Improved Dual Decomposition Based Optimization for DSL Dynamic Spectrum Management

Paschalis Tsiaflakis, *Member, IEEE*, Ion Necoara, Johan A. K. Suykens, *Senior Member, IEEE*, and Marc Moonen, *Fellow, IEEE*

Abstract

Dynamic spectrum management (DSM) has been recognized as a key technology to significantly improve the performance of digital subscriber line (DSL) broadband access networks. The basic concept of DSM is to coordinate transmission over multiple DSL lines so as to mitigate the impact of crosstalk interference amongst them. Many algorithms have been proposed to tackle the nonconvex optimization problems appearing in DSM, many of them relying on a standard subgradient based dual decomposition approach. In practice however, this approach is often found to lead to extremely slow convergence or even no convergence at all, one of the reasons being the very difficult tuning of the stepsize parameters. In this paper we propose a novel improved dual decomposition approach inspired by recent advances in mathematical programming. It uses a smoothing technique for the Lagrangian combined with an optimal gradient based scheme for updating the Lagrange multipliers. The stepsize parameters are furthermore selected optimally removing the need for a tuning strategy. With this approach we show how the convergence of current state-of-the-art DSM algorithms based on iterative convex approximations (SCALE, CA-DSB) can be improved by one order of magnitude. Furthermore we apply the improved dual decomposition approach to other DSM algorithms (OSB, ISB, ASB, (MS)-DSB, MIW) and propose further improvements to obtain fast and robust DSM algorithms. Finally, we demonstrate the effectiveness of the improved dual decomposition approach for a number of realistic multi-user DSL scenarios.

EDICS: SPC-TDLS, SPC-MULT, MSP-APPL

Index Terms

Digital subscriber line (DSL), dual decomposition, dynamic spectrum management, interference channel, multi-carrier, MIMO, multi-agent, optimization.

Copyright (c) 2008 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

This research work was carried out at the ESAT Laboratory of the Katholieke Universiteit Leuven, in the frame of K.U. Leuven Research Council: CoE EF/05/006, GOA AMBioRICS, FWO project G.0235.07('Design and evaluation of DSL systems with common mode signal exploitation'), FWO project G.0226.06, Belgian Federal Science Policy Office IUAP DYSCO. A portion of this paper has appeared in the Proceedings of the 17th European Signal Processing Conference (EUSIPCO),August 2009 [1].

P. Tsiaflakis, J.A.K. Suykens and M. Moonen are with the Department of Electrical Engineering, Katholieke Universiteit Leuven (K.U. Leuven), ESAT/SISTA, B-3001 Leuven-Heverlee, Belgium (e-mail: Paschalis.Tsiaflakis@esat.kuleuven.be; Johan.Suykens@esat.kuleuven.be; Marc.Moonen@esat.kuleuven.be). I. Necoara is with the Department of Automation and Systems Engineering, University Politehnica Bucharest, 060042 Bucharest, Romania (e-mail: ion.necoara@esat.kuleuven.be).

I. INTRODUCTION

Digital subscriber line (DSL) technology refers to a family of technologies that provide digital broadband access over the local telephone network. It is currently the dominating broadband access technology with more than 66% of all broadband access subscribers worldwide using DSL to access the Internet. It is forecasted that the number of DSL subscribers will even rise to 331 million in 2012 with DSL access revenues reaching $136.4 billion in 2012 [2]. The major obstacle for further performance improvement in modern DSL networks is the so-called crosstalk, i.e. the electromagnetic interference amongst different lines in the same cable bundle. Different lines (i.e. users) indeed interfere with each other, leading to a very challenging interference environment where proper management of the resources is required to prevent a huge performance degradation.

Dynamic spectrum management (DSM) has been recognized as a key technology to significantly improve the performance of DSL broadband access networks [3]. The basic concept of DSM is to coordinate transmission over multiple DSL lines so as to mitigate the impact of crosstalk interference amongst them. There are two types of coordination referred to as spectrum level and signal level coordination. Here, we will focus on spectrum level coordination, also referred to as spectrum management, spectrum balancing or multi-carrier power control. Spectrum management aims to allocate transmit spectra, i.e. transmit powers over all available frequencies (tones), to the different users so as to achieve some design objective. This generally corresponds to an optimization problem, where typically a weighted sum of user data rates is maximized subject to power constraints [4]–[6], which will be referred to as "constrained weighted rate sum maximization (cWRS)". Recently this has been extended to other design objectives as well, such as power driven designs (green DSL [7], [8], [9], [10]) and other utility driven designs [11], [12]. As shown in [7], the key component to these designs is an efficient solution for the cWRS problem. Therefore we will mainly focus on this problem and aim to find a robust and efficient solution for it.

The cWRS problem is known to be an NP-hard, separable nonconvex optimization problem, that can have many locally optimal solutions [11] [13]. Even for moderately sized problems (with 5-20 users and 200-4000 tones), finding the globally optimal solution is computationally prohibitive. In [5] and [14] the authors proposed to use a dual decomposition approach with a standard subgradient based updating of the Lagrange multipliers. Many DSM algorithms [4], [6], [13]–[19] have been proposed recently, that use the standard subgradient based dual decomposition approach. In practice, however, this approach is often found to lead to extremely slow convergence or even no convergence at all, especially so for large DSL scenarios with large crosstalk. One of the reasons is the very difficult tuning of the stepsize parameters so as to guarantee fast convergence. We would like to remark here that the subgradient based dual decomposition algorithms are not the only algorithms in use for spectrum management and that a

number of alternative approaches exist, e.g. the ellipsoid dual update approach proposed in [5], Spectrum Balancing Levin-Campello (SBLC) proposed in [20], [21], etc.

In this paper we propose a novel improved dual decomposition approach inspired by recent advances in mathematical programming, more specifically the proximal center based decomposition method recently proposed in [22]. This method uses a smoothing technique [23] for the Lagrangian and combines it with an accelerated scheme for smooth optimization [24]. Moreover, the stepsize parameter is determined automatically so as to obtain fast convergence, removing the need for a stepsize tuning strategy. However, the proximal center based method is proposed for separable convex problems, it is only derived for two users under equality constraints, and it is presented from a rather high-level point of view, where the concrete steps of the algorithms are general optimization problems to be solved (see Algorithm 3.2 in [22]). DSM optimization problems are, however, highly nonconvex problems with multiple user scenarios and multiple tones under inequality constraints. In this paper we extend the proximal center based decomposition method to a concrete improved dual decomposition approach for particular application in the context of DSM. More specifically, it consists of an extension of the proximal center based method to the nonconvex cWRS problem, where it is derived for multiple users and multiple tones under inequality constraints and with concrete efficient implementations for specific types of DSM problems. With the proposed approach, we show how the convergence of current state-of-the-art DSM algorithms based on iterative convex approximations (e.g. successive convex approximation for low-complexity (SCALE) [16], convex approximation distributed spectrum balancing (CA-DSB) [13]) can be speed up by one order of magnitude, without increasing the computational complexity for each iteration. Furthermore we apply the improved dual decomposition approach to other DSM algorithms (optimal spectrum balancing (OSB) [4], modified prismatic branch-and-bound algorithm (PBnB) [15], iterative spectrum balancing (ISB) [14], autonomous spectrum balancing (ASB) [18], (multiple starting point) distributed spectrum balancing ((MS-)DSB) [13], modified iterative water-filling (MIW) [17], branch-and-bound optimal spectrum balancing (BB-OSB) [6]), again leading to much faster converging DSM algorithms. Then we demonstrate an important pitfall of applying the dual decomposition approach to nonconvex DSM problems and propose an effective solution that further improves the robustness of current DSM algorithms. Finally we demonstrate the effectiveness of the improved dual decomposition approach for a number of realistic multi-user DSL scenarios.

This paper is organized as follows. In Section II, the system model is introduced for the DSL multi-user environment. In Section III, the basic cWRS problem is described and existing DSM algorithms for this problem are reviewed, that rely on a subgradient based dual decomposition approach. In Section IV-A an improved dual decomposition approach is proposed for DSM algorithms based on iterative convex approximations. The improved dual decomposition approach is furthermore applied to other

DSM algorithms in Section IV-B. In Section V, the problem of obtaining a primal solution from the dual solution is described and an effective solution for it is proposed. Finally in Section VI, simulation results are shown.

II. System Model

We consider a system consisting of $\mathcal{N} = \{1, \ldots, N\}$ interfering DSL users (i.e., lines, modems) with standard synchronous discrete multi-tone (DMT) modulation with $\mathcal{K} = \{1, \ldots, K\}$ tones (i.e., frequencies or carriers). The transmission can be modeled independently on each tone k by

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{z}_k.$$

The vector $\mathbf{x}_k = [x_k^1, \ldots, x_k^N]^T$ contains the transmitted signals on tone k, where x_k^n refers to the signal transmitted by user n on tone k. Vectors $\mathbf{z}_k$ and $\mathbf{y}_k$ have similar structures; $\mathbf{z}_k$ refers to the additive noise on tone k, containing thermal noise, alien crosstalk, radio frequency interference (RFI), etc, and $\mathbf{y}_k$ refers to the received signals on tone k. $\mathbf{H}_k$ is an $N \times N$ channel matrix with $[\mathbf{H}_k]_{n,m} = h_k^{n,m}$ referring to the channel gain from transmitter m to receiver n on tone k. The diagonal elements are the direct channels and the off-diagonal elements are the crosstalk channels.

The transmit power of user n on tone k, also referred to as transmit power spectral density, is denoted as $s_k^n \triangleq \Delta_f E\{|x_k^n|^2\}$, where Δ_f refers to the tone spacing. The vector $\mathbf{s}_k \triangleq \{s_k^n, n \in \mathcal{N}\}$ denotes the transmit powers of all users on tone k. The vector $\mathbf{s}^n \triangleq \{s_k^n, k \in \mathcal{K}\}$ denotes the transmit powers of user n on all tones. The received noise power by user n on tone k, also referred to as noise spectral density, is denoted as $\sigma_k^n \triangleq \Delta_f E\{|z_k^n|^2\}$.

Note that we assume no signal coordination at the transmitters and at the receivers, and that the interference is treated as additive white Gaussian noise. Under this standard assumption the bit loading for user n on tone k, given the transmit spectra $\mathbf{s}_k$ of all users on tone k, is

$$b_k^n \triangleq b_k^n(\mathbf{s}_k) \triangleq \log_2\left(1 + \frac{1}{\Gamma}\frac{|h_k^{n,n}|^2 s_k^n}{\sum_{m \neq n} |h_k^{n,m}|^2 s_k^m + \sigma_k^n}\right) \text{ bits/Hz}, \tag{1}$$

where Γ denotes the SNR-gap to capacity, which is a function of the desired BER, the coding gain and noise margin [25]. The DMT symbol rate is denoted as f_s. The achievable total data rate for user n and the total transmit power used by user n are equal to, respectively:

$$R^n \triangleq f_s \sum_{k \in \mathcal{K}} b_k^n \quad \text{and} \quad P^n \triangleq \sum_{k \in \mathcal{K}} s_k^n. \tag{2}$$

III. Dynamic spectrum management

A. Dynamic spectrum management problem

The basic goal of DSM through spectrum level coordination is to allocate the transmit powers dynamically in response to physical channel conditions (channel gains and noise) so as to pursue certain design

objectives and/or satisfy certain constraints. The constraints are mostly per-user total power constraints and so-called spectral mask constraints, i.e.

$$\begin{aligned} P^n &\leq P^{n,\mathrm{tot}}, && n \in \mathcal{N}, \\ 0 \leq s_k^n &\leq s_k^{n,\mathrm{mask}}, && n \in \mathcal{N}, k \in \mathcal{K}, \end{aligned} \tag{3}$$

where $P^{n,\mathrm{tot}}$ refers to the total available transmit power budget for user n and $s_k^{n,\mathrm{mask}}$ refers to the spectral mask constraint for user n on tone k. The user total power constraints can also be written in vector notation as $\mathbf{P} \leq \mathbf{P}^{\mathrm{tot}}$, where $\mathbf{P} = [P^1, \ldots, P^N]^T$ and $\mathbf{P}^{\mathrm{tot}} = [P^{1,\mathrm{tot}}, \ldots, P^{N,\mathrm{tot}}]^T$, and where '$\leq$' denotes a component-wise inequality.

The set of all possible data rate allocations that satisfy the constraints (3) can be characterized by the achievable rate region $\mathcal{R}$:

$$\mathcal{R} = \Big\{ (R^n : n \in \mathcal{N}) | R^n = f_s \sum_{k \in \mathcal{K}} b_k^n(\mathbf{s}_k), \text{ s.t. } (3) \Big\}.$$

A typical design objective is to achieve some Pareto optimal allocation of the data rates R^n [4], [6], [13]–[19], [26], [27]. This results in the following typical DSM optimization problem, which will be referred as the constrained weighted rate sum maximization (cWRS) formulation, where w_n is the weight given to user n:

$$\begin{aligned} \max_{\{\mathbf{s}^n, n \in \mathcal{N}\}} \quad & \sum_{n \in \mathcal{N}} w_n R^n \\ \text{s.t.} \quad & P^n \leq P^{n,\mathrm{tot}}, && n \in \mathcal{N}, && (\mathrm{cWRS}) \\ & 0 \leq s_k^n \leq s_k^{n,\mathrm{mask}}, && n \in \mathcal{N}, k \in \mathcal{K}. \end{aligned} \tag{4}$$

However, many other DSM formulations are possible. We refer to [7] containing a collection of other relevant DSM formulations. As shown in [7], the key component to tackling these is an efficient solution for cWRS problem (4). Therefore we will focus on this problem and aim to find a robust and efficient solution for it.

B. Dynamic spectrum management algorithms

cWRS problem (4) is an NP-hard separable nonconvex optimization problem [11]. The number of optimization variables is equal to KN, where the number of users N ranges between 2-100 and the number of tones K can go up to 4000. Depending on the specific values of the channel and noise parameters, there can be many locally optimal solutions, that can differ significantly in value, as shown in [13]. In [5] the authors show that strong duality holds for the continuous (frequency range) formulation, and in [11] the authors prove asymptotic strong duality for the discrete (frequency range) formulation, i.e. the duality gap goes to zero as $K \to \infty$. These results suggest that a Lagrange dual decomposition approach is a viable way to reach approximate optimality for the discrete formulation (4), if the frequency

range is finely discretized, as it is indeed the case in practical DSL scenarios where K is large [5]. Many dual decomposition based DSM algorithms [6], [13]–[19] have been proposed for solving (4), that use a standard subgradient based updating of the Lagrange multipliers.

The dual problem formulation of (4) consists of two subproblems, namely a master problem

$$\begin{aligned} \min_{\boldsymbol{\lambda}} \quad & g(\boldsymbol{\lambda}) \\ \text{s.t.} \quad & \boldsymbol{\lambda} \geq 0 \end{aligned} \tag{5}$$

where $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_N]^T$, and a slave problem defined by the dual function $g(\boldsymbol{\lambda})$:

$$\begin{aligned} g(\boldsymbol{\lambda}) = & \left\{ \begin{array}{ll} \max\limits_{\{\mathbf{s}_k, k\in\mathcal{K}\}} & \mathcal{L}(\boldsymbol{\lambda}, \mathbf{s}_k, k \in \mathcal{K}) \\ \text{s.t.} & 0 \leq s_k^n \leq s_k^{n,\text{mask}}, \qquad n \in \mathcal{N}, k \in \mathcal{K}, \end{array} \right. \\ \text{with} \quad & \mathcal{L}(\boldsymbol{\lambda}, \mathbf{s}_k, k \in \mathcal{K}) = \sum_{n\in\mathcal{N}} w_n R^n - \sum_{n\in\mathcal{N}} \lambda_n (P^n - P^{n,\text{tot}}) \end{aligned} \tag{6}$$

where $\mathcal{L}(\boldsymbol{\lambda}, \mathbf{s}_k, k \in \mathcal{K})$ is the Lagrangian. This can be reformulated as:

$$\begin{aligned} g(\boldsymbol{\lambda}) = & \left\{ \begin{array}{ll} \max\limits_{\{\mathbf{s}_k, k\in\mathcal{K}\}} & \sum\limits_{k\in\mathcal{K}} \Big\{ b_k(\mathbf{s_k}) - \sum\limits_{n\in\mathcal{N}} \lambda_n s_k^n + \sum\limits_{n\in\mathcal{N}} \lambda_n P^{n,\text{tot}}/K \Big\} \\ \text{s.t.} & 0 \leq s_k^n \leq s_k^{n,\text{mask}}, \qquad n \in \mathcal{N}, k \in \mathcal{K} \end{array} \right. \\ \text{with} \quad & b_k(\mathbf{s}_k) = \sum_{n\in\mathcal{N}} w_n f_s b_k^n(\mathbf{s}_k) \end{aligned} \tag{7}$$

The slave optimization problem (7) can then be decomposed into K independent nonconvex subproblems (dual decomposition):

$$\begin{aligned} & g(\boldsymbol{\lambda}) = \sum_{k\in\mathcal{K}} g_k(\boldsymbol{\lambda}) \\ & \text{with } g_k(\boldsymbol{\lambda}) = \left\{ \begin{array}{ll} \max\limits_{\mathbf{s}_k} & b_k(\mathbf{s}_k) - \sum\limits_{n\in\mathcal{N}} \lambda_n s_k^n + \sum\limits_{n\in\mathcal{N}} \lambda_n P^{n,\text{tot}}/K \\ \text{s.t.} & 0 \leq s_k^n \leq s_k^{n,\text{mask}}, \quad n \in \mathcal{N} \end{array} \right. \end{aligned} \tag{8}$$

The master problem (5), also called the dual problem, is a convex optimization problem. Its objective function, i.e. the dual function $g(\boldsymbol{\lambda})$, is however non-differentiable. The reason for this non-differentiability is that the underlying slave optimization problem (6) can have multiple globally optimal solutions for some values of the Lagrange multipliers $\boldsymbol{\lambda}$. In [5] [14] a subgradient updating approach is proposed for this dual master problem, where the subgradient is defined as,

$$dg(\boldsymbol{\lambda}) \triangleq \sum_{k\in\mathcal{K}} \mathbf{s}_k(\boldsymbol{\lambda}) - \mathbf{P}^{\text{tot}} \tag{9}$$

with $\mathbf{s}_k(\boldsymbol{\lambda})$ referring to the optimal solution of (8) for given Lagrange multipliers $\boldsymbol{\lambda}$, also called dual variables, and the corresponding subgradient update is:

$$\boldsymbol{\lambda} = \Big[\boldsymbol{\lambda} + \delta (\sum_{k\in\mathcal{K}} \mathbf{s}_k(\boldsymbol{\lambda}) - \mathbf{P}^{\text{tot}}) \Big]^+, \tag{10}$$

where $[\mathbf{x}]^+$ denotes the projection of $\mathbf{x} \in \mathcal{R}^N$ onto $\mathcal{R}_+^N$, and where the stepsize δ can be tuned using different procedures [5], [6], e.g. $\delta = q$ or $\delta = q/i$ where q is the initial stepsize and i is the iteration counter. By iteratively applying (10) and (8), convergence to an optimal solution of (5) can be achieved, i.e. $\boldsymbol{\lambda} \rightarrow \boldsymbol{\lambda}^*$, for which the complementary conditions, $\lambda_n(\sum_{k\in\mathcal{K}} s_k^n(\boldsymbol{\lambda}) - P^{n,\mathrm{tot}}) = 0,\ n \in \mathcal{N}$, are satisfied when strong duality "holds" ($K \rightarrow \infty$). This general standard subgradient based dual decomposition approach is visualized in Figure 1.

Note that the per-tone subproblems (8) are nonconvex optimization problems. Many existing DSM algorithms differ only in the way these subproblems are solved, where strategies are proposed such as exhaustive discrete search (OSB) [4], branch and bound search (PBnB [15], BB-OSB [6]), coordinate descent discrete search (ISB) [14] [26], solving the KKT system (DSB [13], MIW [17], MS-DSB [13]), and heuristic approximation (ASB [18], ASB2 [13]).

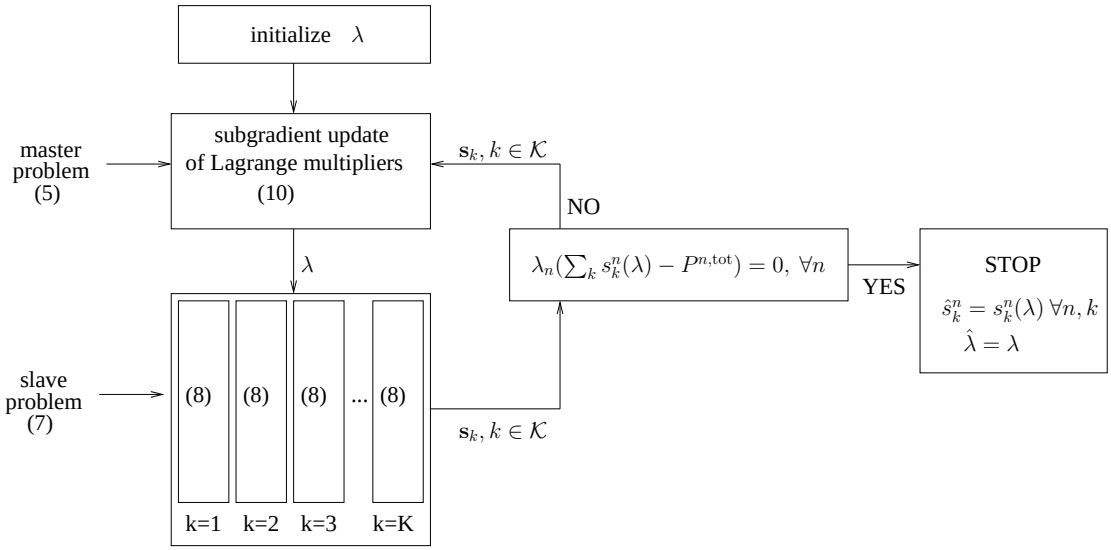


Fig. 1. General structure of subgradient based dual decomposition approach for DSM

An alternative approach is based on iterative convex approximations such as in SCALE [16] and CA-DSB [13]. This approach basically consists of iteratively executing the following two steps: (i) approximating the nonconvex cWRS problem (4) by a separable convex optimization problem F_{cvx}, and (ii) solving this convex approximation by using a subgradient based dual decomposition approach. Note that under some conditions on the approximation, described in [28], iteratively executing these steps results in asymptotic convergence to a locally optimal solution of cWRS (4). The convex approximations used by CA-DSB and SCALE both satisfy these conditions. This approach is visualized in Figure 2, where $F_{k,\mathrm{cvx}}$ refers to the per-tone convex problem obtained from the convex approximation F_{cvx}. We emphasize that these DSM algorithms also use a subgradient based dual decomposition approach to solve

a convex optimization problem in each iteration. This step requires the major part of the computational cost.

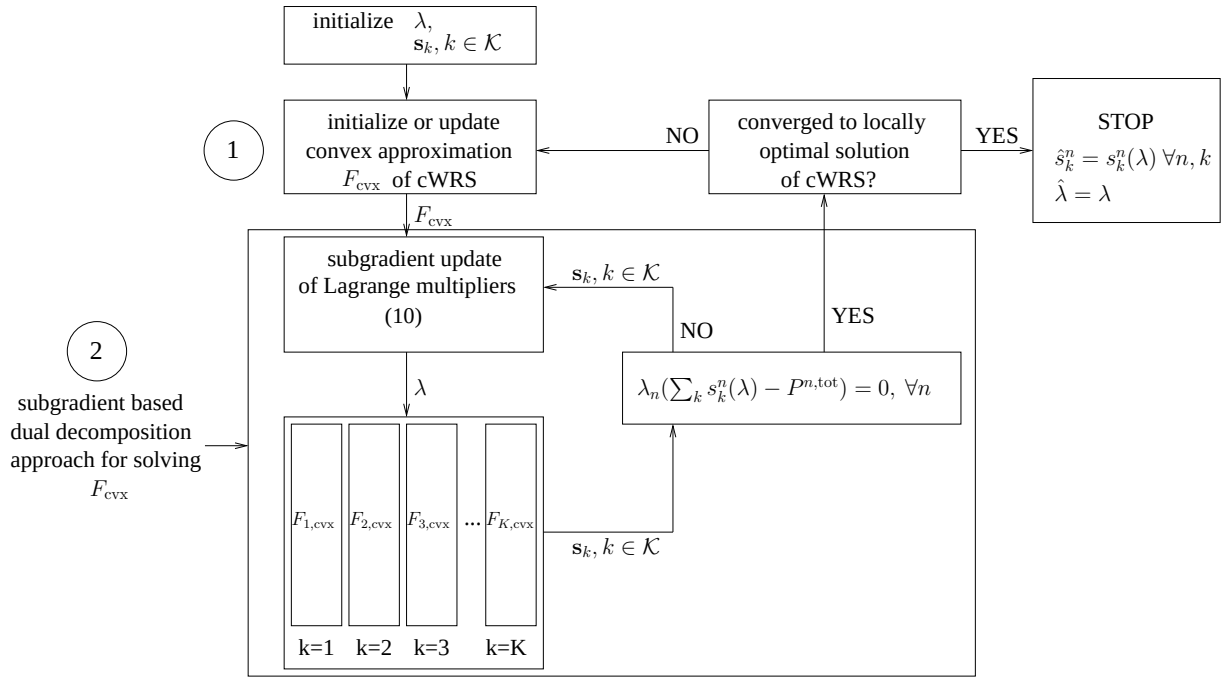


Fig. 2. Structure of iterative convex approximation approach for DSM

IV. Improved dual decomposition

In practice, the standard subgradient based dual decomposition approach is often found to lead to extremely slow convergence or even no convergence at all, especially so for large DSL scenarios (6-20 users) with large crosstalk (VDSL(2)). This is because of different reasons: (i) subgradient methods are generally known not to be efficient, i.e. showing worst case convergence of order $\mathcal{O}(\frac{1}{\epsilon^2})$ with ϵ referring to the required accuracy of the approximation of the optimum [24], (ii) the stepsize used by subgradient methods is quite difficult to tune in order to guarantee fast convergence, (iii) the nonconvex nature of the problem implies that special care should be taken in obtaining the optimal primal variables from the optimal dual variables.

Several alternative dual decomposition approaches have been proposed such as the alternating direction method [29], proximal method of multipliers [30], partial inverse method [31], etc. However these approaches only apply to separable convex problems, i.e. with a separable convex objective function and convex coupling constraints. Furthermore they destroy the separability of the problem, they cannot deal with inequality constraints in general, and they are sensitive to the chosen parameter values. Here, we focus on a recently proposed dual decomposition approach in [22], referred to as the proximal center based

decomposition method. This method shows interesting properties, namely it preserves the separability of the problem, it uses an optimal gradient based scheme, and it uses an optimal stepsize, thus removing the need for a tuning strategy. However, this method is proposed for convex separable problems, it is mainly derived for two users under equality constraints, and it is presented from a rather high-level point of view, where the steps consist of general optimization problems to be solved (see Algorithm 3.2 in [22]). In this section we extend this method to a concrete improved dual decomposition approach for solving the non-convex problem cWRS (4). This approach will be used first in Section IV-A to improve the convergence speed of DSM algorithms using iterative convex approximations (SCALE, CA-DSB) with one order of magnitude. In Section IV-B this will be extended to other DSM algorithms such as OSB, ISB, PBnB, BB-OSB, ASB, (MS-)DSB, MIW, etc. We will refer to these DSM algorithms that are not based on iterative convex approximations as "direct DSM algorithms".

A. An improved dual decomposition approach for iterative convex approximation based DSM algorithms

Two state-of-the-art DSM algorithms that are based on iterative convex approximations are SCALE and CA-DSB. These basically consist of two steps as explained in Section III-B, which are iteratively executed. In this section we will propose an improved dual decomposition approach for solving the convex optimization problem in the second step. We will elaborate this for CA-DSB and proof that its convergence speed is improved by one order of magnitude, i.e. from $\mathcal{O}(\frac{1}{\epsilon^2})$ to $\mathcal{O}(\frac{1}{\epsilon})$, with similar computational complexity with respect to the subgradient based dual decomposition approach. The improved dual decomposition approach can similarly be applied to the SCALE algorithm to obtain a similar speed up, but requires more complicated notation because of the inherent exponential transformation of variables. The content of this section has also appeared in [1].

For CA-DSB, the convex approximation in each iteration is obtained by reformulating the objective of cWRS, as a sum of a concave part and a convex part, and then approximating the convex part by a first order Taylor expansion. The resulting convex approximation, its dual formulation, dual function, and Lagrangian are given in (11), (12), (13), and (14), respectively.

$$f^*_{\text{cvx}} = \Big\{ \max_{\{\mathbf{s}_k \in \mathcal{S}_k, k \in \mathcal{K}\}} \sum_{k \in \mathcal{K}} b_{k,\text{cvx}}(\mathbf{s}_k) \;\; \text{s.t.} \sum_{k \in \mathcal{K}} s_k^n \leq P^{n,\text{tot}}, \;\; n \in \mathcal{N} \Big\} \tag{11}$$

$$\min_{\boldsymbol{\lambda} \geq 0} \; g_{\text{cvx}}(\boldsymbol{\lambda}) \tag{12}$$

$$g_{\text{cvx}}(\boldsymbol{\lambda}) = \max_{\{\mathbf{s}_k \in \mathcal{S}_k, k \in \mathcal{K}\}} \mathcal{L}_{\text{cvx}}(\mathbf{s}_k, k \in \mathcal{K}, \boldsymbol{\lambda}) \tag{13}$$

$$\mathcal{L}_{\text{cvx}}(\mathbf{s}_k, k \in \mathcal{K}, \boldsymbol{\lambda}) = \sum_{k \in \mathcal{K}} b_{k,\text{cvx}}(\mathbf{s}_k) - \sum_{k \in \mathcal{K}} \sum_{n \in \mathcal{N}} \lambda_n s_k^n + \sum_{n \in \mathcal{N}} \lambda_n P^{n,\text{tot}} \tag{14}$$

where $\mathcal{S}_k = \{\mathbf{s}_k \in \mathcal{R}^N : 0 \leq s_k^n \leq s_k^{n,\max}, n \in \mathcal{N}\}$ is a compact convex set with $s_k^{n,\max} :=$

$\min(s_k^{n,\max}, P^{n,\max})$ and $P^{n,\max} < \infty$, and where $b_{k,\text{cvx}}(\mathbf{s}_k)$ is concave and given as:

$$b_{k,\text{cvx}}(\mathbf{s}_k) = \sum_{n\in\mathcal{N}} w_n f_s \log_2(\sum_{m\in\mathcal{N}} |\tilde{h}_k^{n,m}|^2 s_k^m + \Gamma\sigma_k^n) - \sum_{n\in\mathcal{N}} w_n f_s (\sum_{m\neq n} a_k^{m,n} s_k^m + c_k^n), \tag{15}$$

with $a_k^{n,m}, c_k^n, \forall n, m, k$ constant approximation parameters, obtained by a closed-form formula in the approximation step [13], and with

$$|\tilde{h}_k^{n,m}|^2 \begin{cases} = \Gamma |h_k^{n,m}|^2, & n \neq m \\ = |h_k^{n,m}|^2, & n = m. \end{cases} \tag{16}$$

The convex problem (11) has a separable structure and so the standard way to solve it is by focusing on the dual problem (12) and using a subgradient update approach for the dual variables. This subgradient based dual decomposition approach is however known [24] to have a convergence speed of order $\mathcal{O}(\frac{1}{\epsilon^2})$, where ϵ is the required accuracy for the approximation of the optimum. In the sequel, it will be shown how the "proximal center based decomposition" method from [22] can be adapted for solving the convex approximation, leading to a scheme with convergence speed of order $\mathcal{O}(\frac{1}{\epsilon})$, i.e. one order of magnitude faster but with the same computational complexity. The basic steps in this result are as follows. First an approximated (smoothed) dual function $\bar{g}_{\text{cvx}}(\boldsymbol{\lambda})$ is defined that can be chosen to be arbitrarily close to the original dual function $g_{\text{cvx}}(\boldsymbol{\lambda})$. Then it is proven that this smoothed dual function $\bar{g}_{\text{cvx}}$ is differentiable and has a Lipschitz continuous gradient. Finally an optimal gradient scheme is applied to the smoothed dual problem.

We introduce the following functions $d_k(\mathbf{s}_k), k \in \mathcal{K}$, which are called prox-functions in [22] and are defined as follows:

Definition 1: A prox-function $d_k(\mathbf{s}_k)$ has the following properties:

- $d_k(\mathbf{s}_k)$ is a non-negative continuous and strongly convex function[1] with convexity parameter $\sigma_{\mathcal{S}_k}$
- $d_k(\mathbf{s}_k)$ is defined for the compact convex set $\mathcal{S}_k$

An example of a valid prox-function is $d_k(\mathbf{s}_k) = \frac{1}{2}\|\mathbf{s}_k\|^2$, which is also used in our concrete implementations (see Section VI). As many other valid prox-functions exist, and in order not to loose generality, we continue with $d_k(\mathbf{s}_k)$. Since $\mathcal{S}_k, k \in \mathcal{K}$, are compact and $d_k(\mathbf{s}_k)$ are continuous, we can choose finite and positive constants such that

$$D_{\mathcal{S}_k} \geq \max_{\mathbf{s}_k \in \mathcal{S}_k} d_k(\mathbf{s}_k), k \in \mathcal{K}. \tag{17}$$

[1]A continuously differentiable function $f(x)$ is called strongly convex on $\mathbb{R}^N$ if there exists a constant μ, called the convexity parameter of f such that for any $x, y \in \mathbb{R}^N$ we have $f(y) \geq f(x) + \nabla f(x)^T(y-x) + \frac{1}{2}\mu\|y-x\|^2$ [24].

This upper bound for $d_k(\mathbf{s}_k)$ can be easily computed. For instance, for the choice of prox-function $d_k(\mathbf{s}_k) = \frac{1}{2}\|\mathbf{s}_k\|^2$, $D_{\mathcal{S}_k}$ can be computed in closed-form as $D_{\mathcal{S}_k} = \frac{1}{2}\sum_{n\in\mathcal{N}}(s_k^{n,\max})^2$.

The prox-functions can be used to smoothen the dual function $g_{\text{cvx}}(\boldsymbol{\lambda})$ to obtain a smoothed dual function $\bar{g}_{\text{cvx}}(\boldsymbol{\lambda})$ as follows:

$$\bar{g}_{\text{cvx}}(\boldsymbol{\lambda}) = \max_{\{\mathbf{s}_k\in\mathcal{S}_k, k\in\mathcal{K}\}} \sum_{k\in\mathcal{K}} \Big\{ b_{k,\text{cvx}}(\mathbf{s}_k) - \sum_{n\in\mathcal{N}} \lambda_n (s_k^n - P^{n,\text{tot}}/K) - c d_k(\mathbf{s}_k) \Big\}, \tag{18}$$

where c is a positive smoothness parameter that will be defined in closed-form in Theorem 2 later this section. The value of this parameter c is defined sufficiently small, so as to make the smoothed dual function arbitrarily close to the original dual function.

One useful property of the particular choice of prox-functions is that they do not destroy the separability of the objective function in (18), i.e.

$$\bar{g}_{\text{cvx}}(\boldsymbol{\lambda}) = \sum_{k\in\mathcal{K}} \Big\{ \max_{\mathbf{s}_k\in\mathcal{S}_k} b_{k,\text{cvx}}(\mathbf{s}_k) - \sum_{n\in\mathcal{N}} \lambda_n (s_k^n - P^{n,\text{tot}}/K) - c d_k(\mathbf{s}_k) \Big\}. \tag{19}$$

Denote by $\bar{\mathbf{s}}_{k,\text{cvx}}(\boldsymbol{\lambda}), k\in\mathcal{K}$, the optimal solution of the maximization problem in (19). The following theorem describes the properties of the smoothed dual function $\bar{g}_{\text{cvx}}(\boldsymbol{\lambda})$:

Theorem 1 ([22]): The function $\bar{g}_{\text{cvx}}(\boldsymbol{\lambda})$ is convex and continuously differentiable at any $\boldsymbol{\lambda}\in\mathcal{R}^N$. Moreover, its gradient $\nabla\bar{g}_{\text{cvx}}(\boldsymbol{\lambda}) = \sum_{k\in\mathcal{K}} \bar{\mathbf{s}}_{k,\text{cvx}}(\boldsymbol{\lambda}) - \mathbf{P}^{\text{tot}}$ is Lipschitz continuous with Lipschitz constant $L_c = \sum_{k\in\mathcal{K}} \frac{1}{c\sigma_{\mathcal{S}_k}}$. The following inequalities also hold:

$$\bar{g}_{\text{cvx}}(\boldsymbol{\lambda}) \le g_{\text{cvx}}(\boldsymbol{\lambda}) \le \bar{g}_{\text{cvx}}(\boldsymbol{\lambda}) + c\sum_{k\in\mathcal{K}} D_{\mathcal{S}_k} \qquad \boldsymbol{\lambda}\in\mathcal{R}^N \tag{20}$$

The addition of the prox-functions thus leads to a convex differentiable dual function with Lipschitz continuous gradient. Now instead of solving the original dual problem (12), we focus on the problem:

$$\min_{\boldsymbol{\lambda}\ge 0} \bar{g}_{\text{cvx}}(\boldsymbol{\lambda}) \tag{21}$$

Note that, by defining c sufficiently small in (19), the solution of (21) can be brought arbitrarily close to the solution of (12). Taking the particular structure of (21) into account, i.e. a differentiable objective function with Lipschitz continuous gradient, we propose the optimal gradient based scheme given in Algorithm 1, derived from [22], for solving (11). This algorithm will be referred to as the *improved dual decomposition algorithm* for solving the convex approximation of CA-DSB (11). The specific value for the parameters c, $D_{\mathcal{S}_k}$, L_c and $\sigma_{\mathcal{S}_k}$ are fully defined by the choice of the prox-function $d_k(\mathbf{s}_k)$ and also the required accuracy ϵ, that depends on the application. For instance, for the choice of prox-function $d_k(\mathbf{s}_k) = \frac{1}{2}\|\mathbf{s}_k\|^2$, the following simple closed-form expressions can be derived for the parameters

$$D_{\mathcal{S}_k} = \tfrac{1}{2}\sum_{n\in\mathcal{N}}(s_k^{n,\max})^2, \quad c = \frac{\epsilon}{\frac{1}{2}\sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}}(s_k^{n,\max})^2}, \quad \sigma_{\mathcal{S}_k} = 1, \quad L_c = \tfrac{K}{c}. \tag{22}$$

Note that lines 6, 7 and 8 of Algorithm 1 constitute the major part of the computational complexity, and these computations are also done by subgradient based dual decomposition algorithms. Lines 9, 10 and 11 are the extra lines of the improved approach with respect to the subgradient approach, but the computational complexity of these lines is negligible with respect to that of lines 6, 7 and 8. So in terms of complexity per Lagrange multiplier update, we can state that these are similar for both approaches.

Algorithm 1 Improved dual decomposition algorithm for solving (11) for CA-DSB

1: $i := 0$, tmp $:= 0$
2: initialize $i_{\max}$, $\boldsymbol{\lambda}^i$
3: initialize required application accuracy ϵ (= upper bound on the dual gap)
4: $c := \frac{\epsilon}{\sum_{k\in\mathcal{K}} D_{\mathcal{S}_k}}$, $L_c := \sum_{k\in\mathcal{K}} \frac{1}{c\ \sigma_{\mathcal{S}_k}}$
5: **for** $i = 0 \dots i_{\max}$ **do**
6: $\quad \forall k : \mathbf{s}_k^{i+1} = \underset{\{\mathbf{s}_k\in\mathcal{S}_k\}}{\operatorname{argmax}}\ b_{k,\mathrm{cvx}}(\mathbf{s}_k) - \sum_{n\in\mathcal{N}} \boldsymbol{\lambda}_n^i s_k^n - c d_k(\mathbf{s}_k)$
7: $\quad d\bar{g}_c^{i+1} = \sum_{k\in\mathcal{K}} \mathbf{s}_k^{i+1} - \mathbf{P}^{\mathrm{tot}}$
8: $\quad \mathbf{u}^{i+1} = [\frac{d\bar{g}_c^{i+1}}{L_c} + \boldsymbol{\lambda}^i]^+$
9: $\quad \mathrm{tmp} := \mathrm{tmp} + \frac{i+1}{2} d\bar{g}_c^{i+1}$
10: $\quad \mathbf{v}^{i+1} = [\frac{\mathrm{tmp}}{L_c}]^+$
11: $\quad \boldsymbol{\lambda}^{i+1} = \frac{i+1}{i+3}\mathbf{u}^{i+1} + \frac{2}{i+3}\mathbf{v}^{i+1}$
12: $\quad i := i + 1$
13: **end for**
14: Build $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}^{i_{\max}+1}$ and $\hat{\mathbf{s}}_k = \sum_{i=0}^{i_{\max}} \frac{2(i+1)}{(i_{\max}+1)(i_{\max}+2)} \mathbf{s}_k^{i+1}$

The remaining issue is to prove that $\hat{\mathbf{s}}_k, k \in \mathcal{K}$, i.e. the result of Algorithm 1, converges to an ϵ-optimal solution in $i_{\max}$ iterations where $i_{\max}$ is of the order $\mathcal{O}(\frac{1}{\epsilon})$. For this we define the following lemmas that will be used in the sequel.

Lemma 1: For any $\mathbf{y} \in \mathcal{R}^N$ and $\mathbf{z} \geq 0$, the following inequality holds[2]:

$$\mathbf{y}^T\mathbf{z} \ \leq \|[\mathbf{y}]^+\|\|\mathbf{z}\| \tag{23}$$

Proof: Let us define the index sets $\mathcal{I}^- = \{i \in \{1\dots n\} : y_i < 0\}$ and $\mathcal{I}^+ = \{i \in \{1\dots n\} : y_i \geq 0\}$.

[2]For the sake of an easy exposition we consider in the paper only the Euclidian norm $\|\cdot\|$, although other norms can also be used (see [22] for a detailed exposition).

Then,

$$\mathbf{y}^T\mathbf{z} = \sum_{i\in\mathcal{I}^-} y_i z_i + \sum_{i\in\mathcal{I}^+} y_i z_i \leq \sum_{i\in\mathcal{I}^+} y_i z_i = ([\mathbf{y}]^+)^T\mathbf{z} \leq \|[\mathbf{y}]^+\|\|\mathbf{z}\|.$$

■

The following lemma provides a lower bound for the primal gap, $f^*_{\rm cvx} - \sum_{k\in\mathcal{K}} b_{k,\rm cvx}(\hat{\mathbf{s}}_k)$, of (11):

Lemma 2: Let $\boldsymbol{\lambda}^*$ be any optimal Lagrange multiplier, then for any $\hat{\mathbf{s}}_k \in \mathcal{S}_k, k \in \mathcal{K}$, the following lower bound on the primal gap holds:

$$f^*_{\rm cvx} - \sum_{k\in\mathcal{K}} b_{k,\rm cvx}(\hat{\mathbf{s}}_k) \geq -\|\boldsymbol{\lambda}^*\|\|[\sum_{k\in\mathcal{K}} \hat{\mathbf{s}}_k - \mathbf{P}^{\rm tot}]^+\| \tag{24}$$

Proof: From the assumptions of the lemma we have

$$f^*_{\rm cvx} = \max_{\{\mathbf{s}_k\in\mathcal{S}_k, k\in\mathcal{K}\}} \sum_{k\in\mathcal{K}} b_{k,\rm cvx}(\mathbf{s}_k) - \boldsymbol{\lambda}^{*T}(\sum_{k\in\mathcal{K}} \mathbf{s}_k - \mathbf{P}^{\rm tot}) \geq \sum_{k\in\mathcal{K}} b_{k,\rm cvx}(\hat{\mathbf{s}}_k) - \boldsymbol{\lambda}^{*T}(\sum_{k\in\mathcal{K}} \hat{\mathbf{s}}_k - \mathbf{P}^{\rm tot}) \tag{25}$$

and then (24) is obtained by applying Lemma 1.

■

From Lemma 2 it follows that if $\|[\sum_{k\in\mathcal{K}} \hat{\mathbf{s}}_k - \mathbf{P}^{\rm tot}]^+\| \leq \epsilon_c$, then the primal gap is bounded, i.e. for all $\hat{\boldsymbol{\lambda}} \in \mathcal{R}^N_+$

$$-\epsilon_c\|\boldsymbol{\lambda}^*\| \leq f^*_{\rm cvx} - \sum_{k\in\mathcal{K}} b_{k,\rm cvx}(\hat{\mathbf{s}}_k) \leq g_{\rm cvx}(\hat{\boldsymbol{\lambda}}) - \sum_{k\in\mathcal{K}} b_{k,\rm cvx}(\hat{\mathbf{s}}_k). \tag{26}$$

Therefore, if we are able to derive an upper bound ϵ for the dual gap, namely $g_{\rm cvx}(\hat{\boldsymbol{\lambda}}) - \sum_{k\in\mathcal{K}} b_{k,\rm cvx}(\hat{\mathbf{s}}_k)$, and an upper bound ϵ_c for the coupling constraints for some given $\hat{\boldsymbol{\lambda}}$ (≥ 0) and $\hat{\mathbf{s}}_k \in \mathcal{S}_k, \forall k$, then we can conclude that $\hat{\mathbf{s}}_k$ is an (ϵ, ϵ_c)-solution for (11) (since in this case $-\epsilon_c\|\boldsymbol{\lambda}^*\| \leq f^*_{\rm cvx} - \sum_{k\in\mathcal{K}} b_{k,\rm cvx}(\hat{\mathbf{s}}_k) \leq \epsilon$). The next theorem derives these upper bounds for Algorithm 1 and provides a concrete value for c.

Theorem 2: Let $\boldsymbol{\lambda}^*$ be an optimal Lagrange multiplier, taking $c = \frac{\epsilon}{\sum_{k\in\mathcal{K}} D_{\mathcal{S}_k}}$ and $i_{\max} + 1 = 2\sqrt{(\sum_k \frac{1}{\sigma_{\mathcal{S}_k}})(\sum_k D_{\mathcal{S}_k})}\frac{1}{\epsilon}$, then after $i_{\max}$ iterations Algorithm 1 obtains an approximate solution $\hat{\mathbf{s}}_k, k \in \mathcal{K}$, to the convex approximation (11) with a duality gap less than ϵ, i.e.

$$g_{\rm cvx}(\hat{\boldsymbol{\lambda}}) - \sum_{k\in\mathcal{K}} b_{k,\rm cvx}(\hat{\mathbf{s}}_k) \leq \epsilon, \tag{27}$$

and the constraints satisfy

$$\|[\sum_k \hat{\mathbf{s}}_k - \mathbf{P}^{\rm tot}]^+\| \leq \epsilon(\|\boldsymbol{\lambda}^*\| + \sqrt{\|\boldsymbol{\lambda}^*\|^2 + 2}) \tag{28}$$

Proof: Using a similar reasoning as in Theorem 3.4 in [22] we can show that for any c the following inequality holds:

$$\bar{g}_{\rm cvx}(\hat{\boldsymbol{\lambda}}) \leq \min_{\boldsymbol{\lambda}\geq 0} \big\{ \frac{2L_c}{(i_{\max}+1)^2}\|\boldsymbol{\lambda}\|^2 + \sum_{i=0}^{i_{\max}} \frac{2(i+1)}{(i_{\max}+1)(i_{\max}+2)}[\bar{g}_{\rm cvx}(\boldsymbol{\lambda}^i) + (\nabla\bar{g}_{\rm cvx}(\boldsymbol{\lambda}^i))^T(\boldsymbol{\lambda} - \boldsymbol{\lambda}^i)] \big\}$$

Replacing $\bar{g}_{\text{cvx}}(\boldsymbol{\lambda}^i)$ and $\nabla\bar{g}_{\text{cvx}}(\boldsymbol{\lambda}^i)$ by their expressions given in (18) and Theorem 1, respectively, and taking into account that the functions $b_{k,\text{cvx}}$ are concave, we obtain the following inequality:

$$\begin{aligned} g_{\text{cvx}}(\hat{\boldsymbol{\lambda}}) - \sum_{k\in\mathcal{K}} b_{k,\text{cvx}}(\hat{\mathbf{s}}_k) &\leq c(\sum_{k\in\mathcal{K}} D_{\mathcal{S}_k}) + \min_{\boldsymbol{\lambda}\geq 0}\big\{\frac{2L_c}{(i_{\max}+1)^2}\|\boldsymbol{\lambda}\|^2 - \langle\boldsymbol{\lambda}, \sum_k \hat{\mathbf{s}}_k - \mathbf{P}^{\text{tot}}\rangle\big\} \\ &= c(\sum_{k\in\mathcal{K}} D_{\mathcal{S}_k}) - \frac{(i_{\max}+1)^2}{8L_c}\|[\sum_k \hat{\mathbf{s}}_k - \mathbf{P}^{\text{tot}}]^+\|^2 \leq c(\sum_{k\in\mathcal{K}} D_{\mathcal{S}_k}). \end{aligned}$$

By taking $c = \frac{\epsilon}{\sum_{k\in\mathcal{K}} D_{\mathcal{S}_k}}$, we obtain (27). For the constraints using Lemma 2 and the previous inequality we get that $\|[\sum_k \hat{\mathbf{s}}_k - \mathbf{P}^{\text{tot}}]^+\|$ satisfies the second order inequality in y: $\frac{(i_{\max}+1)^2}{8L_c}y^2 - \|\boldsymbol{\lambda}^*\|y - \epsilon \leq 0$. Therefore, $\|[\sum_k \hat{\mathbf{s}}_k - \mathbf{P}^{\text{tot}}]^+\|$ must be less than the largest root of the corresponding second-order equation, i.e.

$$\|[\sum_k \hat{\mathbf{s}}_k - \mathbf{P}^{\text{tot}}]^+\| \leq \big(\|\boldsymbol{\lambda}^*\| + \sqrt{\|\boldsymbol{\lambda}^*\|^2 + \frac{\epsilon(i_{\max}+1)^2}{2L_c}}\big)\frac{4L_c}{(i_{\max}+1)^2}.$$

By taking $i_{\max} = 2\sqrt{(\sum_k \frac{1}{\sigma_{\mathcal{S}_k}})(\sum_k D_{\mathcal{S}_k})\frac{1}{\epsilon}} - 1$, we obtain (28). ■

From Theorem 2 we can conclude that by taking $c = \frac{\epsilon}{\sum_{k\in\mathcal{K}} D_{\mathcal{S}_k}}$, Algorithm 1 converges to a solution with duality gap less than ϵ and the constraints violation satisfy $\|[\sum_k \hat{\mathbf{s}}_k - \mathbf{P}^{\text{tot}}]^+\| \leq \epsilon(\|\boldsymbol{\lambda}^*\| + \sqrt{\|\boldsymbol{\lambda}^*\|^2 + 2})$ after $i_{\max} = 2\sqrt{(\sum_k \frac{1}{\sigma_{\mathcal{S}_k}})(\sum_k D_{\mathcal{S}_k})\frac{1}{\epsilon}} - 1$ iterations, i.e. the convergence speed is of the order $\mathcal{O}(\frac{1}{\epsilon})$.

Note that Algorithm 1 provides a fully automatic approach. Once the required application accuracy ϵ and the particular prox-function $d_k(\mathbf{s}_k)$ are defined, all parameters are fixed. The algorithm then automatically updates its stepsize so as to converge fast to the optimal dual value within the specified accuracy. It does not require any stepsize tuning, which otherwise is known to be a very difficult and crucial process. Finally note that combining this algorithm with an outer loop that iteratively updates the convex approximations leads to an overall procedure that converges to a local maximizer of the nonconvex problem cWRS [28] [13]. The extension of CA-DSB with the improved dual decomposition approach will be referred to as Improved CA-DSB (I-CA-DSB).

A final remark on Algorithm 1 is that the independent convex per-tone problems (line 6 of Algorithm 1) are slightly modified with respect to the standard per-tone problems for CA-DSB. This is a consequence of the addition of the extra prox-function term. One can use state-of-the-art iterative methods (e.g. Newton's method) to solve this convex subproblem with guaranteed convergence. An alternative consists in using an iterative fixed point update approach, which is shown to work well, with very small complexity, and is easily extended to distributed implementation by using a protocol [16] [13]. We propose a modified fixed point update formula for the transmit powers s_k^n used by CA-DSB, so as to take the extra prox-term into account. Following the same procedure as explained in [13], consisting of a fixed point reformulation

of the corresponding KKT stationarity condition of (12), and for the choice of prox-function $d_k(\mathbf{s}_k) = \frac{1}{2}\|\mathbf{s}_k\|^2$, we obtain the following transmit power update formula, that only differs in the presence of the term PROX:

$$s_k^n = \left[\frac{w_n f_s / \log(2)}{\lambda_n + \underbrace{2cs_k^n}_{\text{PROX}} + \sum_{m \neq n} \omega_m f_s a_k^{n,m} - \sum_{m \neq n} \frac{w_m f_s \Gamma |h_k^{m,n}|^2 / \log(2)}{\sum_p |\tilde{h}_k^{m,p}|^2 s_k^p + \Gamma \sigma_k^m}} - \frac{\sum_{m \neq n} \Gamma |h_k^{n,m}|^2 s_k^m + \Gamma \sigma_k^n}{|h_k^{n,n}|^2} \right]_0^{s_k^{n,\max}} . \quad (29)$$

Providing convergence conditions for this type of iterative fixed point updates is outside the scope of this paper. In [13], [18], [19], convergence is proven under certain conditions, and demonstrated for realistic DSL scenarios. This leads to an alternative and fast way of implementing line 6 of Algorithm 1, as specified in Algorithm 2. The number of iterations in line 2 is typically fixed at 3.

Algorithm 2 Iterative fixed point update approach for solving line 6 of Algorithm 1

```
1: for k = 1...K do
2:   for iterations do
3:     for n = 1...N do
4:       s_k^n =(29)
5:     end for
6:   end for
7: end for
```

As mentioned, although the improved dual decomposition approach has been elaborated for CA-DSB, it can similarly be applied to other DSM algorithms based on iterative convex approximations, like for instance SCALE, with a similar speed up of convergence. In this case the prox-function can be taken as $d_k(\mathbf{s}_k) = \frac{1}{2}\|\mathbf{s}_k\|^2$, resulting in concrete values for c, $i_{\max}$ and L_c. The extension of SCALE with the improved dual decomposition approach will be referred to as Improved SCALE (I-SCALE).

Finally, we would like to remark that the idea of adding a strictly convex term to a non-strictly convex objective to improve the sensitivity is a known technique that has been proposed before [32]–[35]. Algorithm 1 however extends this with automatic tuning strategies, concrete convergence orders, and an optimal gradient based method by following the approach of [22], which is here particularly elaborated for DSL DSM.

B. An improved dual decomposition approach for direct DSM algorithms

In this section we extend the improved dual decomposition approach to direct DSM algorithms such as OSB, ISB, ASB, (MS-)DSB, MIW, etc, corresponding to the structure visualized in Figure 1. Using

a similar trick as in Section IV-A, we define a smoothed dual function $\bar{g}(\boldsymbol{\lambda})$ as follows

$$\bar{g}(\boldsymbol{\lambda}) = \max_{\{\mathbf{s}_k \in \mathcal{S}_k, k \in \mathcal{K}\}} \sum_{k\in\mathcal{K}} b_k(\mathbf{s}_k) - \sum_{k\in\mathcal{K}}\sum_{n\in\mathcal{N}} \lambda_n s_k^n + \sum_{n\in\mathcal{N}} \lambda_n P^{n,\mathrm{tot}} - \sum_{k\in\mathcal{K}} c d_k(\mathbf{s}_k) \tag{30}$$

where $d_k(\mathbf{s}_k)$ is a prox-function and c is a positive smoothness parameter.

Note that by defining parameter c to a sufficiently small value, the smoothed dual function $\bar{g}(\boldsymbol{\lambda})$ can be brought arbitrarily close to the original dual function $g(\boldsymbol{\lambda})$, i.e. $\bar{g}(\boldsymbol{\lambda}) \approx g(\boldsymbol{\lambda})$. Based on the obtained smoothed dual function $\bar{g}(\boldsymbol{\lambda})$, we propose the *improved dual decomposition approach for direct DSM algorithms* as shown in Algorithm 3, where line 6 corresponds to solving the following optimization problem:

$$\begin{aligned} \tilde{\mathbf{s}}_k(\boldsymbol{\lambda}) = \quad & \operatorname*{argmax}_{\mathbf{s}_k} \quad b_k(\mathbf{s}_k) - \sum_{n\in\mathcal{N}} \lambda_n s_k^n - c d_k(\mathbf{s}_k) \\ & \text{s.t.} \qquad 0 \le s_k^n \le s_k^{n,\mathrm{mask}}, \quad n \in \mathcal{N}. \end{aligned} \tag{31}$$

For the concrete choice of $d_k(\mathbf{s}_k) = \frac{1}{2}\|\mathbf{s}_k\|^2$, the values for c, L_c, D_{S_k} and σ_{S_k} can be computed by the same simple closed-form expressions as given in (22).

Algorithm 3 Improved dual decomposition approach for direct DSM algorithms

1: $i := 0$, tmp $:= 0$
2: initialize $\boldsymbol{\lambda}^i$ and ϵ_a (desired accuracy on per-user total powers)
3: initialize required application accuracy ϵ
4: $c := \frac{\epsilon}{\sum_{k\in\mathcal{K}} D_{\mathcal{S}_k}}$, $L_c := \sum_{k\in\mathcal{K}} \frac{1}{c\ \sigma_{\mathcal{S}_k}}$
5: **while** $\exists n : (\mathrm{abs}(\lambda_n^i(\sum_{k\in\mathcal{K}} s_k^n - P^{n,\mathrm{tot}})) \ge \epsilon_a)$ **do**
6: $\quad \forall k : \mathbf{s}_k^{i+1} = \tilde{\mathbf{s}}_k(\boldsymbol{\lambda}^i)$ obtained by solving (31)
7: $\quad dg^{i+1} = \sum_{k\in\mathcal{K}} \mathbf{s}_k^{i+1} - \mathbf{P}^{\mathrm{tot}}$
8: $\quad \mathbf{u}^{i+1} = [\frac{dg^{i+1}}{L_c} + \boldsymbol{\lambda}^i]^+$
9: $\quad \mathrm{tmp} := \mathrm{tmp} + \frac{i+1}{2} dg^{i+1}$
10: $\quad \mathbf{v}^{i+1} = [\frac{\mathrm{tmp}}{L_c}]^+$
11: $\quad \boldsymbol{\lambda}^{i+1} = \frac{i+1}{i+3}\mathbf{u}^{i+1} + \frac{2}{i+3}\mathbf{v}^{i+1}$
12: $\quad i := i + 1$
13: **end while**
14: Build $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}^i$ and $\hat{\mathbf{s}}_k = \mathbf{s}_k^i, \forall k \in \mathcal{K}$

Algorithm 3 uses a similar optimal gradient based scheme on the smoothed dual function as in Algorithm 1. Again no stepsize tuning is needed. Besides the improved updating procedure for the Lagrange multipliers (lines 7-11), it involves a slightly different decomposed per-tone problem (31)

(line 6). This can be solved by using a discrete exhaustive search similar to OSB, a discrete coordinate descent method similar to ISB, or a KKT system approach similar to DSB/MIW/MS-DSB using (29), where $a_k^{n,m} = \frac{\Gamma|h_k^{m,n}|^2/\log(2)}{\sum_{p\neq m}\Gamma|h_k^{m,p}|^2 s_k^p + \Gamma\sigma_k^m}$ [13]. One can also use a virtual reference length approach similar to ASB, ASB2. Note that for ASB, and when using $d_k(\mathbf{s}_k) = \|\mathbf{s}_k\|^2$, this increases the complexity as a polynomial equation of degree 4 is then to be solved instead of a cubic equation. Depending on the choice of the algorithm for solving the per-tone problem, there will be a trade-off in complexity versus performance [13]. We will again add the prefix 'I-' to refer to these algorithms using the improved dual decomposition approach, i.e. I-OSB,I-ISB, I-DSB/MIW, I-MS-DSB, I-ASB.

The main difference of Algorithm 3 is that line 6 now involves K nonconvex optimization problems, while line 6 of Algorithm 1 involves K (strong) convex optimization problems. As a consequence, the smoothed dual function $\bar{g}(\boldsymbol{\lambda})$ is not necessarily differentiable and its gradient is not necessarily Lipschitz continuous. More specifically, this is the case when $\bar{g}(\boldsymbol{\lambda})$ has multiple globally optimal solutions for a given Lagrange multiplier $\boldsymbol{\lambda}$. This non-uniqueness problem becomes a true problem only for a particular type of scenarios, namely symmetric DSL scenarios with large crosstalk, where multiple adjacent tones have multiple globally optimal solutions. This will be analyzed and discussed in more detail in Section V. For these particular scenarios the worst case convergence of order $\mathcal{O}(\frac{1}{\epsilon})$ can not be guaranteed, as in Theorem 2, but still we can expect an improved convergence behaviour with respect to the standard subgradient approach. Except for these specific cases, and so for most practical DSL scenarios, the smoothed dual function $\bar{g}(\boldsymbol{\lambda})$ will be differentiable and Lipschitz continuous, and so a worst case convergence speed of $\mathcal{O}(\frac{1}{\epsilon})$ is guaranteed. For instance, in [36] conditions on the channel and noise parameters were given under which cWRS can be "convexified". For these conditions, differentiability and Lipschitz continuity holds for $\bar{g}(\boldsymbol{\lambda})$ and so application of Algorithm 3 will provide a worst case convergence of $\mathcal{O}(\frac{1}{\epsilon})$.

V. An interleaving procedure for recovering the primal solution from the dual solution

The subgradient based dual decomposition approach for solving problem cWRS (4) as well as the improved dual decomposition approach presented in Sections IV-A and IV-B, converge to the optimal dual variables. However, because of the nonconvex nature of cWRS, extra care must be taken when recovering the optimal primal solution, i.e. optimal transmit powers $\mathbf{s}_k^*, k \in \mathcal{K}$, for (4), from the optimal dual variables $\boldsymbol{\lambda}^*$, as was also mentioned in [5] [32]. The fact that the objective function of cWRS is not strictly concave, can result in cases where the optimal $\mathbf{s}_k(\boldsymbol{\lambda}^*), k \in \mathcal{K}$, that solves (7) is not unique, leading to multiple solutions $\mathbf{s}_k(\boldsymbol{\lambda}^*), k \in \mathcal{K}$, for given optimal dual variables $\boldsymbol{\lambda}^*$. Formally this can be

expressed as follows:

$$\{\mathbf{s}_k(\boldsymbol{\lambda}^*), k \in \mathcal{K}\} \in \mathcal{B} = \{(\tilde{\mathbf{s}}_{k,1}, k \in \mathcal{K}), \ldots, (\tilde{\mathbf{s}}_{k,|\mathcal{B}|}, k \in \mathcal{K})\}$$
$$\text{with } \tilde{\mathbf{s}}_{k,m} \in \mathcal{S}_k, k \in \mathcal{K}, \text{ and } \mathcal{L}(\tilde{\mathbf{s}}_{k,m}, k \in \mathcal{K}, \boldsymbol{\lambda}^*) = \max_{\{\mathbf{s}_k \in \mathcal{S}_k, k \in \mathcal{K}\}} \mathcal{L}(\mathbf{s}_k, k \in \mathcal{K}, \boldsymbol{\lambda}^*), \quad m \in \{1, \ldots, |\mathcal{B}|\}, \tag{32}$$

where the cardinality of set $\mathcal{B}$ is larger than 1, i.e. $|\mathcal{B}| > 1$. It is important to note that the elements of $\mathcal{B}$ are not necessarily solutions to (4), i.e. they do not necessarily satisfy the user total power constraints (3). However, there exists at least one element in set $\mathcal{B}$ that does satisfy the total power constraints [5]. In order to obtain convergence to a primal optimal solution for (4) in the case that $|\mathcal{B}| > 1$, the dual decomposition approach has to be extended with an extra procedure that chooses an element out of set $\mathcal{B}$ that satisfies the user total power constraints.

A simple example may be given to clarify this issue; suppose we have a DSL scenario consisting of two users ($N = 2$) and two tones ($K = 2$), where the channel matrices (direct and crosstalk components) and noise components for the two tones are the same, i.e. $\mathbf{H}_1 = \mathbf{H}_2$ and $\sigma_1^n = \sigma_2^n,\ n \in \mathcal{N}$, and the weights are also the same $w_1 = w_2$. Furthermore suppose the crosstalk components are very large. In this case, there will be only one user active on each tone [37]. Finally suppose that the optimal dual variables λ_1^*, λ_2^*, where $\lambda_1^* = \lambda_2^*$, are given and the total power constraints are $P^n \leq \text{ON}$, where ON is a fixed power level. For this setup there will be 4 possible solutions to (7), namely $\{s_1^1 = \text{ON}, s_2^1 = \text{ON}, s_1^2 = 0, s_2^2 = 0\}, \{s_1^1 = 0, s_2^1 = 0, s_1^2 = \text{ON}, s_2^2 = \text{ON}\}, \{s_1^1 = \text{ON}, s_2^1 = 0, s_1^2 = 0, s_2^2 = \text{ON}\}, \{s_1^1 = 0, s_2^1 = \text{ON}, s_1^2 = \text{ON}, s_2^2 = 0\}$. Note that all these solutions correspond to exactly the same objective value but only the last two solutions are primal optimal solutions as they satisfy the user total power constraints. Typical DSM algorithm implementations, however, have a fixed exhaustive search order or iteration order over tones so that one of the two first solutions may be selected and, as a consequence, these algorithms will not provide the primal optimal solutions of (4). To obtain convergence to the optimal primal variables of (4) an extra procedure should be added to the dual decomposition approach.

Note that the above problem is practically only relevant when the phenomenon of non-unique globally optimal solutions $\mathbf{s}_k(\boldsymbol{\lambda}^*)$ occurs at many tones. This is the case for DSL scenarios that have a subset of strong symmetric crosstalkers with equal line lengths, i.e. lines that generate the same interference to their environment over multiple tones k, with equal weights w_n and user total power constraints $P^{n,\text{tot}}$. Here, we can have many adjacent tones with multiple globally optimal solutions, namely where only one of the subset of strong crosstalkers is active [37]. If no special care is taken when recovering the primal transmit powers, this can lead to extremely slow convergence or even no convergence at all for these scenarios. More specifically, a fixed exhaustive search order or iteration order in typical DSM algorithm implementations will choose the same strong crosstalker over all competing tones, instead of equally dividing the resources over the competing users.

To overcome this problem we propose a very simple, but effective, interleaving procedure, that can be combined with Algorithm 3. More specifically this solution consists of alternatingly on a per-tone basis, giving priority to the globally optimal solution that corresponds to a different active strong crosstalker of the symmetric subset. This interleaving procedure replaces line 6 of Algorithm 3 with the following:

$$\forall k : \begin{cases} \mathcal{C}_k & = \{\text{all globally optimal solutions } \tilde{\mathbf{s}}_{\mathrm{k}}(\boldsymbol{\lambda}) \text{ of } (31) \text{ for given } \boldsymbol{\lambda}\}, \\ & = \{\mathcal{C}_k(1), \ldots, \mathcal{C}_k(|\mathcal{C}_k|)\}, \\ \text{index} & = \mathrm{rem}(k, |\mathcal{C}_k|) + 1, \\ \mathbf{s}_k^{i+1} & = \mathcal{C}_k(\text{index}), \end{cases} \tag{33}$$

where 'rem($k, |\mathcal{C}_k|$)' refers to the remainder after dividing k by $|\mathcal{C}_k|$. As the suggested solution requires that all globally optimal solutions in the first step of (33) actually be computed, it should be combined with algorithms for the per-tone nonconvex problem that indeed compute all these solutions such as OSB with a fixed order exhaustive search for all tones or a multiple starting point approach such as MS-DSB with a fixed iteration order for all tones.

In the simulation Section VI, it will be demonstrated how the usage of (33) significantly improves the robustness of the dual decomposition approach for cWRS.

Remark: The above mentioned non-uniqueness also has an impact on the Lipschitz continuity condition of the smoothed gradient. More specifically this condition reduces to [22]:

$$\|\sum_{k\in\mathcal{K}} \tilde{\mathbf{s}}_k(\boldsymbol{\lambda}) - \sum_{k\in\mathcal{K}} \tilde{\mathbf{s}}_k(\boldsymbol{\mu})\|^2 \leq L_c \|\boldsymbol{\lambda} - \boldsymbol{\mu}\|^2 \text{ with } L_c < \infty \tag{34}$$

For the above two-user two-tone symmetric strong crosstalk example, this condition does not hold. This can be shown as follows. Let us compare two cases: (1) optimal dual variables $(\lambda_1^*, \lambda_2^* + \mu)$ corresponding to primal variables $\{s_1^1 = \mathrm{ON}, s_2^1 = \mathrm{ON}, s_1^2 = 0, s_2^2 = 0\}$, (2) optimal dual variables $(\lambda_1^* + \mu, \lambda_2^*)$ corresponding to primal variables $\{s_1^1 = 0, s_2^1 = 0, s_1^2 = \mathrm{ON}, s_2^2 = \mathrm{ON}\}$, where $\mu \geq 0$. For very small μ these two cases have only slightly different dual variables but completely different primal variables. So a small change in Lagrange multipliers can lead to a large change in primal variables. This means that for these specific cases Lipschitz continuity (34) is not satisfied and so the convergence speed will be worse than $O(\frac{1}{\epsilon})$. However adding the interleaving trick alleviates this problem, as will be demonstrated in Section VI.

Remark: In [38], a randomized LP-based algorithm is proposed for recovering the primal variables from the dual optimum. This algorithm is however designed for the cWRS problem with extra FDMA constraint, which is a simpler problem that can be solved using polynomial time algorithms. Furthermore the algorithm assumes the application of time-sharing.

VI. Simulation results

In this section, simulation results are shown that compare the performance of the improved dual decomposition approach with respect to the subgradient based dual decomposition approach. More specifically, in Section VI-A we demonstrate for a DSM algorithm based on iterative convex approximations (CA-DSB) the very fast convergence of the improved dual decomposition approach with respect to the subgradient approach with different stepsize tuning strategies. In Section VI-B we demonstrate how the improved dual decomposition approach in combination with a direct DSM algorithm (MS-DSB) succeeds in providing much faster convergence than with the subgradient based dual decomposition approach. Furthermore the convergence improvement for the interleaving procedure presented in Section V is demonstrated.

The following parameter settings are used for the simulated DSL scenarios. The twisted pair lines have a diameter of 0.5 mm (24 AWG). The maximum per-user total transmit power is 11.5 dBm for the VDSL scenarios and 20.4 dBm for the ADSL scenarios. The SNR gap Γ is 12.9 dB, corresponding to a coding gain of 3 dB, a noise margin of 6 dB, and a target symbol error probability of 10^{-7}. The tone spacing Δ_f is 4.3125 kHz. The DMT symbol rate f_s is 4 kHz. Furthermore the prox-function $d_k(\mathbf{s}_k) = \frac{1}{2}\|\mathbf{s}_k\|^2$ (with convexity parameter equal to 1) is used for all simulations, which is an appropriate prox-function for box constraints [24].

A. Convergence speed up for iterative convex approximation based DSM

A first DSL scenario is shown in Figure 3. This is a so-called near-far scenario which is known to be challenging, where DSM can make a substantial difference. For this scenario, we compare the convergence behaviour for the improved approach for CA-DSB (Algorithm 1) and the standard subgradient based dual decomposition approach for CA-DSB with different stepsize updating strategies, where the stepsize is δ in (10). The first stepsize update strategy is one that is guaranteed to converge, namely $\delta = q/i$, where q is the initial stepsize and i is the iteration counter [5]. We will refer to this with 'decreasing step' with some value for q. The second stepsize update strategy is the fixed stepsize, namely $\delta = q$. Note that this scheme is not guaranteed to converge for large values of q. We will refer to this update strategy with 'fixed step'. Furthermore the target accuracy on the dual gap is specified as 0.5%, which has to be seen as an application requirement and not as a tuning parameter. Finally note that we use an iterative fixed point update approach with the same number of inner iterations to solve the per-tone problems for both the improved and the subgradient approaches. The results are shown in Figures 4 and 5. It can be observed that different initial stepsizes lead to a different convergence behaviour for the subgradient approaches, and this is generally difficult to tune. The subgradient scheme with decreasing stepsize is generally much slower in convergence. The subgradient approach with fixed stepsize is better but can become instable for large values of q as shown in Figure 5. Stepsize tuning is crucial for these schemes.

In contrary, the improved scheme automatically tunes its stepsize and converges very rapidly in only 40 iterations. Finally note that the curve for the subgradient scheme with fixed step and $q = 2500$ initially has a steeper curve with respect to the improved scheme, but as it approaches the optimal dual value its slope decreases fast and it converges only after 90 iterations.

Finally, we remark that different values for the application accuracy ϵ, i.e. the upper bound on the dual gap, lead to a different number of iterations to converge to that accuracy. For Figure 4, we set ϵ to correspond to 0.5% accuracy, which is sufficiently accurate in practice. In Table I, we also show the relation between the specified accuracy (which is actually defined by the application) and the number of iterations so as to converge to that accuracy for the improved scheme. One can see that for very small specified accuracies the improved scheme requires a very small number of iterations to converge to that accuracy, e.g. in 84 iterations it converges to a specified accuracy of 0.1% of the optimum.

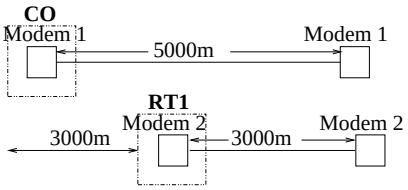


Fig. 3. 2-user near-far ADSL downstream scenario

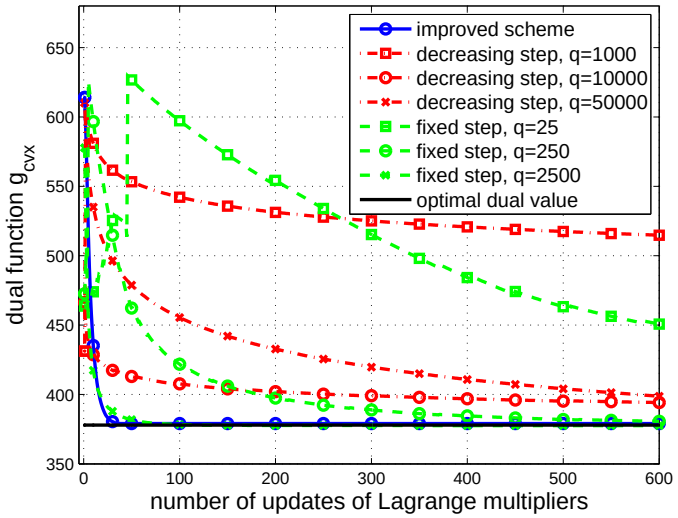


Fig. 4. Comparison of convergence behaviour between subgradient dual decomposition approach, with different stepsize update strategies, and the improved dual decomposition approach, for CA-DSB

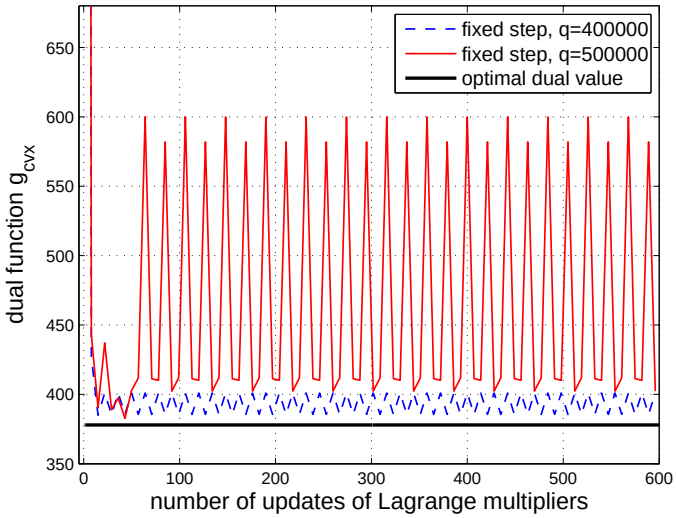


Fig. 5. Convergence behaviour of subgradient dual decomposition approach, with large fixed stepsizes, for CA-DSB

TABLE I

THE RELATION BETWEEN THE ACCURACY ON THE DUAL GAP AND THE NUMBER OF ITERATIONS SO AS TO CONVERGE TO THAT ACCURACY FOR THE IMPROVED DUAL DECOMPOSITION APPROACH

Accuracy on dual gap	Number of Iterations
0.60%	22
0.39%	39
0.10%	84
0.05%	110
0.01%	250

B. Convergence speed up for direct DSM

It was shown in [6] that for direct DSM algorithms the subgradient based dual decomposition approach with a particular stepsize selection procedure works well for ADSL scenarios, i.e. there are typically only 50-100 subgradient iterations needed to converge to the optimal dual variables. However for multi-user VDSL scenarios, which use a much larger frequency range and have to cope with significantly more crosstalk interference, existing subgradient approaches [5] [6] are found to have significant convergence problems. We will focus on such VDSL scenarios and demonstrate how the improved approach succeeds in providing much faster convergence.

The different VDSL scenarios are shown in Figures 6, 7 and 9, i.e. four-user VDSL upstream, six-user

VDSL upstream, and six-user VDSL upstream scenario with a subset of strong symmetric crosstalkers, respectively. The weights w_n are chosen equal for all users n, namely $w_n = 1/N$. Note that we used the multiple starting point procedure MS-DSB to solve the nonconvex per-tone problems for the subgradient based dual decomposition approach as well as the improved dual decomposition approach using (29). In [13] it was shown that this procedure provides globally optimal performance for practical ADSL and VDSL scenarios.

The first scenario, shown in Figure 6, is a four-user upstream VDSL scenario, consisting of two far-users with line length 1200 m and two near-users with line length 300 m. In the higher frequency range, there is a significant crosstalk coupling. This is a near-far scenario where spectrum management is crucial as to avoid significant performance degradation for the far-end users. Note that the near-end users form a subset of strong symmetric crosstalkers, in the high frequency range. As mentioned in Section V, this can cause significant convergence problems for the dual decomposition approach. In fact, simulations show that the subgradient methods in [6] and [5] fail to converge to the dual variables, i.e. after 20000 iterations the complementarity conditions for some users are far from being satisfied. The main problem is that the stepsize selection procedure, which is a crucial component for fast convergence, is difficult to tune. For decreasing and fixed step sizes as proposed in [5], with different initial stepsizes, the procedure does not converge. For adaptive stepsizes, as proposed in [6], very small stepsizes are selected resulting in a very slow convergence (> 20000 iterations). It is observed that for some users there is a fast convergence to the corresponding complementarity conditions whereas for other users convergence is very slow. The presence of the subset of strong symmetric crosstalkers, can lead to large changes in primal variables for small changes in dual variables, as discussed in Section V, if stepsizes are not tuned carefully. The improved approach of Algorithm 3, in contrary, converges very fast to the optimal dual and primal variables. In only 100 iterations convergence is obtained, within an accuracy of 0.05%.

The second VDSL upstream scenario, shown in Figure 7, consists of six users with different line lengths. Also for this large crosstalk scenario, the standard subgradient approaches [6] [5] fail to converge to the optimal dual variables, i.e. after 10000 iterations the complementarity conditions are far from being satisfied. Similarly to the scenario of Figure 6, one can observe very different convergence behaviour for the different users to the corresponding complementarity conditions, where typically for a few users convergence is very slow. The improved dual decomposition approach however converges to the optimal dual and primal variables in only 150 iterations, within an accuracy of 0.05%. The optimal transmit powers are shown in Figure 8 for illustration.

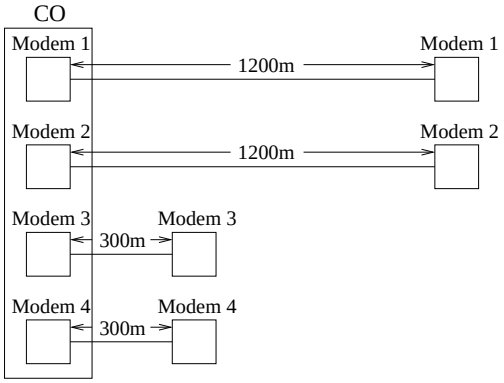


Fig. 6. 4-user VDSL upstream scenario

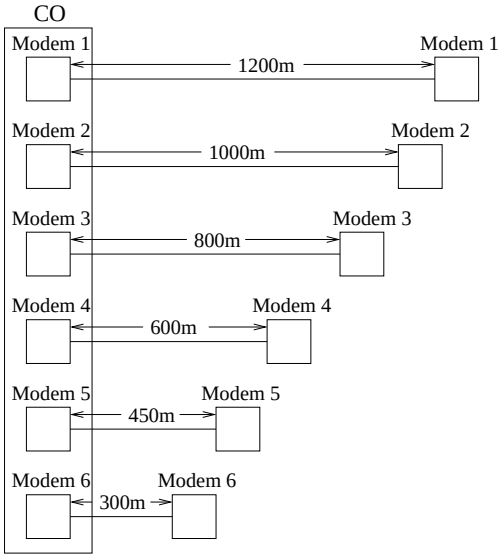


Fig. 7. 6-user VDSL upstream scenario

The VDSL upstream scenario of Figure 9 consists of a six-line cable bundle with a subset of three strong symmetric crosstalkers, namely the set of lines with length 300m. The standard subgradient approaches [6] [5] fail to converge to the optimal dual variables. The presence of the strong symmetric crosstalkers significantly slows down the convergence, as it can lead to multiple globally optimal solutions for particular values of the dual variables. Here, stepsize selection is very crucial as a small change in dual variables can lead to a large change in primal variables, as also explained in Section V. The improved dual decomposition approach converges to the optimal dual variables in only 150 iterations, but does not succeed in obtaining the primal optimal variables, because of the existence of multiple globally optimal solutions (i.e. optimal transmit powers) for optimal dual variables that do not satisfy the user total power constraints. More specifically for this scenario, for the obtained optimal dual variables, the obtained

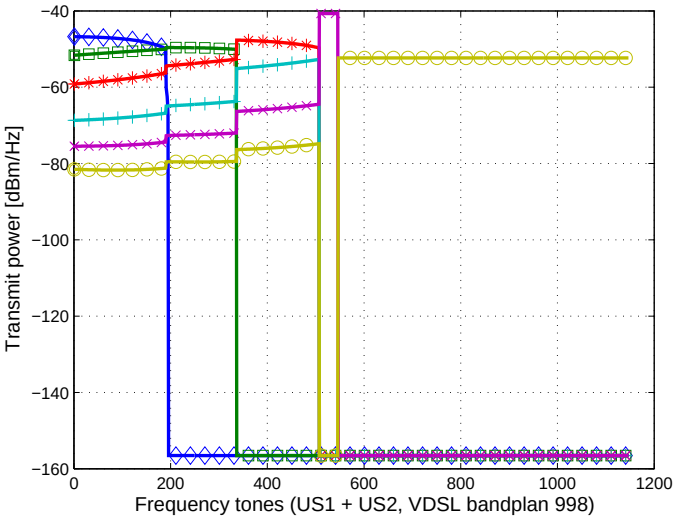


Fig. 8. Optimal transmit powers for DSL scenario of Fig. 7 obtained using the improved dual decomposition approach. Blue diamond, green square, red asterisk, cyan plus, magenta cross and yellow circle curves correspond to transmit powers of users with line length 1200m, 1000m, 800m, 600m, 450m and 300m respectively.

transmit powers jump to different solutions, with total powers $\{P^1, P^2, P^3\} = \{P^{1,\text{tot}}, P^{2,\text{tot}}, P^{3,\text{tot}}\}$, and $\{P^4, P^5, P^6\} \in \big\{\{3P^{\text{tot}}, A, A\}, \{A, 3P^{\text{tot}}, A\}, \{A, A, 3P^{\text{tot}}\}\big\}$, with A being very small. These primal solutions are shown in Figures 10(a), 10(b) and 10(c) . One can observe that in the low and medium frequency range (used tones 1-727), the users with line lengths 1200 m, 900 m and 600 m are active. In this frequency range the strong crosstalkers will back-off and transmit at small similar transmit powers corresponding to a total power equal to A. However in the high frequency range (used tones 727-1147) where the users with line lengths 1200 m, 900 m and 600 are switched off, the three strong crosstalkers will compete, where only one user can be active in each tone k because of the significant crosstalk interference [37]. As explained in Section V, typical DSM algorithm implementations will select the same active user for each of these tones, namely the user that corresponds to the smallest dual variable, where the dual variable can be seen as a penalty. So instead of dividing the total power over the three users equally, which would lead to a primal solution satisfying the per-user total power constraints, one user gets all power, leading to $P^n = 3P^{n,\text{tot}}$ for user n and $P^m = A$ for users $m \neq n$. Note that this prevents convergence to the optimal primal variables satisfying the per-user total power constraints.

However, when applying the proposed interleaving procedure (33), as proposed in Section V, together with the improved dual decomposition approach, we can observe a very fast convergence both in primal and dual variables. Convergence is achieved in only 150 iterations, within an accuracy of 0.05%. The obtained optimal transmit powers are shown in Figure 11. In the frequency range between tone 728 and

tone 1147, one can observe the interleaving effect. In Figure 12 this is zoomed in for tones 970 up to 975.

Remark: In the practical implementation the first step of the interleaving procedure is changed to 'all best solutions that are 99.9% close to each other'. This is to prevent that the procedure is only active when the dual variables are exactly the same. The overall effect of this is a negligible noise on the transmit powers as can be seen in Figure 11.

Remark: Note that applying the interleaving procedure combined with the improved dual decomposition approach for the scenarios in Figures 6 and 7, also leads to a faster convergence in both dual and primal variables.

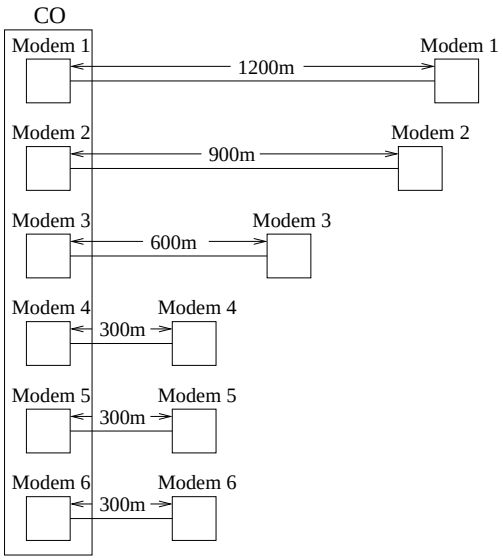
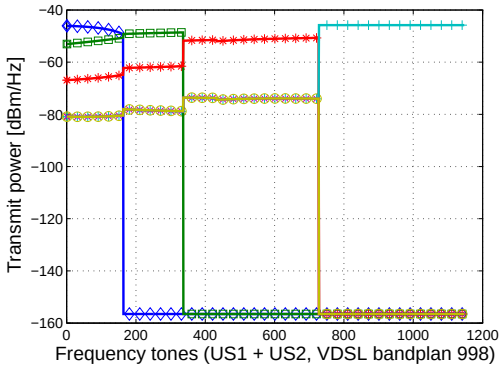


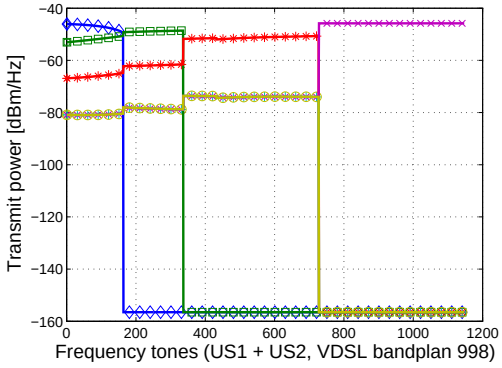
Fig. 9. 6-user VDSL upstream scenario with subset of strong symmetric crosstalkers

VII. Conclusion

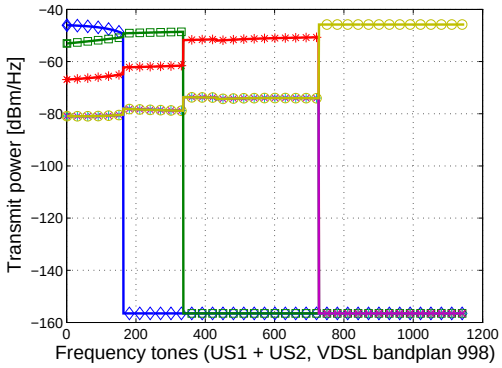
Dynamic spectrum management has been recognized as a key technology to significantly improve the performance of DSL broadband access networks by mitigating the impact of crosstalk interference. Many existing DSM algorithms use a standard subgradient based dual decomposition approach to tackle the corresponding nonconvex optimization problems. However, this standard approach is often found to lead to extremely slow convergence or even no convergence at all. Especially for multiuser VDSL scenarios with subsets of strong symmetric crosstalkers significant convergence problems are observed because (1) the stepsize selection procedure of the subgradient updates is very critical, and (2) because special care must be taken when recovering the optimal transmit powers from the optimal dual solution. This paper proposes an improved dual decomposition approach, which consists of an optimal gradient based scheme with an automatic optimal stepsize selection removing the need for a tuning strategy. With this



(a)



(b)



(c)

Fig. 10. Optimal transmit power allocations for DSL scenario of Fig. 9 for optimal dual variables $\boldsymbol{\lambda}^*$, where for subfigure 10(a) $\{P^1, P^2, P^3, P^4, P^5, P^6\} = \{P^{1,\mathrm{tot}}, P^{2,\mathrm{tot}}, P^{3,\mathrm{tot}}, 3P^{4,\mathrm{tot}}, A, A\}$ with $A << P^{4,\mathrm{tot}}$, for subfigure 10(b) $\{P^1, P^2, P^3, P^4, P^5, P^6\} = \{P^{1,\mathrm{tot}}, P^{2,\mathrm{tot}}, P^{3,\mathrm{tot}}, A, 3P^{5,\mathrm{tot}}, A\}$ with $A << P^{5,\mathrm{tot}}$, and for subfigure 10(c) $\{P^1, P^2, P^3, P^4, P^5, P^6\} = \{P^{1,\mathrm{tot}}, P^{2,\mathrm{tot}}, P^{3,\mathrm{tot}}, A, A, 3P^{6,\mathrm{tot}}\}$ with $A << P^{6,\mathrm{tot}}$. Blue diamond, green square, red asterisk, cyan plus, magenta cross, yellow circle curves correspond to transmit powers of users 1,2,3,4,5 and 6 respectively.

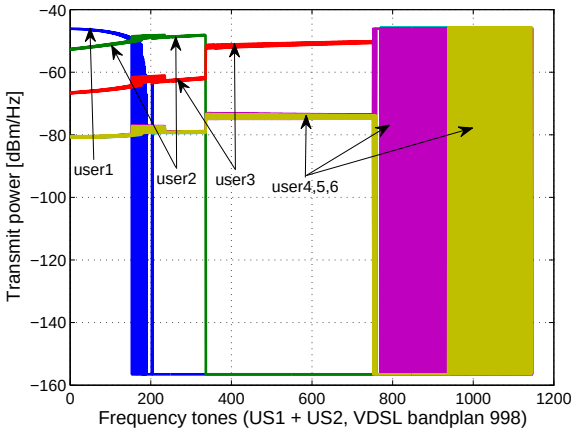


Fig. 11. Transmit powers for scenario of DSL scenario of Fig. 9 obtained using improved dual decomposition approach with the interleaving procedure (33). Blue, green, red, cyan, magenta, yellow curves correspond to transmit powers of users 1,2,3,4,5 and 6 respectively.

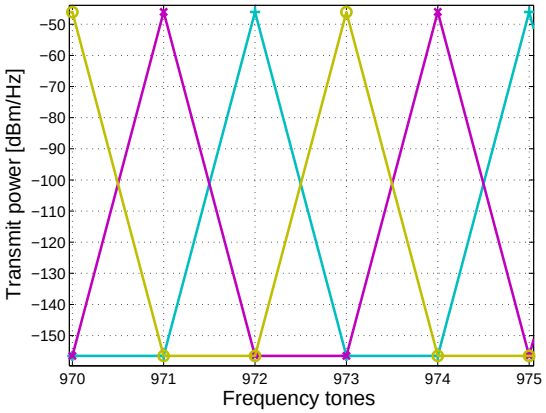


Fig. 12. Zoom in on tones 970 to 975 of Fig. 11.

approach it is proven that the convergence of current state-of-the-art DSM algorithms based on iterative convex approximations (CA-DSB, SCALE), is improved by one order of magnitude. The improved dual decomposition approach is also extended for other DSM algorithms (OSB, ISB, ASB, (MS)-DSB, MIW). The addition of an extra interleaving procedure for recovering the optimal transmit powers from the dual optimal solution furthermore improves the convergence of the proposed approach. Simulation results demonstrate that significant convergence speed ups are obtained for practical DSL scenarios. The proposed improved dual decomposition approach makes an important step towards obtaining numerically fast and effective DSM algorithms.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful comments and suggestions.

REFERENCES

[1] P. Tsiaflakis, I. Necoara, J. Suykens, and M. Moonen, "An improved dual decomposition approach to DSL dynamic spectrum management," in *17th European Signal Processing Conference (EUSIPCO)*, Glasgow, Scotland, Aug. 2009.

[2] IE Market Research Corp. (http://www.iemarketresearch.com/), "Global DSL Subscribers and Access Revenues Forecast, 2009-2012: Global DSL subscribers to rise to 331 million in 2012 with DSL access revenues reaching $136.4 billion in 2012," Tech. Rep., Feb. 2009, internet:http://www.researchandmarkets.com/reports/686876/.

[3] K. Song, S. Chung, G. Ginis, and J. M. Cioffi, "Dynamic spectrum management for next-generation DSL systems," *IEEE Communications Magazine*, vol. 40, no. 10, pp. 101–109, Oct. 2002.

[4] R. Cendrillon, W. Yu, M. Moonen, J. Verlinden, and T. Bostoen, "Optimal multiuser spectrum balancing for digital subscriber lines," *IEEE Transactions on Communications*, vol. 54, no. 5, pp. 922–933, May 2006.

[5] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Transactions on Communications*, vol. 54, no. 7, Jul. 2006.

[6] P. Tsiaflakis, J. Vangorp, M. Moonen, and J. Verlinden, "A low complexity optimal spectrum balancing algorithm for digital subscriber lines," *Signal Processing*, vol. 87, no. 7, pp. 1735–1753, Jul. 2007.

[7] P. Tsiaflakis, Y. Yi, M. Chiang, and M. Moonen, "Green DSL: Energy-Efficient DSM," in *IEEE International Conference on Communications (ICC 2009)*, Jun. 2009.

[8] ——, "Fair greening for DSL broadband access," in *Proc. of the GreenMetrics Workshop in conjunction with ACM Sigmetrics/Performance*, Seattle, Washington, Jun. 2009, pp. 1–5.

[9] M. Wolkerstorfer, D. Statovci, and T. Nordstrom, "Dynamic spectrum management for energy-efficient transmission in DSL," in *The Eleventh IEEE International Conference on Communications Systems (ICCS 2008), China*, Nov. 2008.

[10] J. M. Cioffi, S. Jagannathan, W. Lee, H. Zou, A. Chowdhery, W. Rhee, G. Ginis, and P. Silverman, "Greener copper with dynamic spectrum management," in *AccessNets, Las Vegas, NV, USA*, Oct. 2008.

[11] Z. Q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 57–73, Feb. 2008.

[12] P. Tsiaflakis, Y. Yi, M. Chiang, and M. Moonen, "Throughput and delay of DSL dynamic spectrum management with dynamic arrivals," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Nov. 2008, pp. 1–5.

[13] P. Tsiaflakis, M. Diehl, and M. Moonen, "Distributed spectrum management algorithms for multiuser DSL networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 4825–4843, Oct. 2008.

[14] R. Cendrillon and M. Moonen, "Iterative spectrum balancing for digital subscriber lines," in *IEEE International Conference on Communications (ICC)*, vol. 3, no. 3, May 2005, pp. 1937–1941.

[15] Y. Xu, T. Le-Ngoc, and S. Panigrahi, "Global concave minimization for optimal spectrum balancing in multi-user DSL networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. pp. 2875–2885, Jul. 2008.

[16] J. Papandriopoulos and J. S. Evans, "SCALE: A low-complexity distributed protocol for spectrum balancing in multiuser DSL networks," *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3711–3724, Aug. 2009.

[17] W. Yu, "Multiuser water-filling in the presence of crosstalk," in *Information Theory and Applications (ITA)*, Feb. 2007.

[18] R. Cendrillon, J. Huang, M. Chiang, and M. Moonen, "Autonomous spectrum balancing for digital subscriber lines," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4241–4257, Aug. 2007.

[19] W. Yu, G. Ginis, and J. M. Cioffi, "Distributed multiuser power control for digital subscriber lines," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 5, pp. 1105–1115, Jun. 2002.

[20] S. Jagannathan and J. M. Cioffi, "Distributed adaptive bit-loading for spectrum optimization in multi-user multicarrier systems," *Elsevier Physical Communication*, vol. 1, no. 1, pp. pp. 40–59, 2008.

[21] ——, "DSM level 2 band-preference using penalty tables," ATIS Contribution NIPP-NAI 2007-062, Minneapolis, Tech. Rep., May 2007, internet: http://jsumanth.googlepages.com/.

[22] I. Necoara and J. A. K. Suykens, "Application of a smoothing technique to decomposition in convex optimization," *IEEE Transactions on Automatic Control*, vol. 11, no. 53, pp. 2674–2679, 2008.

[23] Y. Nesterov, "Smooth minimization of non-smooth functions," *Mathematical Programming (A)*, vol. 103, no. 1, pp. 127–152, May 2005.

[24] ——, *Introductory Lectures on Convex Optimization: A Basic Course*, Kluwer, Ed. Boston, 2004.

[25] T. Starr, J. M. Cioffi, and P. J. Silverman, *Understanding digital subscriber lines*. Prentice Hall, 1999.

[26] R. Lui and W. Yu, "Low-complexity near-optimal spectrum balancing for digital subscriber lines," in *IEEE International Conference on Communications*, vol. 3, no. 3, May 2005, pp. 1947–1951.

[27] W. Lee, Y. Kim, M. H. Brady, and J. M. Cioffi, "Band-preference dynamic spectrum management in a DSL environment," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Nov. 2006, pp. 1–5.

[28] M. Chiang, C. W. Tan, D. P. Palomar, D. O'Neill, and D. Julian, "Power control by geometric programming," *IEEE Transactions on Wireless Communications*, vol. 6, no. 7, pp. 2640–2651, Jul. 2007.

[29] S. Kontogiorgis, R. D. Leone, and R. Meyer, "Alternating direction splitings for block angular parallel optimization," *Journal of Optimization Theory and Applications*, vol. 90, no. 1, pp. 1–29, 1996.

[30] G. Chen and M. Teboulle, "A proximal-based decomposition method for convex minimization problems," *Mathematical Programming (A)*, vol. 64, pp. 81–101, 1994.

[31] J. E. Spingarn, "Applications of the method of partial inverses to convex programming: decomposition," *Mathematical Programming (A)*, vol. 32, pp. 199–223, 1985.

[32] L. Xiao, M. Johansson, and S. Boyd, "Simultaneous routing and resource allocation via dual decomposition," *IEEE Transactions on Communications*, vol. 52, no. 7, pp. 1136–1144, Jul. 2004.

[33] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Athena Scientific, 1999.

[34] K. C. Kiwiel, "Approximations in bundle methods and decomposition of convex programs," *J. Optimiz. Theory Applicat.*, vol. 84, pp. 529–548, 1995.

[35] R. T. Rockafellar, "Augmented Lagrangians and applications of the proximal point algorithms in convex programming," *Math. Oper. Res.*, vol. 1, pp. 97–116, 1976.

[36] P. Tsiaflakis, C. W. Tan, Y. Yi, M. Chiang, and M. Moonen, "Optimality certificate of dynamic spectrum management in multi-carrier interference channels," in *IEEE International Symposium on Information Theory*, Jul. 2008, pp. 1298–1302.

[37] S. Hayashi and Z. Q. Luo, "Dynamic spectrum management: When is FDMA sum-rate optimal?" in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, Apr. 2007, pp. III–609 – III–612.

[38] Z. Q. Luo and S. Zhang, "Duality gap estimation and polynomial time approximation for optimal spectrum management," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2675–2689, Jul. 2009.



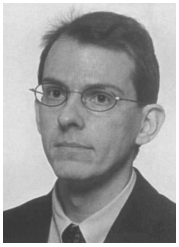
Paschalis Tsiaflakis (S'06-M'09) received the M.Eng. degree (summa cum laude) in electrical engineering from the Katholieke Hogeschool Limburg, Belgium, in 2001, and the Master degree (magna cum laude) in electrical engineering and the Ph.D. degree in electrical engineering from the Katholieke Universiteit Leuven (K.U.Leuven), Belgium, in 2004 and 2009, respectively.

He is a currently a Postdoctoral Research Associate with the Department of Electrical Engineering, K.U.Leuven, Belgium. Since 2004, he has been involved in several industrial projects in cooperation with Alcatel-Lucent, Antwerp, Belgium. In 2007, he was a Visiting Scholar at Princeton University, NJ, USA, under the supervision of Prof. M. Chiang. His research interests include signal processing and optimization of digital communication systems with special emphasis on DSL wired broadband access networks.

Dr. Tsiaflakis was funded by the Research Foundation - Flanders (FWO) from 2004 until 2008. He received the Best Multimedia Master Thesis prize awarded by PIMC in 2001, IEEE Travel Grants in 2006 and 2007, an FWO Aspirant scholarship for the period 2004-2008, an FWO scholarship for a Visiting Research Collaboration at Princeton University in 2007, and a PDMK postdoc grant in 2009.



Ion Necoara received the B.Sc. degree in mathematics from the University of Bucharest, Faculty of Mathematics, and the M.Sc. degree in optimization and statistics from the same faculty in 2000 and 2002, respectively. After graduating he worked as a Ph.D. student at the Delft Center for Systems and Control, Delft University of Technology, The Netherlands. In 2005 he visited for 3 months the Control Group, University of Cambridge. For the period 2006-2009 he completed a post-doctoral fellowship at Katholieke Universiteit Leuven, Belgium. Since 2009, he has been an Assistant Professor at the University Politehnica Bucharest, Romania. His research interests are distributed optimization with applications in control of complex systems and embedded optimization algorithms with a focus on structure exploiting.



Johan A.K. Suykens (M02-SM04) was born in Willebroek Belgium, May 18 1966. He received the degree in Electro-Mechanical Engineering and the Ph.D. degree in Applied Sciences from the Katholieke Universiteit Leuven, in 1989 and 1995, respectively.

In 1996 he has been a Visiting Postdoctoral Researcher at the University of California, Berkeley. He is currently a Professor (Hoogleraar) with K.U.Leuven. He is author of the books ”Artificial Neural Networks for Modelling and Control of Non-linear Systems” (Kluwer Academic Publishers) and ”Least Squares Support Vector Machines” (World Scientific), co-author of the book ”Cellular Neural Networks, Multi-Scroll Chaos and Synchronization” (World Scientific) and editor of the books ”Nonlinear Modeling: Advanced Black-Box Techniques” (Kluwer Academic Publishers) and ”Advances in Learning Theory: Methods, Models and Applications” (IOS Press).

Prof. Suykens has served as associate editor for the IEEE Transactions on Circuits and Systems (1997-1999 and 2004-2007) and since 1998 he is serving as associate editor for the IEEE Transactions on Neural Networks. He received an IEEE Signal Processing Society 1999 Best Paper (Senior) Award and several Best Paper Awards at International Conferences. He is a recipient of the International Neural Networks Society INNS 2000 Young Investigator Award for significant contributions in the field of neural networks. He has served as a Director and Organizer of the NATO Advanced Study Institute on Learning Theory and Practice (Leuven 2002), as a program co-chair for the International Joint Conference on Neural Networks 2004 and the International Symposium on Nonlinear Theory and its Applications 2005, and as an organizer of the International Symposium on Synchronization in Complex Networks 2007.



Marc Moonen (M’94, SM’06, F’07) received the electrical engineering degree and the PhD degree in applied sciences from Katholieke Universiteit Leuven, Belgium, in 1986 and 1990 respectively.

Since 2004 he is a Full Professor at the Electrical Engineering Department of Katholieke Universiteit Leuven, where he is heading a research team working in the area of numerical algorithms and signal processing for digital communications, wireless communications, DSL and audio signal processing.

Prof. Moonen received the 1994 K.U.Leuven Research Council Award, the 1997 Alcatel Bell (Belgium) Award (with Piet Vandaele), the 2004 Alcatel Bell (Belgium) Award (with Raphael Cendrillon), and was a 1997 “Laureate of the Belgium Royal Academy of Science”. He received a journal best paper award from the IEEE Transactions on Signal Processing (with Geert Leus) and from Elsevier Signal Processing (with Simon Doclo). He was chairman of the IEEE Benelux Signal Processing Chapter (1998-2002), and is currently Past-President of EURASIP (European Association for Signal, Speech and Image Processing) and a member of the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications. He has served as Editor-in-Chief for the “EURASIP Journal on Applied Signal Processing” (2003-2005), and has been a member of the editorial board of “IEEE Transactions on Circuits and Systems II” (2002-2003) and “IEEE Signal Processing Magazine” (2003-2005) and “Integration, the VLSI Journal”. He is currently a member of the editorial board of “EURASIP Journal on Applied Signal Processing”, “EURASIP Journal on Wireless Communications and Networking”, and “Signal Processing”.