm{DR}}(x_*) \leq \varphi_\gamma^{\mathrm{DR}}(x^0)$ and $\operatorname{dist}(0, \hat{\partial}\varphi(x_*)) \leq \varepsilon$.

1: **for** $k = 0, 1, \ldots$ **do**
2: $\quad u^k = \operatorname{prox}_{\gamma\varphi_1}(s^k)$
3: $\quad v^k \in \operatorname{prox}_{\gamma\varphi_2}(2u^k - s^k)$
4: $\quad$ **if** $\frac{1}{2\gamma}\|u^k - v^k\| \leq \varepsilon$ **then**
5: $\quad\quad$ **return** $x_* = v^k$
6: $\quad s^{k+1} = s^k + \lambda(v^k - u^k)$

5.16. □

Theorem 6.12 (Asymptotic convergence of relaxed DRS)**.** *Suppose that Assumption 6.1 is satisfied, and consider the iterates generated by DRS (Alg. 6.1) with tolerance $\varepsilon = 0$. The following hold:*

(i) The Douglas-Rachford residual $(\|u^k - v^k\|)_{k\in\mathbb{N}}$ is square-summable; in particular, $\min_{j\leq k}\|u^j - v^j\| \in O(1/\sqrt{k})$.

(ii) $(u^k)_{k\in\mathbb{N}}$ and $(v^k)_{k\in\mathbb{N}}$ have the same cluster points, all of which are stationary for φ and on which φ has the same value, this being the limit of $(\varphi_\gamma^{\mathrm{DR}}(s^k))_{k\in\mathbb{N}}$. In fact, the set ω of accumulation points of $(v^k)_{k\in\mathbb{N}}$ (and of $(u^k)_{k\in\mathbb{N}}$) satisfies $\omega \subseteq \operatorname{fix}\mathcal{T}_\gamma^{\mathrm{FB}} \subseteq \operatorname{zer}\hat{\partial}\varphi$.

(iii) If φ is level bounded, then $(s^k, u^k, v^k)_{k\in\mathbb{N}}$ is bounded, and ω is a nonempty, compact and connected set satisfying $\operatorname{dist}(v^k, \omega) \to 0$ as $k \to \infty$.

(iv) $\varphi_\gamma^{\mathrm{DR}} \equiv \varphi$ on $(\mathrm{id} + \gamma\nabla\varphi_1)(\omega)$, the value being the limit of the (decreasing) sequence $(\varphi_\gamma^{\mathrm{DR}}(s^k))_{k\in\mathbb{N}}$ (or, equivalently, of $(\varphi(v^k))_{k\in\mathbb{N}}$).

Proof. Follows from Thm. 3.22 in light of Cor. 6.10. □

Since both $\operatorname{prox}_{\gamma\varphi_1}$ and its inverse are (continuous and) strongly monotone, it is easy to verify that $\varphi_\gamma^{\mathrm{DR}}$ has the KL property iff so does $\varphi_\gamma^{\mathrm{FB}}$; the same arguments extend to the Lojasiewicz property, in which case the exponent is also preserved.

Theorem 6.13 (Global convergence of relaxed DRS)**.** *Suppose that Assumption 6.1 is satisfied, and consider the iterates generated by DRS (Alg. 6.1) with tolerance $\varepsilon = 0$. Suppose further that the following hold:*

A1 *φ is level bounded.*

A2 *All accumulation points of the sequence are prox-regular, in the sense of Definition 5.10 (with $f = \varphi_1$ and $g = \varphi_2$).*

A3 *$\varphi_\gamma^{\mathrm{DR}}$ (or, equivalently, $\varphi_\gamma^{\mathrm{FB}}$) has the KL property.*

Then, the following hold:

(i) $(v^k)_{k\in\mathbb{N}}$ converges to a point $x_\star \in \operatorname{fix} T_\gamma^{\mathrm{FB}} \subseteq \operatorname{zer} \hat{\partial}\varphi$.

(ii) The Douglas-Rachford residual $(\|u^k - v^k\|)_{k\in\mathbb{N}}$ is summable, and in particular $\min_{j\le k} \|u^j - v^j\| \in O(1/k)$.

Proof. Follows from Thm. 3.23, in light of Cor. 6.10 and Thm. 5.11*(ii)* (since $\varphi_\gamma^{\mathrm{FB}}$ is the $\mathcal{M}$-envelope of DRS). □

Theorem 6.14 (Linear convergence of relaxed DRS)**.** *Suppose that Assumption 6.1 is satisfied, and consider the iterates generated by DRS (Alg. 6.1) with tolerance $\varepsilon = 0$. Suppose further that the following hold:*

A1 *φ is level bounded;*

A2 *All accumulation points of the sequence are prox-regular, in the sense of Definition 5.10 (with $f = \varphi_1$ and $g = \varphi_2$).*

A3 *$\varphi_\gamma^{\mathrm{DR}}$ (or, equivalently, $\varphi_\gamma^{\mathrm{FB}}$) has the KL property and the KL function is of the form $\psi(s) = cs^\vartheta$ for some $c > 0$ and $\vartheta \ge 1/2$.*

Then, the sequences $(v^k)_{k\in\mathbb{N}}$ and $(\operatorname{dist}(0, \mathcal{R}_\gamma^{\mathrm{DR}}(s^k)))_{k\in\mathbb{N}}$ are R-linearly convergent.

Proof. Follows from Thm. 3.24, in light of Cor. 6.10 and Thm. 5.11*(ii)* (since $\varphi_\gamma^{\mathrm{FB}}$ is the $\mathcal{M}$-envelope of DRS). □

6.4.1 Tightness of the ranges

When both φ_1 and φ_2 are convex and $\varphi_1 + \varphi_2$ attains a minimum, well-known results of monotone operator theory guarantee that for any $\lambda \in (0, 2)$ and $\gamma > 0$ the residual $u^k - v^k$ generated by DRS iterations vanishes (see *e.g.*, [10, Cor. 28.3]). In fact, the whole sequence $(u^k)_{k\in\mathbb{N}}$ converges and φ_1 needs not even be differentiable in this case. On the contrary, when φ_2 is nonconvex then the bound $\gamma < 1/L_{\varphi_1}$ plays a crucial role, as the next example shows.

Theorem 6.15 (Necessity of $\gamma < 1/L_{\varphi_1}$)**.** *For any $L > 0$ and $\sigma \in [-L, L]$ there exist $\varphi_1, \varphi_2 : \mathbb{R}^p \to \overline{\mathbb{R}}$ satisfying the following properties*

P1 *φ_1 is L-smooth and σ-hypoconvex;*

P2 *φ_2 is proper and lsc;*

P3 $\operatorname{argmin}(\varphi_1 + \varphi_2) \neq \emptyset$*;*

P4 *for all $s^0 \in \mathbb{R}^p$, $\gamma \geq 1/L$, and $\lambda > 0$, the sequence $(s^k)_{k\in\mathbb{N}}$ generated by DRS iterations with stepsize γ and relaxation λ starting from s^0 satisfies $\|s^k - s^{k+1}\| \not\to 0$ as $k \to \infty$.*

Proof. Fix $t > 1$, and let $\varphi = \varphi_1 + \varphi_2$, where $\varphi_2 = \delta_{\{\pm 1\}}$ and

$$\varphi_1(x) = \begin{cases} \frac{L}{2}x^2 & \text{if } x \leq t, \\ \frac{L}{2}x^2 - \frac{L-\sigma}{2}(x-t)^2 & \text{otherwise.} \end{cases} \tag{6.22}$$

Notice that $\operatorname{dom}\varphi = \{\pm 1\}$, and therefore ± 1 are the unique stationary points of φ (in fact, they are also global minimizers). It can be easily verified that φ_1 and φ_2 satisfy properties 6.15.P1, 6.15.P2 and 6.15.P3. Moreover, $\operatorname{prox}_{\gamma\varphi_1}$ is well defined iff $\gamma < 1/[\sigma]_-$, in which case

$$\operatorname{prox}_{\gamma\varphi_1}(s) = \begin{cases} \frac{s}{1+\gamma L} & \text{if } s \leq t(1+\gamma L), \\ \frac{s-\gamma(L-\sigma)t}{1+\gamma\sigma} & \text{otherwise,} \end{cases} \tag{6.23}$$

$\operatorname{prox}_{\gamma\varphi_2} = \operatorname{sgn}$, where $\operatorname{sgn}(0) = \{\pm 1\}$. Let now $s^0 \in \mathbb{R}^p$, $1/L \leq \gamma < 1/[\sigma]_-$, and $\lambda > 0$ be fixed, and consider a sequence $(s^k)_{k\in\mathbb{N}}$ generated by DRS with stepsize γ and relaxation λ, starting at s^0. To arrive to a contradiction, suppose that $\|s^k - s^{k+1}\| = \lambda\|u^k - v^k\| \to 0$ as $k \to \infty$. For any $k \in \mathbb{N}$ we have $v^k = -\operatorname{sgn}(s^k)$ if $s^k \leq t(1+\gamma L)$, resulting in

$$u^k - v^k \in \begin{cases} \frac{s^k}{1+\gamma L} + \operatorname{sgn}(s^k) & \text{if } s^k \leq t(1+\gamma L), \\ \frac{s^k}{1+\gamma\sigma} - \frac{\gamma(L-\sigma)t}{1+\gamma\sigma} - v^k & \text{otherwise,} \end{cases}$$

where v^k is either 1 or -1 in the second case. Since $u^k - v^k \to 0$, then

$$\min\left\{\left|\tfrac{s^k}{1+\gamma L} + \operatorname{sgn}(s^k)\right|, \left|\tfrac{s^k}{1+\gamma\sigma} - \tfrac{L-\sigma}{1+\gamma\sigma}\gamma t - 1\right|, \left|\tfrac{s^k}{1+\gamma\sigma} - \tfrac{L-\sigma}{1+\gamma\sigma}\gamma t + 1\right|\right\} \to 0.$$

Notice that the first element in the set above is always larger than 1, and therefore eventually s^k will be always close to either $(L-\sigma)\gamma t + (1+\gamma\sigma)$ or $(L-\sigma)\gamma t - (1+\gamma\sigma)$, both of which are strictly smaller than $t(1+\gamma L)$ (since $t > 1$). Therefore, eventually $s^k \leq t(1+\gamma L)$ and the residual will then be $u^k - v^k = \frac{s^k}{1+\gamma L} + \operatorname{sgn}(s^k)$ which is bounded away from zero, contradicting the fact that $u^k - v^k \to 0$. □

Theorem 6.16 (Necessity of $0 < \lambda < 2(1+\gamma\sigma)$)**.** *For any $L > 0$ and $\sigma \in [-L, L]$ there exist $\varphi_1, \varphi_2 : \mathbb{R}^p \to \overline{\mathbb{R}}$ satisfying the following properties*

P1 *φ_1 is L-smooth and σ-hypoconvex;*

P2 *φ_2 is proper, lsc, and strongly convex;*

P3 $\operatorname{arg\,min}(\varphi_1 + \varphi_2) \neq \emptyset$*;*

P4 *for all s^0, $0 < \gamma < 1/L$, and $\lambda > 2(1+\gamma\sigma)$, the sequence $(s^k)_{k\in\mathbb{N}}$ generated by DRS with stepsize γ and relaxation λ starting from s^0 satisfies $\|s^k - s^{k+1}\| \not\to 0$ as $k \to \infty$ (unless s^0 is a fixed point for DRS).*

Proof. Let $\varphi = \varphi_1 + \varphi_2$, where φ_1 is as in (6.22) with $t = 1$, and $\varphi_2 = \delta_{\{p\}}$ for some $p > 1$. Let $\gamma < 1/L$, $\lambda \geq 2(1+\gamma\sigma)$. Starting from $s^0 \neq (1+\gamma\sigma)p + \gamma(L-\sigma)$ (so that $u^0 \neq p$), consider DRS with stepsize γ and relaxation λ. To arrive to a contradiction, suppose that the residual vanishes. Since $v^k = \operatorname{prox}_{\gamma\varphi_2}(2u^k - s^k) = p$, necessarily $u^k \to p$; therefore, eventually $u^k > 1$ and in particular

$$u^{k+1} + \gamma\tfrac{L-\sigma}{1+\gamma\sigma} = \tfrac{1}{1+\gamma\sigma}s^{k+1} = \tfrac{1}{1+\gamma\sigma}(s^k + \lambda(p-u^k)) = u^k + \gamma\tfrac{L-\sigma}{1+\gamma\sigma} + \tfrac{\lambda}{1+\gamma\sigma}(p-u^k),$$

where the identity $s^k = (1+\gamma\sigma)u^k + \gamma(L-\sigma)$ was used, cf. (6.23). Therefore,

$$\left|u^{k+1} - p\right| = \left|1 - \tfrac{\lambda}{1+\gamma\sigma}\right|\left|u^k - p\right| \geq \left|u^k - p\right|,$$

where the inequality is due to the fact that $\lambda \geq 2(1+\gamma\sigma)$. Since $u^0 \neq p$ due to the choice of s^0, apparently $(u^k)_{k\in\mathbb{N}}$ is bounded away from p, hence the contradiction. □

Let us draw some conclusions:

- The nonsmooth function φ_2 is (strongly) convex in Theorem 6.16, therefore even for fully convex formulations the bound $0 < \lambda < 2(1 + \gamma\sigma_{\varphi_1})$ needs be satisfied.
- If $\lambda > 2$ (which is feasible only if φ_1 is strongly convex, *i.e.*, if $\sigma_{\varphi_1} > 0$), then, regardless of whether also φ_2 is (strongly) convex or not, we obtain that *the stepsize must be lower bounded as* $\gamma > \frac{\lambda-2}{2\sigma_{\varphi_1}}$. In the more general setting of σ-strongly monotone operators in Hilbert spaces, hence $\sigma \geq 0$, the similar bound $\lambda < \min\{2(1+\gamma\sigma), 2+\gamma\sigma+{}^1\!/_{\gamma\sigma}\}$ has been recently established in [82].
- Combined with the bound $\gamma < {}^1\!/_{L_{\varphi_1}}$ shown in Theorem 6.15, we infer that (at least when φ_2 is nonconvex) necessarily $0 < \lambda < 2(1 + {}^{\sigma_{\varphi_1}}\!/_{L_{\varphi_1}})$ and consequently $\lambda \in (0, 4)$.

Theorem 6.17 (Tightness)**.** *Unless the generality of Assumption 6.I is sacrificed, when $\lambda \in (0,2)$ or φ_1 is not strongly convex, $\gamma < \min\left\{\frac{1}{L_{\varphi_1}}, \frac{2-\lambda}{2[\sigma_{\varphi_1}]_-}\right\}$ is a tight bound for ensuring convergence of DRS. Similarly, PRS (i.e., DRS with $\lambda = 2$) is ensured to converge iff φ_1 is strongly convex and $\gamma < {}^1\!/_{L_{\varphi_1}}$.*

6.5 A quasi-Newton DRS

Theorem 6.18 (CLyD-DRS (nonmonotone): subsequential convergence)**.** *Suppose that Assumption 6.I is satisfied. Then, the following hold for the iterates generated by CLyD-DRS (Alg. 6.2) with tolerance $\varepsilon = 0$:*

(i) The residual $(\|u^k - v^k\|)_{k\in\mathbb{N}}$ is square-summable; in particular, it vanishes with rate $\min_{j\leq k} \operatorname{dist}(s^j, \mathcal{F}^{\mathrm{DR}\lambda}_\gamma(s^j)) \in O({}^1\!/_{\sqrt{k}})$.

(ii) The set ω of accumulation points of $(u^k)_{k\in\mathbb{N}}$ satisfies $\omega \subseteq \operatorname{fix}\mathcal{T}^{\mathrm{DR}}_\gamma \subseteq \operatorname{zer}\hat{\partial}\varphi$.

If, additionally, $\|d^k\| \to 0$ as $k \to \infty$, then the following also hold:

(iii) If φ is level bounded, then $(u^k)_{k\in\mathbb{N}}$ and $(v^k)_{k\in\mathbb{N}}$ are bounded, and ω is a nonempty, compact and connected set satisfying $\operatorname{dist}(v^k, \omega) \to 0$ as $k \to \infty$.

(iv) $\varphi^{\mathrm{DR}}_\gamma \equiv \varphi$ on ω, the value being the limit of the (decreasing) sequence $(\varphi^{\mathrm{DR}}_\gamma(s^k))_{k\in\mathbb{N}}$.

Algorithm 6.2. CLyD-DRS

REQUIRE
- stepsize γ, relaxation λ, and sufficient decrease constant c as in Thm. 6.9
- scaling factor $\alpha \in (0,1)$ for sufficient decrease constant
- initial iterate $s^0 \in \mathbb{R}^n$
- tolerance $\varepsilon > 0$

PROVIDE x_* with $\operatorname{dist}\big(0, \partial\varphi(x_*)\big) \leq \varepsilon$

1: **for** $k = 0, 1, 2, \ldots$ **do**
2: Do one nominal DR-step:
$$u^k = \operatorname{prox}_{\gamma\varphi_1}(s^k)$$
$$v^k \in \operatorname{prox}_{\gamma\varphi_2}(2u^k - s^k)$$
$$\bar{s}^k = s^k + \lambda(v^k - u^k)$$
3: **if** $\frac{1}{2\gamma}\|u^k - v^k\| \leq \varepsilon$ **then return** $x_* = v^k$
4: Select an update direction $d^k \in \mathbb{R}^n$ at s^k
5: Let $\tau_k \in \{2^{-i} \mid i \in \mathbb{N}\}$ be the largest such that

$$\varphi_\gamma^{\mathrm{DR}}(s^{k+1}) \leq \varphi_\gamma^{\mathrm{DR}}(s^k) - \tfrac{\alpha c \lambda^2}{(1+\gamma L_{\varphi_1})^2}\|u^k - v^k\|^2, \tag{6.24}$$

where $s^{k+1} := (1-\tau_k)\bar{s}^k + \tau_k(s^k + d^k)$

All the claims remain valid if the linesearch condition (6.24) is replaced by the following nonmonotone version:

$$\varphi_\gamma^{\mathrm{DR}}(s^{k+1}) \leq \bar{\mathcal{L}}_k - \tfrac{\alpha c \lambda^2}{(1+\gamma L_{\varphi_1})^2}\|u^k - v^k\|^2, \tag{6.25}$$

where, for any sequence $(p_k)_{k\in\mathbb{N}} \subseteq [0,1]$ bounded away from 0, $\bar{\mathcal{L}}_k$ are recursively defined as follows:

$$\bar{\mathcal{L}}_k := \begin{cases} \varphi_\gamma^{\mathrm{DR}}(s^0) & \text{if } k = 0, \\ (1-p_k)\bar{\mathcal{L}}_{k-1} + p_k\varphi_\gamma^{\mathrm{DR}}(s^k) & \text{otherwise.} \end{cases}$$

Proof. Follows from Theorem 4.1. □

6.5.1 Global and (super)linear convergence

Theorem 6.19 (Global convergence)**.** *Additionally to Assumption 6.I, suppose that the following hold for the iterates generated by CLyD-DRS (Alg. 6.2) with*

tolerance $\varepsilon = 0$.

A1 φ *is level bounded.*

A2 *All accumulation points of the sequence* $(v^k)_{k\in\mathbb{N}}$ *are prox-regular, in the sense of Definition 5.10.*

A3 *The DRE* $\varphi_\gamma^{\mathrm{DR}}$ *has the KL property.*

A4 *There exists* $D > 0$ *such that* $\|d^k\| \leq D\|u^k - v^k\|$ *for all* k*'s.*

Then, the following hold:

(i) $(v^k)_{k\in\mathbb{N}}$ *converges to a point* $x_\star \in \operatorname{fix} \mathcal{T}_\gamma^{\mathrm{DR}} \subseteq \operatorname{zer} \hat{\partial}\varphi$.

(ii) The residual is summable and in particular $\min_{j\leq k} \operatorname{dist}(s^j, \mathcal{F}_\gamma^{\mathrm{DR}\lambda}(s^j)) \in O(1/k)$.

Proof. Follows from Theorems 5.11*(ii)* and 4.2 (since $\varphi_\gamma^{\mathrm{FB}}$ is the $\mathcal{M}$-envelope of DRS). □

Theorem 6.20 (Linear convergence)**.** *Suppose that the assumptions of Theorem 6.19 are satisfied, and that the KL function can be taken of the form* $\psi(s) = cs^\vartheta$ *for some* $c > 0$ *and* $\vartheta \geq 1/2$*. Then, the sequences* $(s^k)_{k\in\mathbb{N}}$, $(v^k)_{k\in\mathbb{N}}$, *and* $\operatorname{dist}(s^k, \mathcal{F}_\gamma^{\mathrm{DR}\lambda}(s^k))$ *are R-linearly convergent.*

Proof. Follows from Theorem 4.3. □

Theorem 6.21 (Acceptance of the unit stepsize and superlinear convergence)**.** *Consider the iterates generated by CLyD-DRS (Alg. 6.2). Additionally to Assumption 6.I, suppose that the following hold:*

A1 $(u^k)_{k\in\mathbb{N}}$ *converges to a strong local minimum* $u_\star$ *of* φ.

A2 $(d^k)_{k\in\mathbb{N}}$ *are superlinearly convergent directions with respect to* $(s^k)_{k\in\mathbb{N}}$.

A3 $\gamma \neq \Gamma^{\mathrm{DR}}(u_\star)$.

Then, there exists $\bar{k} \in \mathbb{N}$ *such that*

$$\varphi_\gamma^{\mathrm{DR}}(s^k + d^k) \leq \varphi_\gamma^{\mathrm{DR}}(s^k) - \tfrac{\alpha c \lambda^2}{(1+\gamma L_{\varphi_1})^2}\|u^k - v^k\|^2 \quad \text{for all } k \geq \bar{k}.$$

In particular, eventually the iterates reduce to $s^{k+1} = s^k + d^k$ *and converge superlinearly.*

Proof. Follows from Theorem 4.5. □

Theorem 6.22 (Dennis-Moré condition)**.** *Consider the iterates generated by CLyD-DRS (Alg. 6.2). Additionally to Assumption 6.I, suppose that the following hold:*

A1 $(u^k)_{k\in\mathbb{N}}$ *converges to a strong local minimum* $u_\star$ *at which Assumption 5.II is (strictly) satisfied (with* $f = \varphi_1$ *and* $g = \varphi_2$*).*

A2 *The Dennis-Moré condition holds:*

$$\lim_{k\to\infty} \frac{\|\mathcal{R}_\gamma^{\mathrm{DR}}(s^k) + J\mathcal{R}_\gamma^{\mathrm{DR}}(x_\star)d^k\|}{\|d^k\|} = 0. \tag{6.26}$$

Then, $(d^k)_{k\in\mathbb{N}}$ *are superlinearly convergent directions with respect to* $(s^k)_{k\in\mathbb{N}}$*.*

Proof. Follows from the same arguments of Thm. 5.24. □

Theorem 6.23 (Superlinear convergence with Broyden directions)**.** *Consider the iterates generated by CLyD-DRS (Alg. 6.2) with directions* d^k *selected with Broyden method* (4.5)*. Additionally to Assumption 6.I, suppose that the following hold:*

A1 $(u^k)_{k\in\mathbb{N}}$ *converges to a point* $u_\star$ *at which* $\mathcal{R}_\gamma^{\mathrm{DR}}$ *is Lipschitz-continuously semidifferentiable and with nonsingular Jacobian* $J\mathcal{R}_\gamma^{\mathrm{DR}}(u_\star)$ *(in particular,* $\mathcal{R}_\gamma^{\mathrm{DR}}$ *is strictly differentiable there).*

Then, the Dennis-Moré condition (6.26) *is satisfied, and in particular all the claims of Theorem 6.22 hold.*

Proof. Follows from the same arguments of Thm. 5.25. □

Chapter 7

Alternating direction method of multipliers

7.1 Introduction

Closely related to DRS and possibly even more popular is the ALTERNATING DIRECTION METHOD OF MULTIPLIERS (ADMM), first appeared in [51, 47], see also [50] for a recent historical overview. ADMM addresses linearly constrained optimization problems

$$\underset{(x,z)\in\mathbb{R}^m\times\mathbb{R}^n}{\text{minimize}}\; f(x)+g(z) \quad \text{subject to } Ax+Bz=b, \tag{7.1}$$

where $f:\mathbb{R}^m\to\overline{\mathbb{R}}$, $g:\mathbb{R}^n\to\overline{\mathbb{R}}$, $A\in\mathbb{R}^{p\times m}$, $B\in\mathbb{R}^{p\times n}$, and $b\in\mathbb{R}^p$. Starting from some $z\in\operatorname{dom} g$ and $y\in\mathbb{R}^p$, one ADMM iteration amounts to the following steps:

$$\begin{cases} x^+ \in \operatorname{arg\,min} \mathscr{L}_\beta(\,\cdot\,, z, y) \\ y^+ = y+\beta(Ax^+ + Bz - b) \\ z^+ \in \operatorname{arg\,min} \mathscr{L}_\beta(x^+, \,\cdot\,, y^+). \end{cases} \tag{7.2}$$

Here, the PENALTY parameter $\beta>0$ plays the role of a stepsize, and

$$\mathscr{L}_\beta(x,z,y) \coloneqq f(x)+g(z)+\langle y, Ax+Bz-b\rangle + \tfrac{\beta}{2}\|Ax+Bz-b\|^2 \tag{7.3}$$

is the β-augmented Lagrangian of (7.1) with $y\in\mathbb{R}^p$ as Lagrange equality multiplier. For convex problems ADMM is DRS applied to a dual formulation [46], and its convergence properties for arbitrary penalty parameters $\beta>0$ are well documented in the literature, see *e.g.*, [27]. ADMM can be seen as fixed-point iterations on the Lagrange multiplier y, with x and z serving as intermediate variables as u and v do in DRS iterations, and for this reason the y-update is usually the last. For reasons that will soon be clear, as well as

to preserve the alphabetical order of the variables, we consider this "shifted" version; the classical ADMM is recovered by simply starting the update from the z-update.[1]

Relaxing ADMM with some parameter λ requires the introduction of an in-between variable $y^{+/2}$. For the sake of completeness we will thus study the following more general formulation:

$$\begin{cases} y^{+/2} = y - \beta(1-\lambda)(Ax + Bz - b) \\ x^+ \in \operatorname{arg\,min} \mathscr{L}_\beta(\,\cdot\,, z, y^{+/2}) \\ y^+ = y^{+/2} + \beta(Ax^+ + Bz - b) \\ z^+ \in \operatorname{arg\,min} \mathscr{L}_\beta(x^+, \,\cdot\,, y^+). \end{cases} \tag{ADMM}$$

Clearly, when $\lambda = 1$ (that is, in absence of relaxation) one has $y^{+/2} = y$ and the scheme reduces to the unrelaxed ADMM version (7.2).

As detailed in the next section, the numerous attempts to extend the applicability of ADMM to nonconvex problems brought forth a patchwork of standalone results, possibly involving implicit constants and burdened with non-trivial assumptions. However, hidden in the convex setting the recent work [125] established a universal primal equivalence of ADMM and DRS: under no assumptions, one ADMM update can be retrieved by one of DRS applied to an equivalent problem reformulation. This is the milestone of our approach, as the analysis of ADMM can be simplified down to that of DRS, well covered in the previous chapter.

7.1.1 Overview on nonconvex ADMM

Before proceeding with our analysis, let us briefly summarize some related work on nonconvex ADMM.

A primal equivalence of DRS and ADMM has been observed in [11, Rem. 3.14] when $A = -B = \mathrm{I}$ and $\lambda = 1$. In [125, Thm. 1] the equivalence is extended to arbitrary matrices; although limited to convex problems, the result is easily extendable. Our generalization to any relaxation parameter (and nonconvex problems) is largely based on this result and uses the same problem reformulation proposed therein. The relaxation considered in this paper corresponds to that introduced in [42]; it is worth mentioning that another type of relaxation

[1] Conventionally, updates follow the order (x, z, y), whereas the shifted version of ADMM here proposed would update z first, then x, and lastly y. Of course, it is simply a matter of swapping the *primal* variables x and z and the functions f and g in problem formulation (7.1), hence there is really no loss of generality in the update order adopted here.

has been proposed, corresponding to $\lambda = 1$ in (ADMM) but with a different steplength for the y-update: that is, with β replaced by $\theta\beta$ for some $\theta > 0$. The known convergence results for $\theta \in (0, \frac{1+\sqrt{5}}{2})$ in the convex case, see [49, §5], were recently extended to nonconvex problems and for $\theta \in (0, 2)$ in [52].

In [123] convergence of ADMM is studied for problems of the form

$$\underset{\boldsymbol{x}=(x_0 \dots x_p),z}{\text{minimize}} \; g(\boldsymbol{x}) + \textstyle\sum_{i=0}^{p} f_i(x_i) + h(z) \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} + Bz = 0.$$

Although addressing a more general class of problem than (7.1), when specialized to the standard two-function formulation analyzed in this paper it relies on numerous assumptions. These include Lipschitz continuous minimizers of all ADMM subproblems (in particular, uniqueness of their solution), whereas we allow for multiple minimizers and make almost no requirement on the nonsmooth function. For instance, the requirements rule out interesting cases involving discrete variables or rank constraints.

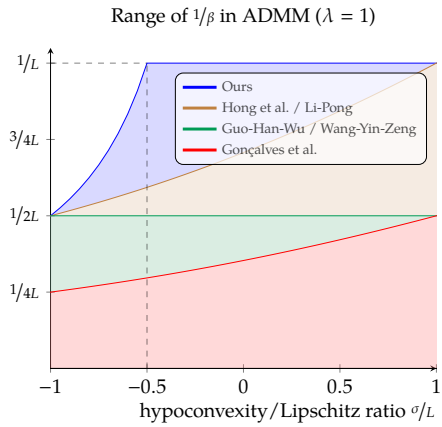


Figure 7.1: *Maximum inverse of the penalty paramenter $1/\beta$ ensuring convergence of ADMM; comparison between our bounds (blue plot) and [52, 53, 58, 70, 123]. On the x-axis the ratio between hypoconvexity and smoothness moduli of the image function (Af). The analysis is made for a common framework: 2-block ADMM with no Bregman or proximal terms, A invertible and B identity. Notice that, due to the proved analogy of DRS and ADMM, our theoretical bounds coincide in Fig. 6.1a and 7.1.*

In [58] a class of nonconvex problems with more than two functions is presented and variants of ADMM with deterministic and random updates are discussed. The paper provides a nice theory and explicit bounds for the penalty paramenter in ADMM, which agree with ours in worst- and best-case scenarios but are more restrictive otherwise (cf. Fig. 7.1 for a more detailed comparison). The main limitation of the proposed approach, however, is that the theory only allows for functions either convex or smooth, differently from ours where the nonsmooth term can basically be any function. Once again, many interesting applications are not covered.

The work [70] studies a proximal ADMM where a possible Bregman divergence

term in the second block update is considered. By discarding the Bregman term so as to recover the original ADMM scheme, the same bound on the stepsize as in [58] is found. Another proximal variant is proposed in [52], under less restrictive assumptions related to the concept of smoothness relative to a matrix that we will introduce in Definition 7.7. When matrix B has full-column rank, the proximal term can be discarded and their method reduces to the classical ADMM.

The problem addressed in [53] is fully covered by our analysis, as they consider ADMM for (7.1) where f is L-Lipschitz continuously differentiable and B is the identity matrix. Their bound $\beta > 2L$ for the penalty parameter is more conservative than ours; in fact, the two coincide only in worst-case scenarios.

7.2 A universal equivalence of ADMM and DRS

In this section we show step by step how to express ADMM as DRS on an equivalent problem, under no convexity assumptions and for arbitrary penalty relaxation parameters. We will pattern the arguments in [125, Thm. 1]; in doing so, it will be important to understand how variables in both algorithms are related.

7.2.1 An unconstrained problem reformulation

We start by eliminating the linear coupling between x and z and bring the problem into the form (P) addressed by DRS. To this end, let us introduce a slack variable $s \in \mathbb{R}^p$ and rewrite (7.1) as

$$\underset{x,z,s}{\text{minimize}} \;\; f(x) + g(z) \quad \text{subject to} \;\; Ax = s, \;\; Bz = b - s.$$

Since the problem is independent of the order of minimization [106, Prop. 1.35] we may minimize first with respect to (x, z) to arrive to

$$\underset{s \in \mathbb{R}^p}{\text{minimize}} \;\; \inf_{x \in \mathbb{R}^m} \{f(x) \mid Ax = s\} + \inf_{z \in \mathbb{R}^n} \{g(z) \mid Bz = b - s\}.$$

The two parametric infima define two image functions, cf. Definition 1.16: indeed, ADMM problem formulation (7.1) can be expressed as

$$\underset{s \in \mathbb{R}^p}{\text{minimize}} \;\; (Af)(s) + (Bg)(b - s), \tag{7.4}$$

which is exactly (P) with $\varphi_1 = (Af)$ and $\varphi_2 = (Bg)(b - \cdot)$. Apparently, unless A and B are injective the correspondence between variable s in (7.4) and variables x, z in (7.1) may fail to be one to one, as s is associated to sets of variables $x \in X(s)$ and $z \in Z(s)$ defined as

$$X(s) \coloneqq \operatorname*{arg\,min}_{x \in \mathbb{R}^m} \{f(x) \mid Ax = s\}$$

and

$$Z(s) \coloneqq \operatorname*{arg\,min}_{z \in \mathbb{R}^n} \{g(z) \mid Bz = b - s\}.$$

7.2.2 From ADMM to DRS

To show the claimed equivalence, it remains to show that DRS applied to

$$\operatorname*{minimize}_{s \in \mathbb{R}^p} \ \varphi_1(s) + \varphi_2(s) \tag{7.5}$$

with $\varphi_1 = (Af)$ and $\varphi_2 = (Bg)(b - \cdot)$ is equivalent to ADMM applied to the original formulation (7.1).

Theorem 7.1 (Primal equivalence of DRS and ADMM)**.** *Starting from a triplet $(x, y, z) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^n$, consider an ADMM-update applied to problem (7.1) with relaxation λ and large enough penalty $\beta > 0$ so that any ADMM minimization subproblem has solutions. Let*

$$\begin{cases} s \coloneqq Ax - y/\beta \\ u \coloneqq Ax \\ v \coloneqq b - Bz \end{cases} \quad \text{and, similarly,} \quad \begin{cases} s^+ \coloneqq Ax^+ - y^+/\beta \\ u^+ \coloneqq Ax^+ \\ v^+ \coloneqq b - Bz^+. \end{cases} \tag{7.6}$$

Then, the variables are related as follows:

$$\begin{cases} s^+ = s + \lambda(v - u) \\ u^+ \in \operatorname{prox}_{\gamma\varphi_1}(s^+) \\ v^+ \in \operatorname{prox}_{\gamma\varphi_2}(2u^+ - s^+), \end{cases} \quad \text{where} \quad \begin{cases} \varphi_1 \coloneqq (Af) \\ \varphi_2 \coloneqq (Bg)(b - \cdot) \\ \gamma \ \coloneqq 1/\beta. \end{cases}$$

Moreover,

(i) $\varphi_1(u^+) = (Af)(Ax^+) = f(x^+)$,

(ii) $\varphi_2(v^+) = (Bg)(Bz^+) = g(z^+)$,

(iii) $-y^+ \in \hat{\partial}\varphi_1(u^+) = \hat{\partial}(Af)(Ax^+)$,

(iv) $-A^\top y^+ \in \hat{\partial} f(x^+)$, *and*

(v) $\operatorname{dist}(-B^\top y^+, \hat{\partial} g(z^+)) \le \beta \|B\| \|Ax^+ + Bz^+ - b\|$.

If, additionally, A has full row rank, $\varphi_1 \in C^{1,1}(\mathbb{R}^p)$ is L_{φ_1}-smooth, and $\beta > L_{\varphi_1}$, then it also holds that

(vi) $\varphi_\gamma^{\mathrm{DR}}(s^+) = \mathscr{L}_\beta(x^+, z^+, y^+)$.

Proof. Observe first that, as shown in Prop. 1.17*(iii)*, it holds that

$$\operatorname{prox}_{\gamma\varphi_1} \supseteq A \operatorname{argmin}\left\{ f + \tfrac{1}{2\gamma}\|A\cdot - s\|^2 \right\}. \tag{7.7a}$$

Similarly, with a simple change of variable one obtains that

$$\operatorname{prox}_{\gamma\varphi_2} \supseteq b - B \operatorname{argmin}\left\{ g + \tfrac{1}{2\gamma}\|B\cdot + s - b\|^2 \right\}. \tag{7.7b}$$

Let (s, u, v) and (s^+, u^+, v^+) be as in (7.6). We have

$$\begin{aligned} s + \lambda(v - u) &= Ax - \tfrac{1}{\beta} y - \lambda(Ax + Bz - b) \\ &= Ax - \tfrac{1}{\beta} y^{+/2} - (Ax + Bz - b) \\ &= -\tfrac{1}{\beta} y^+ + Ax^+ = s^+, \end{aligned}$$

where in the second and third equality the ADMM update rule for $y^{+/2}$ and y^+, respectively, was used. Moreover,

$$\begin{aligned} u^+ = Ax^+ &\in A \operatorname{argmin} \mathscr{L}_\beta(\,\cdot\,, z, y^{+/2}) \\ &\overset{(7.7a)}{\subseteq} \operatorname{prox}_{\varphi_1/\beta}(b - Bz - y^{+/2}/\beta) = \operatorname{prox}_{\varphi_1/\beta}(s^+), \end{aligned}$$

where the last equality uses the identity $b - Bz - y^{+/2}/\beta = v - \gamma y + (1-\lambda)(u - v) = s + \lambda(v - u) = s^+$. Next, observe that $2u^+ - s^+ = 2Ax^+ - (Ax^+ - y^+/\beta) = Ax^+ + y^+/\beta$, hence

$$\begin{aligned} v^+ = b - Bz^+ &\in b - B \operatorname{argmin} \mathscr{L}_\beta(x^+, \,\cdot\,, y^+) \\ &\overset{(7.7b)}{\subseteq} \operatorname{prox}_{\varphi_2/\beta}(Ax^+ + y^+/\beta) = \operatorname{prox}_{\varphi_2/\beta}(2u^+ - s^+). \end{aligned}$$

Let us now show the numbered claims.

♠ 7.1*(i)* & 7.1*(ii)*. Follow from Prop. 1.17*(ii)*.

♠ 7.1*(iii)*. Since $u^+ \in \operatorname{prox}_{\gamma\varphi_1}(s^+)$ and $-y^+ = \frac{1}{\gamma}(s^+ - u^+)$, the claim follows from Prop. 1.12*(vi)*.

♠ 7.1*(iv)*. This follows from the optimality conditions of x^+ in the ADMM-subproblem defining the x-update. Alternatively, the claim can also be deduced from 7.1*(iii)* and Prop. 1.18.

♠ 7.1*(v)*. The optimality conditions in the ADMM-subproblem defining the z-update read

$$0 \in \hat{\partial}_z \mathscr{L}_\beta(x^{k+1}, z^{k+1}, y^{k+1}) = B^\top(Ax^{k+1} + Bz^{k+1} - b + y^{k+1}/\beta),$$

and the claim readily follows.

♠ 7.1*(vi)*. Suppose now that φ_1 is L_{φ_1}-smooth (hence A is surjective, for otherwise φ_1 has not full domain), and that $\beta > L_{\varphi_1}$. Due to smoothness, the inclusion in 7.1*(iii)* can be strengthened to $\nabla\varphi_1(u^+) = -y^+$. We may then invoke the expression (6.8) of the DRE to obtain

$$\begin{aligned}\varphi_\gamma^{\mathrm{DR}}(s^+) &= \varphi_1(u^+) + \varphi_2(v^+) + \langle \nabla\varphi_1(u^+), v^+ - u^+\rangle + \tfrac{1}{2\gamma}\|v^+ - u^+\|^2 \\ &= f(x^+) + g(z^+) + \langle y^+, Ax^+ + Bz^+ - b\rangle + \tfrac{\beta}{2}\|Ax^+ + Bz^+ - b\|^2 \\ &= \mathscr{L}_\beta(x^+, z^+, y^+).\end{aligned}$$

□

7.3 Convergence results

In order to extend the theory developed for DRS to ADMM we shall impose that φ_1 and φ_2 as in (7.4) comply with Assumption 6.I.

Assumption 7.I (Requirements for the ADMM formulation (7.1))**.** The following hold for problem (7.1):

A1 $f : \mathbb{R}^m \to \mathbb{R}$ and $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ are proper and lsc.

A2 A is surjective, and β is large enough so that the ADMM subproblems have solution.

A3 $\varphi_1 := (Af) \in C^{1,1}(\mathbb{R}^p)$ is $L_{(Af)}$-smooth, hence $\sigma_{(Af)}$-hypoconvex with $|\sigma_{(Af)}| \le L_{(Af)}$.

A4 $\varphi_2 := (Bg)$ is lsc.

Algorithm 7.1. ADMM WITH RELAXATION

REQUIRE
- initial triplet $(x^0, z^0, y^0) \in \mathbb{R}^m \times \operatorname{dom} g \times \mathbb{R}^n$
- tolerance $\varepsilon > 0$
- stepsize and relaxation $\gamma, \lambda > 0$ as follows:

$$\begin{cases} \beta > \max\left\{\frac{2[\sigma_{\varphi_1}]_-}{2-\lambda}, L_{\varphi_1}\right\} & \text{and} \quad \lambda \in (0,2), \quad \text{or} \\ \frac{4\sigma_{\varphi_1}}{p\lambda+\delta} < \beta < \frac{4\sigma_{\varphi_1}}{p\lambda-\delta} & \text{and} \quad 2 \le \lambda < \frac{4}{1+\sqrt{1-p}}, \end{cases}$$

where $\varphi_1 = (Af)$, $\varphi_2 = (Bg)(b - \cdot\,)$,
$p := \sigma_{\varphi_1}/L_{\varphi_1}$, and $\delta := \sqrt{(p\lambda)^2 - 8p(\lambda-2)}$.

PROVIDE KKT-suboptimal (x_*, z_*, y_*), in the sense that
- $\|Ax_* + Bz_* - b\| \le \frac{\varepsilon}{\beta}$,
- $-A^\top y_* \in \hat{\partial} f(x_*)$, and
- $\operatorname{dist}(-B^\top y_*, \hat{\partial} g(z_*)) \le \|B\|\varepsilon$.

1: **for** $k = 0, 1, \ldots$ **do**
2: $\quad y^{k+1/2} = y^k - \beta(1-\lambda)(Ax^k + Bz^k - b)$
3: $\quad x^{k+1} \in \operatorname{arg\,min}_x \mathscr{L}_\beta(x, z^k, y^{k+1/2})$
4: $\quad y^{k+1} = y^{k+1/2} + \beta(Ax^{k+1} + Bz^k - b)$
5: $\quad z^{k+1} \in \operatorname{arg\,min}_z \mathscr{L}_\beta(x^{k+1}, z, y^{k+1})$
6: $\quad$ **if** $\beta\|Ax^{k+1} + Bz^{k+1} - b\| \le \varepsilon$ **then**
7: $\quad\quad$ **return** $(x_*, z_*, y_*) = (x^{k+1}, z^{k+1}, y^{k+1})$

A5 Problem (7.1) has a solution: $\operatorname{arg\,min} \Phi \neq \emptyset$, where $\Phi(x,z) := f(x) + g(z) + \delta_{Ax+Bz=b}$.

Theorem 7.2 (Finite termination of relaxed ADMM)**.** *Suppose that Assumption 7.1 holds. Then, the iterates generated by ADMM (Alg. 7.1) satisfy*

$$\mathscr{L}_\beta(x^{k+1}, z^{k+1}, y^{k+1}) \le \mathscr{L}_\beta(x^k, z^k, y^k) - \frac{c\lambda^2}{(1+\gamma L_{\varphi_1})^2}\|Ax^k + Bz^k - b\|^2,$$

where $c > 0$ is as in Theorem 6.9 with $\gamma = 1/\beta$. In particular, the algorithm terminates in a finite number of iterations and yields a triplet (x_*, z_*, y_*) satisfying

- $\|Ax_* + Bz_* - b\| \le \frac{\varepsilon}{\beta}$,
- $-A^\top y_* \in \hat{\partial} f(x_*)$, *and*

⋆ $\operatorname{dist}(-B^\top y_*, \hat{\partial} g(z_*)) \le \|B\|\varepsilon$.

Proof. Follows from Thm. 6.12 and the equivalence provided in Thm. 7.1. □

Theorem 7.3 (Asymptotic convergence of ADMM)**.** *Suppose that Assumption 7.1 is satisfied, and let φ_1, φ_2, and Φ be as defined therein. Starting from $(x^{-1}, y^{-1}, z^{-1}) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^n$, consider a sequence $(x^k, y^k, z^k)_{k\in\mathbb{N}}$ generated by ADMM with penalty $\beta = 1/\gamma$ and relaxation λ, where γ and λ are as in Theorem 6.9. The following hold:*

(i) $\mathscr{L}_\beta(x^{k+1}, z^{k+1}, y^{k+1}) \le \mathscr{L}_\beta(x^k, z^k, y^k) - \frac{c\lambda^2}{(1+\gamma L_{\varphi_1})^2}\|Ax^k + Bz^k - b\|^2$, *where c is as in Theorem 6.9, and the residual $(Ax^k + Bz^k - b)_{k\in\mathbb{N}}$ vanishes with $\min_{i\le k}\|Ax^i + Bz^i - b\| = o(1/\sqrt{k})$.*

(ii) all cluster points (x, z, y) of $(x^k, z^k, y^k)_{k\in\mathbb{N}}$ satisfy the KKT conditions

- $-A^\top y \in \partial f(x)$
- $-B^\top y \in \partial g(z)$
- $Ax + Bz = b$,

and attain the same cost $f(x) + g(z)$, this being the limit of the sequence $(\mathscr{L}_\beta(x^k, z^k, y^k))_{k\in\mathbb{N}}$.

(iii) the sequence $(Ax^k, y^k, Bz^k)_{k\in\mathbb{N}}$ is bounded provided that the cost function Φ is level bounded. If, additionally, $f \in C^{1,1}(\mathbb{R}^m)$, then the sequence $(x^k, y^k, z^k)_{k\in\mathbb{N}}$ is bounded.

Proof. Let $s^0 := Ax^0 - y^0/\beta$, and consider the sequence $(s^k, u^k, v^k)_{k\in\mathbb{N}}$ generated by DRS applied to (7.5), with stepsize γ, relaxation λ, and starting from s^0. Then, for all $k \in \mathbb{N}$ it follows from Thm. 7.1 that the variables are related as

$$\begin{cases} s^k = Ax^k - y^k/\beta \\ u^k = Ax^k \\ v^k = b - Bz^k, \end{cases}$$

and satisfy

$$\begin{cases} \varphi_1(u^k) = f(x^k) \\ \varphi_2(v^k) = g(z^k) \\ \varphi_\gamma^{\mathrm{DR}}(s^k) = \mathscr{L}_\beta(x^k, z^k, y^k) \end{cases} \quad \text{and} \quad \begin{cases} y^k = -\nabla\varphi_1(u^k) \\ -A^\top y^k \in \hat{\partial} f(x^k) \\ \operatorname{dist}(-B^\top y^k, \hat{\partial} g(z^k)) \to 0. \end{cases}$$

♠ 7.3*(i)*. Readily follows from Thm. 6.12.

♠ 7.3*(ii)*. Suppose that for some $K \subseteq \mathbb{N}$ the subsequence $(x^k, y^k, z^k)_{k\in K}$ converges to (x, y, z); then, necessarily $Ax + Bz = b$. Moreover,

$$(Af)(Ax) \le f(x) \le \liminf_{K\ni k\to\infty} f(x^k) = \liminf_{K\ni k\to\infty} (Af)(Ax^k) = (Af)(Ax),$$

where the second inequality is due to the fact that f is lsc, and the last one to the fact that (Af) is continuous. Therefore, $f(x^k) \to f(x)$, and the inclusion $-A^\top y^k \in \hat\partial f(x^k)$ in light of the definition of subdifferential results in $-A^\top y \in \partial f(x)$. In turn, since $\varphi_1(u^k) + \varphi_1(v^k)$ converges to $\varphi_1(Ax) + \varphi_2(b - Bz) = (Af)(Ax) + (Bg)(Bz)$ as it follows from Thm. 6.12*(ii)*, a similar reasoning shows that $g(z^k) \to g(z)$ as $K \ni k \to \infty$. Thus, since $\operatorname{dist}(-B^\top y^k, \hat\partial g(z^k)) \to 0$, g-attentive outer semicontinuity of ∂g, see [106, Prop. 8.7], implies that $-B^\top y \in \partial g(z)$. Finally, that $f(x)+g(z)$ equals the limit of the whole sequence $(\mathscr{L}_\beta(x^k, z^k, y^k))_{k\in\mathbb{N}}$ then follows from Thm. 6.12*(ii)* through the identity $\varphi_\gamma^{\mathrm{DR}}(s^k) = \mathscr{L}_\beta(x^k, z^k, y^k)$.

♠ 7.3*(iii)*. Once we show that $\varphi = \varphi_1 + \varphi_2$ is level bounded, boundedness of $(Ax^k, Bz^k, y^k)_{k\in\mathbb{N}}$ will follow from Thm. 6.12*(iii)*. For $\alpha \in \mathbb{R}$ we have

$$\begin{aligned}
\operatorname{lev}_{\le\alpha}\varphi &= \Big\{s \mid \inf_x \{f(x) \mid Ax = s\} + \inf_z \{g(z) \mid Bz = b - s\} \le \alpha\Big\} \\
&= \Big\{s \mid \inf_{x,z} \{f(x) + g(z) \mid Ax = s,\ Bz = b - s\} \le \alpha\Big\} \\
&= \{Ax \mid f(x) + g(z) \le \alpha,\ \exists z : Ax + Bz = b\} \\
&= \{Ax \mid (x, z) \in \operatorname{lev}_{\le\alpha}\Phi,\ \exists z\}.
\end{aligned}$$

Since $\|Bz\| \le \|B\|\|z\| \le \|B\|\|(x, z)\|$ for any x, z, it follows that if $\operatorname{lev}_{\le\alpha}\Phi$ is bounded, then so is $\operatorname{lev}_{\le\alpha}\varphi$. Suppose now that $f \in C^{1,1}(\mathbb{R}^n)$ is L_f-smooth, and for all $k \in \mathbb{N}$ let $\xi^k := x^k - A^\top(AA^\top)^{-1}(Ax^k + Bz^k - b)$. Then, $A\xi^k = b - Bz^k$, hence $f(\xi^k) + g(z^k) = \Phi(\xi^k, z^k)$, and $\xi^k - x^k \to 0$ as $k \to \infty$. We have

$$\begin{aligned}
|\Phi(\xi^k, z^k) - (f(x^k) + g(z^k))| &= \big|f(\xi^k) - f(x^k)\big| \\
&\le |\langle \nabla f(x^k), \xi^k - x^k\rangle| + \tfrac{L_f}{2}\|\xi^k - x^k\|^2 \\
&\le \big|\langle y^k, Ax^k - A\xi^k\rangle\big| \\
&\quad + \tfrac{L_f}{2}\|A^\top(AA^\top)^{-1}\|^2\|Ax^k + Bz^k - b\|^2,
\end{aligned}$$

where in the second inequality the identity $\nabla f(x^k) = -A^\top y^k$ was used, cf. Thm. 7.1*(iv)*. In particular, $f(\xi^k) - f(x^k) \to 0$ as $k \to \infty$, and therefore $\Phi(\xi^k, z^k)$

converges to a finite quantity (the limit of $\mathscr{L}_\beta(x^k, z^k, y^k)$). Since Φ is level bounded, necessarily $(\xi^k, z^k)_{k\in\mathbb{N}}$ is bounded, hence so is $(x^k)_{k\in\mathbb{N}}$. □

As a consequence of the Tarski-Seidenberg theorem, functions $\varphi_1 \coloneqq (Af)$ and $\varphi_2 \coloneqq (Bg)(b - \cdot\,)$ are semialgebraic provided f and g are, see *e.g.*, [23]. In fact, (Af) is the result of the *parametric minimization* of $F(s, x)$ over variable x, where $F(s, x) = f(x) + \delta_{\{0\}}(Ax - s)$, and as such

$$\operatorname{epi}(Af) = \operatorname{cl}(\operatorname{epi}(Af)) = \operatorname{cl}(\Pi_{\mathbb{R}^p} \operatorname{epi} F).$$

Here, the first equality is due to the assumption of lsc of (Af), and the second follows from [106, Prop. 1.18]. Then, since F is semialgebraic if so is f, and since the closure of a semialgebraic set is still semialgebraic, we conclude that (Af) is semialgebraic. Clearly, the same arguments hold for $(Bg)(b - \cdot\,)$.

Therefore, sufficient conditions for global convergence of ADMM (Alg. 7.1) follow from the similar result for DRS (Alg. 6.1), through the primal equivalence of the algorithms illustrated in Theorem 7.1. We should emphasize, however, that the equivalence identifies $u^k = Ax^k$ and $v^k = b - Bz^k$, and thus only convergence of $(Ax^k)_{k\in\mathbb{N}}$ and $(Bz^k)_{k\in\mathbb{N}}$ can be deduced (as opposed to that of $(x^k)_{k\in\mathbb{N}}$ and $(z^k)_{k\in\mathbb{N}}$).

Theorem 7.4 (Global convergence of relaxed ADMM)**.** *Consider the iterates generated by ADMM (Alg. 7.1) with tolerance $\varepsilon = 0$. Suppose that Assumption 7.1 is satisfied, and let Φ be as defined therein. Suppose further that the following hold:*

A1 *Φ is level bounded.*

A2 *f and g are semialgebraic.*

A3 *All accumulation points of the sequence $(Ax^k)_{k\in\mathbb{N}}$ are prox-regular, in the sense of Definition 5.10 (with $f \leftarrow (Af)$ and $g \leftarrow (Bg)(b - \cdot\,)$).*

Then, the following hold:

(i) The sequence $(Ax^k, y^k, Bz^k)_{k\in\mathbb{N}}$ is convergent.

(ii) The ADMM residual $(\|Ax^k + Bz^k - b\|)_{k\in\mathbb{N}}$ is summable, and in particular $\min_{j\le k} \|Ax^j + Bz^j - b\| \in O(1/k)$.

7.4 Sufficient conditions

In this section we provide some sufficient conditions on f and g ensuring that Assumption 7.I is satisfied.

7.4.1 Lower semicontinuity

Proposition 7.5 (Lsc of (Bg))**.** *Suppose that Assumptions 7.IA1 and 7.IA2 are satisfied. Then, (Bg) is proper. Moreover, it is also lsc provided that for all $\bar{z} \in \operatorname{dom} g$ the set $Z(s) := \operatorname{argmin}_z \{g(z) \mid Bz = s\}$ is nonempty and $\operatorname{dist}(0, Z(s))$ is bounded for all $s \in B \operatorname{dom} g$ close to $B\bar{z}$.*

Proof. Properness is shown in Prop. 1.17*(i)*. Let $(s_k)_{k\in\mathbb{N}} \subseteq \operatorname{lev}_{\leq\alpha}(Bg)$ for some $\alpha \in \mathbb{R}$ and suppose that $s_k \to \bar{s}$. Then, due to the characterization of [106, Thm. 1.6] it suffices to show that $\bar{s} \in \operatorname{lev}_{\leq\alpha}(Bg)$. The assumption ensures the existence of a bounded sequence $(z_k)_{k\in\mathbb{N}}$ such that eventually $Bz_k = s_k$ and $(Bg)(s_k) = g(z_k)$. By possibly extracting, $z_k \to \bar{z}$ and necessarily $B\bar{z} = \bar{s}$. Then,

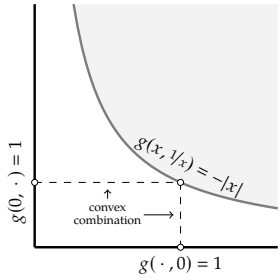
$$(Bg)(\bar{s}) \leq g(\bar{z}) \leq \liminf_{k\to\infty} g(z_k) = \liminf_{k\to\infty} (Bg)(s_k) \leq \alpha,$$

hence $\bar{s} \in \operatorname{lev}_{\leq\alpha}(Bg)$. □

The requirement in Proposition 7.5 is weaker than Lipschitz continuity of the map $s \mapsto Z(s)$, which is the standing assumption in [123] for the analysis of ADMM. In fact, no uniqueness or boundedness of the sets of minimizers is required, but only the existence of minimizers not arbitrarily far.

The pathological behavior occurring when this condition is not met can be well visualized by considering $g : \mathbb{R}^2 \to \mathbb{R}$ defined as

$$g(x, y) = \begin{cases} -|x| & \text{if } |xy| \geq 1, \\ 1 - q(|xy|)(1 + |x|) & \text{otherwise,} \end{cases} \tag{7.8}$$



where $q(t)$ is any function such that $q(0) = 0 < q(t) < 1 = q(1)$ for all $t \in (0, 1)$. In the picture, a graphical representation of the piecewise definition on the positive orthant of $\mathbb{R}^2$ (the function is mirrored in all other orthants). On the axes, f achieves its maximum value, that is, 1. In the gray region $|xy| \geq 1$,

$f(x,y) = -|x|$. In the white portion, f is extended by means of a convex combination of 1 and $-|x|$. g and $B := [1\ 0]$ are ADMM-feasible, meaning that $\operatorname{argmin}_{w\in\mathbb{R}^2} \left\{ g(w) + \frac{\beta}{2}\|Bw - s\|^2 \right\} \neq \emptyset$ for all $s \in \mathbb{R}$ and β large enough (in fact, for all $\beta > 0$, being $g(\,\cdot\,, y) + \frac{\beta}{2}\|\,\cdot\, - s\|^2$ coercive for any $y \in \mathbb{R}$). However, $(Bg)(s) = -|s|$ if $s \neq 0$ while $(Bg)(0) = 1$, resulting in the lack of lsc at $s = 0$. Along $\ker B = \{0\} \times \mathbb{R}$, by keeping x constant g attains a minimum at $\{(x,y) \mid xy \geq 1\}$ for $x \neq 0$, which escapes to infinity as $x \to 0$, and $g(x, x^{-1}) = -|x| \to 0$. However, if instead $x = 0$ is fixed (as opposed to $x \to 0$), then the pathology comes from the fact that $g(0,\,\cdot\,) \equiv 1 > 0$. The *interpolating* function q simply models the transition from a constant function on the axes and a linear function in the regions delimited by the hyperbolae. For any $k \in \mathbb{N}$ it can thus be chosen such that g is k times continuously differentiable; the choice $q(t) = \frac{1}{2}(1 - \cos \pi t)$, for instance, makes $g \in C^1(\mathbb{R}^2)$. In particular, (high-order) continuous differentiability is not enough a requirement for (Bg) to be lsc.

The next result provides necessary and sufficient conditions ensuring the image function (Bg) to inherit lower semicontinuity from that of g. It will be evident that pathological cases such as the one depicted in (7.8) may only occur due to the behavior of g at infinity.

Theorem 7.6. *For any lsc function $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $B \in \mathbb{R}^{p\times n}$, the image function (Bg) is lsc iff*

$$\liminf_{\substack{\|d\|\to\infty \\ Bd\to 0}} g(\bar z + d) \geq \inf_{d\in\ker B} g(\bar z + d) \qquad \forall \bar z \in \operatorname{dom} g. \tag{7.9}$$

In particular, if g is level bounded then (Bg) is lsc.

Proof. Observe first that the right-hand side in (7.9) is $(Bg)(B\bar z)$. Suppose now that (7.9) holds, and given $\bar s \in \operatorname{dom}(Bg)$ consider a sequence $(s_k)_{k\in\mathbb{N}} \subseteq \operatorname{lev}_{\leq\alpha}(Bg)$ for some $\alpha \in \mathbb{R}$ and such that $s_k \to \bar s$. Then, it suffices to show that $\bar s \in \operatorname{lev}_{\leq\alpha}(Bg)$. Let $(z_k)_{k\in\mathbb{N}}$ be such that $Bz_k = s_k$ and $g(z_k) \leq (Bg)(s_k) + 1/k$ for all $k \in \mathbb{N}$. If, up to possibly extracting, there exists z such that $z^k \to z$ as $k \to \infty$, then the claim follows with a similar reasoning as in the proof of Prop. 7.5. Suppose, instead, that $t_k := \|z_k\| \to \infty$ as $k \to \infty$, and let $d_k := z_k - \bar z$, where $\bar z \in \operatorname{dom} g$ is any such that $B\bar z = s$ (such a $\bar z$ exists, being $\bar s \in \operatorname{dom}(Bg) = B \operatorname{dom} g$). Since $Bd_k = B(z_k - \bar z) = s_k - \bar s \to 0$, we have

$$\begin{aligned}(Bg)(\bar s) = \inf_{d\in\ker B} g(\bar z + d) &\leq \liminf_{k\to\infty} g(\bar z + d_k) = \liminf_{k\to\infty} g(z_k) \\ &\leq \liminf_{k\to\infty} (Bg)(s_k) + \tfrac{1}{k} \leq \alpha,\end{aligned}$$

proving that $\bar{s} \in \operatorname{lev}_{\leq \alpha}(Bg)$.

To show the converse implication, suppose that (7.9) does not hold. Thus, there exist $\bar{z} \in \operatorname{dom} g$ and $(d^k)_{k\in\mathbb{N}} \subset \mathbb{R}^n$ such that $Bd^k \to 0$ as $k \to \infty$, and such that, for some $\varepsilon > 0$,

$$g(\bar{z} + d^k) + \varepsilon \leq \inf_{d \in \ker B} g(\bar{z} + d) = (Bg)(B\bar{z}) \quad \forall k.$$

Then, $s_k := B(\bar{z} + d^k)$ satisfies $s_k \to B\bar{z}$ as $k \to \infty$, and

$$(Bg)(B\bar{z}) \geq \liminf_{k\to\infty} g(\bar{z} + d^k) + \varepsilon \geq \liminf_{k\to\infty} (Bg)(s^k) + \varepsilon,$$

hence (Bg) is not lsc at $B\bar{z}$. □

The *asymptotic function* $g_\infty(\bar{d}) := \liminf_{d\to\bar{d},\, t\to\infty} \frac{g(td)}{t}$ is a tool used in [8] to analyze the behavior of g at infinity and derive sufficient properties ensuring lsc of (Bg). These all ensure that the set of minimizers $Z(s)$ as defined in Proposition 7.5 is nonempty, although this property is not necessary as long as lower semicontinuity is concerned. To see this, it suffices to modify (7.8) as follows

$$g(x, y) = \begin{cases} -|x| & \text{if } |xy| \geq 1, \\ e^{-y^2} - q(|xy|)(e^{-y^2} + |x|) & \text{otherwise,} \end{cases}$$

that is, by replacing the constant value 1 on the y axis with e^{-y^2}. Then, the epi-composition $(Bg)(s) = -|s|$ is lsc, but the set of minimizers $\arg\min_w \{g(w) \mid Bw = 0\} = \{0\} \times \arg\min_y e^{-y^2}$ is empty at $s = 0$.

7.4.2 Smoothness

We now turn to the smoothness requirement of (Af). To this end, we introduce the notion of *smoothness with respect to a matrix*, as follows

Definition 7.7 (Smoothness relative to a matrix)**.** *We say that a function $h : \mathbb{R}^n \to \mathbb{R}$ is* SMOOTH RELATIVE TO A MATRIX $C \in \mathbb{R}^{p\times n}$*, and we write $h \in C^{1,1}_C(\mathbb{R}^n)$, if h is differentiable and ∇h satisfies the following Lipschitz condition: there exist $L_{h,C}$ and $\sigma_{h,C}$ with $|\sigma_{h,C}| \leq L_{h,C}$ such that*

$$\sigma_{h,C}\|C(x - y)\|^2 \leq \langle \nabla h(x) - \nabla h(y), x - y\rangle \leq L_{h,C}\|C(x - y)\|^2 \tag{7.10}$$

whenever $\nabla h(x), \nabla h(y) \in \operatorname{range} C^\top$.

This condition is similar to that considered in [52], where $\Pi_{\operatorname{range} A^\top} \nabla f$ is required to be Lipschitz. The paper analyzes convergence of a proximal ADMM; standard

ADMM can be recovered when matrix A is invertible, in which case both conditions reduce to Lipschitz differentiability of f. In general, our condition applies to a smaller set of points only, as it can be verified with $f(x,y) = \frac{1}{2}x^2y^2$ and $A = [1\ 0]$. In fact, $\Pi_{\operatorname{range} A^\top} \nabla f(x,y) = \binom{xy^2}{0}$ is *not* Lipschitz continuous; however, $\nabla f(x,y) \in \operatorname{range} A^\top$ iff $xy = 0$, in which case $\nabla f \equiv 0$. Then, f is smooth relative to A with $L_{f,A} = 0$.

To better understand how this notion of regularity comes into the picture, notice that if f is differentiable, then $\nabla f(x) \in \operatorname{range} A^\top$ on some domain $\mathcal{U}$ if there exists a differentiable function $q : A\mathcal{U} \to \mathbb{R}$ such that $f(x) = q(Ax)$. Then, it is easy to verify that f is smooth relative to A if the local "reparametrization" q is smooth (on its domain). From an a posteriori perspective, if (Af) is smooth, then due to the relation $A^\top \nabla(Af)(Az_s) = \nabla f(z_s)$ holding for $z_s \in \operatorname{argmin}_{z:Az=s} f(z)$ (cf. Prop. 1.18), it is apparent that q serves as (Af). Therefore, smoothness relative to A is somewhat a minimal requirement for ensuring smoothness of (Af).

Theorem 7.8 (Smoothness of (Af))**.** *Let $A \in \mathbb{R}^{p\times n}$ be surjective and $f : \mathbb{R}^n \to \mathbb{R}$ be lsc. Suppose that there exists $\beta \geq 0$ such that the function $f + \frac{\beta}{2}\|A\cdot - s\|^2$ is level bounded for all $s \in \mathbb{R}^p$. Then, the image function (Af) is smooth on $\mathbb{R}^p$, provided that either*

(i) $f \in C^{1,1}_A(\mathbb{R}^n)$, in which case $L_{(Af)} = L_{f,A}$ and $\sigma_{(Af)} = \sigma_{f,A}$,

(ii) or $f \in C^{1,1}(\mathbb{R}^n)$, and $X(s) \coloneqq \operatorname{argmin}\{f(x) \mid Ax = s\}$ is single valued and Lipschitz continuous with modulus M, in which case

$$L_{(Af)} = L_f M^2 \quad \text{and} \quad \sigma_{(Af)} = \begin{cases} \sigma_f/\|A\|^2 & \text{if } \sigma_f \geq 0, \\ \sigma_f M^2 & \sigma_f < 0; \end{cases}$$

(iii) or $f \in C^{1,1}(\mathbb{R}^n)$ is convex, in which case $L_{(Af)} = \frac{L_f}{\sigma_+(A^\top A)}$ and $\sigma_{(Af)} = \sigma_f/\|A\|^2$.

Proof. As shown in Prop. 1.17*(i)*, (Af) is proper. The surjectivity of A and the level boundedness condition ensure that for all $\alpha \in \mathbb{R}$ and $s \in \mathbb{R}^p$ the set $\{x \mid f(x) \leq \alpha, \|Ax - s\| < \varepsilon\}$ is bounded for some $\varepsilon > 0$ (in fact, for all $\varepsilon > 0$). Then, we may invoke [106, Thm. 1.32] to infer that (Af) is lsc, that the set $X(s) \coloneqq \operatorname{argmin}_x \{f(x) \mid Ax = s\}$ is nonempty for all $s \in \mathbb{R}^p$, and that the function $H(x,s) \coloneqq f(x) + \delta_{\{0\}}(Ax - s)$ is uniformly level bounded in x locally uniformly in s, in the sense of [106, Def. 1.16]. Moreover, since f is differentiable, observe that $\partial^\infty H(x,Ax) = \operatorname{range}\binom{A^\top}{1}$ for all $x \in \mathbb{R}^m$. Hence, for all $s \in \mathbb{R}^p$ it

holds that

$$\partial^\infty(Af)(s) \subseteq \bigcup_{x\in X(s)} \{y \mid (0,y) \in \partial^\infty H(x,s)\} = \ker A^\top = \{0\},$$

where the inclusion follows from [106, Thm. 10.13]. By virtue of [106, Thm. 9.13], we conclude that (Af) is strictly continuous and has nonempty subdifferential on $\mathbb{R}^p$. Fix $s_i \in \mathbb{R}^p$ and $y_i \in \partial(Af)(s_i)$, $i = 1, 2$, and let us proceed by cases.

♠ 7.8*(i)* and 7.8*(ii)*. It follows from Prop. 1.18 and continuous differentiability of f that $A^\top y_i \in \partial f(x_i) = \{\nabla f(x_i)\}$, for some $x_i \in X(s_i)$, $i = 1, 2$. We have

$$\begin{aligned}\langle y_1 - y_2, s_1 - s_2\rangle &= \langle y_1 - y_2, Ax_1 - Ax_2\rangle = \langle A^\top y_1 - A^\top y_2, x_1 - x_2\rangle \\ &= \langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2\rangle. \end{aligned} \tag{7.11}$$

If 7.8*(i)* holds, since $\nabla f(x_i) = A^\top y_i \in \operatorname{range} A^\top$, $i = 1, 2$, smoothness of f relative to A implies

$$\begin{aligned}\sigma_{f,A}\|s_1 - s_2\|^2 &= \sigma_{f,A}\|Ax_1 - Ax_2\|^2 \\ &\le \langle y_1 - y_2, s_1 - s_2\rangle \le L_{f,A}\|Ax_1 - Ax_2\|^2 = L_{f,A}\|s_1 - s_2\|^2\end{aligned}$$

for all $s_i \in \mathbb{R}^p$ and $y_i \in \partial(Af)(s_i)$, $i = 1, 2$.

Otherwise, if 7.8*(ii)* holds, then

$$\sigma_f\|x_1 - x_2\|^2 \le \langle y_1 - y_2, s_1 - s_2\rangle \le L_f\|x_1 - x_2\|^2$$

and from the bound $\frac{1}{\|A\|}\|s_1 - s_2\| \le \|x_1 - x_2\| \le M\|s_1 - s_2\|$ we obtain

$$\sigma_{(Af)}\|s_1 - s_2\|^2 \le \langle y_1 - y_2, s_1 - s_2\rangle \le L_{(Af)}\|s_1 - s_2\|^2$$

with the constants $\sigma_{(Af)}$ and $L_{(Af)}$ as in the statement. The claimed smoothness and hypoconvexity then follow by invoking Lem. 1.9.

♠ 7.8*(iii)*. It follows from [57, Thm. D.4.5.1 and Cor. D.4.5.2] that (Af) is a convex and differentiable function satisfying $\nabla(Af)(s) = y$, where y satisfies $A^\top y = \nabla f(x)$ and x is any element of $X(s)$. For $y_i = \nabla(Af)(s_i)$ and $x_i \in X(s_i)$, $i = 1, 2$, the equalities in (7.11) hold. In turn,

$$\begin{aligned}\langle s_1 - s_2, y_1 - y_2\rangle &\ge \tfrac{1}{L_f}\|A^\top(y_1 - y_2)\|^2 \ge \tfrac{\sigma_+(A^\top A)}{L_f}\|\Pi_{\operatorname{range} A}(y_1 - y_2)\|^2 \\ &= \tfrac{\sigma_+(A^\top A)}{L_f}\|y_1 - y_2\|^2,\end{aligned}$$

where the first inequality is due to $1/L_f$-cocoercivity of ∇f, see [84, Thm. 2.1.5], the second inequality is a known fact (see *e.g.*, [52, Lem. A.2]), and the equality is due to the fact that A is surjective. We may again invoke [84, Thm. 2.1.5] to infer the claimed $\frac{L_f}{\sigma_+(A^\top A)}$-smoothness of (Af). Since (Af) is convex (thus 0-hypoconvex), if $\sigma_f = 0$ there is nothing more to show. The case $\sigma_f > 0$ follows from Prop. 1.19. □

Notice that the condition in Theorem 7.8*(ii)* covers the case when $f \in C^{1,1}(\mathbb{R}^n)$ and A has full column rank (hence is invertible), in which case $M = 1/\sigma_+(A)$. This is somehow trivial, since necessarily $(Af)(s) = f \circ A^{-1}$ in this case.

7.5 A quasi-Newton ADMM

In light of the equivalence shown in Theorem 7.1 we can directly translate CLyD-DRS (Alg. 6.2) into a corresponding ADMM enhancement. The following result, easily deducible from the proof of the Theorem 7.1, shows how to convert a DRS update $s \mapsto (u, v)$ in terms of an ADMM update.

Lemma 7.9. *Starting from a point $\bar{s} \in \mathbb{R}^p$, consider a DRS-update $\bar{s} \mapsto (\bar{u}, \bar{v})$ with stepsize $\gamma = 1/\beta$ for $\varphi_1 = (Af)$ and $\varphi_2 = (Bg)(b - \cdot)$. Let $y_0 \in \mathbb{R}^p$ and $z_0 \in \operatorname{dom} g$ be any such that*

$$\bar{s} = b - Bz_0 - \tfrac{1}{\beta} y_0.$$

Then, $\varphi_\gamma^{\mathrm{DR}}(\bar{s}) = \mathscr{L}_\beta(\bar{x}, \bar{z}, \bar{y})$, where

$$\begin{cases} \bar{x} = \operatorname{argmin}_x \mathscr{L}(x, z_0, y_0) \\ \bar{y} = y_0 + \beta(A\bar{x} + Bz_0 - b) \\ \bar{z} \in \operatorname{argmin}_z \mathscr{L}(\bar{x}, z, \bar{y}). \end{cases}$$

In fact, $\bar{u} = A\bar{x}$ and $\bar{v} = b - B\bar{z}$.

With some algebraic manipulations on CLyD-DRS (Alg. 6.2) using this result, one obtains the ADMM variant CLyD-ADMM (Alg. 7.2), that inherits the convergence guarantees shown in the previous chapter. For the sake of describing the nonmonotone linesearch variant, we repropose the subsequential convergence statement.

Theorem 7.10 (CLyD-ADMM (nonmonotone): subseq convergence)**.** *Suppose that Assumption 7.I is satisfied, and let Φ be as defined therein. Then, the fol-*

Algorithm 7.2. CLyD-ADMM

REQUIRE
- penaly β and relaxation λ as in ADMM (Alg. 7.1)
- sufficient decrease constant c as in Thm. 7.3
- scaling factor $\alpha \in (0,1)$ for sufficient decrease constant
- initial triplet (x^{-1}, y^{-1}, z^{-1})
- tolerance $\varepsilon > 0$

PROVIDE KKT-suboptimal (x_*, z_*, y_*), in the sense that
- $\|Ax_* + Bz_* - b\| \le \frac{\varepsilon}{\beta}$,
- $-A^\top y_* \in \hat{\partial} f(x_*)$, and
- $\operatorname{dist}(-B^\top y_*, \hat{\partial} g(z_*)) \le 3\|B\|\varepsilon$.

(for brevity, $r^k \coloneqq Ax^k + Bz^k - b$)

1: Do a nominal ADMM-step:
$$\begin{aligned} y^{-1/2} &= y^{-1} - \beta(1-\lambda)r^{-1} \\ x^0 &= \arg\min_x \mathscr{L}_\beta(x, z^{-1}, y^{-1/2}) \\ y^0 &= y^{-1/2} + \beta(Ax^0 + Bz^{-1} - b) \\ z^0 &\in \arg\min_z \mathscr{L}_\beta(x^0, z, y^0) \end{aligned}$$

2: **for** $k = 0, 1, 2, \ldots$ **do**

3: **if** $\beta\|r^k\| \le \varepsilon$ **then**

4: **return** $(x_*, y_*, z_*) = (x^k, y^k, z^k)$

5: Select an update direction $d^k \in \mathbb{R}^p$ at y^k

6: Let $\tau_k \in \{2^{-i} \mid i \in \mathbb{N}\}$ be the largest such that

$$\mathscr{L}_\beta(x^{k+1}, z^{k+1}, y^{k+1}) \le \mathscr{L}_\beta(x^k, z^k, y^k) - \tfrac{\alpha c \lambda^2}{(1+\gamma L_{\varphi_1})^2}\|r^k\|^2, \qquad (7.12)$$

where $y^{k+1/2} = y^k - \beta\big[\big(1-\lambda(1-\tau_k)\big)r^k + \tau_k d^k\big]$ and $(x^{k+1}, z^{k+1}, y^{k+1})$ comes from a nominal ADMM-step:
$$\begin{aligned} x^{k+1} &= \arg\min_x \mathscr{L}_\beta(x, z^k, y^{k+1/2}) \\ y^{k+1} &= y^{k+1/2} + \beta(Ax^{k+1} + Bz^k - b) \\ z^{k+1} &\in \arg\min_z \mathscr{L}_\beta(x^{k+1}, z, y^{k+1}). \end{aligned}$$

lowing hold for the iterates generated by CLyD-ADMM (Alg. 7.2) with tolerance $\varepsilon = 0$*:*

(i) The residual $(\|r^k\|)_{k\in\mathbb{N}}$ *is square-summable; in particular, it vanishes with rate* $\min_{j\le k}\|r^j\| \in O(1/\sqrt{k})$.

(ii) all cluster points (x, z, y) *of* $(x^k, z^k, y^k)_{k\in\mathbb{N}}$ *satisfy the KKT conditions*

- $-A^\top y \in \partial f(x)$,
- $-B^\top y \in \partial g(z)$,
- $Ax + Bz = b$.

If, additionally, $\|d^k\| \to 0$ *as* $k \to \infty$*, then the following also hold:*

(iii) the sequence $(Ax^k, y^k, Bz^k)_{k\in\mathbb{N}}$ *is bounded provided that the cost function* Φ *is level bounded. If, additionally,* $f \in C^{1,1}(\mathbb{R}^m)$*, then the sequence* $(x^k, y^k, z^k)_{k\in\mathbb{N}}$ *is bounded.*

(iv) all cluster points (x, z, y) *of* $(x^k, z^k, y^k)_{k\in\mathbb{N}}$ *attain the same cost* $f(x) + g(z)$*, this being the limit of the sequence* $(\mathscr{L}_\beta(x^k, z^k, y^k))_{k\in\mathbb{N}}$.

All the claims remain valid if the linesearch condition (7.12) is replaced by the following nonmonotone version:

$$\mathscr{L}_\beta(x^{k+1}, z^{k+1}, y^{k+1}) \le \bar{\mathcal{L}}_k - \tfrac{\alpha c \lambda^2}{2(1+\gamma L_{\varphi_1})^2}\|r^k\|^2, \tag{7.13}$$

where, for any sequence $(p_k)_{k\in\mathbb{N}} \subseteq [0, 1]$ *bounded away from* 0*,* $\bar{\mathcal{L}}_k$ *are recursively defined as follows:*

$$\bar{\mathcal{L}}_k := \begin{cases} \mathscr{L}_\beta(x^0, z^0, y^0) & \text{if } k = 0, \\ (1 - p_k)\bar{\mathcal{L}}_{k-1} + p_k \mathscr{L}_\beta(x^k, z^k, y^k) & \text{otherwise.} \end{cases}$$

7.6 Simulations

7.6.1 Sparse principal component analysis

Given a data matrix $W \in \mathbb{R}^{m\times n}$, the goal of sparse principal component analysis (SPCA) is to explain as much variability in the data by using only few variables, say, $k \ll n$. Denoting $\Sigma := W^\top W$ the covariance matrix of W, this can be done

by solving the following problem:

$$\underset{z\in\mathbb{R}^n}{\text{maximize}}\ \langle z,\Sigma z\rangle \quad \text{subject to}\ \|z\|=1,\ \|z\|_0\le k,$$

where the ℓ_0-quasi-norm $\|z\|_0$ denotes the number of nonzero elements of vector z. Although the constraint $\|z\|=1$ can be convexified to $\|z\|\le 1$ without affecting the solution, the problem is still inherently nonconvex due to the ℓ_0-constraint and the concavity of the cost function, being the maximization of $\langle z,\Sigma z\rangle$ equal to the minimization of $-\langle z,\Sigma z\rangle$. Complying with Assumption 6.I, DRS can be readily applied to this problem. However, as the problem size grows a big limitation is the need to store and operate with large matrices. To account for this issue, we consider the following *consensus formulation*: having fixed a number of *agents* $N\ge 1$, decompose matrix W into N row blocks $W_1,\dots,W_N$ so that $W^\top=[W_1^\top\ \cdots\ W_N^\top]$ and $\langle z,\Sigma z\rangle=\sum_{i=1}^N\|W_iz\|^2$, introduce N copies $x_1,\dots,x_N$ of z (stacked in a vector $\boldsymbol{x}\in\mathbb{R}^{nN}$), and solve

$$\underset{\boldsymbol{x}\in\mathbb{R}^{Nn},z\in\mathbb{R}^n}{\text{minimize}}\ -\sum_{i=1}^N\|W_ix_i\|^2 \quad \text{subject to}\ \begin{array}{r}\|z\|=1,\ \|z\|_0\le k,\\ x_i=z,\ i=1\dots N.\end{array}$$

Denoting

$$\mathcal{Z}\coloneqq\{z\in\mathbb{R}^n\mid\|z\|=1,\ \|z\|_0\le k\}$$

the feasible domain, the problem can be cast in ADMM form as

$$\underset{\boldsymbol{x}\in\mathbb{R}^{Nn},z\in\mathbb{R}^n}{\text{minimize}}\ \underbrace{\sum_{i=1}^N-\|W_ix_i\|^2}_{f(\boldsymbol{x})}+\underbrace{\delta_{\mathcal{Z}}(z)}_{g(z)} \quad \text{subject to}\ \boldsymbol{x}=\begin{pmatrix}\mathrm{I}\\ \vdots\\ \mathrm{I}\end{pmatrix}z. \tag{7.14}$$

Apparently, the ADMM matrix $B\in\mathbb{R}^{nN\times n}$ is the vertical stacking of N many $n\times n$ (negative) identity matrices, A is the $nN\times nN$ identity matrix and b is the zero $\mathbb{R}^{nN}$ vector. Notice that Assumption 7.I is satisfied, as $\operatorname{range}A=\mathbb{R}^{nN}$ and $(Af)=f$ has Lipschitz-continuous gradient with modulus $L_{(Af)}=L_f=\max_{i=1\dots N}\|W_i\|\le\|W\|$.

Notice that the z-update as prescribed by ADMM comes at negligible cost, since

$$\operatorname*{arg\,min}_{z\in\mathbb{R}^n}\left\{\delta_{\mathcal{Z}}(z)+\tfrac{\beta}{2}\|\boldsymbol{x}-Az\|^2\right\}=\Pi_{\mathcal{Z}}\left(\tfrac{1}{N}\textstyle\sum_{i=1}^N x_i\right)\quad\forall\boldsymbol{x}\in\mathbb{R}^{nN},$$

and $\Pi_{\mathcal{Z}}(z)$ amounts to setting to zero the $n-k$ smallest components of z (in absolute value), and then projecting on the ℓ_2-sphere by simply dividing by the norm. The $\boldsymbol{x}$-update, instead, amounts to solving (in parallel) a (small) linear

system for $i = 1 \dots N$:

$$\begin{aligned}\operatorname*{arg\,min}_{x_i \in \mathbb{R}^n} \left\{ -\|W_i x_i\|^2 + \tfrac{\beta}{2}\|x_i - z\|^2 \right\} &= (\beta \mathrm{I} - W_i^\top W_i)^{-1} \beta z \\ &= z + W_i^\top (\beta \mathrm{I} - W_i W_i^\top)^{-1} W_i z,\end{aligned}$$

where the second equality uses the Woodbury identity. The Cholesky factors of the $m_i \times m_i$ matrix $\beta \mathrm{I} - W_i W_i^\top$ (where m_i denotes the number of rows of the block W_i), $i = 1 \dots N$, can be computed offline to efficiently solve the linear systems at each $\boldsymbol{x}$-update, resulting in $O\big(\sum_{i=1}^N m_i^2\big)$ memory requirement, as opposed to $O(N^2) = O\big(\sum_{i=1}^N m_i\big)^2$ (let alone the operational cost) needed for the original single-agent problem expression.

This consensus reformulation, however, increases the problem size and thus the ill conditioning, and for moderate values of m, n and N the convergence of plain ADMM is already prohibitively slow, cf. Figure 7.2. On the contrary, the adoption of L-BFGS directions in CLyD-ADMM (Alg. 7.2) robustifies the performance at the negligible cost of few scalar products per iteration.

Figure 7.2 shows the result of a random simulation. We considered a randomly generated data matrix $W \in \mathbb{R}^{200 \times 4000}$ with sparsity 0.2, and we split W in N blocks of equal size as in (7.14) with $N \in \{5, 10, 25, 50\}$. In each experiment the penalty parameter in both CLyD-ADMM (Alg. 7.2) and the nominal ADMM was set to $\beta = 2.1 L_{(Af)}$. We selected L-BFGS directions with memory 10, and $\sigma = 10^{-4}$ as sufficient decrease parameter (largely below the maximum value for all instances). Both algorithms were started at the same randomly generated initial point, and the tolerance was set to $\varepsilon = 10^{-6}$.

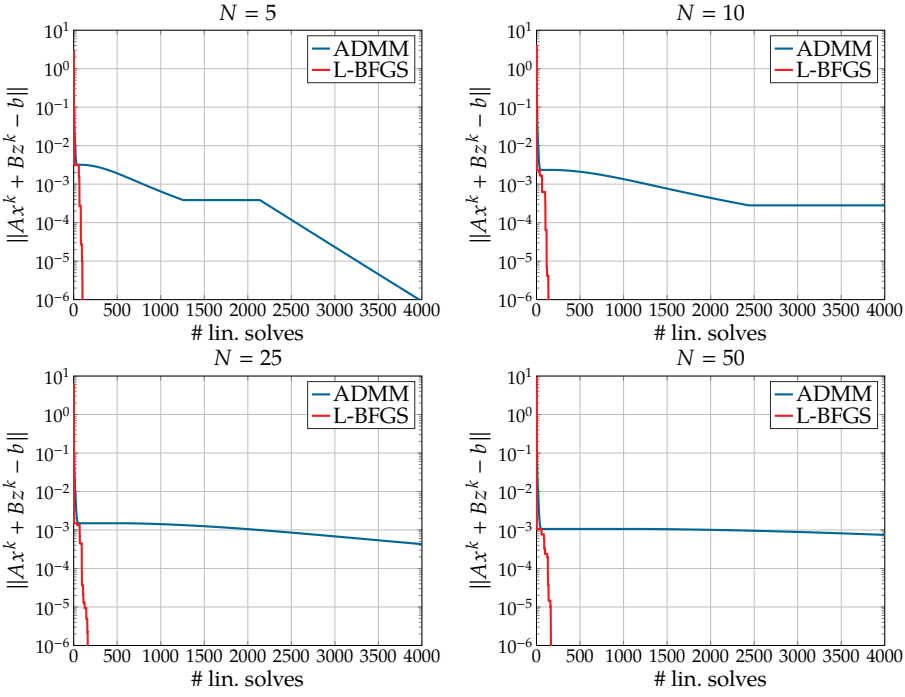


Figure 7.2: *Comparison between ADMM (in blue) and the L-BFGS enhancement (in red) for the consensus SPCA problem (7.14) for different number of agents $N = 5, 10, 25, 50$. On the x-axis, the number of linear systems solved (needed for the $\boldsymbol{x}$-update), which in the case of plain ADMM coincides with the number of iterations. This is the unique expensive operation, as the z-update is negligible. Coordinate (x, y) in the plot indicates the minimum ADMM residual y achieved after x many solutions of linear systems. Apparently, ADMM is severly affected by N, whereas the great performance of the L-BFGS enhancement through CLyD-ADMM (Alg. 7.2) remains stable.*

Chapter 8

SuperMann

A universal CLyD framework for convex splitting algorithms

8.1 Introduction

After the in-depth analysis on nonconvex problems carried out so far, in this final chapter we investigate what more we can do when the problem at hand is instead convex. In doing so, we will stick to the objective of deriving fast methods that preserve operation and iteration complexity as plain splitting algorithms. The result will be a universal scheme that globalizes Newton-type methods of *most splitting algorithms* defined on real Hilbert spaces. Admittedly with an intended pun, since it exhibits *super*linear convergence rates and generalizes the Krasnosel'skiĭ-*Mann* iterations we name our algorithm *SuperMann*. Furthermore, we show that the modified Broyden method discussed in Section 4.3.2 fits into this framework and enables superlinear asymptotic convergence rates. One of the most appealing properties of *SuperMann* is that, contrary to the envelope-based approaches, achieving superlinear convergence does not necessitate nonsingularity of the Jacobian at the solution, but the milder property of *metric subregularity*. This relaxation considerably widens the range of problems which can be solved efficiently, in that, for instance, the solutions need not be isolated for superlinear convergence to take place.

To some extent, SuperMann can be identified as an *"approximate"*-CLyD globalization framework, where the continuous Lyapunov potential is the (implicit and unknown) function $\mathcal{L} = \operatorname{dist}(\,\cdot\,, \operatorname{fix}\mathcal{F})^2$. Given an arbitrary update direction d at s, a novel hyperplane projection step ensures that for stepsizes τ small enough the update $s^+ = s^+(\tau; d)$ satisfies a sufficient decrease on $\mathcal{L}$. Therefore, although the true value of $\mathcal{L}$ remains unknown, the sufficient decrease can be suitably lower bounded, hence the interpretation as an *approximate-*

continuous-Lyapunov descent algorithm. Most importantly, we will show that also SuperMann is robust against the Maratos effect, as anytime the directions are superlinear,[1] unit stepsize is eventually accepted.

8.1.1 Contributions

The contributions can be summarized as follows:

(1) In Section 8.4 we design a universal algorithmic framework (Algorithm 8.1) for finding fixed points of nonexpansive operators, which generalizes the classical Krasnosel'skiĭ-Mann (KM) scheme and possessess its same global and local convergence properties.

(2) In Section 8.5 we introduce a novel separating hyperplane projection tailored for nonexpansive mappings; based on this, in Definition 8.11 we then propose a generalized KM iteration (GKM).

(3) We define a line search based on the novel projection, suited for any nonexpansive operator and update direction (Theorem 8.12).

(4) In Section 8.6 we combine these ideas and derive the *SuperMann scheme (Alg. 8.2)*, an algorithm that

- globalizes the convergence of Newton-type methods for finding fixed points of nonexpansive operators (Theorem 8.13);
- reduces to the local method $x_{k+1} = x_k + d_k$ when the directions d_k are *superlinear*, as it is the case for the modified Broyden scheme of Section 4.3.2 (Theorems 8.16 and 8.19);
- has superlinear convergence guarantees without the usual requirement of nonsingularity of the Jacobian at the limit point, but simply under metric subregularity; in particular, the solution need not be unique!

8.1.2 Chapter organization

The chapter is organized as follows. Section 8.2 serves as an informal introduction to highlight the known limitations of fixed-point iterations and to motivate our interest in Newton-type methods with some toy examples. The formal presentation begins in Section 8.3 with the introduction of some basic notation and known facts. In Section 8.4 we define the problem at hand and propose

[1] The definition of superlinear directions meant here is slightly different from the one given in Definition 4.4. The intended notion will be given in Definition 8.14.

a general abstract algorithmic framework for solving it. In Section 8.5 we provide a generalization of the classical KM iterations that is key for the global convergence and performance of *SuperMann*, an algorithm which is presented and analyzed in Section 8.6. Finally, in Section 8.7 we show how the theoretical findings are backed up by promising numerical simulations, where *SuperMann* dramatically improves classical splitting schemes.

8.2 Motivating examples

Given a nonexpansive operator $T : \mathbb{R}^n \to \mathbb{R}^n$, consider the problem of finding a fixed point, *i.e.*, a point $x_\star \in \mathbb{R}^n$ such that $x_\star = Tx_\star$. The independent works of Krasnosel'skiĭ and Mann [62, 79] provided a very elegant solution which is simply based on recursive iterations $x^+ = (1-\alpha)x+\alpha Tx$ with $\alpha \in (0, \bar{\alpha})$ for some $\bar{\alpha} \geq 1$. The method, known as Krasnosel'skiĭ-Mann scheme or KM scheme for short, has been studied intensively ever since, also because it generalizes a plethora of optimization algorithms. It is well known that the scheme is globally convergent with square-summable and monotonically decreasing residual $R = \mathrm{id} - T$ (in norm), and also locally Q-linearly convergent if R is METRICALLY SUBREGULAR at the limit point $x_\star$. Metric subregularity basically amounts to requiring the distance from the set of solutions to be upper bounded by a multiple of the norm of R for all points sufficiently close to $x_\star$; it is quite mild a requirement — for instance, it does not entail $x_\star$ to be an isolated solution — and as such linear convergence is quite frequent in practice. However, the major drawback of the KM scheme is its high sensitivity to ill conditioning of the problem, and cases for which convergence is prohibitively slow in practice despite the theoretical (sub)linear rate are also abundant. Illustrative examples can be easily constructed for the problem of finding a point in the intersection of two closed convex sets C_1 and C_2 with $C_1 \cap C_2 \neq \emptyset$. The problem can be solved by means of fixed-point iterations of the (nonexpansive) ALTERNATING PROJECTIONS operator $T = \Pi_{C_2} \circ \Pi_{C_1}$.

In Figure 8.1 we consider the case of two polyhedral cones, namely

$$C_1 = \left\{x \in \mathbb{R}^2 \mid 0.1x_1 \leq x_2 \leq 0.2x_1\right\}$$

and

$$C_2 = \left\{x \in \mathbb{R}^2 \mid 0.3x_1 \leq x_2 \leq 0.35x_1\right\}.$$

Alternating projections is then linearly convergent (to the unique intersection point 0) due to the fact that $R = \mathrm{id} - T$ is piecewise affine and hence globally metrically subregular. However, the convergence is extremely slow due to the pathological small angle between the two cones, as it is apparent in Figure 8.1.

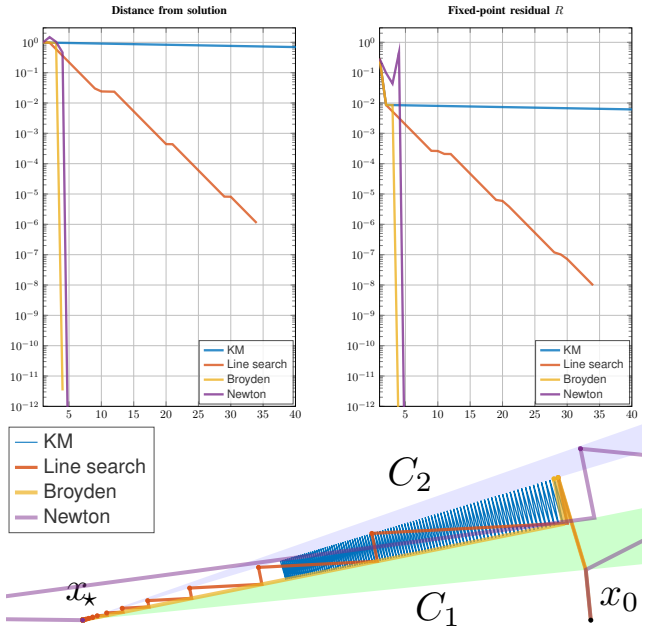


Figure 8.1: Alternating projections on polyhedral cones.
$R = \mathrm{id} - \Pi_{C_2} \circ \Pi_{C_1}$ *is globally metrically subregular, however the Q-linear convergence of the KM scheme is very slow.*

As an attempt to overcome this frequent phenomenon, [48] proposes a foretracking linesearch heuristic which is particularly effective when subsequent fixed-point iterations proceed along almost parallel directions. Iteration-wise, in such instances the line search does yield a considerable improvement upon the plain KM scheme; however, each foretrack prescribes extra evaluations of T and unless T has a specific structure the computational overhead might outweight the advantages. Moreover, its asymptotic convergence rates do not improve upon the plain KM scheme. Figure 8.2 illustrates this fact relative to

$$C_1 = \big\{x \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1\big\} \quad \text{and} \quad C_2 = \big\{x \in \mathbb{R}^2 \mid x_1 = 1\big\}.$$

Despite a good performance on early iterations, the line search cannot improve the asymptotic sublinear rate of the plain KM scheme due to the fact that the residual is not metrically subregular at the (unique) solution $x_\star = (1, 0)$. In particular, it is evident that medium-to-high accuracy cannot be achieved in a reasonable number of iterations with either methods.

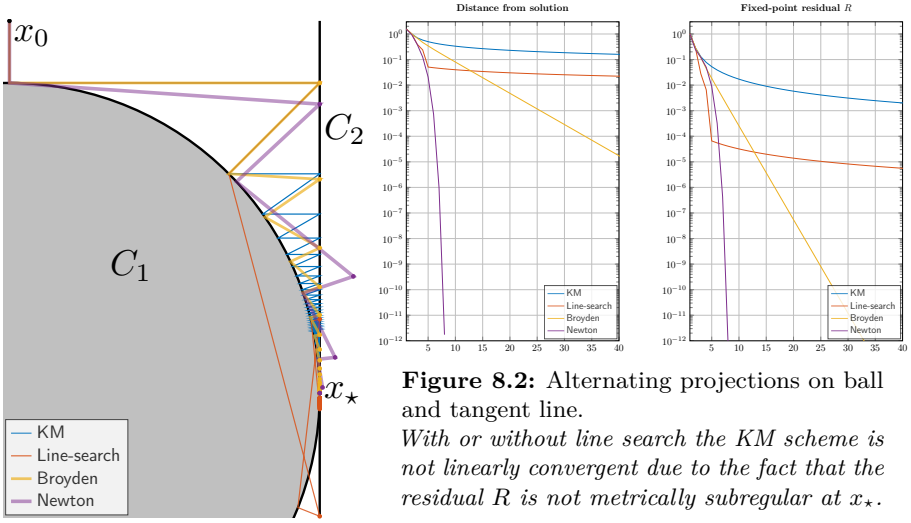


Figure 8.2: Alternating projections on ball and tangent line.
With or without line search the KM scheme is not linearly convergent due to the fact that the residual R is not metrically subregular at $x_\star$.

In response to this limitation there comes the need to include some "first-order-like information". Specifically, the problem of finding a fixed point of T can be rephrased in terms of solving the system of nonlinear (monotone) equations $Rx = 0$, which could *possibly* be solved efficiently with Newton-type methods. In the toy simulations of this section, the purple lines correspond to the semismooth Newton iterations

$$x^+ = x - G^{-1}Rx \quad \text{for some} \quad G \in \partial_C Rx,$$

where $\partial_C R$ is the Clarke generalized Jacobian of R (the convex hull of the Bouligand subdifferential, see [44, Def. 7.1.1]). Interestingly, in the proposed simulations this method exhibits fast convergence even when the limit point is a non isolated solution, as in the case of the second-order cone $C_1 = \left\{x \in \mathbb{R}^3 \mid x_3 \geq 0.1\sqrt{x_1^2 + x_2^2}\right\}$ and the tangent plane $C_2 = \left\{x \in \mathbb{R}^3 \mid x_3 = 0.1x_2\right\}$ considered in Figure 8.3.

However, computing the generalized Jacobian might be too demanding and require extra information not available in closed form. For this reason we focus on *quasi-Newton* methods

$$x^+ = x - HRx,$$

where the linear operator H is progressively updated with only evaluations of R and direct linear algebra in such a way that the vector HRx is asymptotically a *good approximation* of a Newton direction $G^{-1}Rx$. The yellow lines in the simulations of this section correspond to H being selected with the Broyden

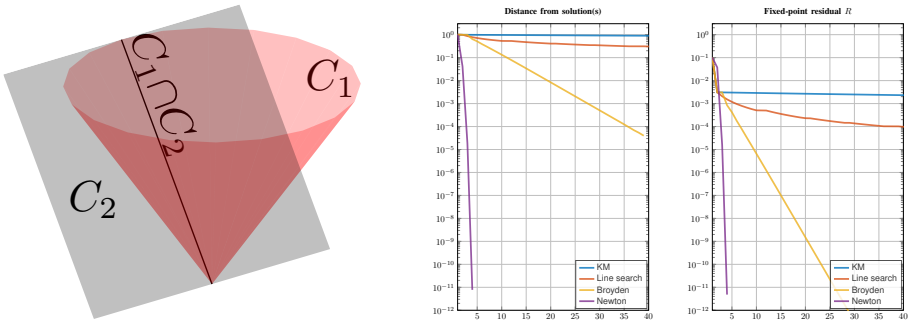


Figure 8.3: Alternating projections on second-order cone and tangent plane. *In contrast with the slow sublinear rate of KM both with and without line search, and despite the non isolatedness of any solution, Broyden scheme exhibits an appealing linear convergence rate.*

quasi-Newton method.

The crucial issue is *convergence* itself. Though in these trivial simulations it is not the case, it is well known that Newton-type methods in general converge only when close to a solution, and may even diverge otherwise. In fact, globalizing the convergence of Newton-type methods is a key challenge in optimization, as the dedicated recent book [61] confirms.

In this chapter we provide the *SuperMann scheme*, a globalization strategy for Newton-type methods (or any local scheme in general) that applies to any (nonsmooth) monotone equation deriving from fixed-point iterations of nonexpansive operators. Our method covers almost all splitting schemes in convex optimization, such as the forward-backward and Douglas-Rachford splittings, ADMM, and the versatile three-term Vũ-Condat splitting discussed more in detail in §8.7.3. We also provide sufficient conditions at the limit point under which the method reduces to the local scheme and converges superlinearly.

8.3 Notation and known results

8.3.1 Hilbert spaces and bounded linear operators

Throughout the chapter, $\mathcal{H}$ is a real separable Hilbert space endowed with an inner product $\langle \cdot\,, \cdot \rangle$ and with induced norm $\|\cdot\|$. The Euclidean norm and

scalar product are denoted as $\|\cdot\|_2$ and $\langle\cdot,\cdot\rangle_2$, respectively.

Given $(x_k)_{k\in\mathbb{N}} \subset \mathcal{H}$ and $x \in \mathcal{H}$ we write $x_k \rightharpoonup x$ to denote weak convergence of $(x_k)_{k\in\mathbb{N}}$ to x. The set of weak sequential cluster points of $(x_k)_{k\in\mathbb{N}}$ is indicated as $\mathcal{W}(x_k)_{k\in\mathbb{N}}$.

The set of bounded linear operators $\mathcal{H} \to \mathcal{H}$ is denoted as $\mathcal{B}(\mathcal{H})$. The adjoint operator of $L \in \mathcal{B}(\mathcal{H})$ is indicated as L^*, *i.e.*, the unique operator in $\mathcal{B}(\mathcal{H})$ such that $\langle Lx, y\rangle = \langle x, L^*y\rangle$ for all $x, y \in \mathcal{H}$.

8.3.2 Nonexpansive operators and Fejér sequences

We now briefly recap some known definitions and results of nonexpansive operator theory that will be used in the chapter.

Definition 8.1. *An operator $T : \mathcal{H} \to \mathcal{H}$ is said to be*

(i) NONEXPANSIVE (NE) *if $\|Tx - Ty\| \le \|x - y\|$ for all $x, y \in \mathcal{H}$;*

(ii) AVERAGED *if it is α-*AVERAGED *for some $\alpha \in (0, 1)$, i.e., if there exists a nonexpansive operator $S : \mathcal{H} \to \mathcal{H}$ such that $T = (1 - \alpha)\mathrm{id} + \alpha S$;*

(iii) FIRMLY NONEXPANSIVE (FNE) *if it is $1/2$-averaged.*

Clearly, for any NE operator T the residual $R = \mathrm{id} - T$ is monotone, in the sense that $\langle Rx - Ry, x - y\rangle \ge 0$ for all $x, y \in \mathcal{H}$; if T is additionally FNE, then not only is R monotone, but it is FNE as well. For notational convenience we extend the definition of α-averagedness to the case $\alpha = 1$ which reduces to plain nonexpansiveness.

For $\lambda \in \mathbb{R}$ we indicate the λ-averaging of T as

$$T_\lambda := (1 - \lambda)\mathrm{id} + \lambda T.$$

Notice that

$$\mathrm{id} - T_\lambda = \lambda(\mathrm{id} - T) \qquad \text{for all } \lambda \in \mathbb{R}, \tag{8.1}$$

and therefore $\operatorname{fix} T_\lambda = \operatorname{fix} T$ for all $\lambda \neq 0$. Moreover, if T is α-averaged and $\lambda \in (0, 1/\alpha]$, then

$$T_\lambda \text{ is } \alpha\lambda\text{-averaged} \tag{8.2}$$

[10, Prop. 4.40] and in particular $T_{1/2\alpha}$ is FNE.

Definition 8.2. *Relative to a nonempty set $S \subseteq \mathcal{H}$, a sequence $(x_k)_{k\in\mathbb{N}} \subset \mathcal{H}$ is*

(i) FEJÉR (-MONOTONE) *if $\|x_{k+1} - s\| \le \|x_k - s\|$ for all $k \in \mathbb{N}$ and $s \in S$;*

(ii) QUASI-FEJÉR (MONOTONE) *if for all* $s \in S$ *there exists a summable sequence* $(\varepsilon_k(s))_{k\in\mathbb{N}}$ *such that*

$$\|x_{k+1} - s\|^2 \leq \|x_k - s\|^2 + \varepsilon_k(s) \quad \forall k \in \mathbb{N}.$$

This definition of quasi-Fejér monotonicity is taken from [31] where it is referred to as *of type III*, and generalizes the classical definition [43].

Theorem 8.3. *Let* $T : \mathcal{H} \to \mathcal{H}$ *be an NE operator with* $\operatorname{fix} T \neq \emptyset$*, and suppose that* $(x_k)_{k\in\mathbb{N}} \subset \mathcal{H}$ *is quasi-Fejér with respect to* $\operatorname{fix} T$*. If* $(x_k - Tx_k)_{k\in\mathbb{N}} \to 0$*, then there exists* $x_\star \in \operatorname{fix} T$ *such that* $x_k \rightharpoonup x_\star$*.*

Proof. From [31, Prop. 3.7(i)] we have $\mathcal{W}(x_k)_{k\in\mathbb{N}} \neq \emptyset$; in turn, from [10, Cor. 4.28] we infer that $\mathcal{W}(x_k)_{k\in\mathbb{N}} \subseteq \operatorname{fix} T$. The claim then follows from [31, Thm. 3.8]. □

8.4 General abstract framework

Although this chapter analyses fixed-point iterations, differently from the rest of the thesis the fixed point operator $\mathcal{F}$, here rather denoted as T, is single valued and Lipschitz continuous with modulus 1. More specifically, the following will be assumed throughout the chapter.

Assumption 8.I. $T : \mathcal{H} \to \mathcal{H}$ is an α-averaged operator for some $\alpha \in (0, 1]$ and with $\operatorname{fix} T \neq \emptyset$. With $R \coloneqq \mathrm{id} - T$ we denote its (2α-Lipschitz continuous) fixed-point residual.

We also stick to this notation, so that, whenever mentioned, T, R, and α are as in Assumption 8.I. Our goal is to find a fixed point of T, or, equivalently, a zero of R:

$$\text{find } x_\star \in \operatorname{fix} T = \operatorname{zer} R. \tag{8.3}$$

In this section we introduce Algorithm 8.1, an abstract procedure to solve problem (8.3). The scheme is not implementable in and of itself, as it gives no hint as to how to compute each of the iterates, but it rather serves as a comprehensive ground framework for a class of algorithms with global convergence guarantees. In Section 8.6 we will derive the *SuperMann scheme*, an implementable instance which also enjoys appealing asymptotic properties.

The general framework prescribes three kinds of updates.

Algorithm 8.1 *General framework* for finding a fixed point of the α-averaged operator T with residual $R = \mathrm{id} - T$

REQUIRE $x_0 \in \mathcal{H}$, $c_0, c_1, q \in [0,1)$, $\sigma > 0$
INITIALIZE $\eta_0 = r_{\mathrm{safe}} = \|Rx_0\|$, $k = 0$

1. IF $Rx_k = 0$, THEN STOP.

2. IF $\|Rx_k\| \leq c_0 \eta_k$, THEN set $\eta_{k+1} = \|Rx_k\|$, proceed with a *blind update* x_{k+1} and go to step **4**.

3. Set $\eta_{k+1} = \eta_k$ and select x_{k+1} such that

3(a) EITHER the *safe condition* $\|Rx_k\| \leq r_{\mathrm{safe}}$ holds, and x_{k+1} is *educated*:

$$\|Rx_{k+1}\| \leq c_1 \|Rx_k\|$$

in which case update $r_{\mathrm{safe}} = \|Rx_{k+1}\| + q^k$;

3(b) OR it is *Fejérian* with respect to $\operatorname{fix} T$:

$$\|x_{k+1} - z\|^2 \leq \|x_k - z\|^2 - \sigma \|Rx_k\|^2 \quad \forall z \in \operatorname{fix} T. \tag{8.4}$$

4. Set $k \leftarrow k + 1$ and go to step **1**.

K_0) ***Blind* updates.** Inspired from [30], whenever the residual $\|Rx_k\|$ at iteration k has *sufficiently* decreased with respect to past iterates we allow for an *uncontrolled* update. For an efficient implementation such guess should be somehow reasonable and not completely a "*blind*" guess; however, for the sake of global convergence the proposed scheme is robust to any choice.

K_1) ***Educated* updates.** To encourage favorable updates, similarly to what has been proposed in [61, §5.3.1] and [44, §8.3.2] an *educated guess* x_{k+1} is accepted whenever the candidate residual is *sufficiently* smaller than the current.

K_2) ***Safeguard* (Fejérian) updates.** This last kind of updates is similar to K_1 as it is also based on the goodness of x_{k+1} with respect to x_k. The difference is that instead of checking the residual, what needs be *sufficiently* decreased is the distance from each point in $\operatorname{fix} T$. This is meant in a Fejérian fashion as in Definition 8.2.

Blind K_0- and *educated* K_1-updates are somehow complementary: the former is enabled when enough progress has been made in the past, whereas the

latter when the candidate update yields a sufficient improvement. *Progress* and *improvement* are meant in terms of a linear decrease of (the norm of) the residual; at iteration k, K_0 is enabled if $\|Rx_k\| \leq c_0\|Rx_{\bar{k}}\|$, where $c_0 \in [0,1)$ is a user-defined constant and $\bar{k}$ is the last blind iteration before k; K_1 is enabled if $\|Rx_{k+1}\| \leq c_1\|Rx_k\|$ where $c_1 \in [0,1)$ is another user-defined constant and x_{k+1} is the candidate next iterate. To ensure global convergence, *educated* updates are authorized only if the current residual $\|Rx_k\|$ is not larger than $\|Rx_{\tilde{k}+1}\|$ (up to a linearly decreasing error $q^{\tilde{k}}$); here $\tilde{k}$ denotes the last K_1-update before k.

While *blind* K_0- and *educated* K_1-updates are in charge of the asymptotic behavior, what makes the algorithm convergent are *safeguard* K_2-iterations.

8.4.1 Global weak convergence

To establish a notation, we partition the set of iteration indices $K \subseteq \mathbb{N}$ as $K_0 \cup K_1 \cup K_2$. Namely, relative to Algorithm 8.1, K_0 K_1 and K_2 denote the sets of indices k passing the test at steps **2**, **3**(a) and **3**(b), respectively. Furthermore, we index the sets K_0 and K_1 of *blind* and *educated* updates as

$$K_0 = \{k_1, k_2, \cdots\}, \qquad K_1 = \{k_1', k_2', \cdots\}. \tag{8.5}$$

To rule out trivialities, throughout the chapter we work under the assumption that a solution is not found in a finite number of steps, so that the residual of each iterate is always nonzero. As long as it is well defined, the algorithm therefore produces an infinite number of iterates.

Theorem 8.4 (Global convergence of the general framework Algorithm 8.1)**.** *Consider the iterates generated by Algorithm 8.1 and suppose that for all k it is always possible to find a point x_{k+1} complying with the requirements of either step **2**, **3**(a) or **3**(b), and further satisfying*

$$\|x_{k+1} - x_k\| \leq D\|Rx_k\| \quad \forall k \in K_0 \cup K_1 \tag{8.6}$$

for some constant $D \geq 0$. Then,

(i) $(x_k)_{k\in\mathbb{N}}$ *is quasi-Fejér monotone with respect to* $\operatorname{fix} T$*;*

(ii) $Rx_k \to 0$ *with* $(\|Rx_k\|)_{k\in\mathbb{N}} \in \ell^2$*;*

(iii) $(x_k)_{k\in\mathbb{N}}$ *converges weakly to a point* $x_\star \in \operatorname{fix} T$*;*

(iv) if $c_0 > 0$ *the number of* blind updates *at step **2** is infinite.*

Proof.

♠ 8.4*(i)*. We start by observing that because of (8.6) and the triangular inequality, for all $k \in K_0 \cup K_1$ we have

$$\|x_{k+1} - z\| \leq \|x_k - z\| + D\|Rx_k\| \qquad \forall z \in \operatorname{fix} T \tag{8.7}$$

and since R is 2α-Lipschitz continuous we also have that

$$\|Rx_{k+1}\| \leq \|Rx_k\| + \|Rx_{k+1} - Rx_k\| \leq (1 + 2\alpha D)\|Rx_k\|. \tag{8.8}$$

By combining [31, Prop. 3.2(i)] with (8.4) and (8.7), it follows that in order to prove quasi-Fejér monotonicity it suffices to show that the sequence $(\|Rx_k\|)_{k\in K_0\cup K_1}$ is summable. Let K_0 and K_1 be indexed as in (8.5). Since η_k is kept constant whenever $k \notin K_0$,

$$\eta_{k_\ell} = \|Rx_{k_{\ell-1}}\| \leq c_0\eta_{k_{\ell-1}} \leq \cdots \leq c_0^{\ell-1}\eta_{k_1} = c_0^{\ell-1}\eta_0 \qquad \forall k_\ell \in K_0. \tag{8.9}$$

In particular, $(\|Rx_{k_\ell}\|)_{k_\ell\in K_0}$ is summable (regardless of whether K_0 is finite or not).

As for $k'_\ell \in K_1$, the safeguard parameter r_{safe} ensures that

$$\|Rx_{k'_\ell}\| \leq \|Rx_{k'_{\ell-1}+1}\| + q^{k'_{\ell-1}} \leq c_1\|Rx_{k'_{\ell-1}}\| + q^{k'_{\ell-1}} \leq c_1\|Rx_{k'_{\ell-1}}\| + q^{\ell-1}$$

holds for all $k'_\ell \in K_1$. Iterating the inequality, for any $\rho \in (0,1)$ such that $\rho > \max\{c_1, q\}$ we have

$$\|Rx_{k'_\ell}\| \leq \rho^{\ell-1}\|Rx_{k'_1}\| + \sum_{i=1}^{\ell-1} c_1^{i-1}\rho^{\ell-i} \leq C\rho^\ell, \tag{8.10}$$

where $C \coloneqq \frac{1}{\rho}\left(\|Rx_{k'_1}\| + \sum_{i\in\mathbb{N}} (c_1/\rho)^i\right) < \infty$. In particular, also $(\|Rx_k\|)_{k\in K_1}$ is summable.

♠ 8.4*(ii)*. Due to quasi-Fejér monotonicity, for all $z \in \operatorname{fix} T$ there exists $(\varepsilon_k(z))_{k\in\mathbb{N}} \in \ell_1^+$ such that

$$\|x_{k+1} - z\|^2 \leq \|x_k - z\|^2 + \varepsilon_k(z).$$

By combining this with (8.4) and telescoping the inequalities, we obtain that for all $z \in \operatorname{fix} T$

$$\|x_0 - z\|^2 \geq \sigma \sum_{k\in K_2} \|Rx_k\|^2 \; - \sum_{k\in K_0\cup K_1} \varepsilon_k(z). \tag{8.11}$$

Since the sequence $(\varepsilon_k(z))_{k\in K_0\cup K_1}$ is summable, then so is $(\|Rx_k\|^2)_{k\in K_2}$. In turn, since $(\|Rx_k\|)_{k\in K_0\cup K_1}$ is also summable it follows that the whole sequence of residuals is square-summable.

♠ 8.4*(iii)*. Follows by combining 8.4*(ii)* with Thm. 8.3.

♠ 8.4*(iv)*. Trivially follows from the already proven point 8.4*(ii)*, together with the observation that since η_k is kept constant whenever $k \notin K_0$, the condition $\|Rx_k\| \leq c_0\eta_k$ will be satisfied infinitely often if $c_0 > 0$. □

8.4.2 Local linear convergence

More can be said about the convergence rates if the mapping R possesses METRIC SUBREGULARITY. Differently from (bounded) linear regularity [12], metric subregularity is a local property and as such it is more general. For a (possibly multivalued) operator R, metric subregularity at $\bar{x}$ is equivalent to calmness of R^{-1} at $R\bar{x}$ [39, Thm 3.2], and is a weaker condition than metric regularity and Aubin property. We refer the reader to [106, §9] for an extensive discussion.

Definition 8.5 (Metric subregularity at zeros)**.** *Let $R : \mathcal{H} \to \mathcal{H}$ and $\bar{x} \in \operatorname{zer} R$. R is* METRICALLY SUBREGULAR *at $\bar{x}$ if there exist $\varepsilon, \gamma > 0$ such that*

$$\operatorname{dist}(x, \operatorname{zer} R) \leq \gamma\|Rx\| \qquad \forall x \in \mathrm{B}(\bar{x}; \varepsilon). \tag{8.12}$$

γ and ε are (one) MODULUS *and (one)* RADIUS *of subregularity of R at $\bar{x}$, respectively.*

In finite-dimensional spaces, if R is differentiable at $\bar{x} \in \operatorname{zer} R$ and $\bar{x}$ is isolated in $\operatorname{zer} R$ (*e.g.*, if it is the unique zero), then metric subregularity is equivalent to nonsingularity of $JR\bar{x}$. Metric subregularity is however a much weaker property than nonsingularity of the Jacobian, firstly because it does not assume differentiability, and secondly because it can cope with 'wide' regions of zeros; for instance, any piecewise linear mapping is globally metrically subregular [101].

If the residual $R = \mathrm{id} - T$ of the α-averaged operator T is metrically subregular at $\bar{x} \in \operatorname{zer} R = \operatorname{fix} T$ with modulus γ and radius ε, then

$$\tfrac{1}{\gamma}\operatorname{dist}(x, \operatorname{fix} T) \leq \|Rx\| \leq 2\alpha \operatorname{dist}(x, \operatorname{fix} T) \qquad \forall x \in \mathrm{B}(\bar{x}; \varepsilon). \tag{8.13}$$

Consequently, if $\|Rx_k\| \to 0$ for some sequence $(x_k)_{k\in\mathbb{N}} \subset \mathcal{H}$, so does $\operatorname{dist}(x_k, \operatorname{fix} T)$ with the same asymptotic rate of convergence, and viceversa.

Metric subregularity is the key property under which the residual in the classical KM scheme achieves linear convergence; in Theorem 8.8 we show that this asymptotic behavior is preserved in the general framework of Algorithm 8.1. We first need to prove two lemmas.

Lemma 8.6 (Asymptotic properties of K_0 and K_1)**.** *Suppose the hypotheses of Theorem 8.4 hold and let $(x_k)_{k\in\mathbb{N}}$ be the sequence generated by Algorithm 8.1. Then,*

(i) $(\|Rx_k\|)_{k\in K_0}$ *is Q-linearly convergent;*

(ii) $(\|Rx_k\|)_{k\in K_1}$ *is R-linearly convergent;*

(iii) if $c_0 > 0$ then for some $\varrho \in (0,1]$ and $\beta \in \mathbb{R}$

$$\ell_0(k) \geq \varrho\, \ell_1(k) - \beta \quad \forall k \in \mathbb{N},$$

where $\ell_j(k) := \#\{k' \in K_j \mid k' \leq k\}$, $j = 0,1,2$, is the number of times K_j was visited up to iteration k.

Proof.

♠ 8.6*(i)* and 8.6*(ii)*. Already shown in (8.9) and (8.10).

♠ 8.6*(iii)*. If $c_1 = 0$, then $K_1 = \emptyset$ and the claim trivially holds with $\varrho = 1$ and $\beta = 0$. Otherwise, from (8.10) and due to the definition of $\ell_1(k)$ there exist $C > 0$ and $\rho \in (0,1)$ such that

$$\|Rx_k\| \leq C\rho^{\ell_1(k)} \qquad \forall k \in K_1.$$

If $k \in K_1$, then $\|Rx_k\|$ didn't pass the test at step **2**, therefore

$$C\rho^{\ell_1(k)} \geq \|Rx_k\| \geq \eta_k \overset{(8.9)}{=} \|Rx_0\| c_0^{\ell_0(k)}.$$

The proof now follows by simply taking the logarithm on the outer inequality. □

Lemma 8.7. *Let $(u_k)_{k\in\mathbb{N}} \subset [0,+\infty)$ be a sequence, and let $K_1, K_2 \subseteq \mathbb{N}$ be such that $\mathbb{N} = K_1 \cup K_2$. Let K_1 be indexed as $K_1 = \{k'_0, k'_1 \ldots\}$, and suppose that there exist $a, b > 0$ and $\rho \in (0,1)$ such that*

$$\begin{cases} u_{k+1} \leq a u_k & \text{for all } k \in \mathbb{N}, \\ u_{k'_\ell} \leq b\rho^\ell & \text{for all } k'_\ell \in K_1, \\ u_{k+1} \leq \rho u_k & \text{for all } k \in K_2. \end{cases}$$

Then, there exists $\sigma \in (0,1)$ such that $u_k \leq ab\sigma^k$.

Proof. To exclude trivialities we assume that K_1 and K_2 are both infinite. To arrive to a contradiction, for all $\sigma \in (0,1)$ let $k = k(\sigma)$ be the minimum such that $u_k > ab\sigma^k$. Let $\sigma \geq \rho$ be fixed. If $k-1 \in K_2$, then

$$\rho u_{k-1} \geq u_k > ab\sigma^k \geq ab\rho\sigma^{k-1}$$

and therefore $u_{k-1} > ab\sigma^{k-1}$ which contradicts minimality of k. It follows that necessarily $k-1 \in K_1$, hence $k-1 = k'_\ell \in K_1$ for some $\ell \in \mathbb{N}$. For all $n \in \mathbb{N}$, let $k'_{\ell_n} = k(\rho^{1/n}) - 1$, *i.e.*, the minimum such that $u_{k'_{\ell_n}+1} > ab\rho^{\frac{k'_{\ell_n}+1}{n}}$. Combining with the property of K_1 we obtain

$$ab\rho^{\frac{k'_{\ell_n}+1}{n}} < u_{k'_{\ell_n}+1} \leq a u_{k'_{\ell_n}} \leq ab\rho^{\ell_n} \tag{8.14}$$

and in particular $\ell_n \leq \frac{k'_{\ell_n}}{n}$. This means that up to $k = k'_{\ell_n}$ there are at most k/n elements in K_1, and consequently at least $k - k/n$ in K_2. Therefore,

$$b\rho^{\frac{k+1}{n}} \overset{(8.14)}{<} u_k \leq a^{k/n}\rho^{k-k/n}u_0.$$

Taking the k-th square root on the outer inequality yields

$$(1/\rho)^{1-2/n-1/nk} < (u_0/b)^{1/k} a^{1/n}.$$

By letting $n \to \infty$ (hence $k \to \infty$) we arrive to the contradiction $\rho \geq 1$. □

Theorem 8.8 (Linear convergence of the general framework Algorithm 8.1)**.** *Suppose that the hypotheses of Theorem 8.4 hold, and suppose further that $(x_k)_{k\in\mathbb{N}}$ converges strongly to a point $x_\star$ (this being true if $\mathcal{H}$ is finite dimensional) at which R is metrically subregular.*

Then, $(x_k)_{k\in\mathbb{N}}$ and $(Rx_k)_{k\in\mathbb{N}}$ are R-linearly convergent.

Proof. Letting $e_k := \operatorname{dist}(x_k, \operatorname{fix} T)$, because of (8.8) and (8.13) there exists $B > 1$ such that

$$\|Rx_{k+1}\| \leq B\|Rx_k\| \quad \text{and} \quad e_{k+1} \leq Be_k \quad \forall k \in \mathbb{N}. \tag{8.15}$$

Suppose that R is metrically subregular at $x_\star$ with radius $\varepsilon > 0$ and modulus $\gamma > 0$; since $x_k \to x_\star$, up to an index shifting without loss of generality we may assume that $(x_k)_{k\in\mathbb{N}} \subset \mathrm{B}(x_\star; \varepsilon)$. Let $z_k = \Pi_{\operatorname{fix} T}\, x_k$, so that $e_k = \|x_k - z_k\|$; by combining (8.4) and (8.13) we obtain

$$e_{k+1}^2 \leq \|x_{k+1} - z_k\|^2 \leq \|x_k - z_k\|^2 - \sigma\|Rx_k\|^2 \leq \rho^2 e_k^2 \quad \forall k \in K_2, \tag{8.16}$$

where $\rho := \sqrt{1-\sigma/\gamma^2} \in (0,1)$. By possibly enlarging ρ we assume $\rho \geq \max\{c_0, c_1\}$.

If $c_0 = 0$, then $K_0 = \emptyset$ and by using Lem. 8.6*(ii)* and (8.15) we may invoke Lem. 8.7 to infer R-linear convergence of the sequence $(e_k)_{k\in\mathbb{N}}$ and conclude the proof.

Therefore, let us suppose that $c_0 > 0$, so that due to Thm. 8.4*(iv)* the set K_0 contains infinite many indices. We now show that there exists $n \in \mathbb{N}$ such that every n consecutive indices at least one is in K_0. Let $k \in K_0$ be fixed and suppose that $k+1 \dots k+n+1 \notin K_0$.

- If $c_1 = 0$ then $K_1 = \emptyset$ and all such indices belong to K_2. Then,

$$\|Rx_{k+n+1}\| \overset{(8.13)}{\leq} 2\alpha e_{k+n+1} \overset{(8.16)}{\leq} 2\alpha\rho^n e_{k+1} \overset{(8.15)}{\leq} 2\alpha B\rho^n e_k \overset{(8.13)}{\leq} 2\alpha\gamma B\rho^n \|Rx_k\|.$$

  Since $k+n+1 \notin K_0$, then $\|Rx_{k+n+1}\|$ failed the test at step **3** and therefore

$$c_0\|Rx_k\| = c_0\eta_{k+n+1} < \|Rx_{k+n+1}\| \leq 2\alpha\gamma B\rho^n\|Rx_k\|,$$

  which proves that n cannot be arbitrarily large.

- If instead $c_1 > 0$, let n_1 be the number of indices among $k+1 \dots k+n$ that belong to K_1, and $n_2 = n - n_1$ those belonging to K_2. Then, from iteration $k+1$ to $k+n+1$ the distance from the fixed set has reduced n_2 times (at least) by a factor ρ and, due to (8.15), increased at most by a factor B the remaining n_1 times:

$$\|Rx_{k+n+1}\| \overset{(8.13)}{\leq} 2\alpha e_{k+n+1} \leq 2\alpha\rho^{n_2}B^{n_1}e_{k+1} \overset{(8.15)}{\leq} 2\alpha\rho^{n_2}B^{n_1+1}e_k$$
$$\overset{(8.13)}{\leq} 2\alpha\gamma\rho^{n_2}B^{n_1+1}\|Rx_k\|.$$

  Again, since $k+n+1 \notin K_0$ we have $c_0\|Rx_k\| < 2\alpha\gamma\rho^{n_2}B^{n_1+1}\|Rx_k\|$, and therefore

$$n_1 > \tfrac{\ln c_0/2\alpha\gamma}{\ln B} - 1 + \tfrac{\ln 1/\rho}{\ln B} n_2.$$

  In particular, for large n the number n_1 of indices in K_1 grows proportionally with respect to n, and from Lem. 8.6*(iii)* we conclude once again that n cannot be arbitrarily large (since the number of visits to K_0 does not change from $k+1$ to $k+n$).

So far we proved that there exists $n \in \mathbb{N}$ such that every n indices at least one belongs to K_0. In particular, indexing $K_0 = \{k_0, k_1 \cdots\}$ we have that $k_\ell \leq n\ell$,

hence

$$\|Rx_{k_\ell}\| \leq c_0^\ell \|Rx_0\| \leq \left(c_0^{1/n}\right)^{k_\ell} \|Rx_0\| \quad \forall k_\ell \in K_0. \tag{8.17}$$

Moreover, any $k \in \mathbb{N}$ is at most $n-1$ indices away from the nearest previous index $k_\ell \in K_0$; combined with (8.17) and by invoking (8.15) we obtain

$$\|Rx_k\| \leq B^{n-1}\|Rx_0\|\left(c_0^{1/n}\right)^{k_\ell} \leq B^{n-1}\|Rx_0\|\left(c_0^{1/n}\right)^{k}$$

proving the sought R-linear convergence of $(\|Rx_k\|)_{k\in\mathbb{N}}$. It follows that for some $b > 0$ and $r \in (0,1)$ we have $\|Rx_k\| \leq br^k$ for all $k \in \mathbb{N}$; then,

$$\|x_k - x_\star\| \leq \sum_{j\geq k} \|x_{j+1} - x_j\| \overset{(8.6)}{\leq} D \sum_{j\geq k} \|Rx_j\| \leq bD \sum_{j\geq k} r^j = \frac{bD}{1-r} r^k,$$

where in the second inequality we used the bound (8.6), which also holds for $k \in K_2$ (up to possibly enlarging D) due to the fact that for $k \in K_2$ under metric subregularity we have

$$\|x_{k+1} - x_k\| \leq \|x_{k+1} - z_k\| + \|x_k - z_k\| \overset{(8.13)}{\leq} 2e_k \leq 2\gamma\|Rx_k\|.$$

This shows that $(x_k)_{k\in\mathbb{N}}$ is R-linearly convergent too. □

8.4.3 Main idea

Being interested in solving the nonlinear equation (8.3), one could think of implementing one of the many existing *fast* methods for nonlinear equations that achieve fast asymptotic rates, such as Newton-type schemes. At each iteration, such schemes compute an update direction d_k and prescribe steps of the form $x_{k+1} = x_k + \tau_k d_k$, where $\tau_k > 0$ is a stepsize that needs to be sufficiently small in order for the method to enjoy global convergence; on the other hand, fast asymptotic rates are ensured if $\tau_k = 1$ is eventually always accepted. The stepsize is a crucial feature of fast methods, and a feasible τ_k is usually backtracked with a line search on a smooth merit function. Unfortunately, in meaningful applications of the problem at hand arising from fixed-point theory the residual mapping R is nonsmooth, and the typical merit function $x \mapsto \|Rx\|^2$ does not meet the necessary smoothness requirement.

What we propose in this chapter is a hybrid scheme that allows for the employment of any (fast) method for solving nonlinear equations, with global convergence guarantees that do not require smoothness, but which is based only on the nonexpansiveness of T. Once *fast* directions d_k are selected, Algorithm 8.1 can be specialized as follows:

1) *blind* updates as in step **2** shall be of the form $x_{k+1} = x_k + d_k$;
2) *educated* updates as in step **3**(a) shall be of the form $x_{k+1} = x_k + \tau_k d_k$, with τ_k small enough so as to ensure the acceptance condition $\|Rx_{k+1}\| \leq c_1\|Rx_k\|$;
3) *safeguard* updates as in step **3**(b) shall be employed as *last resort* both for globalization purposes and for well definedness of the scheme.

Ideally, the scheme should eventually reduce to the local scheme $x_{k+1} = x_k + d_k$ when *good* directions d_k are used.

In Section 8.5 we address the problem of providing explicit *safeguard* updates that comply with the quasi-Fejér monotonicity requirement of step **3**(b). Because of the arbitrarity of the other two updates, once we succeed in this task Algorithm 8.1 will be of practical implementation. In Section 8.6 we will then discuss specific K_0- and K_1-updates to be used at steps **2** and **3**(a) that ensure global *and* fast convergence, yet maintaining the simplicity of fixed-point iterations of T (evaluations of T and direct linear algebra).

8.5 Generalized Mann Iterations

8.5.1 The classical Krasnosel'skiĭ-Mann scheme

Starting from a point $x_0 \in \mathcal{H}$, the classical Krasnosel'skiĭ-Mann scheme (KM) performs the following updates

$$x_{k+1} = T_{\lambda_k} x_k = (1 - \lambda_k)x_k + \lambda_k T x_k \tag{8.18}$$

and converges weakly to a fixed point of T provided that $\lambda_k \in [0, 1/\alpha]$ and $(\lambda_k(1/\alpha - \lambda_k))_{k\in\mathbb{N}} \notin \ell^1$ [10, Thm. 5.14]. The key property of KM iterations is Fejér monotonicity:

$$\|x_{k+1} - z\|^2 \leq \|x_k - z\|^2 - \lambda_k(1/\alpha - \lambda_k)\|Rx_k\|^2 \quad \forall z \in \operatorname{fix} T.$$

In particular, in Algorithm 8.1 KM iterations can be used as *safeguard* updates at step **3**(b). The drawback of such a selection is that it completely discards the hypothetical fast update direction d_k that *blind* and *educated* updates try to enforce. This is particularly penalizing when the local method for computing the directions d_k is a *quasi-Newton* scheme; such methods are indeed very sensitive to past iterations, and discarding directions is neither theoretically sound nor beneficial in practice.

In this section we provide alternative *safeguard* updates that while ensuring the desirable Fejér monotonicity are also amenable to taking into account arbitrary directions. The key idea lies in intepreting KM iterations as projections onto suitable half-spaces (see Fig. 8.4), and then exploiting known properties of projections. These facts are shown in the next result. To this end, let us remark that the projection Π_C onto a nonempty closed and convex set C is FNE [10, Prop. 4.16], and that consequently its λ-averaging $\Pi_{C,\lambda}$ is $\lambda/2$-averaged for any $\lambda \in (0, 2]$, as it follows from (8.2).

Proposition 8.9 (KM iterations as projections)**.** *For $x \in \mathcal{H}$, define*

$$C_x = C_x^{T,\alpha} := \big\{z \in \mathcal{H} \mid \|Rx\|^2 - 2\alpha\langle Rx, x - z\rangle \leq 0\big\}. \tag{8.19}$$

Then,

(i) $x \in C_x$ *iff* $x \in \operatorname{fix} T$*;*

(ii) $\operatorname{fix} T = \bigcap_{x\in\mathcal{H}} C_x$*;*

(iii) for any $\lambda \in [0, 1/\alpha]$ *it holds that* $T_\lambda x = \Pi_{C_x, 2\alpha\lambda} x = (1 - 2\alpha\lambda)x + 2\alpha\lambda \Pi_{C_x} x$*.*

Proof. The set C_x can be equivalently expressed as

$$C_x = \big\{z \in \mathcal{H} \mid \langle x - T_{1/2\alpha}x, z - T_{1/2\alpha}x\rangle \leq 0\big\}.$$

8.9*(i)* is of immediate verification, and 8.9*(ii)* then follows from [10, Cor. 4.25] combined with (8.2).

We now show 8.9*(iii)*. If $Rx = 0$, then $x \in \operatorname{fix} T$ and $C_x = \mathcal{H}$, and the claim is trivial. Otherwise, notice that

$$C_x = \big\{z \in \mathcal{H} \mid \langle Rx, z\rangle \leq \langle Rx, x - \tfrac{1}{2\alpha}Rx\rangle\big\}, \tag{8.20}$$

and the claim can be readily verified using the formula for the projection on a halfspace $H_{v,\beta} := \{z \in \mathcal{H} \mid \langle v, z\rangle \leq \beta\}$, namely

$$\Pi_{H_{v,\beta}} x = x - \frac{[\langle v, x\rangle - \beta]_+}{\|v\|^2} v, \tag{8.21}$$

defined for $v \in \mathcal{H} \setminus \{0\}$ and $\beta \in \mathbb{R}$ [10, Ex. 29.20(iii)]. □

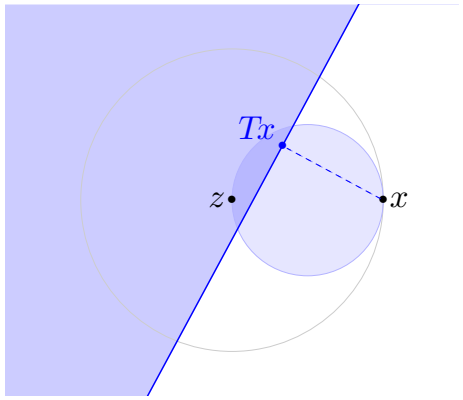


Figure 8.4: Mann iteration of a FNE operator T as projection on C_x *(the blue half-space, as defined in* (8.19) *for* $\alpha = 1/2$*). The outer circle is the set of all possible images of a nonexpansive operator, given that* z *is a fixed point. The inner circle corresponds to the possible images of* firmly *nonexpansive operators. Notice that* C_x *separates* x *from* z *as long as* Tx *is contained in the small circle, which characterizes firm nonexpansiveness.*

8.5.2 Generalized Mann projections

Though particularly attractive for its simplicity and global convergence properties, the KM scheme (8.18) finds its main drawback in its convergence rate, being Q-linear at best and highly sensitive to ill conditioning of the problem. In response to this issue, Algorithm 8.1 allows for the integration of fast local methods still ensuring global convergence properties. The efficiency of the resulting scheme, which will be proven later on, is based on an ad hoc selection of *safeguard* updates for step **3**(b) which is based on the following generalization of Proposition 8.9.

Proposition 8.10. *Suppose that* $x, w \in \mathcal{H}$ *are such that*

$$\rho := \|Rw\|^2 - 2\alpha\langle Rw, w - x\rangle > 0. \tag{8.22}$$

For $\lambda \in [0, 1/\alpha]$ *let*

$$x^+ := x - \lambda\frac{\rho}{\|Rw\|^2}Rw. \tag{8.23}$$

Then, the following hold:

(i) $x^+ = \Pi_{C_w, 2\alpha\lambda}\, x$ *where* $C_w = C_w^{T,\alpha}$ *as in* (8.19)*;*

(ii) $\|x^+ - z\|^2 \leq \|x - z\|^2 - \lambda(1/\alpha - \lambda)\frac{\rho^2}{\|Rw\|^2} \quad \forall z \in \operatorname{fix} T.$

Proof. 8.10*(i)* easily follows from (8.20) and (8.21), since by condition (8.22) the positive part in the formula may be omitted. In turn, 8.10*(ii)* follows from [10, Prop. 4.35(iii)] by observing that $\Pi_{C_w,2\alpha\lambda}$ is $\alpha\lambda$-averaged due to [10, Prop. 4.16] and (8.2), and that $\operatorname{fix} T \subseteq C_w$ as shown in Prop. 8.9*(ii)*. □

Notice that condition (8.22) is equivalent to $x \notin C_w$. Therefore, Proposition 8.10*(ii)* states that whenever a point x lies outside the half-space C_w for some $w \in \mathcal{H}$, since $\operatorname{fix} T \subseteq C_w$ (cf. Prop. 8.9) the projection onto C_w moves closer to $\operatorname{fix} T$. This means that after moving from x along a candidate direction d to the point $w = x + d$, even though w might be farther from $\operatorname{fix} T$ the point $x^+ = \Pi_{C_w} x$ is not. We may then use this projection as a safeguard step to prevent from diverging from the set of fixed points. Based on this, we define a GENERALIZED KM UPDATE ALONG A DIRECTION d.

Definition 8.11 (GKM update)**.** *A* GENERALIZED KM UPDATE (GKM) AT x ALONG d *for the α-averaged operator $T : \mathcal{H} \to \mathcal{H}$ with relaxation $\lambda \in [0, 1/\alpha]$ is*

$$x^+ := \begin{cases} x & \text{if } w \in \operatorname{fix} T \\ x - \lambda \frac{[\rho]_+}{\|Rw\|^2} Rw & \text{otherwise,} \end{cases}$$

where $w = x + d$ and $\rho := \|Rw\|^2 - 2\alpha\langle Rw, w - x\rangle$. In particular, $d = 0$ yields the classical KM update $x^+ = T_\lambda x$.

8.5.3 Line search for GKM

It is evident from Definition 8.11 that a GKM update trivializes to $x^+ = x$ if either $w \in \operatorname{fix} T$ or $\rho \leq 0$. Having $w \in \operatorname{fix} T$ corresponds to having found a solution to problem (8.3), and the case deserves no further investigation. In this section we address the remaining case $\rho \leq 0$, showing how it can be avoided by simply introducing a suitable line search. In order to recover the same global convergence properties of the classical KM scheme we need something more than simply imposing $\rho > 0$. The next result addresses this requirement, showing further that it is achieved for any direction d by sufficiently small stepsizes.

Theorem 8.12. *Let $x, d \in \mathcal{H}$ and $\sigma \in [0, 1)$ be fixed, and consider*

$$\bar{\tau} = \begin{cases} 1 & \text{if } d = 0 \\ \frac{1-\sigma}{4\alpha} \frac{\|Rx\|}{\|d\|} & \text{otherwise.} \end{cases}$$

Then, for all $\tau \in (0, \bar{\tau}]$ the point $w = x + \tau d$ satisfies

$$\rho := \|Rw\|^2 - 2\alpha\langle Rw, w - x\rangle \geq \sigma \|Rw\| \|Rx\|. \tag{8.24}$$

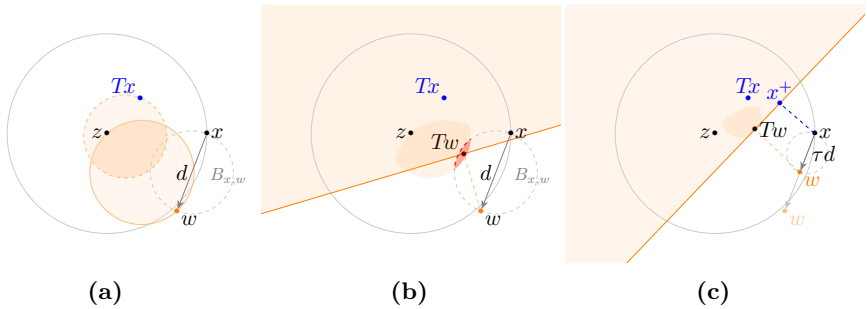


Figure 8.5: SuperMann iteration of a FNE operator T as projection on C_w.
(a) *the darker orange region represents the area in which Tw must lie given the points x, Tx and the fixed point z as prescribed by firm nonexpansiveness of T.*
(b) *if Tw lies (also) in the ball $B_{x,w}$ as in (8.25), then the half-space C_w (shaded in orange) separates x from w, which is to be avoided.*
(c) *when w is close enough to x the feasible region for Tw has empty intersection with $B_{x,w}$ and C_w does not contain x.*

Proof. Let a constant $c \geq 0$ to be determined be such that

$$\tau\|d\| = \|w - x\| \leq c\|Rx\|.$$

Observe that $\rho = 4\alpha^2\langle w - T_{1/2\alpha}w, x - T_{1/2\alpha}w\rangle$, and recall from (8.1) and (8.2) that $T_{1/2\alpha}$ is FNE with residual $\mathrm{id} - T_{1/2\alpha} = \frac{1}{2\alpha}R$. Then,

$$\rho = 4\alpha^2\big(\|w - T_{1/2\alpha}w\|^2 + \langle w - T_{1/2\alpha}w, x - w\rangle\big)$$

using Cauchy-Schwartz inequality,

$$\geq 4\alpha^2\|w - T_{1/2\alpha}w\|\big(\|w - T_{1/2\alpha}w\| - \|x - w\|\big)$$

the bound on $\|x - w\|$,

$$\geq 2\alpha\|Rw\|\big(\|w - T_{1/2\alpha}w\| - 2\alpha c\|x - T_{1/2\alpha}x\|\big)$$

the (reverse) triangular inequality,

$$\geq 2\alpha\|Rw\|\big((1 - 2\alpha c)\|x - T_{1/2\alpha}x\| - \|(\mathrm{id} - T_{1/2\alpha})w - (\mathrm{id} - T_{1/2\alpha})x\|\big)$$

the nonexpansiveness of $\mathrm{id} - T_{1/2\alpha}$

$$\geq 2\alpha\|Rw\|\big(\tfrac{1-2\alpha c}{2\alpha}\|Rx\| - \|w - x\|\big)$$

and again the bound on $\|w - x\|$,

$$\geq (1 - 4\alpha c)\|Rw\|\|Rx\|$$

equating $\sigma = 1 - 4\alpha c$ the assert follows. □

Notice that if $d = 0$, then $\rho = \|Rx\|^2 \geq \sigma\|Rx\|^2$ for any $\sigma \in [0, 1)$, and therefore the line search condition (8.24) is always satisfied; in particular, the classical KM step $x^+ = Tx$ is always accepted regardless of the value of σ.

Let us now observe how a GKM projection extends the classical KM depicted in Figure 8.4 and how the line search works. In the following we use the notation of Theorem 8.12, and for the sake of simplicity we consider $\sigma = 0$ in (8.24) and a FNE operator T. Suppose that the fixed point z and the points x, Tx, and w are as in Figure 8.5a; due to firm nonexpansiveness, the image Tw of w is somewhere inside both orange circles. We want to avoid the unfavorable situation depicted in Figure 8.5b, where the couple (w, Tw) generates a halfspace C_w that contains x, *i.e.*, such that $\rho \leq 0$: in fact, with simple algebra it can be seen that $\rho \leq 0$ iff Tw belongs to the dashed circle of Figure 8.5b:

$$B_{x,w} \coloneqq \{\bar{w} \mid \langle w - \bar{w}, x - \bar{w} \rangle \leq 0\}. \tag{8.25}$$

Since the dashed orange circle (in which Tw must lie) is simply the translation by a vector $Tx - x$ of $B_{x,w}$, both having diameter $\tau\|d\|$, for sufficiently small τ the two have empty intersection, meaning that $\rho > 0$ regardless of where Tw is.

8.6 The SuperMann scheme

In this section we introduce the *SuperMann scheme (Alg. 8.2)*, a special instance of the general framework of Algorithm 8.1 that employs GKM updates as safeguard K_2-steps. While the global worst-case convergence properties of *SuperMann* are the same as for the classical KM scheme, its asymptotic behavior is determined by how *blind* K_0- and *educated* K_1-updates are selected. In Section 8.6.2 we will characterize the "quality" of update directions and the mild requirements under which superlinear convergence rates are attained; in particular, Section 8.6.3 is dedicated to the analysis of quasi-Newton Broyden directions.

The scheme follows the same philosophy of the general abstract framework. The main idea is globalizing a local method for solving the monotone equation $Rx = 0$, in such a way that when the iterates get close enough to a solution the fast convergence of the local method is automatically triggered. Approaching a solution is possible thanks to the generalized KM updates (step **5(b)**), provided enough backtracking is performed, as ensured by Prop. 8.10*(ii)* and Thm. 8.12. When a basin of fast (*i.e.*, superlinear) attraction for the local method is reached, the (norm of) Rx will decrease more than linearly, and the condition triggering the *educated* updates of step **5(a)** (which is checked first) will be verified without performing any backtracking.

Algorithm 8.2 *SuperMann* scheme for solving (8.3), given an α-averaged operator T with residual $R = \mathrm{id} - T$

REQUIRE $x_0 \in \mathcal{H}$, $c_0, c_1, q \in [0,1)$, $\beta, \sigma \in (0,1)$, $\lambda \in (0, 1/\alpha)$.
INITIALIZE $\eta_0 = r_{\text{safe}} = \|Rx_0\|$, $k = 0$

1. IF $Rx_k = 0$, THEN STOP.

2. Choose an update direction $d_k \in \mathcal{H}$

3. **(K_0)** IF $\|Rx_k\| \le c_0\eta_k$, THEN set $\eta_{k+1} = \|Rx_k\|$, proceed with a *blind update* $x_{k+1} = w_k := x_k + d_k$ and go to step **6**.

4. Set $\eta_{k+1} = \eta_k$ and $\tau_k = 1$.

5. Let $w_k = x_k + \tau_k d_k$.

5(a) **(K_1)** IF the *safe condition* $\|Rx_k\| \le r_{\text{safe}}$ holds and w_k is *educated*:

$$\|Rw_k\| \le c_1\|Rx_k\|$$

THEN set $x_{k+1} = w_k$, update $r_{\text{safe}} = \|Rw_k\| + q^k$, and go to step **6**.

5(b) **(K_2)** IF $\rho_k := \|Rw_k\|^2 - 2\alpha\langle Rw_k, w_k - x_k\rangle \ge \sigma\|Rw_k\|\|Rx_k\|$ THEN set

$$x_{k+1} = x_k - \lambda\frac{\rho_k}{\|Rw_k\|^2}Rw_k$$

OTHERWISE set $\tau_k \leftarrow \beta\tau_k$ and go to step **5**.

6. Set $k \leftarrow k + 1$ and go to step **1**.

To discuss its global and local convergence properties we stick to the same notation of the general framework of Algorithm 8.1, denoting the sets of *blind*, *educated*, and *safeguard* updates as K_0, K_1 and K_2, respectively.

8.6.1 Global and linear convergence

To comply with (8.6), we impose the following requirement on the magnitude of the directions (see also Rem. 8.20).

Assumption 8.II. There exists a constant $D \ge 0$ such that the directions $(d_k)_{k\in\mathbb{N}}$ in the *SuperMann scheme (Alg. 8.2)* satisfy

$$\|d_k\| \le D\|Rx_k\| \qquad \forall k \in \mathbb{N}. \tag{8.26}$$

Theorem 8.13 (Global and linear convergence of the *SuperMann scheme*)**.** *Consider the iterates generated by the SuperMann scheme (Alg. 8.2) with* $(d_k)_{k\in\mathbb{N}}$ *selected so as to satisfy Assumption 8.II. Then,*

(i) $(x_k)_{k\in\mathbb{N}}$ *is quasi-Fejér monotone with respect to* $\operatorname{fix} T$*;*

(ii) $\tau_k = 1$ *if* $d_k = 0$*, and* $\tau_k \geq \min\left\{\beta\frac{1-\sigma}{4\alpha D}, 1\right\}$ *otherwise.*

(iii) $Rx_k \to 0$ *with* $(\|Rx_k\|)_{k\in\mathbb{N}} \in \ell^2$*;*

(iv) $(x_k)_{k\in\mathbb{N}}$ *converges weakly to a point* $x_\star \in \operatorname{fix} T$*;*

(v) if $c_0 > 0$ *the number of blind updates at step* ***3*** *is infinite.*

Moreover, if $(x_k)_{k\in\mathbb{N}}$ *converges strongly to a point* $x_\star$ *(this being true if* $\mathcal{H}$ *is finite dimensional) at which* R *is metrically subregular, then*

(vi) $(x_k)_{k\in\mathbb{N}}$ *and* $(Rx_k)_{k\in\mathbb{N}}$ *are R-linearly convergent.*

Proof. Because of Thm. 8.12 we know that for any direction d_k a feasible stepsize τ_k complying with the requirements of step **5(b)** will eventually be found, lower bounded as in 8.13*(ii)* due to Thm. 8.12 and Assumption 8.II. In particular, the scheme is well defined. Moreover, from Prop. 8.10*(ii)* we have that there exists a constant $\underline{\sigma} > 0$ such that

$$\|x_{k+1} - z\|^2 \leq \|x_k - z\|^2 - \underline{\sigma}\|Rx_k\|^2 \quad \text{for all } k \in K_2 \text{ and } z \in \operatorname{fix} T.$$

It follows that the *SuperMann scheme* is a special case of Alg. 8.1 and the proof entirely follows from Thm.s 8.4 and 8.8. □

8.6.2 Superlinear convergence

Though global convercence of the *SuperMann scheme* is independent of the choice of the directions d_k, its performance and tail convergence surely is not. We characterize the *quality* of the directions d_k in terms of the following definition.

Definition 8.14 (Superlinear directions for the *SuperMann scheme*)**.** *Relative to the sequence* $(x_k)_{k\in\mathbb{N}}$ *generated by the SuperMann scheme, we say that* $(d_k)_{k\in\mathbb{N}} \subset \mathcal{H}$ *are* SUPERLINEAR DIRECTIONS *if the following limit holds*

$$\lim_{k\to\infty} \frac{\|R(x_k + d_k)\|}{\|Rx_k\|} = 0.$$

Remark 8.15. Definition 8.14 makes no mention of a limit point $x_\star$ of the sequence $(x_k)_{k\in\mathbb{N}}$, differently from the previously given Definition 4.4, taken from [44, §7.5] that instead requires $\frac{\|x_k+d_k-x_\star\|}{\|x_k-x_\star\|}$ to be vanishing with no mention of R. Due to 2α-Lipschitz continuity of R, whenever the directions d_k are bounded as in (8.26) we have

$$\frac{\|R(x_k+d_k)\|}{\|Rx_k\|} \leq 2\alpha D \frac{\|x_k+d_k-x_\star\|}{\|d_k\|}.$$

Invoking [44, Lem. 7.5.7] it follows that Definition 8.14 is implied by the one in [44] and is therefore more general. □

Theorem 8.16. *Consider the iterates generated by the SuperMann scheme (Alg. 8.2) with either $c_0 > 0$ or $c_1 > 0$, and with $(d_k)_{k\in\mathbb{N}}$ being superlinear directions as in Definition 8.14. Then,*

(i) eventually, stepsize $\tau_k = 1$ is always accepted and safeguard *updates K_2 are deactivated (i.e., the scheme reduces to the local method $x_{k+1} = x_k + d_k$);*

(ii) $(Rx_k)_{k\in\mathbb{N}}$ converges Q-superlinearly;

(iii) if the directions d_k satisfy Assumption 8.II, then $(x_k)_{k\in\mathbb{N}}$ converges R-superlinearly;

(iv) if $c_0 > 0$, then the complement of K_0 is finite.

Proof.

♠ 8.16*(i)* and 8.16*(iv)*. Let $w_k^0 \coloneqq x_k + d_k$. Superlinear convergence of $(d_k)_{k\in\mathbb{N}}$ then reads $\frac{\|Rw_k^0\|}{\|Rx_k\|} \to 0$. In particular, if $c_1 > 0$ then there exists $\bar{k} \in \mathbb{N}$ such that $\|Rw_k^0\| \leq c_1\|Rx_k\|$ for all $k \geq \bar{k}$, *i.e.*, the point $w_k^0 = x_k + d_k$ will always pass condition at step **5(a)** resulting in $x_{k+1} = w_k^0 = x_k + d_k$ for all $k \geq \bar{k}$.

Similarly, if $c_0 > 0$ then K_0 is infinite as shown in Thm. 8.13*(v)*; moreover, for $\ell \in \mathbb{N}$

$$\frac{\|Rx_{k_\ell+1}\|}{\eta_{k_\ell+1}} = \frac{\|Rx_{k_\ell+1}\|}{\|Rx_{k_\ell}\|} = \frac{\|R(x_{k_\ell}+d_{k_\ell})\|}{\|Rx_{k_\ell}\|} \to 0 \quad \text{as} \quad \ell \to \infty$$

and therefore the ratio eventually is always smaller than c_0, resulting in $k_\ell + 1 \in K_0$ for ℓ large enough. Consequently, the sequence will eventually reduce to $x_{k+1} = x_k + d_k$.

♠ 8.16*(ii)* and 8.16*(iii)*. Q-superlinear convergence of $(Rx_k)_{k\in\mathbb{N}}$ follows from the fact that $x_{k+1} = x_k + d_k$ for $k \geq \bar{k}$. In particular, $(\|Rx_k\|)_{k\in\mathbb{N}}$ is summable and there exists a sequence $(\delta_k)_{k\in\mathbb{N}} \to 0$ such that $\|Rx_{k+1}\| \leq \delta_k \|Rx_k\|$ for all k. If $\|d_k\| \leq D\|Rx_k\|$ for some $D > 0$, then

$$\sum_{k\geq\bar{k}} \|x_{k+1} - x_k\| \leq D \sum_{k\geq\bar{k}} \|Rx_k\| < \infty,$$

which implies that $(x_k)_{k\in\mathbb{N}}$ is a Cauchy sequence, and hence converges to a point, be it $x_\star$. Moreover, by possibly enlarging D so as to account for the iterates $k < \bar{k}$, we have

$$\begin{aligned}\|x_k - x_\star\| &\leq \sum_{j\geq k} \|x_{j+1} - x_j\| \leq D \sum_{j\geq k} \|Rx_j\| \\ &\leq D\delta_0\delta_1 \cdots \delta_{k-1} \sum_{j\in\mathbb{N}} \|Rx_j\| =: \Delta_k.\end{aligned}$$

This shows that $(x_k)_{k\in\mathbb{N}}$ is R-superlinearly convergent, since ${}^{\Delta_{k+1}}/_{\Delta_k} = \delta_k \to 0$. □

Theorem 8.16 shows that when the directions d_k are *good*, then eventually the *SuperMann scheme* reduces to the local method $x_{k+1} = x_k + d_k$ and consequently inherits its local convergence properties. The following result specializes to the choice of semismooth Newton directions.

Corollary 8.17 (Superlinear convergence for semismooth Newton directions)**.** *Suppose that $\mathcal{H}$ is finite dimensional, and that R is semismooth. Consider the iterates generated by the SuperMann scheme (Alg. 8.2) with either $c_0 > 0$ or $c_1 > 0$ and directions d_k chosen as solutions of*

$$(G_k + \mu_k \mathrm{id})d_k = -Rx_k \quad \text{for some } G_k \in \partial_C Rx_k, \tag{8.27}$$

where $\partial_C R$ denotes the Clarke generalized Jacobian of R and $0 \leq \mu_k \to 0$. Suppose that the sequence $(x_k)_{k\in\mathbb{N}}$ converges to a point $x_\star$ at which all the elements in $\partial_C R$ are nonsingular.

Then, $(d_k)_{k\in\mathbb{N}}$ are superlinear directions as in Definition 8.14, and in particular all the claims of Theorem 8.16 hold.

Proof. Any $G_k \in \partial_C R$ is positive semidefinite due to the monotonicity of R, and therefore d_k as in (8.27) is well defined for any $\mu_k > 0$. The bound (8.26)

holds due to [44, Thm. 7.5.2]. Moreover,

$$\frac{\|Rx_k + G_k d_k\|}{\|d_k\|} = \mu_k \to 0 \quad \text{as } k \to \infty,$$

and the proof follows invoking [44, Thm. 7.5.8(a)] and Rem. 8.15. □

Notice that since $\partial_C R = \mathrm{id} - \partial T$, nonsingularity of the elements in $\partial_C R(x_\star)$ is equivalent to having $\|G\| < 1$ for all $G \in \partial T(x_\star)$, *i.e.*, that T is a local contraction around $x_\star$.

However, in the same spirit of the previous chapters we are oriented towards choices of directions that (1) are defined for any nonexpansive mapping, regardless of the (generalized) first-order properties, and that (2) require exactly the same black-box oracle as the original KM scheme. Once again we shall thus investigate the employment of quasi-Newton directions.

Theorem 8.18 (Dennis-Moré criterion for superlinear convergence)**.** *Consider the iterates generated by the SuperMann scheme (Alg. 8.2) and suppose that $(x_k)_{k\in\mathbb{N}}$ converges strongly to a point $x_\star$ at which R is strictly differentiable. Suppose further that the update directions $(d_k)_{k\in\mathbb{N}}$ satisfy Assumption 8.II and the Dennis-Moré condition*

$$\lim_{k\to\infty} \frac{\|Rx_k + JR(x_\star)d_k\|}{\|d_k\|} = 0. \tag{8.28}$$

Then, the directions d_k are superlinear as in Definition 8.14. In particular, all the claims of Theorem 8.16 hold.

Proof.

$$\begin{aligned} 0 &\overset{(8.28)}{=} \lim_{k\to\infty} \frac{\big\|Rx_k + JR(x_\star)d_k + \big(R(x_k + d_k) - R(x_k + d_k)\big)\big\|}{\|d_k\|} \\ &= \lim_{k\to\infty} \frac{\big\|R(x_k + d_k)\big\|}{\|d_k\|} \\ &\overset{(8.26)}{\geq} \frac{1}{D} \lim_{k\to\infty} \frac{\big\|R(x_k + d_k)\big\|}{\|Rx_k\|} \end{aligned}$$

where in the second equality we used strict differentiability of R at $x_\star$. □

8.6.3 The modified Broyden scheme

In practical application the Hilbert space $\mathcal{H}$ is finite dimensional, and consequently it can be identified with $\mathbb{R}^n$. Consistently with the discussion in Section 4.3, the computation of quasi-Newton directions d_k in the *SuperMann scheme* amounts to selecting

$$d_k = -H_k R x_k, \tag{8.29}$$

where H_k are linear operators recursively defined with low-rank updates. To avoid notational clashes, we indicate such pairs of vectors as (s_k, y_k) instead of (p_k, q_k) as in (4.4). In particular,

$$\begin{cases} s_k = w_k - x_k \\ y_k = Rw_k - Rx_k. \end{cases} \tag{8.30}$$

Contrary to what experienced with the envelope-based approach, the Broyden scheme seems to be more beneficial than BFGS.

Theorem 8.19 (Superlinear convergence of the *SuperMann scheme* with Broyden directions)**.** *Suppose that $\mathcal{H}$ is finite dimensional. Consider the sequence $(x_k)_{k\in\mathbb{N}}$ generated by the SuperMann scheme (Alg. 8.2), $(d_k)_{k\in\mathbb{N}}$ being selected with the modified Broyden scheme of (4.3.2) for some $\bar{\vartheta} \in (0,1)$ and with pairs as in (8.30).*

Suppose that $(H_k)_{k\in\mathbb{N}}$ remains bounded, and that R is calmly semidifferentiable and metrically subregular at the limit $x_\star$ of $(x_k)_{k\in\mathbb{N}}$. Then, $(d_k)_{k\in\mathbb{N}}$ satisfies the Dennis-Moré condition (8.28). In particular, all the claims of Theorem 8.18 hold.

Proof. The proof is similar to that of Thm. 4.7. Let $G_\star = JRx_\star \in \mathbb{R}^{n\times n}$ and let $\|\cdot\|$ denote the Euclidean norm. From [59, Lem. 2.2] we have that there exist a constant L and a neighborhood $U_{x_\star}$ of $x_\star$ such that

$$\begin{aligned} \frac{\|y_k - G_\star s_k\|}{\|s_k\|} &= \frac{\|Rw_k - Rx_k - G_\star(w_k - x_k)\|}{\|w_k - x_k\|} \\ &\leq L \max\{\|x_k - x_\star\|, \|w_k - x_\star\|\}. \end{aligned}$$

Because of (8.26), the fact that $\tau_k \leq 1$, and the triangular inequality we have $\|w_k - x_\star\| \leq \|x_k - x_\star\| + D\|Rx_k\|$ and consequently

$$\sum_{k\in\mathbb{N}} \frac{\|y_k - G_\star s_k\|}{\|s_k\|} \leq L \sum_{k\in\mathbb{N}} \big(\|x_k - x_\star\| + D\|Rx_k\|\big) < \infty$$

where the last inequality follows from Thm. 8.13*(vi)*.

Let $E_k = B_k - G_\star$ and let $\|\cdot\|_F$ denote the Frobenius norm. With a simple modification of the proofs of [59, Thm. 4.1] and [4, Lem. 4.4] that takes into account the scalar $\vartheta_k \in [\bar\vartheta, 2-\bar\vartheta]$ we obtain

$$\begin{aligned}\|E_{k+1}\|_F &\le \left\|E_k\left(\mathrm{id} - \vartheta_k \tfrac{s_k s_k^\top}{\|s_k\|^2}\right)\right\|_F + \vartheta_k \frac{\|y_k - G_\star s_k\|}{\|s_k\|} \\ &\le \|E_k\|_F - \frac{\bar\vartheta(2-\bar\vartheta)}{2\|E_k\|_F}\frac{\|E_k s_k\|^2}{\|s_k\|^2}\end{aligned}$$

The last term on the right-hand side, be it σ_k, is summable and therefore the sequence $(E_k)_{k\in\mathbb{N}}$ is bounded. Let $\bar E \coloneqq \sup(\|E_k\|_F)_{k\in\mathbb{N}}$, then

$$\|E_{k+1}\|_F - \|E_k\|_F \le \sigma_k - \frac{\bar\vartheta(2-\bar\vartheta)}{2\bar E}\left(\frac{\|(B_k - G_\star)s_k\|}{\|s_k\|}\right)^2.$$

Telescoping the above inequality, summability of σ_k ensures that of the sequence $\frac{\|(B_k-G_\star)s_k\|^2}{\|s_k\|^2}$, proving in particular the claimed Dennis-Moré condition (8.28). □

Remark 8.20. It follows from Theorem 8.13*(iv)* that the *SuperMann scheme* is globally convergent as long as $\|d_k\| \le D\|Rx_k\|$ for some constant D. To enforce it we may select a (large) constant $D > 0$ and as a possible choice truncate $d_k \leftarrow D\frac{\|Rx_k\|}{\|d_k\|}d_k$ whenever d_k does not satisfy (8.26). □

Let us observe that in order to achieve superlinear convergence *the SuperMann scheme does not require nonsingularity of the Jacobian at the solution*. This is the standard requirement for asymptotic properties of quasi-Newton schemes, which is needed to show first that the method converges at least linearly. [4] generalizes this property invoking the concepts of (strong) metric (sub)regularity (see also [39] for an extensive review on these properties). However, if R is strictly differentiable at $x_\star$, then strong subregularity, regularity and strong regularity are equivalent to injectivity, surjectivity and invertibility of $JR(x_\star)$, respectively, these conditions being all equivalent for mappings $\mathcal{H} \to \mathcal{H}$ with $\mathcal{H}$ finite dimensional. In particular, contrary to the *SuperMann scheme* standard approaches require the solution $x_\star$ at least to be isolated, a property that rules out many interesting applications (cf. §8.7.1).

Restarted (modified) Broyden scheme

The Broyden scheme requires storing and operating with $n \times n$ matrices, where n is the dimension of the optimization variable, and is consequently feasible in practice only for small problems. Alternatively, one can restrict the Broyden update rule to only the most recent pairs of vectors (s_i, y_i). As detailed in Algorithm 8.3, this can be done by keeping track of the last vectors s_i and some auxiliary vectors $\tilde{s}_i = \frac{s_i - H_i \tilde{y}_i}{\langle s_i, H_i \tilde{y}_i \rangle_2}$. These are stored in some buffers S and $\tilde{S}$, which are initially empty and can contain up to m vectors. The *memory* m is a small integer typically between 3 and 20; when the memory is full, the buffers are emptied and Broyden scheme is *restarted*. The choice of a *restarted* rather than a *limited-memory* variant obviates the need of a nested for-loop to account for Powell's modification.

8.6.4 Parameters selection in *SuperMann*

As shown in Theorem 8.16, the *SuperMann scheme* makes sense as long as either $c_0 > 0$ or $c_1 > 0$; indeed, *safeguard* K_2-steps are only needed for globalization, while it is *blind* K_0- and *educated* K_1-steps that exploit the quality of the directions d_k. Evidently, K_1-updates are more reliable than K_0-updates in that they take into account the residual of the candidate next point. As such, it is advisable to select c_1 close to 1 and use small values of c_0 if more conservatism and robustness are desired. To further favor K_1-updates, the parameter q used for updating the safeguard r_{safe} at step **5(a)** may be also chosen very close to 1.

As to *safeguard* K_2-steps, a small value of σ makes condition (8.24) easier to satisfy and results in fewer backtrackings; the averaging factor λ may be chosen equal to 1 whenever possible, *i.e.*, if $\alpha \lneqq 1$ (which is the typical case when, *e.g.*, T comes from splitting schemes in convex optimization), or any close value

Algorithm 8.3 Restarted Broyden scheme with memory m

Input: old buffers $S, \tilde{S}$; new pair (s, y); current Rx
Output: new buffers $S, \tilde{S}$; update direction d

1: $d \leftarrow -Rx$, $\tilde{s} \leftarrow y$
2: **for** $i = 1 \dots \#S$ **do**
$\tilde{s} \leftarrow \tilde{s} + \langle s_i, \tilde{s} \rangle_2 \tilde{s}_i$, $d \leftarrow d + \langle s_i, d \rangle_2 \tilde{s}_i$
3: compute ϑ as in (4.5b) with $\gamma = \frac{1}{\|s\|_2^2} \langle \tilde{s}, s \rangle_2$
4: $\tilde{s} \leftarrow \frac{\vartheta}{(1-\vartheta+\vartheta\gamma)\|s\|_2^2}(s - \tilde{s})$, $d \leftarrow d + \langle s, d \rangle_2 \tilde{s}$
5: **if** $\#S = m$ **then** $S, \tilde{S} \leftarrow [\,]$ **else** $S \leftarrow [S, s]$, $\tilde{S} \leftarrow [\tilde{S}, \tilde{s}]$

otherwise. In the simulations of Section 8.7 we used $c_0 = c_1 = q = 0.99$, $\sigma = 0.1$, $\lambda = 1$ and $\beta = 1/2$. For a matter of scaling, we multiplied the summable term q^k by $\|Rx^0\|$ in updating the parameter r_{safe} at step **5(a)**. The directions were computed according to the restarted modified Broyden scheme (Alg. 8.3) with memory $m = 20$ and $\bar{\vartheta} = 0.2$; we applied the truncation rule as in Remark 8.20 with $D = 10^4$. We also imposed a maximum of 8 backtrackings after which a nominal KM iteration would be executed.

8.6.5 Comparisons with other methods

Hybrid global and local phase algorithms

Blind K_0-updates in the *SuperMann scheme* are inspired from [30, Alg. 1], and so is the notation $K_0 = \{k_0, k_1, \ldots\}$.

Educated K_1- and *safeguard* K_2-updates instead play the role of *inner*- and *outer-phases* in the general algorithmic framework described in [61, §5.3] for finding a zero of a candidate merit function φ (e.g. $\varphi(x) = \frac{1}{2}\|Rx\|^2$ in our case). Differently from [61, Alg. 5.16] where all previous inner-phase iterations are discarded as soon as the required sufficient decrease is not met, the *SuperMann scheme* allows for an alternation of phases that eventually stabilizes on the fast local one, provided the solution is sufficiently regular. Our scheme is more in the flavor of [61, Alg. 5.19], although with less conservative requirements for triggering *inner* K_1-updates ($\varphi(x_{k+1})$ is here compared with $\varphi(x_k)$, whereas in the cited scheme with the smallest past value).

Inexact Newton methods for monotone equations

The GKM updates are closely related to the extra-gradient steps described in [109, Alg. 2.1]. This work introduces an inexact Newton algorithm for solving systems of continuous monotone equations $Rx = 0$, where $\mathrm{id} - R$ needs not be nonexpansive. At a given point x, first a direction d is computed as (possibly approximate) solution of $Gd = -Rx$, where G is some positive definite matrix; then, an intermediate point $w = x + \tau d$ is retrieved with a line search on τ that ensures the condition

$$\|Rw\|^2 - \langle Rw, x - Tw\rangle \leq -\sigma\tau\|d\|^2 \tag{8.31}$$

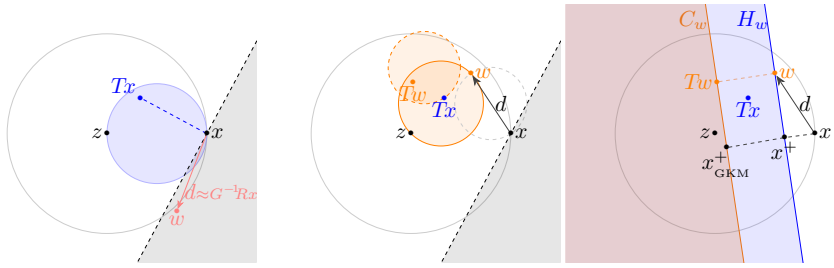


Figure 8.6: *The positive definiteness of G prevents the update directions d in the scheme of [109] to point in the gray-shaded area. As a result, differently from the GKM scheme the cited algorithm is not robust to any choice of direction (e.g., it cannot accept the one as in Figure 8.5). In any case, the half-space C_w onto which x is projected according to the GKM scheme is properly contained in the half-space H_w corresponding to the update of [109]; consequently, the GKM update is always closer to any solution.*

for some $\sigma > 0$; here, we defined $T \coloneqq \mathrm{id} - R$ to highlight the symmetry with (8.19). Finally, the new iterate is given by $x^+ = \Pi_{H_w} x$, where

$$H_w \coloneqq \big\{ z \in \mathcal{H} \mid \|Rw\|^2 - \langle Rw, z - Tw \rangle \geq 0 \big\}. \tag{8.32}$$

Letting C_w be the half-space as in Prop. 8.10, so that $x^+_{\mathrm{GKM}} = \Pi_{C_w} x$ (for simplicity we set $\lambda = 1$), for the half-spaces (8.32) it holds that

$$\operatorname{zer} R \subseteq C_w \subseteq H_w,$$

the last inclusion holding as equality iff $Rw = 0$. This means that in the GKM scheme, the same w yields an iterate x^+_{GKM} which is closer to any $z \in \operatorname{zer} R$ with respect to x^+ (cf. Fig. 8.6). Notice further that the hyperplanes delimiting the two half-spaces are parallel, with $\operatorname{bdry} C_w$ passing by Tw (or $T_{1/2\alpha} w$ for generic α's) and $\operatorname{bdry} H_w$ by w.

The requirement of positive definiteness of matrix G in defining the update direction d is due to the fact that [109] addresses a broader class of monotone operators; we instead exploited at full the nonexpansiveness of $\mathrm{id} - R$ and as a result have complete freedom in selecting d (Fig. 8.6a) and better projections (Fig. 8.6c).

Line-search for KM

The recent work [48] proposes an acceleration of the classical KM scheme for finding a fixed point of an α-averaged operator T based on a line search on the relaxation parameter. Namely, instead of the *nominal* update $\bar{x} = T_\lambda x$ with $\lambda \in [0, 1/\alpha]$ as in (8.18), values $\lambda' > 1/\alpha$ are first tested and the update $x^+ = T_{\lambda'} x$ is accepted as long as long as $\|Rx^+\| \leq c_1 \|R\bar{x}\|$ holds for some constant $c_1 \in (0, 1)$.

In the setting of the *SuperMann scheme*, this corresponds to selecting $d_k = -Rx_k$, discarding *blind* updates (*i.e.*, setting $c_0 = 0$), foretracking *educated* updates and using plain KM iterations as safeguard steps. Convergence can be enhanced and the method is attractive when $T = S_2 \circ S_1$ is the composition of an affine mapping S_1 and a cheap operator S_2, in which case the line search is inexpensive. However, though preserving the same theoretical convergence guarantees of KM (hence of the *SuperMann scheme*), it does not improve its best-case local linear rate.

Although other choices d_k may also be considered, fast directions such as Newton-type ones would be discarded and replaced by nominal KM updates every time the candidate point $x_k + d_k$ does not meet some requirements. Avoiding this take-it-or-leave-it behavior is exactly the primary goal of GKM iterations, so that candidate good directions are never discarded.

8.7 Simulations

We conclude with some numerical examples to give tangible evidence of the robustifying and enhancing effect that the *SuperMann scheme* has on fixed-point iterations. In all simulations we deactivated blind updates by setting $c_0 = 0$, and we selected $\sigma = 10^{-3}$ for safeguard updates and $c_1 = q = 1 - \sigma$ for educated updates. Due to problem size we used restarted Broyden directions with a memory buffer of 20 vectors.

8.7.1 Cone programs

We consider cone problems of the form

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ \langle c, x \rangle \quad \text{subject to } \ Ax + s = b, \ s \in \mathcal{K}, \tag{8.33}$$

where $\mathcal{K}$ is a nonempty closed convex cone. Almost any convex program can be recast as (8.33), and many convex optimization solvers address problems by first translating them into this form. The KKT conditions for optimality of the primal-dual couple $\big((x_\star, s_\star), (y_\star, r_\star)\big)$ are

$$Ax_\star + s_\star = b, \quad s_\star \in K, \quad A^\top y_\star + c = r_\star, \quad r_\star = 0, \quad y_\star \in K^\star, \quad c^\top x_\star + b^\top y_\star = 0$$

where $\mathcal{K}^*$ is the dual cone of $\mathcal{K}$. A recently developed conic solver for (8.33) is SCS [89], which solves the corresponding so-called *homogeneous self-dual embedding*

$$\text{find} \quad u \in \mathcal{C} \quad \text{subject to } Qu \in \mathcal{C}^*, \tag{8.34}$$

where

$$\mathcal{C} = \mathbb{R}^n \times \mathcal{K}^* \times \mathbb{R}_+ \quad \text{and} \quad Q = \begin{bmatrix} 0 & A^\top & c \\ -A & 0 & b \\ -c^\top & -b^\top & 0 \end{bmatrix}.$$

Problem (8.34) can be equivalently reformulated as the variational inequality

$$\text{find } u \in \mathcal{C} \quad \text{s.t.} \quad 0 \in Qu + N_{\mathcal{C}}(u). \tag{8.35}$$

Indeed, for all $u \in \mathcal{C}$ we have

$$\begin{aligned} N_{\mathcal{C}}(u) &= \{y \mid \langle v - u, y\rangle \le 0 \ \ \forall v \in \mathcal{C}\} = \{u\}^\perp \cap \{y \mid \langle v, y\rangle \le 0 \ \ \forall v \in \mathcal{C}\} \\ &= \{u\}^\perp \cap (-\mathcal{C}^*), \end{aligned}$$

where the second equality follows by considering, *e.g.*, $v = \frac{1}{2}u$ and $v = \frac{3}{2}u$, which both belong to $\mathcal{C}$ being it a cone. From this equivalence and the fact that $Qu \in \{u\}^\perp$ for any u due to the skew symmetry of Q, the equivalence of (8.34) and (8.35) is apparent. This leads to the short and elegant interpretation of SCS as Douglas-Rachford splitting (DRS) applied to the splitting $N_{\mathcal{C}} + Q$ in (8.35), which, after a well known change of variables and index shifting, reads

$$\begin{cases} \tilde{u}_{k+1} \approx (I + Q)^{-1}(u_k + v_k) \\ u_{k+1} = \Pi_{\mathcal{C}}(\tilde{u}_{k+1} - v_k) \\ v_{k+1} = v_k - \tilde{u}_{k+1} + u_{k+1}. \end{cases} \tag{8.36}$$

The "$\approx$" symbol refers to the fact that v_k may be retrieved inexactly by means of conjugate gradient (CG) method; see [89] for a detailed discussion.

Here we consider instead DRS applied to the (equivalent) splitting $Q + N_{\mathcal{C}}$ in (8.35), namely

$$\begin{cases} v_{k+1} \approx (I+Q)^{-1}(u_k) \\ w_{k+1} = \Pi_{\mathcal{C}}(2v_{k+1} - u_k) \\ u_{k+1} = u_k + w_{k+1} - v_{k+1}. \end{cases} \tag{8.37}$$

For any initial point u_0, the variable v_k converges to a solution to (8.35) [10, Thm. 26.11]. DRS is a (firmly) nonexpansive operator and as such it can be integrated in the *SuperMann scheme* with $\lambda \in (0, 2)$; in these simulations we set $\lambda = 1$.

We run a cone problem (8.33) of size $m = 487$ and $n = 325$, with density 0.01 and condition number 100, both by solving exactly the linear systems and by adopting the CG technique. $\mathcal{C}$ is the cartesian product of all the primitive cones implemented in SCS solver: positive orthant, second-order, positive semidefinite, (dual) exponential, and (dual) power cones. We reported primal residual, dual residual, and duality gap; consistently with SCS' termination criterion, the algorithm is stopped when all these quantities are below some tolerance [89, §3.5], which we set to 10^{-6}.

Notice that if u solves (8.34), or equivalently (8.35), then so does any multiple tu with $t > 0$. In particular no isolated solution exists, and therefore whenever the residual R of the DRS operator is differentiable at a solution $u_\star$, $JR(u_\star)$ is singular. Fortunately, the *SuperMann scheme* does not necessitate nonsingularity of the Jacobian but merely metric subregularity, the same property that enables linear convergence rate of the original DRS (or equivalently SCS). In particular, whenever the original SCS scheme is linearly convergent, the SuperMann enhancement is provably superlinear provided that R is strictly differentiable at the limit point. However, since *restarted* Broyden directions are implemented instead of the full-memory method, rather than superlinear convergence we expect an "extremely steep" linear convergence.

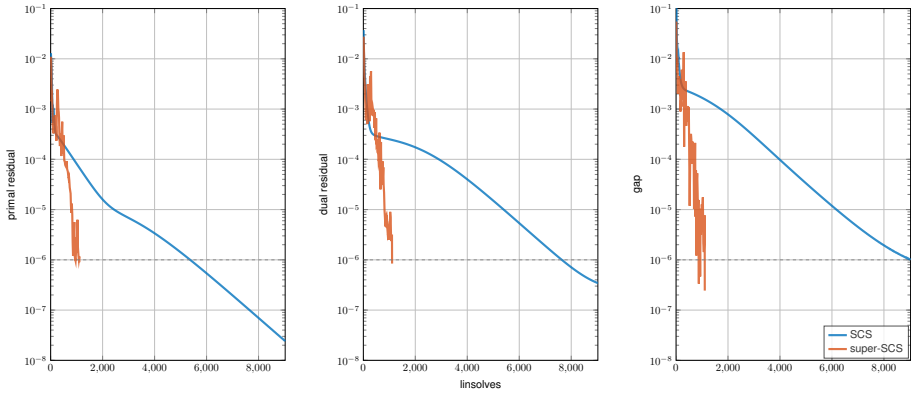
In Figure 8.7 we can observe how the original SCS scheme (blue) converges at a fair linear rate; however, its super-enhancement greatly outperforms it both when solving linear systems exactly and approximately.

8.7.2 Lasso

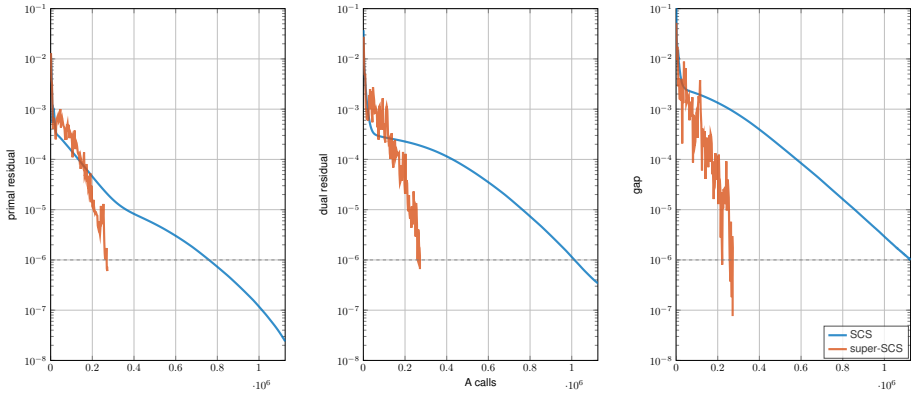
We consider a lasso problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ \tfrac{1}{2}\|Ax - b\|^2 + \nu\|x\|_1$$

Figure 8.7: *Comparison between Splitting Cone Solver [89] (blue) and its enhancement with the SuperMann scheme for solving a cone program (8.33).*



(a) *On the x-axis the number of times a linear system is solved, the most expensive operation, needed for computing the resolvent of Q. SCS performs quite well, however its super-enhancement converges considerably faster in terms of operations.*



(b) *Comparison with respect to the same problem, but with linear systems solved approximately with CG on a reduced system. On the x-axis the number of times the operators A and $A^\top$ are called, which amount to the most expensive operations. Apparently, solving the system inexactly does not affect the comparison between SCS and super-SCS.*

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and $\nu > 0$. In Figure 8.7 the comparison of forward-backward splitting (or proximal gradient, in blue) and its super-enhanced version (red) on a random problem with $m = 1500$ $n = 5000$ and $\nu = 10^{-2}$. On the x-axis the number of matvecs, being them the most expensive operations of FB

and hence of super-FB, and on the y-axis the fixed-point residual. Superlinear convergence cannot be observed due to the fact that a limited-memory method is used for computing directions, however an outstanding speedup is noticeable.

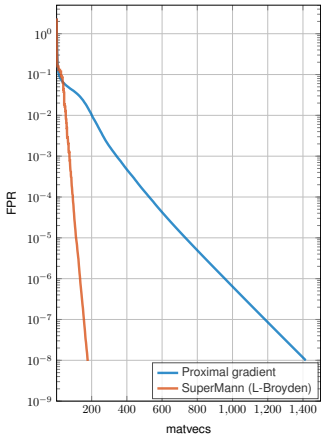


Figure 8.7: *Comparison between FBS and Super-FBS (using modified Broyden limited-memory directions) in a lasso problem.*

8.7.3 Constrained linear optimal control

For matrices A_t and B_t of suitable size, $t = 0, \dots, N-1$, consider a state-input dynamical system

$$x_{t+1} = A_t x_t + B_t u_t, \quad t = 0, \dots, N-1, \tag{8.38a}$$

where $x_0 \in \mathbb{R}^{n_x}$ is given, and the next states $x_t \in \mathbb{R}^{n_x}$ are determined by the user-defined inputs $u_\tau \in \mathbb{R}^{n_u}$, $\tau = 0, \dots, t-1$. States $\boldsymbol{x} = (x_1, \dots, x_N)$ can be expressed in terms of the inputs $\boldsymbol{u} = (u_0, \dots, u_{N-1})$ through a linear operator $L \in \mathbb{R}^{Nn_x \times Nn_u}$ as $\boldsymbol{x} = L\boldsymbol{u} + b$ for some constant $b \in \mathbb{R}^{Nn_x}$. The goal is to choose inputs that minimize a cost

$$\ell(\boldsymbol{u}, \boldsymbol{x}) = \sum_{t=0}^{N-1} \ell_t(u_t, x_t) + \ell_N(x_N) \tag{8.38b}$$

subject to some constraints

$$x_{t+1} \in \mathcal{X}_{t+1}, \quad u_t \in \mathcal{U}_t, \quad t = 0, \dots, N-1. \tag{8.38c}$$

Vũ-Condat splitting

The constraint sets in (8.38c) are typically simple and easy to project onto (boxes, Euclidean balls…). However, while simple input constraints can be easily handled, due to the coupling enforced by the dynamics (8.38a), expressing $\mathcal{X}_{t+1}$ in terms of the optimization variable $\boldsymbol{u}$ results in much more complicated sets (polyhedra, ellipsoids…). To avoid this complication we make use of the extremely versatile algorithm that Vũ-Condat three-term splitting offers [35, Alg. 3.1]. In its general form, the algorithm addresses problems of the form

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \;\; f(x)+g(x)+h(Lx) \tag{8.39}$$

where $f:\mathbb{R}^n\to\mathbb{R}$ is convex with L_f-Lipschitz continuous gradient, $g:\mathbb{R}^n\to\overline{\mathbb{R}}$ and $h:\mathbb{R}^m\to\overline{\mathbb{R}}$ are convex, and $L\in\mathbb{R}^{n\times m}$, by iterating the following steps:

$$\begin{cases} x^+ = \operatorname{prox}_{\tau g}\big(x-\tau(\nabla f(x)+L^\top y)\big) \\ y^+ = \operatorname{prox}_{\tau h^*}\big(y+\sigma L(2x^+-x)\big). \end{cases} \tag{8.40}$$

Here, $0<\tau<\frac{2}{L_f}$ and $0<\sigma<\frac{1}{\|L\|^2}\big(\frac{1}{\tau}-\frac{L_f}{2}\big)$ are stepsizes, and $y\in\mathbb{R}^m$ is a Lagrange multiplier. Vũ-Condat splitting is a primal-dual method that generalizes FBS by allowing an extra nonsmooth term h and a linear operator L (by neglecting h and L one recovers the proximal gradient iterations of FBS).

The optimal control problem (8.38) can be cast into Vũ-Condat splitting form (8.39) by simply letting $f(\boldsymbol{u})=\ell(\boldsymbol{u},L\boldsymbol{u})$, $g=\delta_{\mathcal{U}}$ and $h=\delta_{\mathcal{X}}(\,\cdot\,+b)$, where $\mathcal{U}=\mathcal{U}_0\times\cdots\times\mathcal{U}_{N-1}$ and $\mathcal{X}=\mathcal{X}_1\times\cdots\times\mathcal{X}_N$ (in particular, $n=Nn_u$ and $m=Nn_x$). Then, $\operatorname{prox}_{\tau g}=\Pi_{\mathcal{U}}$ and $\operatorname{prox}_{\sigma h^*}(y)=y-\sigma\,\Pi_{\mathcal{X}}(\sigma^{-1}y+b)+b$. Notice that $\Pi_{\mathcal{U}}$ and $\Pi_{\mathcal{X}}$ are fully decoupled as the projection of each input and state onto the corresponding constraint set. Moreover, the full matrix L needs not be computed, as both L and $L^\top$ can be treated as abstract operators that simulate forward and backward dynamics.

Apparently, the appeal of Vũ-Condat splitting in addressing the optimal control problem lies in the extreme simplicity of its operations and low memory requirements, making it particularly suited for medium-to-large-scale problems in which traditional interior point algorithms fail. However, like all first-order methods it is extremely sensitive to ill conditioning, which gets worse as the problem size increases. Fortunately, this splitting fits into the *SuperMann* framework. The operator T that maps (x,y) into (x^+,y^+) as in (8.40) is averaged in the Hilbert space $\mathcal{H}_P$, where $\mathcal{H}_P$ is defined as $\mathbb{R}^n\times\mathbb{R}^m$ equipped with the scalar product $\langle z,z'\rangle_P\coloneqq\langle z,Pz'\rangle$, where $P\coloneqq\begin{pmatrix}\tau^{-1}I & -L^\top\\ -L & \sigma^{-1}I\end{pmatrix}$ [35, proof of Thm. 3.1].

Oscillating masses experiment

We tried this approach on the benchmark problem of controlling a chain of oscillating masses connected by springs and with both ends attached to walls. The chain is composed of $2K$ bodies of unit mass subject to a viscous friction of 0.1, the springs have elastic constant 1 and no damping, and the system is controlled through K actuators, each being a force acting on a pair of masses, as depicted in Figure 8.8. Therefore $n_x = 4K$ (the states are the displacement from the rest position and velocity of each mass) and $n_u = K$. The inputs are constrained in $[-2, 2]$, while the position and velocity of each mass is constrained in $[-5, 5]$.

The continuous-time system was discretized with a sampling time $T_s = 0.1s$. We considered quadratic stage costs $\frac{1}{2}x^\top Qx$ for the states and $\frac{1}{2}u^\top u$ for the inputs, where Q is diagonal positive definite with random diagonal entries, and generated a random (feasible) initial state x_0. Notice that a QP reformulation would require the computation of the full cost matrix, differently from the splitting approach where only the small dynamics matrices A and B are needed, as L and $L^\top$ can be abstract operators.

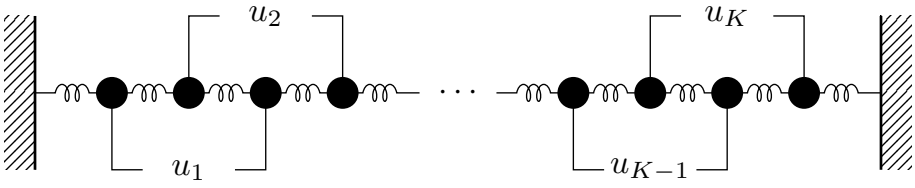


Figure 8.8: *Oscillating masses.*

We simulated different scenarios for all combinations of $K \in \{8, 16\}$ and $N \in \{10, 20, 30, 40, 50\}$. We compared Vu-Condat splitting (VC) with its 'super' enhancement (SuperVC); parameters were set as detailed in Section 8.6.4. Table 8.1 offers an overview of the experiment: SuperVC is roughly 13 times faster on average and 21 times better in worst-case performance than VC algorithm in reaching the termination criterion $\|Rx^k\| \leq 10^{-4}\|Rx^0\|$.

Number of calls to L and $L^\top$ ($\times 10^3$)

$K = 8$	$N = 10$		$N = 20$		$N = 30$		$N = 40$		$N = 50$	
	avg	max	avg	max	avg	max	avg	max	avg	max
VC	19.0	337.1	15.0	174.4	25.0	400+	21.0	136.5	16.0	61.9
SVC	1.0	5.5	1.0	4.3	2.0	19.3	2.0	10.9	2.0	6.6

$K = 16$	$N = 10$		$N = 20$		$N = 30$		$N = 40$		$N = 50$	
	avg	max	avg	max	avg	max	avg	max	avg	max
VC	62.0	400+	30.0	344.9	30.0	400+	65.0	400+	29.0	318.6
SVC	4.0	39.5	2.0	11.6	3.0	46.6	8.0	58.1	3.0	26.1

Table 8.1: *Comparison between Vũ-Condat algorithm (VC) and its "super" enhancement (SuperVC) in solving the oscillating masses problem with* $\|Rx^k\| \leq 10^{-4}\|Rx^0\|$ *as termination criterion. Average and worst performances among 25 simulations with randomly generated starting point* x_0 *for each combination of* $K \in \{8, 16\}$ *and* $N \in \{10, 20, 30, 40, 50\}$*. The tables compare the number of calls to the operators* L *and* $L^\top$*, which are the expensive operations (the rest are projections on boxes). In four problems Vũ-Condat exceeded* $4 \cdot 10^5$ *many calls (corresponding to* 10^5 *iterations) and was stopped prematurely.*

Conclusions

In this thesis we carried out an in-depth analysis of splitting algorithms in the nonconvex setting. The main contribution and novelty of the methodology is twofold:

- We pioneered a general framework where splitting algorithms are represented and identified by two components: an *inner* majorizing model, and an *outer* transformation mapping. The properties of what we defined *"proximal"* models set the ground for building a solid theory of convergence, reminiscent of that of Lyapunov type that ensures stability of dynamical systems. *Proximal envelopes*, here generalized for any splitting algorithm covered by the framework, proved to be suitable such Lyapunov functions.
- Building upon the proposed framework, a new linesearch strategy was proposed. Solely based on the continuity of the Lyapunov functions, property enjoyed by the investigated proximal envelopes, the ***C****ontinuous-****Ly****apunov* ***D****escent* paradigm *(**CLyD**)* allows to customize any proximal algorithm with arbitrary update directions. Once again proximal envelopes prove to be the perfect Lyapunov candidates, as (1) they allow to preserve the operational complexity of the underlying splitting algorithms, and (2) robustify CLyD against the Maratos effect: when *good* directions are selected, unitary stepsize is eventually always accepted and fast convergence thus triggered.

For both the forward-backward and the Douglas-Rachford splittings, it is shown how the solution of elementary algebraic inequalities is enough for obtaining bounds on stepsizes and relaxation parameters so as to ensure convergence. Nevertheless, with more sophisticated analysis of the Douglas-Rachford envelope, tight convergence results were derived. In light of a primal equivalence between the algorithms, as a byproduct tight convergence results for ADMM were easily inferred.

A CLyD-like framework restricted to convex splitting algorithms was also proposed. Although bound to convexity, the *SuperMann scheme* allows to accelerate pretty much any splitting algorithm, including those with a purely primal-dual nature that cannot be captured by proximal envelopes, as it is the case of the recent Vũ-Condat splitting.

Future directions

This thesis already accomplished some of the open research directions advanced in [112], such as a higher-order analysis of the Douglas-Rachford envelope, the derivation of Douglas-Rachford splitting-based Newton-type methods, and the consequent adaptation of the findings to ADMM. Nevertheless, other promising workplans such as the integration with augmented Lagrangian methods therein suggested have not been covered yet. In this perspective, we believe that the CLyD framework constitutes a valid tool for the achievement of such goals.

The analysis of other algorithms such as the proximal ADMM and the Chambolle-Pock splitting [29] in a fully nonconvex setting is already being investigated. Partially presented at an invited workshop of the 2018 European Control Conference, the study is providing further evidence in support of the potential of the proximal framework pioneered in the thesis. It is also worth pointing out that, for fully nonconvex problems and under a smoothness assumption as already investigated in [73], the recent Davis-Yin three-term splitting algorithm [38] falls in the GPMM framework; the convergence results of Corollary 3.19 and Theorem 3.22, as well as quasi-Newton enhancements through the CLyD Algorithm 4.1 are thus readily applicable. Nevertheless, as it was the case of the Douglas-Rachford splitting, a more in-depth analysis may yield tighter results.

Further extensions are also being considered:

- *Bregman-type models.* A challenging yet extremely powerful possible generalization of the framework consists in replacing the quadratic bounds defining the proximal models with a nonsymmetric Bregman distance. Embracing Bregman-type extensions of popular algorithms in a unified framework and consequently simplifying their arduous convergence analysis is an attractive prospect: not only would this considerably widen the range of covered methods, but it would also open the possibility to provide new purely primal interpretations of algorithms that are so far only understood through duality arguments, such as the Vũ-Condat splitting. Thus, similarly to what done in the thesis with the ADMM, nonconvex (and quasi-Newton) primal-dual algorithms would then be possible. In this perspective, due to its high degree of

generality encompassing the asymmetric forward-backward-adjoint (AFBA) splitting proposed in [66] is a particularly appealing goal.

- *Block-coordinate and matrix-free variants.* When facing the "big-data" reality, operating with huge variables and/or matrices may constitute a problem. Due to their amenability to operate on small portions of the problem at a time, incorporating randomized and block-coordinate variants in the investigated framework would create a major impact on modern huge-scale applications such as those arising in machine learning [22, 33, 65, 85]. An efficient management of matrices and even high-order tensors is already being investigated for the funded EOS SeLMA project,[2] and applications on embedded hardware with low memory and computational capabilities have already produced promising accomplishments [3, 107].

- *Higher-order smoothing.* The seemingly impractical use of third-order information in smooth minimization was recently shown to be feasible in some special cases [87]. A question then arises as to whether the proximal framework can be extended so that envelopes may take advantage of higher order smoothness, in such a way that more problems can be solved with the third-order techniques proposed. Motivated by the known smoothing effect that proximal minimization reflects on the Moreau envelope, an intuitive approach would be to analyze higher-order properties of the envelope functions in the attempt to broaden the class of problems for which such a method is "implementable". Alternatively, the investigated framework may be restricted to higher-order MM models, although this option may lead to difficult evaluations of the resulting MM mapping (the inner minimization problems).

Ultimately, it would be desirable to address some technical questions that, although supported by much evidence, so far have only been conjectured. One of these regards the often observed good performance of BFGS directions in the CLyD framework. We suppose that this behavior owes to fact that, despite being nonsymmetric, the involved Jacobians are *similar* to symmetric and positive definite matrices, hence in particular have all strictly positive eigenvalues. We also consider analyzing the iterations from a manifold perspective, whence results on partly smooth functions may prove to be useful [68, 36].

Another still unanswered issue relates to the assumptions needed for ensuring global and linear convergence of proximal algorithms and their CLyD enhancements. In particular, we have reasons to believe that the requirement of prox-regularity needed for the FBE to satisfy the error bound inequality as in

[2]Structured Low-Rank Matrix/Tensor Approximation, https://www.esat.kuleuven.be/stadius/selma/. Fonds de la Recherche Scientifique — FNRS and the Fonds Wetenschappelijk Onderzoek — Vlaanderen under EOS Project 30468160 (SeLMA).

Theorem 5.11*(ii)* could be dropped. As equation (5.10) indicates, answering the conjecture boils down to showing whether or not for arbitrary proper, lsc, and prox-bounded functions g the following *equality* holds

$$\limsup_{\substack{w \to x \\ w \neq x}} \operatorname{prox}_{\gamma g}(w) = \operatorname{prox}_{\gamma g}(x),$$

stronger than the mere inclusion '$\subseteq$' ensured by outer semicontinuity of $\operatorname{prox}_{\gamma g}$. The closest we got to a positive answer is by either assuming the nonsmooth term to be an indicator function, or the proximal mapping to be at most two-valued (around the critical points of interest).

In the same spirit, succeeding in reducing the assumptions to ensure superlinear convergence of the investigated algorithms would also be extremely appealing. A first step in this direction was obtained with the SuperMann scheme, which was shown to achieve superlinear convergence under no nonsingularity requirements, but merely metric subregularity (and suitable regularity assumptions). A similar result was recently achieved in [117], where the semismooth forward-backward truncated-Newton method first proposed in [94] was suitably adapted to the CLyD framework. Other than considerably simplifying the convergence analysis, this modification was the turning point that allowed to drop the nonsingularity assumption and to further reduce the other needed regularity requirements.

Bibliography

[1] Aharon, M., Elad, M., and Bruckstein, A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing 54*, 11 (2006), 4311–4322.

[2] Anderson, D. G. Iterative procedures for nonlinear integral equations. *Journal Of The ACM 12*, 4 (oct 1965), 547–560.

[3] Antonello, N., Stella, L., Patrinos, P., and van Waterschoot, T. Proximal gradient algorithms: Applications in signal processing. *arXiv e-prints* (2018).

[4] Aragón Artacho, F. J., Belyakov, A., Dontchev, A. L., and López-Cerdá, M. A. Local convergence of quasi-Newton methods under metric regularity. *Computational Optimization and Applications 58*, 1 (2014), 225–247.

[5] Attouch, H., Bolte, J., Redont, P., and Soubeyran, A. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. *Mathematics of Operations Research 35*, 2 (2010), 438–457.

[6] Attouch, H., Bolte, J., and Svaiter, B. F. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. *Mathematical Programming 137*, 1 (Feb 2013), 91–129.

[7] Attouch, H., and Peypouquet, J. The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $1/k^2$. *SIAM Journal on Optimization 26*, 3 (2016), 1824–1834.

[8] Auslender, A., and Teboulle, M. *Asymptotic Cones and Functions in Optimization and Variational Inequalities*. Springer Monographs in Mathematics. Springer New York, 2002.

[9] Bauschke, H., and Noll, D. On the local convergence of the Douglas-Rachford algorithm. *Archiv der Mathematik 102*, 6 (Jun 2014), 589–600.

[10] Bauschke, H. H., and Combettes, P. L. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics. Springer, 2017.

[11] Bauschke, H. H., and Koch, V. R. Projection methods: Swiss army knives for solving feasibility and best approximation problems with halfspaces. In *Infinite Products of Operators and Their Applications*, S. Reich and A. J. Zaslavski, Eds., vol. 636. American Mathematical Society, 2015, pp. 1–40.

[12] Bauschke, H. H., Noll, D., and Phan, H. M. Linear and strong convergence of algorithms involving averaged nonexpansive operators. *Journal of Mathematical Analysis and Applications 421*, 1 (2015), 1–20.

[13] Bauschke, H. H., Phan, H. M., and Wang, X. The method of alternating relaxed projections for two nonconvex sets. *Vietnam Journal of Mathematics 42*, 4 (Dec 2014), 421–450.

[14] Beck, A., and Hallak, N. On the minimization over sparse symmetric sets: Projections, optimality conditions, and algorithms. *Math. Oper. Res. 41*, 1 (2016), 196–223.

[15] Beck, A., and Pan, D. Convergence of an inexact majorization-minimization method for solving a class of composite optimization problems. In *Large-Scale and Distributed Optimization*, P. Giselsson and A. Rantzer, Eds. Springer International Publishing, Cham, 2018, pp. 375–410.

[16] Beck, A., and Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences 2*, 1 (2009), 183–202.

[17] Ben-Ameur, W., Bianchi, P., and Jakubowicz, J. Robust distributed consensus using total variation. *IEEE Transactions on Automatic Control 61*, 6 (June 2016), 1550–1564.

[18] Bernstein, D. S. *Matrix mathematics: theory, facts, and formulas with application to linear systems theory*. Princeton University Press, Woodstock, 2009.

[19] Bertsekas, D. P. Constrained optimization and Lagrange multiplier methods. *Computer Science and Applied Mathematics, Boston: Academic Press, 1982* (1982).

[20] Bertsekas, D. P. *Convex Optimization Theory.* Athena Scientific, 2015.

[21] Bertsekas, D. P. *Nonlinear Programming.* Athena Scientific, 2016.

[22] Bianchi, P., Hachem, W., and Iutzeler, F. A coordinate descent primal-dual algorithm and application to distributed asynchronous optimization. *IEEE Transactions on Automatic Control 61*, 10 (2016), 2947–2957.

[23] Bochnak, J., Coste, M., and Roy, M.-F. *Real Algebraic Geometry.* A Series of Modern Surveys in Mathematics. Springer Berlin Heidelberg, 2013.

[24] Bolte, J., Daniilidis, A., and Lewis, A. The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM Journal on Optimization 17*, 4 (2007), 1205–1223.

[25] Bolte, J., Sabach, S., and Teboulle, M. Proximal Alternating Linearized Minimization for nonconvex and nonsmooth problems. *Mathematical Programming 146*, 1–2 (2014), 459–494.

[26] Boţ, R. I., Csetnek, E. R., and László, S. C. An inertial forward-backward algorithm for the minimization of the sum of two nonconvex functions. *EURO Journal on Computational Optimization 4*, 1 (2016), 3–25.

[27] Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn. 3*, 1 (Jan. 2011), 1–122.

[28] Broyden, C. G. A class of methods for solving nonlinear simultaneous equations. *Mathematics of Computation 19*, 92 (1965), 577–593.

[29] Chambolle, A., and Pock, T. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision 40*, 1 (May 2011), 120–145.

[30] Chen, X., and Fukushima, M. Proximal quasi-Newton methods for nondifferentiable convex optimization. *Mathematical Programming 85*, 2 (Jun 1999), 313–334.

[31] Combettes, P. L. Quasi-Fejérian analysis of some optimization algorithms. *Studies in Computational Mathematics 8* (2001), 115–152.

[32] Combettes, P. L., and Pesquet, J.-C. Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. Springer New York, New York, NY, 2011, pp. 185–212.

[33] Combettes, P. L., and Pesquet, J.-C. Stochastic quasi-fejér block-coordinate fixed point iterations with random sweeping. *SIAM Journal on Optimization 25*, 2 (2015), 1221–1248.

[34] Combettes, P. L., and Vũ, B. C. Variable metric forward-backward splitting with applications to monotone inclusions in duality. *Optimization 63*, 9 (2014), 1289–1318.

[35] Condat, L. A primal-dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. *Journal of Optimization Theory and Applications 158*, 2 (2013), 460–479.

[36] Daniilidis, A., Hare, W., and Malick, J. Geometrical interpretation of the predictor-corrector type algorithms in structured optimization problems. *Optimization 55*, 5-6 (2006), 481–503.

[37] Daubechies, I., DeVore, R., Fornasier, M., and Güntürk, C. S. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics 63*, 1 (2010), 1–38.

[38] Davis, D., and Yin, W. A three-operator splitting scheme and its optimization applications. *Set-Valued and Variational Analysis 25*, 4 (Dec 2017), 829–858.

[39] Dontchev, A. L., and Rockafellar, R. T. Regularity and conditioning of solution mappings in variational analysis. *Set-Valued Analysis 12*, 1-2 (2004), 79–109.

[40] Douglas, J., and Rachford, H. H. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American Mathematical Society 82*, 2 (1956), 421–439.

[41] Drusvyatskiy, D., and Lewis, A. S. Error bounds, quadratic growth, and linear convergence of proximal methods. *Mathematics of Operations Research* (2018).

[42] Eckstein, J., and Bertsekas, D. P. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming 55*, 1 (Apr 1992), 293–318.

[43] Ermol'ev, Y. M., and Tuniev, A. D. Random Fejér and quasi-Fejér sequences. *Theory of Optimal Solutions — Akademiya Nauk Ukrainskoı SSR Kiev 2* (1968), 76–83.

[44] Facchinei, F., and Pang, J.-S. *Finite-dimensional variational inequalities and complementarity problems*, vol. II. Springer, 2003.

[45] Fang, H.-r., and Saad, Y. Two classes of multisecant methods for nonlinear acceleration. *Numerical Linear Algebra with Applications 16*, 3 (2009), 197–221.

[46] Gabay, D. Chapter IX applications of the method of multipliers to variational inequalities. In *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, M. F. and R. G., Eds., vol. 15 of *Studies in Mathematics and Its Applications*. Elsevier, 1983, pp. 299–331.

[47] Gabay, D., and Mercier, B. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications 2*, 1 (1976), 17–40.

[48] Giselsson, P., Fält, M., and Boyd, S. Line search for averaged operator iteration. In *2016 IEEE 55th Conference on Decision and Control (CDC)* (Dec 2016), pp. 1015–1022.

[49] Glowinski, R. *Numerical Methods for Nonlinear Variational Problems*. Scientific Computation. Springer, Berlin Heidelberg, 2013.

[50] Glowinski, R. On alternating direction methods of multipliers: A historical perspective. In *Modeling, Simulation and Optimization for Science and Technology*, W. Fitzgibbon, Y. A. Kuznetsov, P. Neittaanmäki, and O. Pironneau, Eds. Springer Netherlands, Dordrecht, 2014, pp. 59–82.

[51] Glowinski, R., and Marrocco, A. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique 9*, R2 (1975), 41–76.

[52] Goncalves, M. L. N., Melo, J. G., and Monteiro, R. D. C. Convergence rate bounds for a proximal ADMM with over-relaxation stepsize parameter for solving nonconvex linearly constrained problems. *ArXiv e-prints* (Feb. 2017).

[53] Guo, K., Han, D., and Wu, T.-T. Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear

constraints. *International Journal of Computer Mathematics 94*, 8 (2017), 1653–1669.

[54] Güler, O. New proximal point algorithms for convex minimization. *SIAM Journal on Optimization 2*, 4 (1992), 649–664.

[55] Hesse, R., and Luke, R. Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems. *SIAM Journal on Optimization 23*, 4 (2013), 2397–2419.

[56] Hesse, R., Luke, R., and Neumann, P. Alternating projections and Douglas-Rachford for sparse affine feasibility. *IEEE Transactions on Signal Processing 62*, 18 (Sept 2014), 4868–4881.

[57] Hiriart-Urruty, J.-B., and Lemaréchal, C. *Fundamentals of Convex Analysis*. Grundlehren Text Editions. Springer Berlin Heidelberg, 2012.

[58] Hong, M., Luo, Z.-Q., and Razaviyayn, M. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. *SIAM Journal on Optimization 26*, 1 (2016), 337–364.

[59] Ip, C.-M., and Kyparisis, J. Local convergence of quasi-Newton methods for B-differentiable equations. *Mathematical Programming 56*, 1-3 (1992), 71–89.

[60] Iutzeler, F., Bianchi, P., Ciblat, P., and Hachem, W. Explicit convergence rate of a distributed alternating direction method of multipliers. *IEEE Transactions on Automatic Control 61*, 4 (April 2016), 892–904.

[61] Izmailov, A. F., and Solodov, M. V. *Newton-type methods for optimization and variational problems*. Springer, 2014.

[62] Krasnosel'skii, M. A. Two remarks on the method of successive approximations. *Uspekhi Matematicheskikh Nauk 10*, 1 (1955), 123–127.

[63] Kurdyka, K. On gradients of functions definable in o-minimal structures. *Annales de l'institut Fourier 48*, 3 (1998), 769–783.

[64] Lange, K. *MM Optimization Algorithms*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2016.

[65] Latafat, P., Freris, N. M., and Patrinos, P. A new randomized block-coordinate primal-dual proximal algorithm for distributed optimization. *ArXiv e-prints* (Jun 2017).

[66] Latafat, P., and Patrinos, P. Asymmetric forward-backward-adjoint splitting for solving monotone inclusions involving three operators. *Computational Optimization and Applications 68*, 1 (Sep 2017), 57–93.

[67] Lemaréchal, C., and Sagastizábal, C. Practical aspects of the Moreau-Yosida regularization: Theoretical preliminaries. *SIAM Journal on Optimization 7*, 2 (1997), 367–385.

[68] Lewis, A. S. Active sets, nonsmoothness, and sensitivity. *SIAM Journal on Optimization 13*, 3 (2002), 702–725.

[69] Li, G., Liu, T., and Pong, T. K. Peaceman–Rachford splitting for a class of nonconvex optimization problems. *Computational Optimization and Applications 68*, 2 (Nov 2017), 407–436.

[70] Li, G., and Pong, T. K. Global convergence of splitting methods for nonconvex composite optimization. *SIAM Journal on Optimization 25*, 4 (2015), 2434–2460.

[71] Li, G., and Pong, T. K. Douglas-Rachford splitting for nonconvex optimization with application to nonconvex feasibility problems. *Mathematical Programming 159*, 1 (Sep 2016), 371–401.

[72] Li, H., and Lin, Z. Accelerated proximal gradient methods for nonconvex programming. In *Advances in Neural Information Processing Systems 28*. 2015, pp. 379–387.

[73] Liu, Y., and Yin, W. An envelope for Davis-Yin splitting and strict saddle point avoidance. *ArXiv e-prints* (Apr 2018), arXiv:1804.08739.

[74] Łojasiewicz, S. Une propriété topologique des sous-ensembles analytiques réels. *Les équations aux dérivées partielles* (1963), 87–89.

[75] Łojasiewicz, S. Sur la géométrie semi- et sous- analytique. *Annales de l'institut Fourier 43*, 5 (1993), 1575–1595.

[76] Luke, D. R., Teboulle, M., and Thao, N. H. Necessary conditions for linear convergence of iterated expansive, set-valued mappings with application to alternating projections. *ArXiv e-prints* (apr 2017).

[77] Luo, Z.-Q., and Tseng, P. Error bounds and convergence analysis of feasible descent methods: a general approach. *Annals of Operations Research 46*, 1 (Mar 1993), 157–178.

[78] Mairal, J. Incremental majorization-minimization optimization with application to large-scale machine learning. *SIAM Journal on Optimization 25*, 2 (2015), 829–855.

[79] Mann, W. R. Mean value methods in iteration. *Proceedings of the American Mathematical Society 4*, 3 (1953), 506–510.

[80] Maratos, N. *Exact penalty function algorithms for finite dimensional and control optimization problems.* PhD thesis, Imperial College London (University of London), 1978.

[81] Martinet, B. Brève communication. Régularisation d'inéquations variationnelles par approximations successives. *Revue française d'informatique et de recherche opérationnelle. Série rouge 4*, R3 (1970), 154–158.

[82] Monteiro, R. D. C., and Sim, C.-K. Complexity of the relaxed Peaceman-Rachford splitting method for the sum of two maximal strongly monotone operators. *Computational Optimization and Applications 70*, 3 (Jul 2018), 763–790.

[83] Nesterov, Y. A method of solving a convex programming problem with convergence rate $o(1/k^2)$. *Soviet Mathematics Doklady 27* (1983).

[84] Nesterov, Y. *Introductory lectures on convex optimization: A basic course*, vol. 87. Springer, 2003.

[85] Nesterov, Y. Efficiency of coordinate descent methods on huge-scale optimization problems. *SIAM Journal on Optimization 22*, 2 (2012), 341–362.

[86] Nesterov, Y. Gradient methods for minimizing composite functions. *Mathematical Programming 140*, 1 (Aug 2013), 125–161.

[87] Nesterov, Y. Implementable tensor methods in unconstrained convex optimization. Tech. rep., UC Louvain, Center for Operations Research and Econometrics (CORE), Belgium, 2018.

[88] Nocedal, J., and Wright, S. *Numerical optimization.* Springer Science & Business Media, 2006.

[89] O'donoghue, B., Chu, E., Parikh, N., and Boyd, S. Conic optimization via operator splitting and homogeneous self-dual embedding. *Journal of Optimization Theory and Applications 169*, 3 (jun 2016), 1042–1068.

[90] Pang, J.-S. Newton's method for B-differentiable equations. *Mathematics of Operations Research 15*, 2 (1990), 311–341.

[91] Parikh, N., and Boyd, S. Proximal algorithms. *Found. Trends Optim. 1*, 3 (Jan. 2014), 127–239.

[92] Patrinos, P., and Bemporad, A. Proximal Newton methods for convex composite optimization. In *52nd IEEE Conference on Decision and Control* (2013), pp. 2358–2363.

[93] Patrinos, P., Stella, L., and Bemporad, A. Douglas-Rachford splitting: Complexity estimates and accelerated variants. In *53rd IEEE Conference on Decision and Control* (Dec 2014), pp. 4234–4239.

[94] Patrinos, P., Stella, L., and Bemporad, A. Forward-backward truncated Newton methods for convex composite optimization. *ArXiv e-prints* (feb 2014).

[95] Poliquin, R. A., and Rockafellar, R. T. Amenable functions in optimization. *Nonsmooth optimization: methods and applications* (1992), 338–353.

[96] Poliquin, R. A., and Rockafellar, R. T. Second-order nonsmooth analysis in nonlinear programming. *Recent advances in nonsmooth optimization* (1995), 322–349.

[97] Poliquin, R. A., and Rockafellar, R. T. Generalized Hessian properties of regularized nonsmooth functions. *SIAM Journal on Optimization 6*, 4 (1996), 1121–1137.

[98] Poliquin, R. A., and Rockafellar, R. T. Prox-regular functions in variational analysis. *Transactions of the American Mathematical Society 348*, 5 (1996), 1805–1838.

[99] Powell, M. A hybrid method for nonlinear equations. In *Numerical Methods for Nonlinear Algebraic Equations*. Gordon and Breach, London, 1970, ch. 6, pp. 87–144.

[100] Powell, M. J. A fast algorithm for nonlinearly constrained optimization calculations. In *Numerical Analysis* (Berlin, Heidelberg, 1978), G. A. Watson, Ed., Springer Berlin Heidelberg, pp. 144–157.

[101] Robinson, S. M. Some continuity properties of polyhedral multifunctions. In *Mathematical Programming at Oberwolfach*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1981, pp. 206–214.

[102] Rockafellar, R. T. *Convex Analysis*. Princeton University Press, 1970.

[103] Rockafellar, R. T. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization 14*, 5 (1976), 877–898.

[104] Rockafellar, R. T. First- and second-order epi-differentiability in nonlinear programming. *Transactions of the American Mathematical Society 307*, 1 (1988), 75–108.

[105] Rockafellar, R. T. Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives. *Mathematics of Operations Research 14*, 3 (1989), 462–484.

[106] Rockafellar, R. T., and Wets, R. J.-B. *Variational analysis*, vol. 317. Springer Science & Business Media, 2011.

[107] Sathya, A. S., Sopasakis, P., Van Parys, R., Themelis, A., Pipeleers, G., and Patrinos, P. Embedded nonlinear model predictive control for obstacle avoidance using PANOC. In *2018 European Control Conference (ECC)* (2018), pp. 1523–1528. to appear.

[108] Scutari, G., and Sun, Y. Parallel and distributed successive convex approximation methods for big-data optimization. In *Multi-agent Optimization: Cetraro, Italy 2014*, F. Facchinei and J.-S. Pang, Eds. Springer International Publishing, Cham, 2018, pp. 141–308.

[109] Solodov, M. V., and Svaiter, B. F. A globally convergent inexact Newton method for systems of monotone equations. In *Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*. Springer, 1998, pp. 355–369.

[110] Sopasakis, P., Themelis, A., Suykens, J., and Patrinos, P. A primal-dual line search method and applications in image processing. In *2017 25th European Signal Processing Conference (EUSIPCO)* (Aug 2017), pp. 1065–1069.

[111] Stathopoulos, G., Shukla, H., Szucs, A., Pu, Y., and Jones, C. N. Operator splitting methods in control. *Foundations and Trends in Systems and Control 3*, 3 (2016), 249–362.

[112] Stella, L. *Proximal Envelopes: Smooth Optimization Algorithms for Nonsmooth Problems*. PhD thesis, KU Leuven, Belgium, 2017.

[113] Stella, L., Themelis, A., and Patrinos, P. Forward-backward quasi-Newton methods for nonsmooth optimization problems. *Computational Optimization and Applications 67*, 3 (Jul 2017), 443–487.

[114] Stella, L., Themelis, A., and Patrinos, P. Newton-type alternating minimization algorithm for convex optimization. *IEEE Transactions on Automatic Control 64*, 2 (2018).

[115] Stella, L., Themelis, A., Sopasakis, P., and Patrinos, P. A simple and efficient algorithm for nonlinear model predictive control. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)* (Dec 2017), pp. 1939–1944.

[116] Sun, Y., Babu, P., and Palomar, D. P. Majorization-minimization algorithms in signal processing, communications, and machine learning. *IEEE Transactions on Signal Processing 65*, 3 (Feb 2017), 794–816.

[117] Themelis, A., Ahookhosh, M., and Patrinos, P. On the acceleration of forward-backward splitting via an inexact Newton method. In *Splitting Algorithms, Modern Operator Theory, and Applications*, R. Luke, H. Bauschke, and R. Burachik, Eds. Springer. To appear https://arxiv.org/abs/1811.02935.

[118] Themelis, A., and Patrinos, P. SuperMann: a superlinearly convergent algorithm for finding fixed points of nonexpansive operators. *ArXiv e-prints* (Sep 2016).

[119] Themelis, A., and Patrinos, P. Douglas-Rachford splitting and ADMM for nonconvex optimization: tight convergence results. *ArXiv e-prints* (Sep 2017).

[120] Themelis, A., Stella, L., and Patrinos, P. Forward-backward envelope for the sum of two nonconvex functions: Further properties and nonmonotone linesearch algorithms. *SIAM Journal on Optimization 28*, 3 (2018), 2274–2303.

[121] Themelis, A., Villa, S., Patrinos, P., and Bemporad, A. Stochastic gradient methods for stochastic model predictive control. In *2016 European Control Conference (ECC)* (June 2016), pp. 154–159.

[122] Tseng, P. On accelerated proximal gradient methods for convex-concave optimization. Tech. rep., 2008. Submitted to SIAM Journal on Optimization.

[123] Wang, Y., Yin, W., and Zeng, J. Global convergence of ADMM in nonconvex nonsmooth optimization. *Journal of Scientific Computing* (Jun 2018).

[124] Xu, Z., Chang, X., Xu, F., and Zhang, H. $l_{1/2}$ regularization: a thresholding representation theory and a fast solver. *IEEE Transactions on neural networks and learning systems 23*, 7 (2012), 1013–1027.

[125] Yan, M., and Yin, W. Self equivalence of the alternating direction method of multipliers. In *Splitting Methods in Communication, Imaging,*

Science, and Engineering, R. Glowinski, S. J. Osher, and W. Yin, Eds. Springer International Publishing, Cham, 2016, pp. 165–194.

[126] Zhang, H., and Hager, W. W. A nonmonotone line search technique and its application to unconstrained optimization. *SIAM Journal on Optimization 14*, 4 (2004), 1043–1056.

[127] Zheng, Y., Fantuzzi, G., Papachristodoulou, A., Goulart, P., and Wynn, A. Fast ADMM for semidefinite programs with chordal sparsity. In *2017 American Control Conference (ACC)* (May 2017), pp. 3335–3340.



FACULTY OF ENGINEERING SCIENCE
DEPARTMENT OF ELECTRICAL ENGINEERING
STADIUS
Kasteelpark Arenberg 10, bus 2446
B-3001 Leuven
andreas.themelis@kuleuven.be
https://www.esat.kuleuven.be/stadius/