Computing the Pareto Front by Polynomial Elimination, With an Application From System Identification*

Hans van Rooij, Christof Vermeersch†, Marie Deferme, and Bart De Moor‡

***Abstract*— We propose a novel numerical approach to compute the Pareto front in multivariate polynomial multi-objective optimization problems. When the objective functions and (equality) constraints are multivariate polynomials, the Pareto front, which describes the efficient points of the multiple (often conflicting) objective functions, can be interpreted as a subset of a positive-dimensional algebraic variety. By combining the objective functions with weights and considering the weights as additional decision variables, we can eliminate all variables except the objective values and obtain one (or multiple) polynomial equation(s) that describes the Pareto front. Unlike sampling-based methods that approximate the Pareto front point-wise, our elimination-based approach yields an explicit algebraic relation between the objective values, representing the Pareto front as a geometric object in the objective space without requiring a predetermined number of sample points. Besides numerical examples illustrating the elimination-based approach, we use elimination on a challenging application that originates from system identification, in which we analyze the trade-off between misfit and latency terms when determining the optimal model parameters from measured data.**

I. Introduction

Due to possibly conflicting and incommensurable objective functions in multi-objective (or vector) optimization problems, it is generally not possible to find a single solution that is optimal for all objectives simultaneously [1]. In that sense, such problems do not admit a unique optimum but rather a set of efficient solutions called *Pareto-efficient points*, for which no objective can be improved without worsening another.[1] The corresponding set of objective values, the *Pareto front*, represents the trade-offs inherent to the optimization problem and is, therefore, central in the decision making process.

When the objective functions and (equality) constraints are multivariate polynomials, the Pareto front can be interpreted as a subset of a positive-dimensional algebraic variety. We present in this paper a novel approach to compute in that case the Pareto front by (numerical) polynomial elimination: eliminating all variables except the objective values yields polynomial equations(s) of which a part of the solution set corresponds to the Pareto front.

Existing methods for computing (parts of) the Pareto front typically encode the relative importance of the objective functions before (*a priori*) or after (*a posteriori*) optimization, for instance through weights, constraints, or ordering on the objective functions. A survey of such methods is given in [3]. A common approach is scalarization, in which the multi-objective optimization problem is reformulated as a family of scalar problems that sample the Pareto front. The (convex) weighted sum method is widely used in scalarization because of its simplicity: a weighted combination of the objective functions is minimized for varying weights and each minimizer corresponds to a distinct point on the Pareto front. However, it has two well-known limitations: (i) a uniform spread of weights does not generally produce a uniform spread on the Pareto front and (ii) non-convex regions of the Pareto front cannot be detected. Moreover, obtaining a finer approximation of the Pareto front increases the computational cost proportionally.

In contrast, our elimination-based approach treats the weights as additional variables rather than fixed parameters. The first-order necessary optimality conditions for the corresponding scalarized, but parametrized, multivariate polynomial multi-objective optimization problem results in a system of multivariate polynomial equations. Unlike sampling-based methods that approximate the Pareto front point-wise, eliminating all variables except the objective values yields an explicit algebraic relation between the objective values; the Pareto front is a subset of this implicitly represented geometric object in the objective space. It is no longer necessary to decide in advance on the number of sampling points. Although our approach is still computationally intensive, it does not suffer from the above-mentioned two limitations.

Multi-objective trade-offs arise naturally in many engineering applications (see [1], [4] and references therein). For example, in system identification, determining model parameters that best describe measured data leads inherently to competing objectives. Here, we apply our elimination-based approach to the misfit-versus-latency model class,

Hans van Rooij, Christof Vermeersch (corresponding author), Marie Deferme, and Bart De Moor are with the Center for Dynamical Systems, Signal Processing, and Data Analytics (STADIUS), Dept. of Electrical Engineering (ESAT), KU Leuven. Kasteelpark Arenberg 10, 3001 Leuven, Belgium. E-mail addresses: {hans.vanrooij, christof.vermeersch, marie.deferme, bart.demoor}@esat.kuleuven.be.

*This work was supported in part by the KU Leuven Research Fund (grants IBOF/23/064, C3/20/117, and C3I/21/00316); in part by the Fonds Wetenschappelijk Onderzoek (grants S005319 and T001919N); in part by the Departement Economie, Wetenschap & Innovatie via the Flanders AI Research Program; in part by the Vlaams Agentschap Innoveren & Ondernemen (grant HBC/2021/0076); and in part by the European Research Council (grant 885682). †The work of Christof Vermeersch was supported in part by a FWO PHC Tournesol mobility grant (grant VS00326N). ‡Bart De Moor is an *IEEE*, *IFAC*, and *SIAM* fellow.

[1]The concept is attributed to both Francis Ysidro Edgeworth and Vilfredo Pareto, two economists from the late 19th, early 20th century. Although Edgeworth presented the idea first, it is named after Pareto, who introduced the notion of economic efficiency in the context of income distribution and welfare economics [2]. His key insight was that a society's resources are optimally allocated when no individual can be made better off without making another worse off—a condition now known as Pareto efficiency.

where the mismatch between measured and model-compliant data, which result from external influences such as measured noise, observed disturbances, and unmodeled effects, are represented by a misfit term and an unknown latent input. The optimal model parameters are obtained by distributing this mismatch between both terms, resulting in a multi-objective optimization problem that minimizes them simultaneously. Although only a limited number of data points is considered, the accompanying numerical example provides both a challenging elimination benchmark (with ten decision variables) and an illustration of the application potential of Pareto fronts in system identification.

Outline and contributions: The remainder of this paper is organized as follows. Firstly, in Section II, we explain how Pareto fronts appear in multi-objective minimization. We state the problem mathematically and give a motivational example. In Section III, we present our elimination-based approach for computing an algebraic description of the Pareto front; *this constitutes the main contribution of the paper*. We consider in Section IV three numerical examples. Afterwards, in Section V, we *apply elimination to parameter estimation in the misfit-versus-latency model class, which forms a secondary contribution*. Finally, Section VI concludes the paper and outlines directions for future work.

II. Pareto Fronts in Minimization

Given m objective functions $f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x})$ in n decision variables $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ with $m \geq 2$ and $f_i(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$, the multi-objective minimization problem denotes the simultaneous minimization of each of these objective functions on a given real-valued[2] feasible decision space $\boldsymbol{X} \subseteq \mathbb{R}^n$,

$$\begin{aligned} &\min_{\boldsymbol{x}} \begin{bmatrix} f_1(\boldsymbol{x}) & f_2(\boldsymbol{x}) & \ldots & f_m(\boldsymbol{x}) \end{bmatrix}^{\mathrm{T}}, \\ &\text{subject to } \boldsymbol{x} \in \boldsymbol{X} \end{aligned} \tag{1}$$

which is defined as finding all Pareto-efficient points $\boldsymbol{x}^\star$ in $\boldsymbol{X}$. The concept of Pareto efficiency (other terms are *admissibility*, *non-inferiority*, and *Pareto optimality*) is used in multi-objective minimization because typically no single global minimum exists, and it is defined as follows.

Definition 1: A point $\boldsymbol{x}^\star \in \boldsymbol{X}$ is *Pareto-efficient* if there exists no feasible $\boldsymbol{x} \in \boldsymbol{X}$ such that $f_i(\boldsymbol{x}) \leq f_i(\boldsymbol{x}^\star)$ for all $i = 1, 2, \ldots, m$, and $f_j(\boldsymbol{x}) < f_j(\boldsymbol{x}^\star)$ for at least one j. □

The set of objective values of all (possibly infinitely many) Pareto-efficient points is the so-called Pareto front. A stricter definition exists that excludes certain degenerate Pareto fronts.

Definition 2: A point $\boldsymbol{x}^\star \in \boldsymbol{X}$ is *properly Pareto-efficient* if it is Pareto-efficient and if there exists a scalar $R > 0$ such that, for each i, we have that

$$\frac{f_i(\boldsymbol{x}) - f_i(\boldsymbol{x}^\star)}{f_j(\boldsymbol{x}^\star) - f_j(\boldsymbol{x})} \leq R, \tag{2}$$

for some j such that $f_j(\boldsymbol{x}) > f_j(\boldsymbol{x}^\star)$ whenever $\boldsymbol{x} \in \boldsymbol{X}$ and $f_i(\boldsymbol{x}) < f_i(\boldsymbol{x}^\star)$. □

The ratio in (2) measures the gain in objective function $f_i(\boldsymbol{x})$ relative to the loss in objective function $f_j(\boldsymbol{x})$. Proper Pareto efficiency requires this ratio to be bounded by the same R. If no finite R exists, which means that there are sequences of feasible points where this ratio grows unbounded, then $\boldsymbol{x}^\star$ is Pareto-efficient but improper.

The (convex) weighted sum method rephrases the multi-objective minimization problem (1) into a scalar minimization problem given by

$$\begin{aligned} &\min_{\boldsymbol{x}} \sum_{i=1}^{m} w_i f_i(\boldsymbol{x}), \\ &\text{subject to } \boldsymbol{x} \in \boldsymbol{X}, \end{aligned} \tag{3}$$

where the weights w_i, for $i = 1, 2, \ldots, m$, satisfy the following constraints:

$$w_i \geq 0 \text{ and } \sum_{i=1}^{m} w_i = 1.$$

For a specific choice of weights $\boldsymbol{w}$, the solution of the weighted minimization problem (3) gives one point on the Pareto front. The proof of the following important theorem is given, for maximization, in [6].

Theorem 1: Let $w_i > 0$ be fixed for $i = 1, 2, \ldots m$ with $\sum_{i=1}^{m} w_i = 1$. If $\boldsymbol{x}^\star$ is a minimum of the (convex) weighted minimization problem (3), then $\boldsymbol{x}^\star$ is properly Pareto-efficient in (1). □

The multi-objective minimization problem potentially has n_g equality constraints $g_k(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$:

$$g_k(\boldsymbol{x}) = 0, \quad k = 1, 2, \ldots, n_g,$$

which determine the feasible decision space $\boldsymbol{X} \subseteq \mathbb{R}^n$.

Definition 3: The *feasible decision space* $\boldsymbol{X} \subseteq \mathbb{R}^n$, also called *constraint set*, is defined as the set $\{\boldsymbol{x} : g_k(\boldsymbol{x}) = 0, k = 1, 2, \ldots, n_g\}$. □

The critical points of the weighted minimization problem (3) can in that case be found by solving the Karush–Kuhn–Tucker (KKT) conditions [7] for a specific choice of weights $\boldsymbol{w}$:

$$\begin{aligned} \nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) &= \boldsymbol{0}, \\ g_k(\boldsymbol{x}) &= 0, \quad k = 1, 2, \ldots, n_g, \end{aligned} \tag{4}$$

where the Lagrangian $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ is given by

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \sum_{i=1}^{m} w_i f_i(\boldsymbol{x}) - \sum_{k=1}^{n_g} \lambda_k g_k(\boldsymbol{x}).$$

One of the critical points obtained by solving the KKT system (4), specifically the global minimum, is a Pareto-efficient point of the multi-objective minimization problem (1).

Since the focus in the paper is on the subclass of multivariate polynomial multi-objective minimization with equality constraints, we restrict ourselves to objective functions and constraints that are multivariate polynomials:

$$f_i(\boldsymbol{x}), g_k(\boldsymbol{x}) \in \mathbb{R}[\boldsymbol{x}].$$

[2] In engineering applications, the decision variables are typically real-valued. For an introductory text that discusses multivariate polynomial minimization with complex-valued decision variables, see [5].

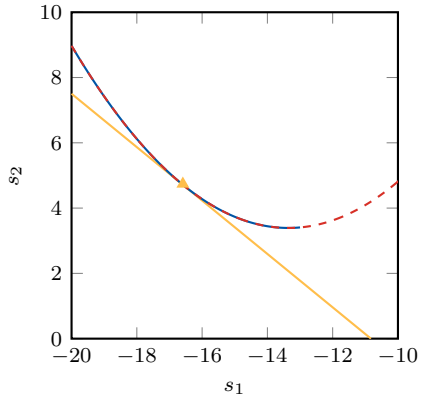


Fig. 1. Visualization of the Pareto front for the portfolio optimization problem in Example 1. The Pareto front (——) is obtained by considering the zero set of the eliminant polynomial (- - -), which is computed via our novel elimination-based approach in Section III-D. Only the points that correspond to nonnegative weights are contained in the Pareto front. A linear level curve (——) allows to recover the weight vector and decision variables that correspond to a specific point on the Pareto front (▲).

The Pareto front can then be described by multivariate polynomial equations and is, therefore, contained in a (positive-dimensional) algebraic variety. Note that when polynomial inequalities constraints are present, it is a subset of a semi-algebraic set, which itself is contained in a variety.

Example 1 (Motivational example): Consider the problem of creating an investment portfolio[3], in which one has to decide the amount of money to invest in three different stock options x_i, $i = 1, 2, 3$. The two conflicting objectives are to maximize revenue, $f_1(\boldsymbol{x}) = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}$, and minimize risk, $f_2(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x}$. Under the constraint that one has a limited amount of money to invest, $g(\boldsymbol{x}) = \sum_{i=1}^{3} x_i - 100 = 0$, the multi-objective minimization problem corresponds to

$$\min_{\boldsymbol{x}} \begin{bmatrix} -\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} \\ \boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x} \end{bmatrix}$$
$$\text{subject to} \sum_{i=1}^{3} x_i - 100 = 0,$$

where the vector $\boldsymbol{a} \in \mathbb{R}^3$ contains the expected revenue per stock and the covariance matrix $\boldsymbol{B}$ describes the risk associated with every stock. Depending on the values of $\boldsymbol{a}$ and $\boldsymbol{B}$, the objective functions can be conflicting, which is the case for

$$\boldsymbol{a} = \begin{bmatrix} 10 \\ 20 \\ 15 \end{bmatrix} \cdot 10^{-2} \text{ and } \boldsymbol{B} = \begin{bmatrix} 5 & 1 & 2 \\ 1 & 10 & 3 \\ 2 & 3 & 7 \end{bmatrix} \cdot 10^{-4}.$$

Fig. 1 displays the Pareto front for these values of $\boldsymbol{a}$ and $\boldsymbol{B}$ as a solid blue line (——). The objective variables s_1 and s_2 denote the negated expected revenue $-\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}$ and expected risk $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x}$, respectively. □

[3]As Pareto fronts have their roots in economics, we introduce a small motivational example from that setting, which we use throughout the paper to illustrate the key concepts and our novel approach. The example is based on the work of Nobel prize winner Harry Markowitz [8].

III. Computational Elimination Approach

We now present our elimination-based approach for computing the Pareto front of multivariate polynomial multi-objective minimization problems. We first describe how the Pareto front can be characterized through a system of multivariate polynomial equations derived from the KKT conditions, and how its computation can be formulated as an elimination problem (Section III-A). Next (in Section III-B), we introduce a numerical elimination algorithm that uses the Macaulay matrix to compute polynomial relations between the objective variables. The interpretation of the resulting eliminant system and corresponding Pareto front are discussed afterwards (Section III-C). Finally, the complete procedure is illustrated on the motivational example (Section III-D).

A. Reformulation of the Pareto front as an eliminant

If each $f_i(\boldsymbol{x}), g_k(\boldsymbol{x})$ is a multivariate polynomial, then the KKT conditions (4) yield a multivariate polynomial system in the variables $(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{\lambda})$. To obtain a description of the Pareto front in the objective space, we introduce m additional variables

$$\boldsymbol{s} = (s_1, s_2, \ldots, s_m),$$

representing the objective values, and impose the relations

$$s_i - f_i(\boldsymbol{x}) = 0, \quad i = 1, 2, \ldots, m. \tag{5}$$

Combining the KKT conditions (4) and objective relations (5) yields the so-called Pareto front (PF) system:

$$\begin{aligned} \nabla_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) &= \boldsymbol{0}, \\ g_k(\boldsymbol{x}) &= 0, \quad k = 1, 2, \ldots, n_g, \\ s_i - f_i(\boldsymbol{x}) &= 0, \quad i = 1, 2, \ldots, m. \end{aligned} \tag{6}$$

Eliminating $n' = n + m + n_g$ variables $(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{\lambda})$ from (6) yields algebraic relations in the objective values $\boldsymbol{s}$. Hence, computing the Pareto front can be formulated as an elimination problem. The *eliminant system*, of which the solutions contain the Pareto front, may consist of a single polynomial equation or a multivariate polynomial system.

The elimination of $(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{\lambda})$ can be done symbolically or numerically. Numerical elimination is a relatively recent development compared to the long-standing symbolic tradition. In contrast to symbolic approaches, numerical formulations rely on techniques from numerical linear algebra at every stage of the computation, allowing floating-point arithmetic [9], [10], [11].

B. Numerical elimination algorithm

Our numerical elimination algorithm uses Macaulay matrices $\boldsymbol{M}_d \in \mathbb{R}^{p_d \times q_d}$ of increasing degree d, which are built from the coefficients of the PF system (6) and monomial multiples up to total degree d [12]. To construct a Macaulay matrix, we take each polynomial in the PF system (6) and multiply it by all monomials of increasing total degree so that the highest total degree does not exceed d, and we collect the coefficients of these polynomials in a structured matrix

with columns labeled by the monomials [12]. Its dimensions $p_d \times q_d$ depend on the degree d.

Each row of the Macaulay matrix $\boldsymbol{M}_d$ represents one polynomial (either an original equation or a monomial multiple) expressed in a fixed monomial basis [12]. Hence, the row space of $\boldsymbol{M}_d$, denoted by $\boldsymbol{\mathcal{R}}_d$, consists of polynomial combinations of the polynomials, $r_i(\boldsymbol{z})$ with total degree d_i, in the PF system (6):

$$\boldsymbol{\mathcal{R}}_d = \left\{ \sum_{i=1}^{n'} c_i(\boldsymbol{z})\, r_i(\boldsymbol{z}) : c_i(\boldsymbol{z}) \in \mathbb{R}[\boldsymbol{z}]_{\leq d-d_i} \right\}.$$

Here, $\mathbb{R}[\boldsymbol{z}]_{\leq d-d_i}$ denotes the set of polynomials in $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{\lambda}, \boldsymbol{s})$ with real coefficients of degree at most $d - d_i$. Thus, the row space $\boldsymbol{\mathcal{R}}_d$ can be interpreted as the space of all polynomial combinations of the PF system with degree bounded by d.

The goal of our numerical elimination algorithm is to remove all variables except the objective variables $\boldsymbol{s}$. To achieve this goal, we construct polynomial combinations of the original system that no longer contain the elimination variables. Since each vector in the row space $\boldsymbol{\mathcal{R}}_d$ represents such a polynomial combination, this amounts to finding vectors whose elements corresponding to monomials involving elimination variables are zero. Equivalently, we seek vectors whose nonzero entries correspond only to monomials in $\boldsymbol{s}$. We construct a so-called *zero pattern vector* $\boldsymbol{t} \in \mathbb{R}^{q_d}$ that "forces" coefficients to zero in the row space. The idea is illustrated in the following example.

Example 2: Let $\boldsymbol{x} = (x_1, x_2, x_3)$ and consider two objectives ($m = 2$) without constraints. The variables are $(x_1, x_2, x_3, w_1, w_2, s_1, s_2)$. We eliminate all decision and weight variables, so that only the objective variables (s_1, s_2) remain. The predetermined zero pattern in the row space corresponds to

$$\boldsymbol{t}^{\mathrm{T}} = \begin{matrix} 1 & x_1 & \cdots & w_2 & s_1 & s_2 & x_1^2 & \\ [\bullet & 0 & 0 & 0 & \bullet & \bullet & 0 & \cdots] \end{matrix},$$

where $\bullet$ denotes an arbitrary real number. This vector represents a linear combination of the rows of $\boldsymbol{M}_d$ in which all coefficients corresponding to monomials containing elimination variables are zero, while coefficients of monomials in only (s_1, s_2) may be nonzero. As a result, the corresponding polynomial(s) depend solely on (s_1, s_2). □

The zero pattern vector $\boldsymbol{t}$ can be parametrized as

$$\boldsymbol{t}^{\mathrm{T}} = \begin{bmatrix} \bullet & \bullet & \bullet & \cdots \end{bmatrix} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}}_{\boldsymbol{E}_d}.$$

The rows of $\boldsymbol{E}_d$ form an orthonormal basis of the elimination vector space $\boldsymbol{\mathcal{E}}_d$. The matrix $\boldsymbol{E}_d$ acts as a selector that keeps only the entries corresponding to monomials in the decision variables $\boldsymbol{s}$ and sets all other entries to zero. Consequently, $\boldsymbol{\mathcal{E}}_d$ consists of all coefficient vectors that have this prescribed zero pattern, i.e., vectors whose nonzero entries correspond only to monomials in $\boldsymbol{s}$.

Consequently, elimination reduces to finding vectors that lie in both $\boldsymbol{\mathcal{R}}_d$ and $\boldsymbol{\mathcal{E}}_d$, i.e., to computing their intersection. Vectors in $\boldsymbol{\mathcal{R}}_d$ correspond to polynomial combinations of the original system, while vectors in $\boldsymbol{\mathcal{E}}_d$ encode the requirement that only monomials in the objective variables $\boldsymbol{s}$ remain. Therefore, vectors in the intersection $\boldsymbol{\mathcal{R}}_d \cap \boldsymbol{\mathcal{E}}_d$ represent polynomial combinations of the PF system (6) in which all elimination variables vanish, yielding relations purely in $\boldsymbol{s}$. A nontrivial intersection requires

$$\dim(\boldsymbol{\mathcal{R}}_d \cap \boldsymbol{\mathcal{E}}_d > 0.$$

Using Grassmann's dimension theorem [13], the dimension of the intersection is given by

$$\dim(\boldsymbol{\mathcal{R}}_d \cap \boldsymbol{\mathcal{E}}_d) = \operatorname{rank}(\boldsymbol{M}_d) - \operatorname{rank}(\boldsymbol{N}_d),$$

where $\boldsymbol{N}_d \in \mathbb{R}^{p_d \times q'_d}$ is a submatrix of the Macaulay matrix $\boldsymbol{M}_d$ that consists of the q'_d columns associated with at least one variable that need to be eliminated.[4] A positive dimension, i.e.,

$$\operatorname{rank}(\boldsymbol{M}_d) > \operatorname{rank}(\boldsymbol{N}_d), \tag{7}$$

guarantees the existence of an eliminant system, meaning that there exist nontrivial combinations of the original polynomials in which all eliminated variables cancel out.

The degree d is increased until the rank condition in (7) is satisfied, since higher degrees include more monomial multiples of the original equations, allowing additional polynomial combinations that can eliminate the elimination variables.

For that degree, a basis $\boldsymbol{V} \in \mathbb{R}^{p_d \times l}$ for the left null space of $\boldsymbol{N}_d$ is computed, such that

$$\boldsymbol{V}^{\mathrm{T}} \boldsymbol{N}_d = \boldsymbol{0},$$

A left multiplication of the Macaulay matrix $\boldsymbol{M}_d$ by $\boldsymbol{V}^{\mathrm{T}}$ thus results in linear combinations of the rows of $\boldsymbol{M}_d$, while eliminating all monomials containing one or more of the variables to be removed. The result, $\boldsymbol{S}^{\mathrm{T}} = \boldsymbol{V}^{\mathrm{T}} \boldsymbol{M}_d \in \mathbb{R}^{l \times q_d}$, only involves the objective variables $\boldsymbol{s}$.

The eliminant system is then constructed by interpreting the l' nonzero columns of $\boldsymbol{S}$ as coefficient vectors of multivariate polynomials

$$t_1(\boldsymbol{s}) = \cdots = t_{l'}(\boldsymbol{s}) = 0. \tag{8}$$

These polynomials form the eliminant system, of which the solution set contains the Pareto front.

Computing the dimension of the intersection between $\boldsymbol{\mathcal{R}}_d$ and $\boldsymbol{\mathcal{E}}_d$ involves two rank computations per degree d, which constitute the main computational complexity of the elimination-based approach. The first one requires $\mathcal{O}(p_d q_d \min(p_d, q_d))$ operations, while the second one has complexity $\mathcal{O}(p_d q'_d \min(p_d, q'_d))$. Once the degree d is found

[4]The presented approach uses the left singular vectors of $\boldsymbol{N}_d$, which is a submatrix of the Macaulay matrix. Although not explained in this paper, an alternative approach via principal angles and directions between $\mathcal{R}_d$ and $\mathcal{E}_d$ is computationally more efficient [11].

for which (7) is fulfilled, an orthogonal basis matrix $\boldsymbol{V}$ for the left null space of $\boldsymbol{N}_d$ is computed, which requires again $\mathcal{O}(p_d q_d' \min(p_d, q_d'))$ operations. Finally, a matrix-matrix multiplication, $\boldsymbol{S}^{\mathrm{T}} = \boldsymbol{V}^{\mathrm{T}} \boldsymbol{M}_d$, yields the vector space that contains the representation of the eliminant system, costing $\mathcal{O}(l p_d q_d)$ operations.

C. Geometric interpretation of the Pareto front

Let $\mathcal{P} \subset \mathbb{R}^m$ denote the Pareto front in the objective space, obtained after elimination as (a subset of) the algebraic variety defined by (8). The Jacobian matrix of the eliminant system is

$$J_t(\boldsymbol{s}^\star) = \begin{bmatrix} \nabla t_1(\boldsymbol{s}^\star)^{\mathrm{T}} \\ \vdots \\ \nabla t_{l'}(\boldsymbol{s}^\star)^{\mathrm{T}} \end{bmatrix},$$

and the tangent space of $\mathcal{P}$ at $\boldsymbol{s}^\star$ is then given by $(J_t(\boldsymbol{s}^\star))$, while the normal space is $\mathrm{row}\,(J_t(\boldsymbol{s}^\star))$, where and row denote the right null space and row space of $J_t(\boldsymbol{s}^\star)$, respectively.

At a Pareto-efficient point $\boldsymbol{s}^\star$, there exists a vector $\boldsymbol{w}^\star \in \mathbb{R}^m$ such that

$$(\boldsymbol{w}^\star)^{\mathrm{T}} \boldsymbol{s} \geq (\boldsymbol{w}^\star)^{\mathrm{T}} \boldsymbol{s}^\star,$$

for all feasible points $\boldsymbol{s}$ in the objective space. The hyperplane $(\boldsymbol{w}^\star)^{\mathrm{T}} \boldsymbol{s} = (\boldsymbol{w}^\star)^{\mathrm{T}} \boldsymbol{s}^\star$ passes through $\boldsymbol{s}^\star$ and all feasible points lie on one side of it. Geometrically, this hyperplane is tangent to the Pareto front at $\boldsymbol{s}^\star$; its normal vector must belong to the normal space of $\mathcal{P}$ at that point. Therefore,

$$\boldsymbol{w}^\star \in \mathrm{row}\,(J_t(\boldsymbol{s}^\star)).$$

In the case where the Pareto front is described by a single polynomial equation, $t_1(\boldsymbol{s}) = 0$, this reduces to checking whether $\boldsymbol{w}^\star$ is parallel to $\nabla t_1(\boldsymbol{s}^\star)$. Alternatively, when the eliminant system (8) consist of multiple polynomials, the weight vector lies in the span of their gradients, i.e.,

$$\boldsymbol{w}^\star = \sum_{i=1}^{l'} \alpha_i \nabla t_i(\boldsymbol{s}^\star), \quad \alpha_i \in \mathbb{R}.$$

Since we consider a multi-objective minimization problem, only weight vectors with strictly positive components are of interest. Not every point of the algebraic variety thus corresponds to a Pareto-efficient point. The Pareto front $\mathcal{P}$ consists of those points $\boldsymbol{s}^\star$ for which the normal space $\mathrm{row}\,(J_t(\boldsymbol{s}^\star))$ contains a vector with only nonnegative components.

For fixed weights $\boldsymbol{w}^\star$, the corresponding decision variables $\boldsymbol{x}^\star$ can be obtained from the first-order optimality conditions given by the KKT system (4). Substituting $\boldsymbol{w}^\star$ results in a multivariate polynomial system that is generically zero-dimensional. Solving the substituted system yields all critical points $\boldsymbol{x}^\star$ and the corresponding Lagrange multipliers $\boldsymbol{\lambda}^\star$.

D. Step-by-step solution of the motivational example

Example 3: The PF system associated with Example 1 takes the form

$$\begin{aligned} -w_1 \boldsymbol{a} + 2w_2 \boldsymbol{B}\boldsymbol{x} + \lambda \mathbf{1} &= \mathbf{0}, \\ x_1 + x_2 + x_3 - 100 &= 0, \\ s_1 + \boldsymbol{x}^{\mathrm{T}} \boldsymbol{a} &= 0, \\ s_2 - \boldsymbol{x}^{\mathrm{T}} B \boldsymbol{x} &= 0, \end{aligned} \tag{9}$$

with $\mathbf{1} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^{\mathrm{T}}$. Since we are using convex weights, w_2 is replaced by $1 - w_1$. We have a system of six equations in seven variables and eliminate the variables x_1, x_2, x_3, w_1, and λ. The elimination-based approach requires a 384×330 Macaulay matrix of degree four and yields a bivariate polynomial relation between s_1 and s_2. The zero set of this eliminant polynomial is shown as the red dashed curve (- - -) in Fig. 1. Every point on the true Pareto front is contained in this zero set. The converse does not hold: not all points in the zero set correspond to Pareto-efficient solutions. A discrepancy can arise because the KKT conditions are necessary but not sufficient, and therefore they also capture stationary points that are not minima. However, this is not the case here. The difference arises from omitting the inequality constraints $w_i \geq 0$.

Suppose that we are interested in the point $\boldsymbol{s}^\star = (-16.59, 4.74)$ on the Pareto front, which is indicated by a yellow triangle (▲) on Fig. 1. The gradient of the eliminant polynomial at this point equals $\begin{bmatrix} -3.12 & -3.82 \end{bmatrix}^{\mathrm{T}} \cdot 10^{-2}$, which yields, after normalization, the weights $\boldsymbol{w}^\star = (0.45, 0.55)$. The level curve $(\boldsymbol{w}^\star)^{\mathrm{T}} \boldsymbol{s} = (\boldsymbol{w}^\star)^{\mathrm{T}} \boldsymbol{s}^\star$ is also shown on Fig. 1 in yellow (——). As can be seen, this curve is tangent to the Pareto front at $\boldsymbol{s}^\star$. Substituting $\boldsymbol{s}^\star$ and $\boldsymbol{w}^\star$ in the PF system in (9), and solving the resulting system, gives the corresponding decision variable $\boldsymbol{x}^\star = (18.18, 50.00, 31.82)$. It can easily be verified that this point indeed corresponds to the objective values in $\boldsymbol{s}^\star$. □

IV. Various Numerical Examples

To evaluate the practical behavior of the proposed elimination-based approach, we present three numerical examples.[5] Although these examples appear simple, they already illustrate how rapidly the Macaulay matrix (and therefore the computational complexity) grows when the degree d increases. For each numerical example, we compute the eliminant system that describes the Pareto front and verify its correctness by substituting points on the sampled Pareto front into the algebraic expression. The characteristics are reported in Table I.

Example 4: Consider the unconstrained multivariate polynomial tri-objective optimization problem

$$\min_{\boldsymbol{x}} \begin{bmatrix} (x_1 - 3)^2 + (x_2 - 2)^2 \\ x_1 + x_2 \\ x_1 + 2x_2 \end{bmatrix}.$$

[5]We use the MACAULAYLAB toolbox to perform our numerical examples, which is a MATLAB implementation of to construct, manipulate, and use Macaulay matrices [14].

The first objective function has a unique minimizer $\boldsymbol{x}^\star = (3, 2)$, while the other two objective functions are linear and favor different directions in $\mathbb{R}^2$. As a result, the three objective functions are conflicting, and the Pareto front is a subset of a two-dimensional surface in the (s_1, s_2, s_3)-space. The numerical elimination algorithm yields an eliminant polynomial in the objective variables whose zero set describes this surface. After rounding the coefficients for readability, the eliminant can be written as $5s_2^2 - 6s_2s_3 + 2s_3^2 - s_1 - 8s_2 + 2s_3 + 13 = 0$. □

The next numerical example illustrates how the Macaulay matrix can grow quickly, even when the underlying optimization problem is low-dimensional and structurally simple.

Example 5: Consider the constrained multivariate polynomial bi-objective optimization problem

$$\min_{\boldsymbol{x}} \begin{bmatrix} -x_1^3 - x_2^3 \\ x_1^2 - x_2^2 \end{bmatrix}, \quad \text{subject to } x_1^2 + (x_2 + 1)^2 - 1 = 0.$$

Applying our elimination-based approach to this problem results in a polynomial of degree six: $s_2^6 - 12s_2^5 - 12s_1s_2^4 + 16s_1^4 + 48s_1^3s_2 + 48s_1^2s_2^2 + 32s_1s_2^3 + 48s_2^4 - 32s_1^3 - 48s_1^2s_2$. The degree of the Macaulay matrix, however, has to be eight, leading to a 3234×3003 matrix. □

Not every Pareto front is contained in the solution set of a single multivariate polynomial.

Example 6: Consider the unconstrained multivariate polynomial tri-objective problem

$$\min_{\boldsymbol{x}} \begin{bmatrix} x^2 \\ (x-1)^2 \\ (x-2)^2 \end{bmatrix},$$

over a one-dimensional decision space. The Pareto front is a one-dimensional curve embedded in $\mathbb{R}^3$. Its algebraic description is given by the two independent polynomial equations

$$\begin{aligned} t_1(s_1, s_2, s_3) &= s_1 - 2s_2 + s_3 - 2 = 0, \\ t_2(s_1, s_2, s_3) &= s_2^2 - 2s_2s_3 + s_3^2 - 2s_2 - 2s_3 + 1 = 0. \end{aligned} \tag{10}$$

The numerical elimination algorithm for a Macaulay matrix of degree three does not produce this system exactly, but instead returns five polynomial equations that have the same zero set. The numerical elimination procedure returns a set of linearly independent polynomials, but it does not enforce algebraic independence. As a consequence, the output may contain more equations than strictly necessary, even though they define the same zero set as the symbolically obtained relations in (10). The residual in Table I is calculated using the numerically obtained eliminant system. □

V. System Identification Problem

We apply elimination to the misfit-versus-latency model class, where finding the model parameters that "best" describe measured data is inherently a minimization problem with two competing objective functions. This example also serves to illustrate the potential of elimination-based techniques for computing Pareto fronts.

TABLE I

Characteristics of the numerical examples. Here, m and n denote the number of objectives and decision variables, respectively; $\#z$ is the total number of variables in the PF system; d is the minimal required degree of the Macaulay matrix; p_d and q_d are the matrix dimensions for that degree; and the final column reports the maximum residual obtained by evaluating the eliminant system at sampled points on the Pareto front.

problem	m	n	$\#z$	d	p_d	q_d	$\max \lVert \boldsymbol{t}(\boldsymbol{s}^\star) \rVert_2$
Example 4	3	2	7	2	19	36	3.16×10^{-15}
Example 5	2	2	6	8	3234	3003	2.79×10^{-12}
Example 6	3	1	6	3	49	84	4.73×10^{-15}

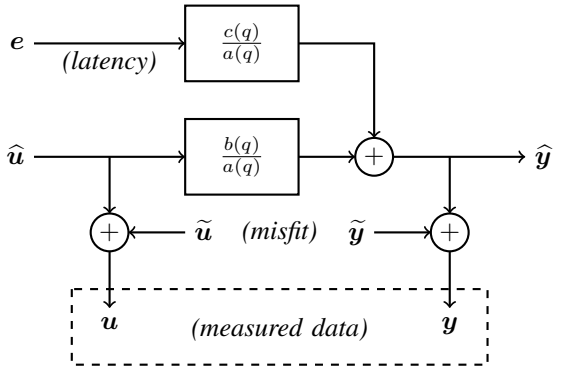


Fig. 2. Schematic representation of the misfit-latency model class: the measured data $(\boldsymbol{u}, \boldsymbol{y})$ is described by a latent input term $(\boldsymbol{e})$ and misfit term $(\widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{y}})$, such that the model-compliant data $(\widehat{\boldsymbol{u}}, \widehat{\boldsymbol{y}})$ satisfies a linear time-invariant model equation of the form $a(q)\widehat{\boldsymbol{y}} = b(q)\widehat{\boldsymbol{u}} + c(q)\boldsymbol{e}$.

We start firstly by giving a short background on the misfit-versus-latency model class (Section V-A). Afterwards, we explain how the Pareto front enters the picture (Section V-B) and solve the identification problem for some given output data (Section V-C). Notice that our approach is at this point not useful for real-life system identification problems with hundreds of data points, due to scale of the multivariate polynomial optimization problem and its corresponding computational complexity.

A. Misfit-versus-latency modeling

Misfit-versus-latency modeling [15] is a system identification framework for (dynamic) linear time-invariant models that combines the so-called "latency" approach in which residuals are introduced to explain the measured data (i.e., the prediction error approach [16] containing models such as autoregressive moving-average), and "misfit" approach that considers an error term between the measured and model-compliant data (i.e., the behavioral approach [17] containing models such as total least squares or errors-in-variables). The motivation for this modeling approach lies in the fact that there is no reason to exclude a misfit of the measurements in the presence of a latency term, and vice versa.

A schematic representation of the misfit-versus-latency model class is given in Fig. 2. It considers model equations

of the form

$$a(q)\widehat{\boldsymbol{y}} = b(q)\widehat{\boldsymbol{u}} + c(q)\boldsymbol{e}, \tag{11}$$

in which $\widehat{\boldsymbol{u}} \in \mathbb{R}^N$ and $\widehat{\boldsymbol{y}} \in \mathbb{R}^N$ are model-compliant input and output data of length N, respectively. The polynomials $a(q)$, $b(q)$, and $c(q)$, respectively of degrees n_a, n_b and n_c, are polynomials in the delay operator q and constitute the model dynamics. The vector $\boldsymbol{e} \in \mathbb{R}^{N-n_a+n_c}$ represents the unkown, latent input. Between the measured input-output data $\boldsymbol{u}$, $\boldsymbol{y}$ and the model-compliant input-output data $\widehat{\boldsymbol{u}}$, $\widehat{\boldsymbol{y}}$, there is a misfit

$$\widetilde{\boldsymbol{y}} = \boldsymbol{y} - \widehat{\boldsymbol{y}}, \quad \widetilde{\boldsymbol{u}} = \boldsymbol{u} - \widehat{\boldsymbol{u}}. \tag{12}$$

When identifying the model parameters, the goal is to minimize both the misfit and latency introduced in the model, leading to the following objective function

$$\frac{1}{2}\left(\alpha \left\|\widetilde{\boldsymbol{y}}\right\|_2^2 + \beta \left\|\widetilde{\boldsymbol{u}}\right\|_2^2 + \gamma \left\|\boldsymbol{e}\right\|_2^2\right), \tag{13}$$

with $\|\cdot\|_2$ denoting the Euclidean norm. The scalars α, β, and γ are nonnegative hyperparameters that determine the relative importance of the misfit and latency terms, and they can be infinity as well. Several well-known model classes arise for particular choices of these hyperparameters (see [15, Table 1]). Together with (11) and (12), the objective function (13) defines a constrained multivariate polynomial optimization problem. Using the KKT conditions, a system of multivariate polynomial equations can be constructed. For fixed values of α, β and γ, this multivariate polynomial system can be solved using any polynomial root-finding method.

B. Relation to the Pareto front

When considering misfit-versus-latency models, we actually have a trade-off between minimizing the output misfit, input misfit, and latency term. Rather than fixing the hyperparameters α, β and γ in advance, we can approach misfit-versus-latency modeling as a multi-objective polynomial minimization problem:

$$\begin{aligned} \min_{\boldsymbol{x}} \frac{1}{2}\begin{bmatrix}\left\|\widetilde{\boldsymbol{y}}\right\|_2^2 & \left\|\widetilde{\boldsymbol{u}}\right\|_2^2 & \left\|\boldsymbol{e}\right\|_2^2\end{bmatrix}^{\mathrm{T}} \\ \text{subject to } a(q)\widehat{\boldsymbol{y}} = b(q)\widehat{\boldsymbol{u}} + c(q)\boldsymbol{e}, \end{aligned} \tag{14}$$

where the decision variables $\boldsymbol{x}$ are $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, $\boldsymbol{e}$, $\widehat{\boldsymbol{y}}$, and $\widehat{\boldsymbol{u}}$. Via elimination, we aim to characterize the Pareto front that describes the optimal misfit and latency for each possible combination of hyperparameters, which constitute, after normalization, the weights $\boldsymbol{w}$.

C. Applied numerical example

As an applied numerical example, we consider a simplified version of the model in (11) that is autonomous and has no moving-average term ($n_c = 0$). The minimization problem in (14) reduces to the following weighted minimization problem:

$$\begin{aligned} \min_{\boldsymbol{a},\widehat{\boldsymbol{y}},\boldsymbol{e}} \ & \frac{1}{2}\left(\alpha \left\|\widetilde{\boldsymbol{y}}\right\|^2 + (1-\alpha)\left\|\boldsymbol{e}\right\|^2\right), \\ \text{subject to } & \widehat{\boldsymbol{Y}}\boldsymbol{a} = \boldsymbol{e} \end{aligned} \tag{15}$$

We only retain one hyperparameter and replace the other one (γ in this case) by $1-\alpha$, resulting in a convex combination. The matrix $\widehat{\boldsymbol{Y}} \in \mathbb{R}^{(N-n_a)\times(n_a+1)}$ is a Hankel matrix constructed from the model-compliant output data $\widehat{\boldsymbol{y}}$ and $\widehat{\boldsymbol{Y}}\boldsymbol{a} = \boldsymbol{e}$ is identical to considering $a(q)\widehat{\boldsymbol{y}} = \boldsymbol{e}$ for all data points. The Lagrangian is equal to

$$\begin{aligned} \mathcal{L}(\boldsymbol{a}, \widehat{\boldsymbol{y}}, \boldsymbol{e}, \boldsymbol{\lambda}) = \frac{1}{2}\left(\alpha \left\|\widetilde{\boldsymbol{y}}\right\|_2^2 + (1-\alpha)\left\|\boldsymbol{e}\right\|_2^2\right) \\ - \boldsymbol{\lambda}^{\mathrm{T}}\left(\widehat{\boldsymbol{Y}}\boldsymbol{a} - \boldsymbol{e}\right), \end{aligned}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^{N-n_a}$ are the Lagrange multipliers for the equality constraint. The corresponding KKT system is

$$\begin{aligned} -(\widehat{\boldsymbol{Y}}')^{\mathrm{T}}\boldsymbol{\lambda} &= \boldsymbol{0}, \\ \alpha(\boldsymbol{y} - \widehat{\boldsymbol{y}}) - \boldsymbol{T}_{\boldsymbol{a}}^{\mathrm{T}}\boldsymbol{\lambda} &= \boldsymbol{0}, \\ (1-\alpha)\boldsymbol{e} + \boldsymbol{\lambda} &= \boldsymbol{0}, \\ \widehat{\boldsymbol{Y}}\boldsymbol{a} - \boldsymbol{e} &= \boldsymbol{0}, \end{aligned} \tag{16}$$

in which $\widehat{\boldsymbol{Y}}'$ denotes the Hankel matrix $\widehat{\boldsymbol{Y}}$ without its rightmost column and $\boldsymbol{T}_{\boldsymbol{a}}$ is a banded Toeplitz matrix with $\boldsymbol{a}$ on its band diagonal. The PF system can be constructed from (16) supplemented with

$$\begin{aligned} s_1 - \left\|\widetilde{\boldsymbol{y}}\right\|_2^2 &= 0, \\ s_2 - \left\|\boldsymbol{e}\right\|_2^2 &= 0. \end{aligned}$$

Eliminating all decision variables except s_1 and s_2 from this multivariate polynomial system with $3N-2n_a+2$ equations yields an algebraic expression that contains the Pareto front of the identification problem.

Example 7: Given $N = 4$ observed output data points,

$$\boldsymbol{y} = \begin{bmatrix}1 & 4 & 2 & 3\end{bmatrix}^{\mathrm{T}},$$

we consider the weighted minimization problem in (15) for a first-order model, i.e., $n_a = 1$. We eliminate all variables in the PF system (which consists of twelve equations) except for s_1 and s_2, leading to one bivariate polynomial equation $t_1(s_1, s_2) = 0$ that contains the Pareto front. Its zero set is visualized in Fig. 3 as a solid blue line (——). We can verify that the obtained zero set contains the Pareto front by using a sampling approach. Some points of the sampled Pareto front are shown as black dots (●). Note that, with the approach in Section III-C, it is now possible to easily retrieve the different weights and model parameters along the Pareto front, providing insights into the models that describe the measured output data.

The two intersection points with the axes correspond to two specific model classes. On the one hand, the intersection with the s_2-axis corresponds to a pure autoregressive model without misfit. The value of s_2 at this intersection is equal to the norm of the latent input $\|\boldsymbol{e}\|_2^2$. When identifying a pure autoregressive model for the given measured output data (with the globally optimal techniques described in [18]), the same model parameters can be obtained. The corresponding objective values are visualized in Fig. 3 with the green dot (●). The intersection with the s_1-axis, on the other hand, resembles an autonomous model with output misfit.

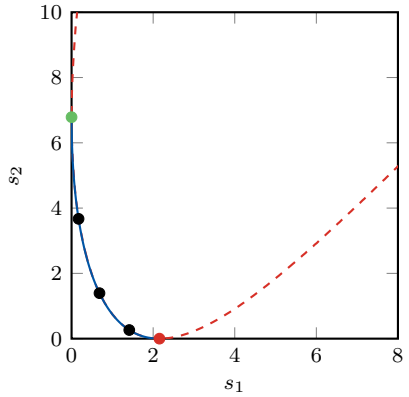


Fig. 3. Visualization of the Pareto front for the misfit-versus-latency modeling problem in Example 7. The Pareto front (——) is a subset of the eliminant (- - -) obtained through the elimination approach described in Section III-B, and it is validated by sampling Pareto-efficient points (●). The two intersection points with the axes correspond to a pure autoregressive model without misfit (●) and a pure autonomous model with output misfit (●), which can also be obtained using dedicated identification algorithms.

Computing the misfit in the case that the latent input is zero yields the same value as when using the globally optimal methods from [19], as is shown with the red dot (●). It is clear that along the Pareto front, the mismatch between the measured output data and model is distributed among these two modifications. □

VI. Conclusions and Future Work

We presented a novel approach to compute the Pareto front in multi-objective minimization for multivariate polynomial optimization problems. Through systematic elimination of all variables except the objective values, the method produces an exact algebraic description that contains the entire Pareto front. In contrast to sampling-based methods, our elimination-based approach does not depend on a sampling distance or *a priori* weight assumptions. When elimination is performed numerically, it handles floating-point coefficients naturally. Several numerical examples confirmed the correctness of the proposed elimination-based approach and illustrate how elimination yields algebraic relations that describe the Pareto front.

We applied elimination also in the context of system identification: computing the Pareto-efficient trade-off between misfit and latency when identifying the model from given output data. This example highlighted an interesting potential application of elimination-based techniques for computing Pareto fronts. The resulting Pareto front confirms the theoretical idea in [15] that a continuum of equivalent models exists, with the pure latency and pure misfit models as limiting cases. Crucially, the Pareto front gives engineers a principled, complete picture of the trade-off landscape, enabling an informed model selection rather than an arbitrary, *a priori* choice of hyperparameters.

Although our elimination-based approach looks very promising, two challenges that popped up during the system identification problem merit attention in future work. Firstly, the computational complexity remains a bottleneck for problems with many variables or high polynomial degrees; a hierarchical elimination strategy could enable scaling to larger problems. Secondly, extending the elimination idea beyond polynomial objective functions is an important open direction that we intend to pursue in the near future.

References

[1] K. Miettinen, *Nonlinear Multiobjective Optimization*, 1st ed. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1999.

[2] V. Pareto, *Cours d'Économie Politique*. Lausanne, Switzerland: F. Rouge, 1896, vol. 1, *in French*.

[3] T. R. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Structural and Multidisciplinary Optimization*, vol. 26, pp. 369–395, 2004.

[4] W. Stadler, Ed., *Multicriteria Optimization in Engineering and in the Sciences*, ser. Mathematical Concepts and Methods in Science and Engineering. New York, NY, USA: Springer, 1988, vol. 1.

[5] C. Vermeersch, S. Lagauw, and B. De Moor, "Multivariate polynomial optimization in complex variables is a (rectangular) multiparameter eigenvalue problem," in *Proc. of the 62nd IEEE Conference on Decision and Control (CDC)*, Marina Bay Sands, Singapore, 2023, pp. 7305–7311.

[6] A. M. Geoffrion, "Proper efficiency and the theory of vector maximization," *Journal of Mathematical Analysis and Applications*, vol. 22, pp. 618–630, 1968.

[7] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed., ser. Springer Series in Operations Research and Financial Engineering. New York, NY, USA: Springer, 2006.

[8] H. Markowitz, "Portfolio selection," *The Journal of Finance*, vol. 7, no. 1, pp. 77–91, 1952.

[9] K. Batselier, P. Dreesen, and B. De Moor, "A geometrical approach to finding multivairate approximate LCMs and GCDs," *Linear Algebra and its Applications*, vol. 438, no. 9, pp. 3618–3628, 2013.

[10] Z. Zeng, "A numerical elimination method for polynomial computations," *Theoretical Computer Science*, vol. 409, no. 2, pp. 318–331, 2008.

[11] H. van Rooij and B. De Moor, "Numerical elimination theory via the Macaulay matrix," KU Leuven, Leuven, Belgium, Tech. Rep., 2026. [Online]. Available: https://ftp.esat.kuleuven.be/stadius/hvanrooi/numerical_elimination_theory.pdf

[12] C. Vermeersch, "The (block) Macaulay matrix," Ph.D. dissertation, KU Leuven, Leuven, Belgium, 2023.

[13] D. Fearnley-Sander, "Hermann Grassmann and the creation of linear algebra," *The American Mathematical Monthly*, vol. 86, no. 10, pp. 809–817, 1979.

[14] C. Vermeersch and B. De Moor, "Solving multivariate polynomial systems and rectangular multiparameter eigenvalue problems with MacaulayLab," KU Leuven, Leuven, Belgium, Tech. Rep., 2025.

[15] P. Lemmerling and B. De Moor, "Misfit versus latency," *Automatica*, vol. 37, no. 12, pp. 2057–2067, 2001.

[16] L. Ljung, *System Identification: Theory for the User*, 2nd ed., ser. Prentice Hall Information and System Science Series. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1999.

[17] J. C. Willems, "From time series to linear systems — Part I. finite dimensional linear time invariant systems," *Automatica*, vol. 22, no. 5, pp. 561–580, 1986.

[18] C. Vermeersch and B. De Moor, "Globally optimal least-squares ARMA model identification is an eigenvalue problem," *IEEE Control Systems Letters*, vol. 3, no. 4, pp. 1062–1067, 2019.

[19] S. Lagauw, L. Vanpoucke, and B. De Moor, "Exact solution to the least squares realization problem as multiparameter eigenvalue problem," *Automatica*, vol. 185, p. 112792, 2026.