

OPTIMAL DESIGNS FOR RATIONAL FUNCTION REGRESSION

DÁVID PAPP

ABSTRACT. We consider optimal non-sequential designs for a large class of (linear and nonlinear) regression models involving polynomials and rational functions with heteroscedastic noise also given by a polynomial or rational weight function. The proposed method treats D-, E-, A-, and Φ_p -optimal designs in a unified manner, and generates a polynomial whose zeros are the support points of the optimal approximate design, generalizing a number of previously known results of the same flavor. The method is based on a mathematical optimization model that can incorporate various criteria of optimality and can be solved efficiently by well established numerical optimization methods. In contrast to previous optimization-based methods proposed for similar design problems, it also has theoretical guarantee of its algorithmic efficiency; in fact, the running times of all numerical examples considered in the paper are negligible. The stability of the method is demonstrated in an example involving high degree polynomials. After discussing linear models, applications for finding locally optimal designs for nonlinear regression models involving rational functions are presented, then extensions to robust regression designs, and trigonometric regression are shown. As a corollary, an upper bound on the size of the support set of the minimally-supported optimal designs is also found. The method is of considerable practical importance, with the potential for instance to impact design software development. Further study of the optimality conditions of the main optimization model might also yield new theoretical insights.

1. INTRODUCTION

This paper is concerned with optimal approximate designs for polynomial and rational regression models with heteroscedastic error modeled by a rational weight function. In our focus is the general linear model

$$y(t) = \sum_{i=1}^m \theta_i f_i(t) + \varepsilon(t), \quad t \in \mathcal{I}, \quad (1)$$

where each f_i is a known rational function defined on \mathcal{I} , and the error $\varepsilon(t)$ is a normally distributed random variable with mean zero and variance $\sigma^2(t) = 1/\omega(t)$, where the known *weight function* ω is a rational function whose numerator and denominator are both positive on \mathcal{I} . We are interested in experiments designed to help estimate the unknown parameters θ_i . The design space \mathcal{I} is the finite union of

Key words and phrases. Optimal design, Approximate design, Rational function regression, Semi-definite programming, Linear matrix inequality.

closed, bounded intervals in \mathbb{R} , also allowing singletons as degenerate intervals. We assume that observations are uncorrelated, and that the f_i are linearly independent.

Our main result is a characterization of the support of the D-, E-, A-, and Φ_p -optimal designs as the optimal solutions of a semidefinite optimization problem. This directly translates to a method to numerically determine the optimal design, using readily available optimization software. The characterization is applicable to every linear model involving polynomials and rational functions with heteroscedastic noise also given by a polynomial or rational weight function. We demonstrate that the method is numerically robust (in the sense that it can handle ill-conditioned problems, such as those involving polynomials of high degree), and has very short running time on problems of practical size.

Optimal designs for Fourier regression models and locally optimal designs for certain nonlinear models can also be found with similar methods.

In many cases the experimenter is interested only in certain linear combinations of the parameter vector $\theta := (\theta_1, \dots, \theta_m)^\top$, which are given by the components of $K^\top \theta$ for some $m \times s$ matrix K . In the presentation of our approach it is convenient to assume that our goal is to estimate the entire parameter vector, that is $K = I_m$ (the $m \times m$ identity matrix), and that the design space contains enough points to make all parameters estimable. (If $K = I_m$, the latter assumption means that there is a design whose information matrix is non-singular, see later.) In Section 6 we show how the proposed method can be generalized to handle problems with general K .

Much attention has been devoted to optimal designs for special cases of model (1). It is well known that when the design space \mathcal{I} is finite, the D-, E-, and A-optimal approximate designs can be found by convex optimization even for arbitrary f_i 's, see, for example [4, Chapter 7], or a generalization of this approach to multi-response experiments in [2]. However, when \mathcal{I} is an interval, considerable difficulties arise, as the finite support of the optimal design also has to be characterized.

A popular approach in the literature is that a polynomial is sought whose roots are the support points of the optimal design. For instance, as discovered by Guest [18] and Hoel [21], the D-optimal design for ordinary polynomial regression, when $f_i = t^i$, and ω is a positive constant, on $\mathcal{I} = [-1, 1]$ is the one that assigns uniform weights to each of the zeros of $t \rightarrow (1 - t^2) \frac{d}{dt} L_m(t)$, where L_m is the Legendre polynomial of degree m . The number of support points had already been determined in [7]. Similar characterizations are known for A- and E-optimal designs for polynomial regression, see, for example the classic monographs [14, 33]. Another common approach is to determine the canonical moments of the optimal design [11, 12]. Further optimality criteria for polynomial models, and closed-form characterizations of the optimal designs for linear and quadratic models, are discussed in [36]. See also [24] for E-optimal designs for linear models with rational functions $f_i(t) = (t - \alpha_i)^{-1}$ with $\alpha_i \notin \mathcal{I}$. The Optimum Experimental Design website [1] also contains a rather comprehensive list of solved models, along with an impressive, and continuously maintained, list of references.

More recently considerable attention has been paid to polynomial models with missing interactions, also called *incomplete* or *improper* polynomial models. Representative results include [8], which gives D-optimal designs when only odd or only even degree terms appear in the model; [22] and [5], which consider D- and E-optimal designs (respectively) for polynomial models with zero constant term; [11], which considers D-optimal designs, also for some multivariate problems, over the unit cube under less restrictive assumptions on the missing terms; and [13], which gives D-optimal designs when only the lowest degree terms, up to a fixed degree m' , are absent. Note that even the union of these methods does not yield a complete solution to incomplete polynomial models, even for univariate regression with homoscedastic error.

Results in the heteroscedastic case are even more scarce and typically less general. For instance, [23] is devoted to D-optimal designs for polynomial regression over $[0, b]$ with the weight function $\omega(t) = t/(1+t)$.

The design space \mathcal{I} is almost always a (closed, bounded) interval, which is probably sufficient for most applications. Imhof and Studden [24] also considered some rational models when \mathcal{I} is the union of two disjoint intervals.

Most of the above results are based on the theory of orthogonal polynomials, canonical moments [12], and Chebyshev systems [25]. They are rather specific in their scope, and generalization of their proofs appears to be difficult. On the other hand, most of them yield numerically very efficient methods for computing numerically optimal designs. The bottleneck in these methods is either polynomial root-finding, which can be carried out in nearly linear time in the degree of the polynomial [32], or the reconstruction of a measure on finite support from its canonical moments, which can also be carried out relatively easily [12]. An exception is the method of [13], which involves finding the global maximum of a *multivariate* polynomial (even though it is concerned with univariate polynomial regression only). This is an NP-hard problem even in very restricted classes of polynomials, and is known to be very difficult to solve in practice even when the number of variables and the degree are rather small [20].

In the pursuit of more widely applicable methods, some of the attention has turned to the numerical solution of optimization models that characterize optimal designs. Pukelsheim's monograph [33] is a comprehensive overview of optimal design problems with an optimization-oriented viewpoint, but it is not concerned with algorithms or numerical computations. Most numerical methods proposed in the literature are variants of the popular *coordinate-exchange method* from [29], which is a variant of the classic Gauss–Seidel method (also known as coordinate descent method) used in derivative-free optimization. These algorithms maintain a finite working set of support points, and iteratively replace one of the support points by another one from \mathcal{I} if the optimal design on the new support set is better than that of the current support set. See [6] for a recent variant of this idea for finding approximate D-optimal designs.

However, this approach has serious drawbacks, and care has to be taken not to abuse them: (i) some variants require that the size of the minimally supported optimal design be known *a priori*; (ii) no bound is known on the number of iterations the algorithm might take; (iii) in fact, the number of iterations of the coordinate descent method is known to be quite high in practice even for some very simple convex optimization problems [31, Chapter 9]; and (iv) the coordinate descent method does not necessarily converge at all if the function being optimized is not continuously differentiable [35]. Hence, these methods can hardly be considered a completely satisfactory solution of most polynomial regression problems, even though some successful numerical experiments have been reported, cf. [6].

This paper proposes a different approach to linear regression models involving polynomials and rational functions. Motivated in part by the approach of [4], it is also based on an optimization model involving linear matrix inequalities, which can be solved efficiently, both in theory and in practice, by readily available optimization software.

The novelty of the proposed method is that it does not work with the support points directly, as existing numerical methods, such as the coordinate-exchange method, do. Instead, it follows some of the previous symbolic approaches by computing the coefficients of a polynomial whose zeros are the support points of the optimal design.

After introducing the problem formally, we derive our main theorems in Section 3 for the estimation of the full parameter vector θ . Illustrative examples are presented in Section 4. Section 6 is concerned with the more general case, when only a subset of the parameters (or their linear combinations) need to be estimated. We then apply these results to finding locally optimal designs for nonlinear models in Section 7. Finally, in Section 8 we give an outlook to models of regression involving other functions than rational functions.

Notation. We will make use of the following, mostly standard, notations: $\deg p$ denotes the degree of the polynomial p , lcm stands for the least common multiple of polynomials. The denominator of a rational function r is denoted by $\text{den}(r)$. The positive part function is denoted by $(\cdot)_+$. The brackets $\langle \cdot, \cdot \rangle$ denote the usual (Frobenius) inner product of vectors and matrices, that is, $\langle A, B \rangle = \sum_{ij} A_{ij}B_{ij}$. Since many decision variables in the paper are matrices, linear constraints on matrices are written in operator form. For example, a linear equality constraint on an unknown matrix X will be written as $A(X) = b$ (where A is a linear operator and b is a vector) to avoid the cumbersome “vec” notation necessary to use matrix-vector products. For the linear operator A , A^* denotes its adjoint. The identity operator is written as id .

The space of $m \times m$ symmetric matrices is denoted by \mathbb{S}^m , the cone of $m \times m$ positive semidefinite real symmetric matrices is \mathbb{S}_+^m . The *Löwner partial order* on \mathbb{S}^m , denoted by \succcurlyeq , is the conic order generated by \mathbb{S}_+^m ; in other words, we write $A \succcurlyeq B$ when $A - B \in \mathbb{S}_+^m$.

2. OPTIMALITY CRITERIA AND THEIR SEMIDEFINITE REPRESENTATIONS

A *design for infinite sample size* (also called *approximate design* or *design* for short) is a finitely supported probability measure ξ on \mathcal{I} . Using the notation $f(t) = (f_1(t), \dots, f_m(t))^T$, the *Fisher information matrix* of θ corresponding to the design ξ is

$$M(\xi) = \int_{\mathcal{I}} f(t)f(t)^T \omega(t) d\xi(t). \quad (2)$$

Of course, this integral simplifies to a finite sum for every design. Note that for every ξ , $M(\xi) \in \mathbb{S}_+^m$. A design $\hat{\xi}$ is considered *optimal* if $M(\hat{\xi})$ is maximal with respect to the Löwner partial order (recall the end of the previous section); see [33, Chapter 4] for detailed statistical interpretation. If Φ is an $\mathbb{S}_+^m \rightarrow \mathbb{R}$ function, the design $\hat{\xi}$ is called *optimal with respect to Φ* , or *Φ -optimal* for short, if $\Phi(M(\hat{\xi}))$ is maximum. Again, only those criteria are interesting which are compatible with the Löwner partial order, that is functions Φ satisfying $\Phi(A) \geq \Phi(B)$ whenever $A \succcurlyeq B \succcurlyeq 0$. Popular choices of Φ include the following.

- (1) When $\Phi(M) = \det(M)$, $\hat{\xi}$ is called *D-optimal*.
- (2) When $\Phi(M) = \lambda_1(M)$, the smallest eigenvalue of M , $\hat{\xi}$ is called *E-optimal*.
- (3) When $\Phi(M) = -\text{tr}(M^{-1})$, where tr denotes matrix trace, $\hat{\xi}$ is called *A-optimal*.
- (4) When $\Phi(M) = (\text{tr}(M^p))^{1/p}$, $\hat{\xi}$ is called *Φ_p -optimal*.

For most purposes of the paper Φ could be an arbitrary concave extended real valued function on \mathbb{S}_+^m with finite values on the interior of \mathbb{S}_+^m . However, to avoid certain technical difficulties, and in order to obtain good characterizations of optimal designs, we will assume that the Φ of our choice is *representable by linear matrix inequalities* (LMIs) or *semidefinite representable*, this includes all of the criteria discussed above. The precise definitions we need are summarized next.

Definition 1. A set $S \subseteq \mathbb{R}^n$ is *semidefinite representable* if for some $k \geq 1$ and $l \geq 0$ there exist affine functions $A : \mathbb{R}^n \rightarrow \mathbb{S}^k$ and $C : \mathbb{R}^l \rightarrow \mathbb{S}^k$ such that the set S can be characterized by a linear matrix inequality in the following way:

$$S = \{s \in \mathbb{R}^n \mid \exists u \in \mathbb{R}^l : A(s) + C(u) \succcurlyeq 0\}.$$

Note that the intersection of semidefinite representable sets is also semidefinite representable, so we could equivalently allow to have a characterization of the above form with $p \geq 1$ inequalities. The motivation behind the idea of semidefinite representable sets is that finding global optima of “nice” functions over them is easy, and a number of numerical methods are available to that in an efficient manner. “Nice” functions include semidefinite representable functions, defined below, in Definition 3.

In this paper we will encounter two important instances of semidefinite representable sets: the coefficient vectors of polynomials that are nonnegative over an interval, and the level sets of the optimality criteria Φ .

Lemma 2 ([25, Chapter 2]). *The set*

$$P_d^{[a,b]} = \left\{ (p_0, \dots, p_d) : \sum_{i=0}^d p_i x^i \geq 0 \quad \forall x \in [a, b] \right\}$$

of coefficient vectors of polynomials of degree d that are nonnegative over the interval $[a, b]$ is a semidefinite representable subset of \mathbb{R}^{d+1} .

This is a reasonably well known theorem in probability and statistics owing to its application in moment problems [12], for completeness we provide a specific representation in the Appendix. The same assertion holds even if the polynomials are represented in another basis, not in the monomial basis, but the actual characterization will, of course, be different.

The next definition is necessary to define the class of optimality criteria our approach can handle.

Definition 3. A function $\Phi : \mathbb{S}_+^m \rightarrow \mathbb{R}$ is *semidefinite representable* if its (closed) upper level sets are semidefinite representable, that is, if for some k_1, \dots, k_p and l there exist linear functions $A_i : \mathbb{S}_+^m \rightarrow \mathbb{S}^{k_i}$, $C_i : \mathbb{R}^l \rightarrow \mathbb{S}^{k_i}$, and matrices $B_i \in \mathbb{S}^{k_i}$, $D_i \in \mathbb{S}^{k_i}$ ($i = 1, \dots, p$) such that for all $X \in \mathbb{S}_+^m$ and $z \in \mathbb{R}$, $\Phi(X) \geq z$ holds if and only if

$$A_i(X) + B_i z + C_i(u) + D_i \succcurlyeq 0 \quad i = 1, \dots, p \quad (3)$$

for some $u \in \mathbb{R}^l$.

As mentioned above, finding the optimal value (and the optimizer) of a semidefinite representable function over a semidefinite representable set is generally easy; optimization problems of this form are called *semidefinite optimization problems* or *semidefinite programs*; see also the beginning of the next section.

We will also need the following (technical) assumption on the relationship between the model (as defined by the functions f_i and ω) and the criterion function Φ . It is only used in the proof of the main theorem.

Definition 4. We say that the semidefinite representable function $\Phi : \mathbb{S}_+^m \rightarrow \mathbb{R}$ is *admissible* with respect to the set $\mathcal{X} \subseteq \mathbb{S}_+^m$ if Φ has a representation (3) for which there exists an $\hat{X} \in \mathcal{X}$ satisfying (3) with strict inequality for some z and u . That is to say that the left-hand side of each of the p inequalities can be made positive definite simultaneously for at least one $\hat{X} \in \mathcal{X}$.

This is a rather technical condition in the sense that most interesting functions Φ are admissible with respect to every non-empty set \mathcal{X} (a sufficient condition for this is that in the semidefinite representation of Φ each B_i be positive or negative definite), or at least with respect to every \mathcal{X} that contains a non-singular matrix.

D-, E-, and A-optimality are all semidefinite representable, or are equivalent to other criteria given by semidefinite representable functions. The same holds for Φ_p -optimality. They are also admissible with respect to every set of Fisher information matrices for which the criteria is well-defined (see below). Note that all semidefinite representable functions are quasi-concave, continuous functions.

Example 5 (E-optimality). For every $M \in \mathbb{S}^m$, $\lambda_1(M) \geq z$ if and only if $M - zI \succcurlyeq 0$, so λ_1 admits a simple semidefinite representation. In this representation $p = 1$, $A_1 = \text{id}$, $B_1 = -I$, $C_1 \equiv 0$, and $D_1 = 0$, hence λ_1 is admissible with respect to every non-empty set of Fisher information matrices.

Example 6 (A-optimality). It follows from Haynsworth's theorem [19] on the inertia of Hermitian block matrices that a symmetric block matrix $\begin{pmatrix} P & Q \\ Q^T & R \end{pmatrix}$ with positive definite block P is positive semidefinite if and only if its *Schur complement*, given by $R - Q^T P^{-1} Q$, is positive semidefinite. Let $M \in \mathbb{S}_+^m$ be invertible, for example an invertible Fisher-information matrix, and fix a $k \in \{1, \dots, m\}$. Plugging in M for P , the k th unit vector e_k for Q^T , and a scalar u for R we have that $(M^{-1})_{k,k} \leq u$ if and only if $\begin{pmatrix} M & e_k \\ e_k^T & u \end{pmatrix} \succcurlyeq 0$. This observation yields a semidefinite representation of A-optimality of the form (3) with $p = m + 1$:

$$\text{tr}(M^{-1}) \leq z \quad \text{iff} \quad \exists u_1, \dots, u_m: z \geq \sum_i u_i, \quad \text{and} \quad \begin{pmatrix} M & e_k \\ e_k^T & u_k \end{pmatrix} \succcurlyeq 0, \quad k = 1, \dots, m.$$

It follows that the A-optimality criterion is admissible with respect to every set of Fisher information matrices that contains at least one non-singular matrix.

Example 7 (D- and Φ_p -optimality). The cases of D-optimality and Φ_p -optimality are more complicated, but can also be fitted in the above framework. Owing to page limitations we can only give the flavor of this result, and pointers to the literature.

D-optimality is equivalent to optimality with respect to the criterion $\Phi(M) = (\det(M))^{1/m}$, where m is the size of M . Note that this is the geometric mean of the eigenvalues of M . Φ_p -optimality is expressed by the matrix mean $\Phi_p(M) = (\text{tr}(M^p))^{1/p} = (\sum_{i=1}^m \lambda_i^p)^{1/p}$, where λ_i is the i th eigenvalue of M . Hence, both criteria are symmetric functions of the eigenvalues of M . Moreover, both the geometric mean and the p -norm, for every rational $p \geq 1$ are also semidefinite representable [3, Section 3.3.1]. Finally, we can invoke [3, Proposition 4.2.1], which states that for every semidefinite representable symmetric $g: \mathbb{R}^m \rightarrow \mathbb{R}$, the function $\Phi(M) = g(\lambda_1(M), \dots, \lambda_m(M))$ is also semidefinite representable.

D- and Φ_p -optimality are also admissible with respect to every set of Fisher information matrices that contains at least one non-singular matrix.

Another interesting optimality criterion, not considered in this paper, is the maximin efficient criterion. Models for which maximin efficient approximately optimal designs can be found using semidefinite programming (this includes polynomial models) can be found in the recent technical report [15].

3. OPTIMAL DESIGNS AND SEMIDEFINITE OPTIMIZATION

First we shall give a very short introduction to semidefinite optimization to summarize the background necessary to keep this paper self-contained. The reader is also encouraged to consult [39]; or [40] for a considerably more in-depth survey to this vast field.

Semidefinite optimization (or semidefinite programming) is a generalization of the familiar linear optimization. A *semidefinite program* (or SDP for short) is the mathematical problem of finding the optimum of a linear function subject to the constraint that an affine combination of matrices is positive semidefinite. In other words, it is an optimization problem of the form

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{minimize}} && \sum_i c_i x_i \\ & \text{subject to} && A_0 + \sum_{i=1}^n A_i x_i \succcurlyeq 0, \end{aligned} \tag{4}$$

where $c \in \mathbb{R}^n$ and $A_i \in \mathbb{S}^m$, ($i = 0, \dots, n$) are given; x_i denotes the i th component of the vector x ; these are the variables.

Constraints of the above form are called *semidefinite constraints* or *linear matrix inequalities*. The format of problem (4) is regarded as a “standard form”, but other, seemingly more general optimization problems that can be converted to the above form are also considered semidefinite programming problems. In particular, multiple semidefinite constraints can be added to the problem, and the constraints can be augmented by linear inequalities and equations, as these translate to constraints on diagonal matrices. Matrices of variables can also be considered, and constrained simultaneously in the form $L(X) \succcurlyeq C$; here X is the matrix of variables, L is a linear operator, and C is a matrix of appropriate size. More generally, the maximization of every semidefinite representable function over every semidefinite representable set (as defined in the previous section) can be cast as an SDP. In this paper we will show that finding the support of the optimal design can be cast as an SDP of this more general form, for every regression model (1).

Semidefinite programs are special convex optimization problems, and the standard duality theory of convex optimization [34, 35] applies to them. Algorithms to numerically compute the optimal solutions of a semidefinite program have been well studied for more than two decades; SDPs involving tens of thousands of variables are routinely solved in the literature [39]. The SDPs of this paper are considerably smaller; they can be solved in a fraction of a second without any numerical issues by commonly used SDP solver software, such as SeDuMi [37], a freely available Matlab toolbox. Additional toolboxes, such as CVX [17] and YALMIP [27], are available to translate “high-level” semidefinite programs involving semidefinite functions such as the optimality criteria mentioned in this paper to the semidefinite programs in the above “standard” form.

3.1. Semidefinite representation of optimal designs. The main result in this section, and of the paper, is that the problem of finding an optimal design with respect to Φ can be equivalently written as a semidefinite programming problem whenever the functions $f_i, i = 1, \dots, m$ and ω are rational functions defined over a finite union \mathcal{I} of closed intervals, and Φ is a semidefinite representable function that

satisfies the mild technical condition that it is admissible with respect to the set of all Fisher information matrices.

As mentioned in the Introduction, this is already known for design spaces \mathcal{I} consisting of finitely many points, even for arbitrary $\{f_i\}$. While it is not stated there in this general form, the following theorem is implicit in [4, Chapter 7]:

Theorem 8 ([4, Chapter 7]). *Let $\mathcal{I} \subset \mathbb{R}$ be finite, and Φ be a semidefinite representable function compatible with the Löwner partial order. Then the Φ -optimal designs for model (1) are characterized as the set of optimal solutions to a semidefinite programming problem.*

In this semidefinite programming problem the support points are fixed parameters, and the variables are the masses the optimal design assigns to the support points; hence Theorem 8 allows us to find the optimal design only once its support is known. Treating the support points as variables would be problematic for two reasons: the number of support points for the optimal design may not be known, and even if it was, the resulting optimization problem would be intractable. Our goal in this paper is to characterize the support of the optimal design as a solution of a semidefinite program. In the optimization problem we are about to define, the variables are the *coefficients of a polynomial whose roots are the support points of the optimal design*.

Our main result, Theorem 9 below, is the characterization of the support of the optimal design as a solution of a semidefinite program. After finding the support, Theorem 8 can be applied to find the weights—by solving another semidefinite program.

Theorem 9. *Suppose that in the linear model (1) \mathcal{I} is a finite union of closed intervals, the functions f_i are rational functions with finite values on \mathcal{I} , and ω is a nonnegative rational function on \mathcal{I} . Let Φ be an admissible semidefinite representable function (with representation (3)) with respect to the set of Fisher information matrices $\mathcal{M} = \text{conv}\{f(t)f(t)^\top\omega(t) \mid t \in \mathcal{I}\}$. Then the support of the Φ -optimal design is a subset of the real zeros of the polynomial π obtained by solving the following semidefinite programming problem:*

$$\begin{array}{l} \text{minimize } y \\ \begin{array}{l} y \in \mathbb{R}, \pi \in \mathbb{R}^d, \\ W_1, \dots, W_p \in \mathbb{S}_+^k \end{array} \end{array} \quad (5a)$$

$$\text{subject to } \sum_{i=1}^p \langle W_i, B_i \rangle = -1, \quad \sum_{i=1}^p C_i^*(W_i) = 0, \quad (5b)$$

$$\pi = \Pi(y, W_1, \dots, W_p), \quad (5c)$$

$$\pi \in P_d^{\mathcal{I}}, \quad (5d)$$

where d is the degree of the polynomial

$$t \rightarrow \text{lcm}(\text{den}(\omega), \text{den}(f_1^2), \dots, \text{den}(f_p^2)) \left(y - \sum_{i=1}^p \langle W_i, A_i(M(\xi_t)) + D_i \rangle \right), \quad (6)$$

whose coefficient vector is denoted by $\Pi(y, W_1, \dots, W_p)$ in (5c) above.

Note that the operator Π in (6) is affine, hence aside from (5d) every constraint in (5) is a linear equation or linear matrix inequality. Furthermore, (5d) can be translated to linear matrix inequalities using Lemma 2. Hence, (5) is indeed a semidefinite program.

Not wanting to defer the discussion of examples and extensions, the proof was moved to the Appendix. Instead, we discuss a few examples.

4. EXAMPLES

We start with two detailed examples demonstrating how E- and A-optimal design problems translate to semidefinite optimization models. Then the numerical robustness of the proposed method is investigated using a high degree polynomial model. Finally, an example with rational models is shown, in which the parameters of point sources emitting radiation are estimated from measurements of total intensity.

All timing results in this paper were obtained using the semidefinite solver SeDuMi [37] running on an ordinary desktop computer with a 2.83GHz processor, using a single core.

Example 10 (E-optimal designs without an intercept). This problem was considered in [5], and we use it here to illustrate the steps of the approach and to verify the correctness of our model in a relatively high degree model that has been solved: $\mathcal{I} = [-1, 1]$, $f_i = t^i$, $i = 1, \dots, m$, and ω is a positive constant. Using the semidefinite representation of E-optimality given in Example 5, the variables in the optimization model of Theorem 9 are the scalar y and the positive semidefinite matrix W_1 of order m . The constraints can be derived as follows: from Example 5 we have $A_1 = \text{id}$, $B_1 = -I$, $C_1 \equiv 0$, and $D_1 = 0$, hence $\langle B_1, W_1 \rangle = -\text{tr}(W_1)$, and $\langle C_1, W_1 \rangle = 0$. Also note that

$$M(\xi_t) = (t, t^2, \dots, t^m)^T (t, t^2, \dots, t^m) = \begin{pmatrix} t^2 & t^3 & \dots & t^{m+1} \\ t^3 & t^4 & \dots & t^{m+2} \\ \vdots & \vdots & \ddots & \vdots \\ t^{m+1} & t^{m+2} & \dots & t^{2m} \end{pmatrix},$$

where the last matrix has t^{i+j} as its (i, j) -th entry.

Hence, the first constraint of (5b) is $\text{tr}(W_1) = 1$, whereas the second constraint of (5b) is simply $0 = 0$, and can be omitted. We have $\deg(\pi) = 2m$, and the correspondence between the entries of W_1 and the coefficients of $\pi(t) = \sum_{i=0}^{2m} p_i t^i$, given in (6), simplifies to the system of equations

$$p_0 = y, \quad p_1 = 0, \quad \text{and} \quad p_k = - \sum_{i+j=k} (W_1)_{ij} \quad \text{for } k = 2, \dots, 2m.$$

```

cvx_begin
    m = 8;
    variable y;
    variable W(m,m) symmetric;
    variable pi(2*m+1);

    minimize y;
    subject to
        W == semidefinite(m);
        trace(W) == 1;
        pi(1) == y;
        pi(2) == 0;
        -pi(3) == W(1,1);
        ...
        -pi(17) == W(8,8);

        pi(end:-1:1) == nonneg_poly_coeffs(2*m, [-1,1]);
cvx_end

```

FIGURE 1. Matlab solution for Example 10 using the CVX toolbox. Note that Matlab indexes vectors starting from 1 instead of 0. The equations defining the coefficients π_4 through π_{16} have been omitted for brevity.

In summary, dropping the subscript from W_1 , we have the optimization problem

$$\begin{aligned}
 & \underset{y \in \mathbb{R}, \pi \in \mathbb{R}^{2m}, W \in \mathbb{S}_+^m}{\text{minimize}} && y \\
 & \text{subject to} && \text{tr}(W) = 1, \\
 & && \pi = (y, 0, S_2, \dots, S_{2m}) \in P^{[-1,1]},
 \end{aligned}$$

where $S_k = -\sum_{i+j=k} W_{ij}$ ($k = 2, \dots, 2m$) are the anti-diagonal sums of the matrix W , and the constraint $\pi = (y, 0, S_2, \dots, S_{2m}) \in P^{[-1,1]}$ can be turned into the system of linear and semidefinite constraints (17) given in the Appendix, plugging in $a = -1, b = 1$.

For practical computations several Matlab toolboxes, such as CVX [17] and YALMIP [27], are available to facilitate the translation of semidefinite programs such as the one above to the so-called “standard form” required by semidefinite solvers. Rather than providing a detailed description or comparison of these programs, we offer a completely self-explanatory example, the formulation of the above problem in the language of the CVX toolbox, in Figure 1. Note that both the trace constraint and the nonnegative polynomial constraint are represented at the same high level in the code as in the mathematical model above. They are translated to a standard form semidefinite program and solved using a semidefinite programming solver automatically by CVX, leaving virtually no work to the user.

For example, solving the resulting problem for $m = 8$, the optimal vector π is the coefficient vector of a degree 16 polynomial whose real roots are: $\{\pm 1, \pm 0.9207, \pm 0.693, \pm 0.3357\}$. It also has two imaginary roots. The eight real roots constitute the support of the E-optimal design. The same numerical example was considered in [5] with, of course, the same conclusion. The running time of SeDuMi in this example was 0.2 seconds.

Example 11 (A heteroscedastic polynomial model). Consider the cubic model $f_i = t^{i-1}$, $i = 1, \dots, 4$, with heteroscedastic noise given by $\omega(t) = 1/(1+t^2)$, over the design space $[-5, 5]$. We chose this arbitrary model because it is one of the simplest among those whose solution appears to not to be characterized.

The A-optimal design is computed as follows. The parameters A_i, B_i, C_i, D_i in the semidefinite representation (3) of A-optimality are determined first from Example 6. Using this representation, the constraints of the semidefinite programming problem in Theorem 9 are compiled in the following way.

- There are 4 semidefinite matrices W_1, \dots, W_4 of order 5, and W_5 is a nonnegative scalar.
- The first constraint of (5b) is simply $W_5 = 1$. The second constraint of (5b) translates to $(W_i)_{5,5} = 1$ for each $i = 1, \dots, 4$.
- We have $\deg(\pi) = 6$, and comparing the coefficients on the two sides of (6), we obtain a linear system of equations and matrix inequalities for (5c) and (5d), along the same lines as in the previous example.

The optimal solution is a polynomial whose real roots are $\{\pm 5, \pm 0.854\}$, this is the support of the A-optimal design.

In the remaining examples we shall refrain from the detailed list of the above steps, and concentrate on the main features of the models and the numerical results.

Example 12 (Polynomial models of high degree). We now consider the problem of designing experiments for very high degree polynomial models in order to test the numerical stability and scalability of our approach. Models involving high degree polynomials are rarely justifiable, but they are good problems to test numerical stability, as they are notoriously ill-conditioned. For example, in the basic polynomial model, when $f_i = t^{i-1}$, $i = 1, \dots, n$, the the Fisher information matrix $M(\xi)$ in (2) becomes a Hankel matrix, which is known to be ill-conditioned [38]. Also note that in the case of rational models, the polynomial defined by (6) might also have a degree that is considerably higher than the degree of the numerators and denominators of the functions f_i , leading to potentially ill-conditioned optimization models. The numerical difficulties can be somewhat alleviated by using an orthogonal polynomial basis in (1). In this example we look for the E-optimal polynomial design in the ordinary polynomial regression model, but using the Legendre polynomial basis: $f_i = P_{i-1}$, the $(i-1)$ -st Legendre polynomial defined by $P_0 = 1, P_1(t) = t$, and $P_{n+1}(t) = (n+1)P_{n+1}(t) + (2n+1)tP_n(t) - nP_{n-1}(t)$ for $n \geq 1$.

The constraints are obtained along the same lines as in Example 10, except that the coefficients of π in (6) need to be changed as the moment matrix $M(\xi_t)$ changes with the change of basis.

We solved the resulting semidefinite program for the degree 20 model; the computation required 0.4 seconds. The optimal polynomial π is a nonnegative polynomial on $[-1, 1]$ with single roots at the endpoints ± 1 , and double real roots at the points

$$\{\pm 0.981, \pm 0.937, \pm 0.872, \pm 0.788, \pm 0.686, \pm 0.568, \pm 0.438, \pm 0.297, \pm 0.150, 0\}.$$

The E-optimal design is supported on these 21 points.

We remark that the use of high degree polynomials can also be circumvented using polynomial splines, which allow for the same large number of parameters without numerical difficulties; this will be the subject of a forthcoming paper.

Finally, we present an example using rational functions.

Example 13 (Measuring radiation parameters). Consider the measurement of total radiation emitted from point sources, whose intensity obeys the inverse square law: $I_i(r) = \theta_i r^{-2}$ where I_i is the intensity of the radiation emitted by source i measured at distance r from the source, for some unknown parameter θ_i . The locations x_i of the sources are known. The response variable in our model (1) is the total radiation. To be estimated are the values θ_i , affected by parameters of sources, shielding between the sources and detector, and several other factors. In this numerical example we consider a simple one-dimensional instance: the locations of the three sources are $x_1 = -2$, $x_2 = 2$, $x_3 = 4$, and we are interested in the effective values of θ_i as measurable in the interval $[-1, 1]$, where the variance of the measurement error and the parameters are assumed to be constant.

The distance of a detector at t from the i th point source is $r_i = |t - x_i|$, so in our model (1) we have $f_i = r_i^{-2} = (t - x_i)^{-2}$, $i = 1, 2, 3$, and $\mathcal{I} = [-1, 1]$. The solution of the semidefinite program, which took 0.2 seconds, yielded a three-point support for the E-optimal design: $\{-1, 0.231, 1\}$.

5. RECONSTRUCTING THE OPTIMAL DESIGN

Once we obtained a non-zero polynomial π from the optimal solution of (19), we can find the optimal design by solving a second semidefinite programming problem, using Theorem 8. But Theorem 9 is only useful if the polynomial π in the optimal solution is not the zero polynomial. As the following example shows, in sufficiently degenerate cases it might be.

Example 14. Consider the E-optimal design problem when $m = 2$, $f(t) = (1, t)^T$, $\omega = 1$, and $\mathcal{I} = [-1, 1]$. Then the corresponding semidefinite programming problem simplifies to

$$\min_{y, W} y \quad \text{s.t. } W \succcurlyeq 0, \quad \text{tr}(W) = 1, \quad \pi = (y - W_{11}, -2W_{12}, -W_{22}) \in P^{[-1, 1]},$$

by essentially the same calculations as in Example 10. It is not hard to see that the set of optimal solutions to this problem is $\{(y, W) \mid y = 1, W_{12} = 0, W_{11} + W_{22} = 1, 0 \leq W_{11} \leq 1\}$. Hence, we have infinitely many solutions, including $W_{11} = 1 - W_{22} = 1$, which corresponds to $\pi(t) = 0$. Choosing any other optimal solution yields a polynomial whose roots are the expected $t = \pm 1$.

Alternatively, we can change f to a different basis of degree one polynomials. This does not really change the model, however, if we choose, for example, $f(t) = (\alpha, t)^T$ for any $\alpha > 1$, the above problem disappears: the resulting semidefinite programming problem has a unique optimal solution, and that solution corresponds to a nonzero polynomial π , with two real roots.

In the rest of the section we list a number of sufficient conditions that ensure that the optimal π in (5) is not the zero polynomial. The first one is perhaps the most obvious one.

Lemma 15. *Let f_1, \dots, f_m and ω in (1) be chosen such that $1 \notin \text{span}\{\omega f_i f_j \mid 1 \leq i \leq j \leq m\}$. Then no solution satisfying the constraints of (5) has $\pi = 0$.*

Special cases covered by this lemma include designs for incomplete polynomial models with no intercept, such as those considered in [22] and [5], and models involving rational functions, such as Example 13 above.

The last observation of Example 14 also generalizes to E-optimal designs for arbitrary polynomial systems.

Lemma 16. *Consider the E-optimal design problem for a polynomial model with at least two parameters to be estimated. By choosing an appropriate basis $\{f_1, \dots, f_m\}$ in (1) it can be guaranteed that no optimal solution of (5) has $\pi = 0$.*

Proof. Let $(\hat{y}, \hat{W}, \hat{\pi})$ be an optimal solution to (5). Then $\hat{W} = Y^T Y$ for some matrix Y , and the polynomial $q: t \rightarrow \langle \hat{W}, M(\xi_t) \rangle$ can be written as $q(t) = z(t)^T z(t)$ with $z(t) = Y f(t)$. Consequently, q can only be a constant (and $\hat{\pi}$ can only be the zero polynomial) if $z(t) = Y f(t)$ is componentwise constant.

If $1 \notin \text{span}\{f_1, \dots, f_m\}$, then this is impossible, because $Y = 0$ is excluded by the constraints (5b), which simplifies to $\text{tr}(W) = 1$ for E-optimal designs.

If $1 \in \text{span}\{f_1, \dots, f_m\}$, then we can assume without loss of generality that $f_1 = 1$. Now q can be a constant only if $\hat{W}_{11} = 1$, and every other entry of \hat{W} is zero, making $q(t) = 1$ and $\hat{y} = 1$. Replacing f_i , $i \geq 2$ by λf_i with a sufficiently small positive λ that satisfies $\lambda |f_i(t)| < 1$ for all $t \in \mathcal{I}$ ensures that this is not the optimal solution to (5). \square

A similar argument applies to A-optimal designs for polynomial models. For brevity we omit the details. As above, one can argue that by scaling the non-constant basis functions, solutions to the semidefinite programming problem that yield constant zero π cannot be optimal.

Lemma 17. *Consider the A-optimal design problem for a polynomial model with at least two parameters to be estimated. By choosing an appropriate basis $\{f_1, \dots, f_m\}$ in (1) it can be guaranteed that no optimal solution of (5) has $\pi = 0$.*

Finally, as a corollary to Theorem 9, we also obtain an upper bound on the size of the support set of the minimally-supported optimal designs.

Corollary 18. *Let n_ω and d_ω be the degree of the numerator and denominator of ω , n_i and d_i be the degree of the numerator and denominator of f_i , and $d_{\text{den}} =$*

$\text{lcm}(d_\omega, d_1^2, \dots, d_p^2)$. Furthermore, suppose that \mathcal{I} is the union of $k_1 + k_2$ disjoint closed intervals, k_1 of which are singletons. (The remaining k_2 intervals have distinct endpoints.) Then for every admissible criterion Φ for which the optimal solution to (5) does not have $\pi = 0$ there is a Φ -optimal design supported on not more than $\min(\frac{1}{2}(k_1 + 2k_2 + \deg \pi), \deg \pi)$ points, where $\deg \pi = d_{\text{den}} + (n_\omega - d_\omega + 2 \max_i (n_i - d_i))_+$.

Proof. We need to count the number of distinct zeros of the polynomial π in (5). On one hand, π cannot have more than $\deg \pi$ roots. On the other hand, since π is nonnegative over \mathcal{I} , each of its zeros must be either an endpoint of an interval constituting \mathcal{I} or a root of multiplicity at least two. Hence the number of distinct zeros of π is at most $\frac{1}{2}(k_1 + 2k_2 + \deg \pi)$. Finally, the expression for $\deg \pi$ comes directly from (6). \square

6. PARAMETER SUBSYSTEMS, ESTIMABILITY

Often the experimenter is not interested in the entire parameter vector θ , but rather in a subset of them, or more generally in $s \leq m$ specific linear combinations of the parameters: $k_j^\top \theta$, $j = 1, \dots, s$. Let K be the matrix whose columns are k_1, \dots, k_s ; so far we have assumed $s = m$ and $K = I$. An application of this more general setting is polynomial regression, when the experimenter needs to test whether the highest degree terms in the model are indeed non-zero.

It can be assumed without loss of generality that K has full (column) rank, and to make the problem meaningful, it must be assumed that the parameters $K^\top \theta$ are *estimable*, that is,

$$\text{range}(K) \subseteq \text{range}(M), \quad (7)$$

see for example [33, Chapter 3]. In this setting the matrix M is replaced by the information matrix $(K^\top M^\dagger K)^{-1}$, where M^\dagger denotes the Moore–Penrose pseudo-inverse of M . In particular, the optimal design is a probability measure $\hat{\xi}$ that maximizes the matrix $(K^\top M^\dagger(\xi)K)^{-1}$, or the function $\xi \rightarrow \Phi((K^\top M^\dagger(\xi)K)^{-1})$ for some criterion function Φ compatible with the Löwner partial order.

The optimization models for this setting can be developed analogously to the model of the previous section. Since Φ is assumed to be compatible with the Löwner partial order, $\max_{M \in \mathcal{M}} \Phi((K^\top M^\dagger(\xi)K)^{-1})$ is equivalent to

$$\max\{\Phi(Y) \mid M \in \mathcal{M}, (K^\top M^\dagger K)^{-1} \succcurlyeq Y \succcurlyeq 0\}. \quad (8)$$

Note that the optimum does not change if we require Y to be positive definite, in which case the last two inequalities are equivalent to $Y^{-1} \succcurlyeq K^\top M^\dagger K$. We shall use now a Schur complement characterization of semidefinite matrices, which is a generalization of the result used in Example 6.

Proposition 19 ([41, Theorem 1.20]). *The symmetric block matrix $\begin{pmatrix} M & K \\ K^\top & Z \end{pmatrix}$ is positive semidefinite if and only if $M \succcurlyeq 0$, $Z \succcurlyeq K^\top M^\dagger K$, and $\text{range}(K) \subseteq \text{range}(M)$.*

By this proposition, (8) is equivalent to

$$\max\{\Phi(Y) \mid M \in \mathcal{M}, \begin{pmatrix} M & K \\ K^\top & Y^{-1} \end{pmatrix} \succcurlyeq 0\}.$$

Using Schur complements again, the inversion from the last inequality can be eliminated, and we obtain the following equivalent optimization problem:

$$\max\{\Phi(Y) \mid M \in \mathcal{M}, M \succcurlyeq KYK^T, Y \succcurlyeq 0\}. \quad (9)$$

Finally, we can simplify this problem essentially identically to how we obtained (5) from (18). Doing so we obtain the following.

Theorem 20. *Consider the linear model (1) and a matrix $K \in \mathbb{R}^{m \times s}$ satisfying $\text{rk}(K) = s$ and the estimability condition (7). Then for every semidefinite representable criterion function Φ a polynomial π whose real zeros contain the support of a Φ -optimal design for the parameter vector $K^T\theta$ is an optimal solution of the following semidefinite program:*

$$\begin{aligned} & \underset{\substack{y \in \mathbb{R}, V \in \mathbb{S}_+^m, \pi \in \mathbb{R}^d, \\ W_1, \dots, W_p \in \mathbb{S}_+^k}}{\text{minimize}} && y \\ & \text{subject to} && K^T V K \succcurlyeq \sum_{i=1}^p A_i^*(W_i), \\ & && \sum_{i=1}^p \langle W_i, B_i \rangle = -1, \quad \sum_{i=1}^p C_i^*(W_i) = 0, \\ & && \pi = \Pi(y, V, W_1, \dots, W_p) \in P^{\mathcal{I}}, \end{aligned}$$

where A_i, B_i, C_i and D_i come from Definition 3, and d is the degree of the polynomial

$$t \rightarrow \text{lcm}(\text{den}(\omega), \text{den}(f_1^2), \dots, \text{den}(f_p^2)) \left(y - \langle V, M(\xi_t) \rangle - \sum_{i=1}^p \langle W_i, D_i \rangle \right),$$

whose coefficient vector is denoted by $\Pi(y, V, W_1, \dots, W_p)$.

We omit the rest of the proof as it is essentially identical to that of Theorem 9, given in the Appendix. The main difference is the appearance of the variable V , which is the dual variable of the constraint $M \succcurlyeq KYK^T$.

7. LOCALLY OPTIMAL DESIGNS FOR NONLINEAR MODELS

In this section we show how to apply Theorems 9 and 20 to find *locally optimal designs* (with respect to various optimality criteria) for nonlinear rational models (see the definition and its motivation below). We consider the general nonlinear model

$$y(t) = f(t; \theta) + N(0, \sigma(t)), \quad t \in \mathcal{I}, \quad (10)$$

where f is a rational function of $(t; \theta)$, θ is an m -vector of unknown parameters. The designs space \mathcal{I} is the union of finitely many closed intervals, as before.

Nonlinear regression models are widely used and researched, but finding optimal designs for nonlinear regression is particularly challenging – so much so that even numerical solutions to simple two- and three-variable models are highly non-trivial to obtain. (See for example [9] for recent results on a number of models used in

dose-finding studies, and [26] for pharmacokinetic models.) Nonlinear rational models (where the response variable is a rational function of the explanatory variable and the unknown parameters) and models involving exponential functions and logarithms are particularly well studied. Imhof and Studden [24] considered E-optimal designs for different classes of rational models. More recently, Dette *et al.* [10] investigated E-optimal designs for a more general family of functions (not only rational functions), under the assumption that some partial derivatives of the model function form a *weak Chebyshev system* [25]. Note that this class of problems is not comparable to the rational models we are considering: the partial derivatives of many non-rational functions satisfy this criterion, but many rational models, for instance, the E_{max} model from Example 21 below, are outside that class.

Perhaps the most fundamental complication in designing non-sequential experiments for nonlinear models is in the formulation of the problem as a meaningful optimization problem. For a nonlinear regression model (10) the Fisher information matrix corresponding to the design ξ is

$$M(\xi, \theta) = \int_{\mathcal{I}} (\partial f(t, \theta) / \partial \theta) (\partial f(t, \theta) / \partial \theta)^T \omega(t) d\xi(t). \quad (11)$$

It is immediate that (unlike in the linear case) the Fisher information matrix for nonlinear models depends on the parameters whose estimation is the purpose of the experiments we are to design. Hence defining the optimal designs as the optimizers of the $M(\xi, \theta)$ is meaningless. Nevertheless, if the experimenter can guess reasonable values of the parameters, it can be useful to design the experiment that would be optimal if the guessed parameters were correct. Some more advanced design methods, such as sequential designs [16] also build on the same concept, often called *locally optimal designs*. (The same ideas can also be used for the estimation of nonlinear functions of the parameters of a linear model.) Before considering the general case, let us look at a simple example that we shall readily generalize below.

Example 21. Consider the three-parameter E_{max} model

$$y(t) = \theta_0 + \frac{\theta_1 t}{t + \theta_2} + N(0, 1), \quad (12)$$

from the dose-finding study [9]. With the notation of (11),

$$\frac{\partial f(t, \theta)}{\partial \theta} = (1, t(t + \theta_2)^{-1}, -\theta_1 t(t + \theta_2)^{-2})^T,$$

so for every fixed value $(\theta_0^*, \theta_1^*, \theta_2^*)$ of θ the integrand in the Fisher information matrix (11) can be written as

$$\begin{pmatrix} 1 & \frac{t}{t + \theta_2^*} & -\frac{\theta_1^* t}{(t + \theta_2^*)^2} \\ \frac{t}{t + \theta_2^*} & \frac{t^2}{(t + \theta_2^*)^2} & -\frac{\theta_1^* t^2}{(t + \theta_2^*)^3} \\ -\frac{\theta_1^* t}{(t + \theta_2^*)^2} & -\frac{\theta_1^* t^2}{(t + \theta_2^*)^3} & \frac{(\theta_1^* t)^2}{(t + \theta_2^*)^4} \end{pmatrix}, \quad (13)$$

which is the same information matrix as the information matrix of the parameter vector $(\alpha_0, \alpha_1, \alpha_2)$ for the linear model

$$y(t) = \alpha_0 + \alpha_1 \frac{t}{t + \theta_2^*} + \alpha_2 \frac{\theta_1^* t}{(t + \theta_2^*)^2} + N(0, 1). \quad (14)$$

Hence, finding locally optimal designs for the E_{max} model (12) is equivalent to finding optimal designs for the linear model (14), which is a linear model with rational functions, hence Theorem 9 is applicable.

A further simplification is possible: we can find an equivalent polynomial model, and use Theorem 20 to find optimal designs. It is easy to verify that the matrix (13) can also be written as

$$K^T \begin{pmatrix} 1 & \chi & \chi^2 \\ \chi & \chi^2 & \chi^3 \\ \chi^2 & \chi^3 & \chi^4 \end{pmatrix} K$$

with $\chi = (t + \theta_2^*)^{-1}$ and $K = \begin{pmatrix} 1 & 1 & 0 \\ 0 & -\theta_2^* & -\theta_1^* \\ 0 & 0 & \theta_1^* \theta_2^* \end{pmatrix}$. Hence, for every fixed θ^* the Fisher information matrix of the design ξ for model (12) is identical to the Fisher information matrix of the design that puts $\xi(t_i)$ mass to the point $\chi_i = (t_i + \theta_2^*)^{-1}$ for the three-parameter linear model

$$y(\chi) = \alpha_0 + \alpha_1 \chi + \alpha_2 \chi^2 + N(0, 1) \quad (15)$$

and the parameter vector $K^T \alpha = (\alpha_0, \alpha_0 - \alpha_1 \theta_2^*, \alpha_2 \theta_1^* \theta_2^* - \alpha_1 \theta_1^*)^T$. Now the problem is reduced to polynomial regression, and Theorem 20 is applicable.

Generally, for a nonlinear regression model (10) with m parameters, the problem of finding a locally optimal design for a given parameter vector θ^* is equivalent to finding the optimal design for the associated linear model of the form (1) with $f_i = (\partial f) / (\partial \theta_i) |_{\theta = \theta^*}$, $i = 1, \dots, m$. If f is a rational function of (t, θ) , then so are its partial derivatives. Hence the equivalent linear model (for every fixed value of θ) is always one with rational functions f_i .

The same observation was used in [10] to derive E-optimal designs for the class of nonlinear regression models where the partial derivatives form a weak Chebyshev system. Now this assumption can be dropped, and other optimality criteria can also be considered.

8. OPTIMAL DESIGNS IN OTHER FUNCTIONAL SPACES

Most of Section 3 applies to every f_i and ω , not only to rational functions; for example, (22) is not specific to polynomials or rational functions. As long as the set of constraints (23) can be expressed by finitely many semidefinite constraints (or in any other computationally tractable manner), the same approach works. Examples include the following (we consider only the homoscedastic case for simplicity):

- (1) $f_i(t) = \cos(it)$ for every $i \in \mathbb{N}$ and t ;
- (2) $f_{2i}(t) = \cos(it)$, $f_{2i+1}(t) = \sin(it)$ for every $i \in \mathbb{N}$, and t ;
- (3) $f_i(t) = \exp(it)$ for every $i \in \mathbb{N}$ and t .

These three examples, however, do not truly generalize the approach of Section 3, since they can also be reduced to the polynomial case by an appropriate change of variables. (We omit the details.)

Our estimate on the number of support points is also valid for some functional spaces other than polynomials. The only property of polynomials that we used were that their degree bounds the number of their roots (counted with multiplicity: roots in the interior of the domain have multiplicity two). Hence, bounds similar to the one in Corollary 18 can be obtained for models where the functions $\{\omega f_i f_j | i, j\}$ form a Chebyshev system.

9. DISCUSSION

Computing optimal designs for linear models involving rational functions is easy when the design space is finite, hence the key difficulty in obtaining optimal designs for infinite design spaces, such as intervals or unions of intervals, is that the finite support of an optimal design has to be determined. Symbolic or closed form solutions are unavailable for most models, and their scope is often limited by assumptions that are neither technical, nor have any statistical interpretation. In this paper, we have presented a method that does not rely on such assumptions. It is an effective method to determine a polynomial whose zeros contain the support of the optimal designs. The method is applicable to every linear regression problem involving only rational functions; it treats D-, A-, E-, and general Φ_p optimal designs in a unified manner, and generalizes to the heteroscedastic case if the variance of the noise is a positive rational function. The design space can be an interval or the union of finitely many intervals.

This level of generality is far greater than what appears to be possible by closed-form approaches. It is achieved at the price of providing numerical, rather than symbolic, solutions: the method generates the (numerical) coefficients of the sought polynomial. The main step of the method is the solution of a semidefinite programming problem, which can be done (to high precision) with readily available software in trivial running time. Unlike other iterative methods previously proposed in the literature, including all of those based on coordinate descent, semidefinite programming algorithms have a theoretically guaranteed low running time, and are guaranteed to find the globally optimal design, rather than a local optimum. This is of considerable practical importance, with the potential for instance to impact design software development.

Further study of the optimality conditions of the main optimization model might also yield new theoretical insights.

Through a number of examples we have demonstrated the flexibility of the proposed method, and we also found that the algorithm is robust enough to handle ill-conditioned problems involving high-degree polynomials, and yields solutions in a fraction of a second for problems of practical size.

A corollary of our main theorem is a bound on the size of the support set, and an analogous optimization model for the estimation of parameter subsystems.

Most results of this paper readily generalize to linear models involving certain exponential families rather than rational functions; these include Fourier regression, where the model is a trigonometric polynomial with unknown coefficients. The method may also be used to find locally optimal designs for nonlinear models. In this area almost no symbolic solutions are available, but model-specific numerical methods are abound. Details are available from the author, and may be subject of a future paper.

A few important questions remain open. The first one is how to extend the results of Section 5. Since the optimal solution to the problem (5) is sensitive to both the representation of the optimality criterion Φ and also to the basis $\{f_i\}$ of the space of regression functions (meaning that equivalent representations of Φ and basis transformations lead to different optimal solutions), one may readily conjecture that for every model (1) and for every admissible optimality criterion one can find an equivalent model (that is, a basis $\{f_i\}$ of the same functional space) and a semidefinite representation (3) for Φ such that the optimal π in every solution of (5) is nonzero.

Another subject of future research may be the generalization of our results to larger classes of functions. Chebyshev systems are natural candidates to look at, but more importantly, the ideas of the paper would generalize word by word to every family (f_1, \dots, f_m) and weight function ω for which functions in the space $\text{span}\{\omega f_i f_j | i, j\}$ are easy to maximize. Hence, identifying such spaces of functions would be particularly important.

Finally, the ability to design experiments in a discontinuous design space is extremely relevant in practice, especially in the multivariate case (e.g., when measurements cannot be taken at inaccessible locations, or are practically impossible very close to signal sources). Existing models with closed-form solutions are not applicable, and most of the current numerical methods cannot address this problem even in the univariate case, aside from sporadic results involving two disjoint intervals for a few concrete models.

The applicability of the proposed method in the multivariate setting also requires further study.

REFERENCES

1. *The optimum experimental design website*, <http://www.optimal-design.org/>, Accessed on April 1, 2011.
2. Ali Babapour Atashgah and Abbas Seifi, *Optimal design of multi-response experiments using semi-definite programming*, *Optimization in Engineering* **10** (2009), no. 1, 75–90.
3. Aharon Ben-Tal and Arkadi Nemirovski, *Lectures on modern convex optimization*, SIAM, Philadelphia, PA, 2001.
4. Stephen P. Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.
5. Fu-Chuen Chang and Berthold Heiligers, *E-optimal designs for polynomial regression without intercept*, *Journal of Statistical Planning and Inference* **55** (1996), no. 3, 371–387.

6. Fu-Chuen Chang and Hung-Ming Lin, *On minimally-supported D-optimal designs for polynomial regression with log-concave weight function*, *Metrika* **65** (2007), no. 2, 227–233.
7. A. de la Garza, *Spacing of information in polynomial regression*, *Annals of Mathematical Statistics* **25** (1954), no. 1, 123–130.
8. Holger Dette, *Optimal designs for a class of polynomials of odd or even degree*, *The Annals of Statistics* **20** (1992), no. 1, 238–259.
9. Holger Dette, Frank Bretz, Andrey Pepelyshev, and José Pinheiro, *Optimal designs for dose-finding studies*, *Journal of the American Statistical Association* **103** (2008), no. 483, 1225–1237.
10. Holger Dette, Viatcheslav B. Melas, and Andrey Pepelyshev, *Optimal designs for a class of nonlinear regression models*, *The Annals of Statistics* **32** (2004), no. 5, 2142–2167.
11. Holger Dette and Ingo Röder, *Optimal product designs for multivariate regression with missing terms*, *Scandinavian Journal of Statistics* **23** (1996), no. 2, 195–208.
12. Holger Dette and William J. Studden, *The theory of canonical moments with applications in statistics, probability, and analysis*, Wiley Interscience, New York, NY, September 1997.
13. Zhide Fang, *D-optimal designs for polynomial regression models through origin*, *Statistics & Probability Letters* **57** (2002), no. 4, 343–351.
14. V. V. Fedorov, *Theory of optimal experiments*, Academic Press, New York, NY, 1972.
15. Lenka Filová and Mária Trnovská, *Computing maximin efficient designs using the methods of semidefinite programming*, Tech. report, Comenius University, Bratislava, Slovakia, May 2010.
16. I. Ford and S. D. Silvey, *A sequentially constructed design for estimating a nonlinear parametric function*, *Biometrika* **67** (1980), no. 2, 381–388.
17. Michael Grant and Stephen Boyd, *CVX: Matlab software for disciplined convex programming (web page and software)*, Stanford University, December 2007, <http://cvxr.com/cvx/>.
18. P. G. Guest, *The spacing of observations in polynomial regression*, *The Annals of Mathematical Statistics* **29** (1958), no. 1, 294–299.
19. Emilie V. Haynsworth, *Determination of the inertia of a partitioned Hermitian matrix*, *Linear Algebra and its Applications* **1** (1968), no. 1, 73–81.
20. Didier Henrion and Jean-Bernard Lasserre, *GloptiPoly: Global optimization over polynomials with Matlab and SeDuMi*, *ACM Transactions on Mathematical Software* **29** (2002), 165–194.
21. Paul G. Hoel, *Efficiency problems in polynomial estimation*, *The Annals of Mathematical Statistics* **29** (1958), no. 4, 1134–1145.
22. Mong-Na Lo Huang, Fu-Chuen Chang, and Weng Kee Wong, *D-optimal designs for polynomial regression without an intercept*, *Statistica Sinica* **5** (1995), no. 2, 441–458.
23. Lorens A. Imhof, O. Krafft, and M. Schaefer, *D-optimal designs for polynomial regression with weight function $x/(1+x)$* , *Statistica Sinica* **8** (1998), no. 4, 1271–1274.
24. Lorens A. Imhof and William J. Studden, *E-optimal designs for rational models*, *The Annals of Statistics* **29** (2001), no. 3, 763–783.
25. Samuel Karlin and William J. Studden, *Chebyshev systems, with applications in analysis and statistics*, Pure and Applied Mathematics, vol. XV, Wiley Interscience, New York, NY, 1966.
26. Gang Li and Dibyen Majumdar, *Some results on D-optimal designs for nonlinear models with applications*, *Biometrika* **96** (2009), no. 2, 487–493.
27. Johan Löfberg, *YALMIP: A toolbox for modeling and optimization in MATLAB*, Proceedings of the CACSD Conference (Taipei, Taiwan), 2004.
28. Franz Lukács, *Verschärfung der ersten Mittelwertsatzes der Integralrechnung für rationale Polynome*, *Mathematische Zeitschrift* **2** (1918), 229–305, Available from <http://www.digizeitschriften.de/>.
29. Ruth K. Meyer and Christopher J. Nachtsheim, *The coordinate-exchange algorithm for constructing exact optimal experimental designs*, *Technometrics* **37** (1995), no. 1, 60–69.

30. Yurii Nesterov, *Squared functional systems and optimization problems*, High Performance Optimization (H. Frenk, K. Roos, T. Terlaky, and S. Zhang, eds.), Appl. Optim., Kluwer Acad. Publ., Dordrecht, 2000, pp. 405–440.
31. Jorge Nocedal and Stephen J. Wright, *Numerical optimization*, 2 ed., Springer, New York, NY, 2000.
32. Victor Y. Pan, *Structured matrices and polynomials: unified superfast algorithms*, Birkhäuser, Boston, MA, 2001.
33. Friedrich Pukelsheim, *Optimal design of experiments*, Wiley Interscience, 1993.
34. Ralph Tyrrell Rockafellar, *Convex analysis*, Princeton University Press, Princeton, NJ, 1970.
35. Andrzej Ruszczyński, *Nonlinear optimization*, Princeton University Press, Princeton, NJ, 2005.
36. Stephen M. Stigler, *Optimal experimental design for polynomial regression*, Journal of the American Statistical Association **66** (1971), no. 334, 311–318.
37. Jos F. Sturm, *Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones*, Optimization Methods and Software **11–12** (1999), no. 1–4, 625–653.
38. Evgenij E. Tyrtshnikov, *How bad are Hankel matrices?*, Numerische Mathematik **67** (1994), no. 2, 261–269.
39. Lieven Vandenbergh and Stephen P. Boyd, *Semidefinite programming*, SIAM Review **38** (1996), no. 1, 49–95.
40. Henry Wolkowicz, Romesh Saigal, and Lieven Vandenbergh (eds.), *Handbook of semidefinite programming: Theory, algorithms, and applications*, Kluwer, Norwell, MA, 2000.
41. Fuzhen Zhang (ed.), *The Schur complement and its applications*, Springer, New York, NY, 2005.

APPENDIX A. THE SEMIDEFINITE REPRESENTABILITY OF POLYNOMIALS OVER INTERVALS

For a $\Delta \subseteq \mathbb{R}$ let \mathcal{P}_n^Δ denote the set of degree n polynomials nonnegative over Δ . The following representation of nonnegative polynomials is well-known:

Proposition 22 ([28]). *For every polynomial p of degree n , $p \in \mathcal{P}_n^{[a,b]}$ if and only if*

$$p(t) = \begin{cases} r^2(t) + (t-a)(b-t)q^2(t) & (\text{if } n = 2k) \\ (t-a)r^2(t) + (b-t)s^2(t) & (\text{if } n = 2k+1) \end{cases}$$

for some polynomials r and s of degree k and q of degree $k-1$.

On the other hand, functions expressible as sums of squares of functions from a given finite dimensional functional space (such as polynomials of a fixed degree) are semidefinite representable; see [30] for a constructive proof of this claim. Applying this construction to part (2) of Proposition 22 yields the following.

Proposition 23 ([30]). *Suppose p is a polynomial of degree $n = 2m+1$, $p(t) = \sum_{k=0}^n p_k t^k$, and let $a < b$ are real numbers. Then $p \in \mathcal{P}_n^{[a,b]}$ if and only if there exist positive semidefinite matrices $X = (x_{ij})_{i,j=0}^m$ and $Y = (y_{ij})_{i,j=0}^m$ satisfying*

$$p_k = \sum_{i+j=k} (-ax_{ij} + by_{ij}) + \sum_{i+j=k-1} (x_{ij} - y_{ij}) \quad (16)$$

for all $k = 0, \dots, 2m+1$.

Similarly, if p is a polynomial of degree $n = 2m$, then $p \in \mathcal{P}_n^{[a,b]}$ if and only if there exist positive semidefinite matrices $X = (x_{ij})_{i,j=0}^m$ and $Y = (y_{ij})_{i,j=0}^{m-1}$ satisfying

$$p_k = \sum_{i+j=k} (x_{ij} - aby_{ij}) + \sum_{i+j=k-1} (a+b)y_{ij} - \sum_{i+j=k-2} y_{ij} \quad (17)$$

for all $k = 0, \dots, 2m$.

This is rather involved (and the details are only important for the purposes of actual computations), but close inspection reveals that this proposition characterizes $\mathcal{P}_n^{[a,b]}$ as a linear image of the Cartesian product of two semidefinite cones, thus, it proves the semidefinite representability of $\mathcal{P}_n^{[a,b]}$ in the sense of Definition 1.

Since the intersection of semidefinite representable sets are also semidefinite representable, it follows that $\mathcal{P}_n^{\mathcal{I}}$ is semidefinite representable for every union of finitely many closed intervals \mathcal{I} .

APPENDIX B. PROOF OF THEOREM 9

Consider the problem of finding $\max\{\Phi(M(\xi)) \mid \xi \in \Xi(\mathcal{I})\}$, where $\Xi(\mathcal{I})$ is the set of probability measures on \mathcal{I} with finite support, and M is the Fisher information matrix defined by (2). Considering the Fisher information as the variable, this can be expressed as a finite dimensional optimization problem:

$$\max\{\Phi(M) \mid M \in \mathcal{M}\}, \quad \text{where } \mathcal{M} = \{M(\xi) \mid \xi \in \Xi(\mathcal{I})\}. \quad (18)$$

Let ξ_t be the probability measure that assigns all of its mass to $t \in \mathcal{I}$. Because \mathcal{I} is assumed to be compact and the mapping $t \rightarrow M(\xi_t)$ is continuous, $\{M(\xi_t) \mid t \in \mathcal{I}\}$ is compact. Hence, $\mathcal{M} = \text{conv}\{M(\xi_t) \mid t \in \mathcal{I}\}$ is a convex compact set, and the optimization problem (18) is well-defined: The maximum is finite, and is attained (for every continuous function Φ).

Now let us assume that Φ is semidefinite representable. Then using the notations of Definition 3, problem (18) may be written as follows.

$$\begin{aligned} & \underset{z \in \mathbb{R}, u \in \mathbb{R}^l, M \in \mathcal{M}}{\text{maximize}} \quad z \\ & \text{subject to} \quad A_i(M) + B_i z + C_i(u) + D_i \succcurlyeq 0, \quad i = 1, \dots, p, \end{aligned} \quad (19)$$

where A_i, B_i, C_i , and D_i are the functions and matrices as in Definition 3.

Because \mathbb{S}_+^k is a closed convex cone, (19) is equivalent to the following Lagrangian relaxation (in which the dual variable W_i is the Lagrange multiplier associated with the i th constraint):

$$\max_{z, u, M \in \mathcal{M}} \inf_{\substack{W_i \succcurlyeq 0 \\ (i=1, \dots, p)}} z + \sum_{i=1}^p \langle W_i, A_i(M) + B_i z + C_i(u) + D_i \rangle \quad (20)$$

Suppose that Φ is admissible with respect to \mathcal{M} . Then the optimization problem (19) has a Slater point, consequently its optimum is equal to optimum of its dual problem [35, Chapter 4], obtained by replacing the “max inf” by “min sup” in the

Lagrangian (20). This dual problem then can be simplified as follows (C_i^* denotes the dual operator of C_i):

$$\begin{aligned}
& \min_{W_1, \dots, W_p \succcurlyeq 0} \sup_{z, u, M \in \mathcal{M}} z + \sum_{i=1}^p \langle W_i, A_i(M) + B_i z + C_i(u) + D_i \rangle \\
&= \min_{W_1, \dots, W_p \succcurlyeq 0} \sup_{z, u, M \in \mathcal{M}} z \left(1 + \sum_{i=1}^p \langle W_i, B_i \rangle \right) + \\
&\quad + \sum_{i=1}^p \langle C_i^*(W_i), u \rangle + \sum_{i=1}^p \langle W_i, A_i(M) + D_i \rangle \\
&= \min_{\substack{W_1, \dots, W_p \succcurlyeq 0 \\ \sum_i \langle W_i, B_i \rangle = -1 \\ \sum_i C_i^*(W_i) = 0}} \sup_{M \in \mathcal{M}} \sum_{i=1}^p \langle W_i, A_i(M) + D_i \rangle \tag{21} \\
&= \min_{\substack{W_1, \dots, W_p \succcurlyeq 0 \\ \sum_i \langle W_i, B_i \rangle = -1 \\ \sum_i C_i^*(W_i) = 0}} \max_{M \in \mathcal{M}} \sum_{i=1}^p \langle W_i, A_i(M) + D_i \rangle.
\end{aligned}$$

(The last equation simply means that the supremum is attained.)

Finally, with the help of a dummy variable y the optimization problem in the last line can be conveniently written as:

$$\begin{aligned}
& \text{minimize } y \\
& \quad y \in \mathbb{R}, \\
& \quad W_1, \dots, W_p \in \mathbb{S}^k \\
& \text{subject to } W_i \succcurlyeq 0 \quad i = 1, \dots, p, \\
& \quad \sum_{i=1}^p \langle W_i, B_i \rangle = -1, \quad \sum_{i=1}^p C_i^*(W_i) = 0, \tag{22} \\
& \quad y \geq \sum_{i=1}^p \langle W_i, A_i(M) + D_i \rangle \quad \forall M \in \mathcal{M}.
\end{aligned}$$

Aside from the last set of constraints, which is an uncountably infinite collection of linear inequalities, every constraint is either a linear equality or a linear matrix inequality on the variables W_i . Using that $\mathcal{M} = \text{conv}\{M(\xi_t) \mid t \in \mathcal{I}\}$, the last set of constraints can also be simplified to

$$y - \sum_{i=1}^p \langle W_i, A_i(M(\xi_t)) + D_i \rangle \geq 0 \quad \forall t \in \mathcal{I}. \tag{23}$$

Since $M(\xi_t)$ is a matrix whose entries are rational functions of t , this inequality expresses the nonnegativity of a rational function (over \mathcal{I}) that lives in the space

$$V = \text{span}(\{\omega f_i f_j \mid i, j = 1, \dots, m\} \cup \{1\}),$$

with variable coefficients. Multiplying both sides with the least common denominator of the functions $\omega f_i f_j$ (which is positive on \mathcal{I}) turns (23) to the equivalent inequality (5d) with Π defined in (6), giving us (5).

Finally, suppose $(\hat{y}, \hat{\pi}, \hat{W}_1, \dots, \hat{W}_p)$ is an optimal solution to (5). Then, since (5), (21), and (22) are equivalent, $(\hat{y}, \hat{W}_1, \dots, \hat{W}_p)$ is also an optimal solution to (22), and because the optimum in (21) is attained, there exists an $\hat{M} \in \mathcal{M}$ that satisfies the last constraint of (22) with inequality. The way we obtained (22) from (18) ensures that this \hat{M} is also an optimal solution to our original problem (18). Suppose $\hat{M} = \hat{M}(\hat{\xi})$ for some measure $\hat{\xi} \in \Xi(\mathcal{I})$ that is concentrated on $\{t_1, \dots, t_k\} \subseteq \mathcal{I}$ and assigns weight λ_i to t_i , $i = 1, \dots, k$. Then with the optimal \hat{y} and $\hat{W}_1, \dots, \hat{W}_p$ each of these t_i must satisfy (23) with equality. Consequently each t_i is a root of $\hat{\pi}$. \square

DÁVID PAPP, NORTHWESTERN UNIVERSITY, DEPARTMENT OF INDUSTRIAL ENGINEERING AND MANAGEMENT SCIENCES, TECHNOLOGICAL INSTITUTE, 2145 SHERIDAN RD, C210, EVANSTON, IL 60208. EMAIL: DPAPP@IEMS.NORTHWESTERN.EDU