MATHEMATICAL ENGINEERING
TECHNICAL REPORTS

Symbolic Optimization of Algebraic Functions

Masaaki KANNO, Kazuhiro YOKOYAMA,
Hirokazu ANAI, and Shinji HARA

(Communicated by Kazuo MUROTA)

METR 2008–19 April 2008

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.keisu.t.u-tokyo.ac.jp/research/techrep/index.html

Symbolic Optimization of Algebraic Functions

Masaaki KANNO*, Kazuhiro YOKOYAMA†,
Hirokazu ANAI‡, and Shinji HARA§

April 21st, 2008

Abstract

This report attempts to establish a new framework of symbolic optimization of algebraic functions that is relevant to possibly a wide variety of practical application areas. The crucial aspects of the framework are (i) the suitable use of algebraic methods coupled with the discovery and exploitation of structural properties of the problem in the conversion process into the framework, and (ii) the feasibility of algebraic methods when performing the optimization. As an example an algebraic approach is developed for the discrete-time polynomial spectral factorization problem that illustrates the significance and relevance of the proposed framework. A numerical example of a particular control problem is also included to demonstrate the development.

Keywords: Parametric optimization, Gröbner basis, quantifier elimination, polynomial spectral factorization

1 Introduction

Whilst numerical computation based on floating-point arithmetic is prevalent in the applied science and engineering fields, algebraic methods and algebraic algorithms have been attracting much attention from those application areas due to computed results of different quality such approaches can potentially offer. The early days saw results of theoretical interest, but even a tiny toy problem could hardly be solved because of lack of effective algorithms and implementation and also because of the limited computation power available at that time. Algorithmic development, coupled with the ever increasing computation capability, made it possible

*CREST, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi-shi, Saitama, 332-0012, Japan. E-mail: `M.Kanno.99@cantab.net`

†Department of Mathematics, Rikkyo University, 3-34-1 Nishi Ikebukuro, Toshima-ku, Tokyo, 171-8501, Japan. E-mail: `yokoyama@rkmath.rikkyo.ac.jp`

‡Fujitsu Laboratories Ltd, 4-1-1 Kamikodanaka, Nakahara-ku, Kawasaki, 211-8588, Japan. E-mail: `anai@jp.fujitsu.com`

§Department of Information Physics and Computing, Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan. E-mail: `Shinji_Hara@ipc.i.u-tokyo.ac.jp`

to find solutions for (sometimes more than) toy problems. The achievement further stimulated both computer algebraists and people on the application side. Solution for larger, more realistic problems has been envisaged and formulation of a wider class of practical problems in the algebraic framework pursued.

This report attempts to establish a new framework of symbolic optimization which has relevance to practical problems. Emphasized in the development is the significance of the appropriate use of algebraic methods and of the discovery of structural properties inherent in original problems in the application areas. More specifically algebraic methods play a vital rôle in converting the original problem into the suggested optimization framework and moreover in solving the resulting optimization problem. As an example we consider a particular control problem which falls into the framework, but the framework in fact encompasses a wide class of problems in signal processing and control. The authors believe that a large number of problems from the application side can be dealt with in a similar fashion and that the framework indeed embraces an extensive area of problems of practical significance. It is thus hoped that this report will stimulate the research in various fields of application and also the computer algebra community into the search of latent structures and exciting algorithmic improvements.

The rest of the report is organized as follows. Section 2 is devoted to the formulation of the new framework this report proposes, and some discussions are given on how to achieve the framework and to solve the formulated optimization problem effectively are given. Section 3 takes up an example of the discrete-time polynomial spectral factorization problem, which can help convert various problems in signal processing and control into the suggested framework and thus indicates the relevance of the framework in the context of the solution of problems arising from application fields. Then a numerical example of a control problem is given in Section 4. Some concluding remarks are made in Section 5.

2 New Framework of Symbolic Optimization

Two algebraic techniques most commonly relied upon to solve problems from applied fields may be Gröbner bases and quantifier elimination (QE). After the formal establishment of the notion of Gröbner bases, endeavours based on Gröbner basis theory have been made for the solution of various problems. For instance the power of Gröbner basis theory is exerted to observe the structure of the solution set of algebraic equations [1], and also to reduce the number of variables (or elucidate free variables) for parametrization of solutions [2]. The technique of QE, originated by Tarski in 1950s, has a long history of application examples, but more recently benefits from the algorithmic development [3, 4, 5, 6] and easy-to-use implementation [7] and also from the computation capability of modern computers. Typical usages of QE include computation of algebraic expressions for feasible regions of parameters [8, 9, 10, 11, 12] and (possibly non-convex) optimization of a cost function which is rational in parameters under some algebraic constraints on parameters [13, 14, 15]. Possibilities of other methods are explored as well and new application examples have been constantly appearing [16].

However, in order to make such algebraic methods more useful in practice, it is of crucial importance to greatly reduce computation time and moreover to allow treatment of a wider class of problems that are relevant to practical applications. For the reduction of the computation burden, it is never sufficient to solely expect improvements in algorithms. Extensive efforts are also to be made to pursue methods that exploit the structural properties of the problems that are being solved, and furthermore the discovery of better representations of the problems that reveal desirable structures is to be made. More importantly several techniques need to be organically blended so that sundry problems of practical significance may be formulated appropriately in the algebraic framework.

In this report a new framework of symbolic optimization is proposed that may overcome current limitations and proceed in the direction mentioned above. The purpose of the report is twofold:

- to show the significance of combining various algebraic techniques in reducing a practical problem to a new framework of symbolic optimization suggested in this report and further performing optimization; and
- to indicate that the suggested framework has relevance to practical problems by giving a particular example in control.

The suggested formulation is written as

$$\begin{cases} \text{maximize/minimize } \phi(\mathbf{p}) \\ \text{subject to } \mathcal{F}(\mathbf{p}) \end{cases} \tag{1}$$

where $\mathbf{p} = (p_1, p_2, \ldots)$ is a vector of real parameters, $\phi(\mathbf{p}) \in L$ with $L/\mathbb{Q}(\mathbf{p})$ being a finite algebraic extension of $\mathbb{Q}(\mathbf{p})$, and $\mathcal{F}(\mathbf{p})$ is a set of algebraic constraints on $\mathbf{p}$. In this report we call $\phi(\mathbf{p})$ a real algebraic function for simplicity (and hence the title of the report). Namely, $\phi(\mathbf{p})$ is a rational function in ζ and $\mathbf{p}$, where ζ is some algebraic number defined as a particular root of a minimal polynomial over $\mathbb{Q}(\mathbf{p})$. By $\mathcal{F}$, the feasible region of parameters are specified algebraically, e.g., $-1 \leq p_1 < 2$ or $p_1^2 + 2p_2^2 \leq 1$. Furthermore, as a special case of (1), we consider the situation where the cost function $\phi(\mathbf{p})$ is the *largest real root* of a polynomial whose coefficients are polynomials/rational functions in $\mathbf{p}$.

The problem formulation (1) may be in fact contained in a general framework stated in [13], but there are two distinctive points in our framework: (i) formulation in this way and (ii) solution of this optimization problem. It can hardly be expected that a realistic optimization problem arising from a practical application is readily given in the form of (1). Indeed it is often the case that a practical problem is tackled by solving a series of equations and thus that the input to an equation depends on the output (i.e., solution) of the preceding equations. With parameters the situation is more problematic. Algebraic methods can assist reformulation of the original optimization problem in the form of (1). We will see as an example that sophisticated Gröbner basis techniques turn a control problem into this form.

It goes without saying that the computation cost is always the issue when an algebraic method is utilized. We further point out the significance of the exploitation of structural properties of the problem and of the pursuit of more desirable parametrization of variables for accomplishing this part, which cannot be overstressed.

Once converted into (1), various optimization algorithms, both numerical and algebraic [17], may be applicable, and extensive efforts have been made for improvement of the efficacy of such algorithms. It is indicated that this optimization can also be performed by means of the QE approach [13, 14, 15]. The optimization problem in (1) is in general non-convex and thus it is not an easy task for a numerical optimization algorithm to find the global optimum. Algebraic methods based on, e.g., QE, have an advantage that they can in principle find the global optimum. This report shows that a wide class of realistic practical problems can be formulated as in (1) and thus that QE-based optimization has immense significance in the application field.

More specifically, write $\phi(\mathbf{p}) = \bar{\phi}(\zeta, \mathbf{p})$, where ζ is defined as a particular real root of a polynomial over $\mathbb{Q}(\mathbf{p})$ and $\bar{\phi}(\zeta, \mathbf{p})$ is a rational function in ζ and $\mathbf{p}$. That ζ is a particular real root can be stated in an equivalent condition on ζ that there are a particular number of real roots between ζ and $+\infty$. Therefore, ζ can be specified algebraically by way of the defining polynomial and the condition from the Sturm-Habicht sequence [18]; see [14, 19] for more details. The optimization problem is thus stated as a QE problem:

$$\begin{aligned}\exists\zeta\,\exists\mathbf{p}\,(\,\eta = \bar{\phi}(\zeta, \mathbf{p})\ \wedge\ \mathcal{F}(\mathbf{p})\\ \wedge\ [\,\text{Condition from the Sturm-Habicht sequence}\,]\,)\,.\end{aligned}$$

After the elimination of variables ζ and $\mathbf{p}$, the condition only on η results, which indicates the range of the values $\bar{\phi}(\zeta, \mathbf{p})$ (equivalently, $\phi(\mathbf{p})$) can take when parameters $\mathbf{p}$ change subject to $\mathcal{F}(\mathbf{p})$. It is immediate to find the minimum/maximum of $\phi(\mathbf{p})$ from the result.

We point out that, in this part, further exploitation of structural properties is also possible for speeding up optimization when $\phi(\mathbf{p})$ is defined as the largest real root of a polynomial over $\mathbb{Q}(\mathbf{p})$ [20]. This further expands the practical applicability of algebraic approaches.

It is emphasized here that all the points stated above have a significant meaning to practical applications and are indispensable for devising effective algebraic approaches. Indeed control problems considered in [14, 21] are suitably transformed in the form of (1) and also solved via the QE-based optimization approach. In Section 3, we see another example of a control problem that falls into the framework suggested in this section. It is believed that these problems are convincing evidences of the versatility of this framework.

3 Discrete-time Polynomial Spectral Factorization

The problem of polynomial spectral factorization is of crucial importance in signal processing [22] and control [23] for finite dimensional linear systems. The pursuit of its solution approaches thus has a long history [24]. Early results includes approaches using constant matrix factorization or the Newton-Raphson method. A typical approach in recent times may be to reduce the problem to an algebraic Riccati equation and solve it by means of numerical linear algebra. The authors of this report pointed out an intriguing and useful relationship between the *continuous-time* polynomial spectral factorization and Gröbner basis theory [25] and discussed an algebraic approach to parametric polynomial spectral factorization based on this discovery [14, 19]. The key ingredient in this approach is a quantity called the sum of roots (SoR), and it is further indicated that this quantity also has an essential meaning in control theory [21].

This section develops a similar algebraic approach to the *discrete-time* case. Since modern signal processing and control systems utilize digital computers to fulfil severe requirements for high-level performances and accomplish complicated tasks, analysis and synthesis of discrete-time systems have become of practical significance [26]. It is thus beneficial to develop mathematical tools for discrete-time systems.

In general, given a result for the *continuous-time* case, a conceivable approach is to make use of the bilinear transform (Tustin transform) [26], which is a conformal mapping that transforms the *continuous-time* representation to the *discrete-time* one and vice versa. Indeed this approach works for the polynomial spectral factorization problem, and parametric *discrete-time* polynomial spectral factorization is feasible via the approach to the *continuous-time* counterpart. There is however a drawback in such an approach because it may obscure a fundamental quantity in the discrete-time domain.

The aim of this section is thus to develop a more direct method for parametric *discrete-time* polynomial spectral factorization which preserves a quantity that has an essential meaning in control theory, just like the SoR in the *continuous-time* case. The development turns out to be analogous to the continuous-time case [14, 19] and the approach is fairy effective. Nevertheless it is emphasized that this analogy is achieved by means of a novel parametrization. The development is thus believed to serve as a persuasive evidence for the importance of seeking nice representations.

3.1 Problem Formulation

Consider the polynomial[1] of degree $2n$ in $\mathbb{R}[x]$ of the following form:

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \\ + \frac{a_1}{x} + \cdots + \frac{a_{n-1}}{x^{n-1}} + \frac{a_n}{x^n}\ , \ a_n \neq 0\ , \quad (2)$$

where $a_i \in \mathbb{R}$, $i = 0, 1, \ldots, n$. Here, for the brevity of the exposition, we first assume that the coefficients of $f(x)$ are real constants and do not contain parameters. The discussion on the parametric case is deferred until Subsection 3.3. The polynomial $f(x)$ is called *self-reciprocal* since $f(x)$ and its reciprocal polynomial[2], $f\left(\frac{1}{x}\right)$, are coincident:

$$f(x) = f\left(\frac{1}{x}\right) .$$

Suppose that $f(x)$ has no roots of unit modulus, i.e., has no roots on the unit circle. Due to the self-reciprocal property, if α is a root of $f(x)$, then so is $\frac{1}{\alpha}$. All the roots are situated symmetrically about the unit circle and there are n roots each inside and outside the unit circle. The task in the discrete-time polynomial spectral factorization problem is to decompose $f(x)$ as a product of two polynomials, a polynomial that captures all the roots inside the unit circle (namely, 'stable' roots) and its reciprocal.

Definition 1 *The* spectral factorization *of $f(x)$ in* (2) *is a decomposition of $f(x)$ of the following form:*

$$f(x) = g(x)\, g\left(\frac{1}{x}\right)\ , \quad (3)$$

where

$$g(x) = b_n x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0 \in \mathbb{R}[x]\ ,\ b_n > 0\ , \quad (4)$$

and $g(x)$ has roots strictly inside *the unit circle only. The polynomial $g(x)$ is called the* spectral factor *of $f(x)$.*

Some investigation on the structural properties of the problem is made. Let α_i, $i = 1, 2, \ldots, n$, be the n roots of $f(x)$ inside the unit circle (i.e., $|\alpha_i| < 1$). The remaining n roots located outside of the unit circle can then be written as $\frac{1}{\alpha_i}$, $i = 1, 2, \ldots, n$. Using α_i's, we can express $f(x)$ and $g(x)$ as

$$f(x) = a_n \prod_{i=1}^{n} (x - \alpha_i)\left(1 - \frac{1}{\alpha_i x}\right) ,$$

$$g(x) = b_n \prod_{i=1}^{n} (x - \alpha_i)\ . \quad (5)$$

[1] The polynomial (2) is obviously *not* a polynomial, but it can be easily converted to a polynomial: $x^n f(x)$. We thus regard (2) as a polynomial to follow the convention in signal processing and control and also to simplify the notation in the development in this report.

[2] Again, to be precise, this is *not* a polynomial, but we also consider this one as a polynomial.

Unlike the continuous-time case, the leading coefficient b_n of $g(x)$ is not immediately determined from $f(x)$, but, comparing the leading coefficients of the both sides of (3), we can observe the following relationship:

$$a_n = b_n b_0 = b_n^2 \prod_{i=1}^{n}(-\alpha_i) \ . \tag{6}$$

Now, just as the SoR in the continuous-time case, let us introduce the quantity called the *product of roots* (PoR):

$$\pi := (-\alpha_1)(-\alpha_2)\cdots(-\alpha_n) \ . \tag{7}$$

The naming may be obvious since π is the product of roots of $g(x)$ up to sign. As will be seen below, π is a crucial quantity in carrying out discrete-time polynomial spectral factorization. Since $|\alpha_i| < 1$, and also any non-real root of $f(x)$ (resp., $g(x)$) has its complex conjugate as a root of $f(x)$ (resp., $g(x)$) and their product becomes real, the following fact is immediately derived.

Fact 2 *The quantity π is real and its modulus (absolute value) is strictly less than* 1.

A naïve approach to polynomial spectral factorization may be to first calculate the roots α_i's of $f(x)$ inside the unit circle, compute b_n from (6) and then construct $g(x)$ using (5). Our main target is nevertheless parametric $f(x)$ and we pursue an algebraic approach that computes π without explicitly finding α_i's. Also the approach is expected to reduce the polynomial spectral factorization problem in essence to the computation of π, just as in the continuous-time case where the crucial part of the approach is to obtain the SoR. In preparation for the development some polynomials which has π as one of their roots are introduced.

Definition 3 *Let $\mathcal{P} = \big\{(\epsilon_1, \epsilon_2, \ldots, \epsilon_n) \mid \epsilon_i \in \{1, -1\}\big\}$, and $C(\epsilon_1, \epsilon_2, \ldots, \epsilon_n) = (-\alpha_1)^{\epsilon_1} \cdot (-\alpha_2)^{\epsilon_2} \cdots (-\alpha_n)^{\epsilon_n}$ for each $(\epsilon_1, \epsilon_2, \ldots, \epsilon_n)$ in $\mathcal{P}$. The* characteristic polynomial $S_f(y)$ of π *is defined as*

$$S_f(y) = \prod_{(\epsilon_1,\epsilon_2,\ldots,\epsilon_n)\in\mathcal{P}} \big(y - C(\epsilon_1, \epsilon_2, \ldots, \epsilon_n)\big) \ .$$

Notice that

$$\pi = C(1, 1, \ldots, 1) \ ,$$

and the degree of $S_f(y)$ is 2^n. It is further noted that $S_f(y)$ belongs to $\mathbb{R}[y]$. In the case where $f(x)$ belongs to $\mathbb{Q}[x]$, $S_f(y)$ belongs to $\mathbb{Q}[y]$, as well (while $g(x) \notin \mathbb{Q}[x]$ in general).

The following lemma can be deduced in a manner analogous to the continuous-time case.

Lemma 4 *The PoR π defined in* (7) *coincides with the real root of $S_f(y)$ with the smallest modulus. Moreover, under the assumption that $f(x)$ does not have roots on the unit circle, π is always a simple root.*

In the discrete-time case we are dealing with in this section, we need an extra trick to facilitate solution of the polynomial spectral factorization problem. Using π, we can write the relationship in (6) as

$$b_n^2 = \frac{a_n}{\pi} \quad \left(\pi = \frac{a_n}{b_n^2}\right) . \tag{8}$$

Then,

$$\tilde{S}_f(\tilde{y}) := \tilde{y}^{2^{n+1}} S_f\left(\frac{a_n}{\tilde{y}^2}\right) \tag{9}$$

is a polynomial in $\tilde{y}$ of degree 2^{n+1} whose largest real root is equal to b_n (remember that we assume that $b_n > 0$). Also, b_n is always a simple root of $\tilde{S}_f(\tilde{y})$ when $f(x)$ has not roots on the unit circle.

The relationship (8) permits us to treat π and b_n interchangeably. The advantage of dealing with b_n (rather than π) is twofold: it allows us to solve the problem of polynomial spectral factorization effectively by means of Gröbner bases, which we will see in the following subsection; and the quantity b_n is directly related to some essential quantity in control theory [27].

The next subsection develops an algebraic approach to the problem of discrete-time polynomial spectral factorization. Since the structure of roots of $f(x)$ is clear, we can make an extensive use of its properties we have investigated in this subsection. We stress here the importance of the discovery of useful structural properties and the exploitation thereof for acquiring effective algebraic approaches.

3.2 Solution Approach

If we compare the coefficients of the both sides of (3), a system of quadratic polynomial equations in b_i's is obtained. Write as $\bar{g}_k$ the coefficient of the k-th order term of $g(x)g\left(\frac{1}{x}\right) - f(x)$, that is,

$$g(x)g\left(\frac{1}{x}\right) - f(x) = \sum_{k=-n}^{n} \bar{g}_{|k|} x^k . \tag{10}$$

Then, for each k, $k = 0, 1, \ldots, n$, we have

$$\bar{g}_k = \sum_{i=0}^{n-k} b_i b_{i+k} - a_k . \tag{11}$$

The polynomial spectral factorization problem thus reduces to finding a particular zero of an ideal. Considering b_i, $i = 0, 1, \ldots, n$, as variables and letting $\mathbf{B}_0 =$

$\{b_0, b_1, \ldots, b_n\}$, we have an ideal $\langle \mathcal{G}_0 \rangle$ in $\mathbb{R}[\mathbf{B}_0]$ generated by $\mathcal{G}_0 = \{\bar{g}_0, \bar{g}_1, \ldots, \bar{g}_n\}$. Then each zero of $\langle \mathcal{G}_0 \rangle$ corresponds to the sign of b_n and $(\epsilon_1, \epsilon_2, \ldots, \epsilon_n) \in \mathcal{P}$ through the relationships $g(x) = b_n \prod_{i=1}^{n}(x - \alpha_i^{\epsilon_i})$ and $a_n = b_n^2 \prod_{i=1}^{n}(-\alpha_i^{\epsilon_i})$. Hence the ideal $\langle \mathcal{G}_0 \rangle$ is 0 dimensional and has at most 2^{n+1} distinct zeros.

In the *continuous-time* case, the set of polynomials derived in an analogous fashion forms a Gröbner basis of the ideal generated by itself and we can immediately employ various results from Gröbner basis theory [25, 14]. Unfortunately the same is not true for the *discrete-time* case under investigation. However we can show that a different representation of the coefficients of $g(x)$ and some manipulation of the set of polynomials yield a desired Gröbner basis.

Another representation of $g(x)$ we use is

$$g(x) = \beta_n(x+1)^n + \beta_{n-1}(x+1)^{n-1} + \cdots + \beta_0 \,. \tag{12}$$

Notice that b_i and β_j are related as

$$\begin{aligned} b_i &= \sum_{j=i}^{n} \binom{j}{i} \beta_j \\ \beta_j &= \sum_{i=j}^{n} \binom{i}{j} (-1)^{i-j} b_i \,, \quad j = 0, 1, \ldots, n \,, \end{aligned} \tag{13}$$

where $\binom{j}{i}$ is the binomial coefficient for $i, j \in \mathbb{N}$. Denote $\{\beta_0, \beta_1, \ldots, \beta_n\}$ by $\mathbf{B}$. Then the conversion between $\mathbf{B}_0$ and $\mathbf{B}$ is linear and there is a one-to-one relationship. Also notice that $\beta_n = b_n (> 0)$, and we develop our approach around β_n.

We can transform each $\bar{g}_k(\mathbf{B}_0)$ to a polynomial in $\mathbf{B}$, which we denote by $\bar{g}_k(\mathbf{B})$. The set of polynomials $\{\bar{g}_0, \bar{g}_1, \ldots, \bar{g}_n\}$ is still not a Gröbner basis, but a Gröbner basis can be computed in a very simple manner from $\bar{g}_k$ without resorting to algorithms such as Buchberger's algorithm. Let $c_{k,\ell}$, $k = 0, 1, \ldots, n$, $k \le \ell \le n$, be

$$\begin{cases} c_{k,k} = 1 & k = 0, 1, \ldots, n \,, \\ c_{0,\ell} = (-1)^{\ell} 2 & \ell = 1, 2, \ldots, n \,, \\ c_{k,\ell} = (-1)^{k+\ell} \cdot \frac{2}{(2k)!} \cdot \frac{(k+\ell-1)!}{(\ell-k)!} \cdot \ell & \begin{cases} k = 1, 2, \ldots, n \,, \\ k < \ell \le n \,. \end{cases} \end{cases}$$

Lemma 5 *Let*

$$\bar{\bar{g}}_k := \sum_{\ell=k}^{n} c_{k,\ell}\, \bar{g}_\ell \ \left(= \bar{g}_k + \sum_{\ell=k+1}^{n} c_{k,\ell}\, \bar{g}_\ell \right) \,, \ k = 0, 1, \ldots, n \,. \tag{14}$$

Then the set of polynomials

$$\mathcal{G} := \{\bar{\bar{g}}_0, \bar{\bar{g}}_1, \ldots, \bar{\bar{g}}_n\} \,, \tag{15}$$

forms the reduced Gröbner basis of the ideal generated by $\{\bar{g}_0, \bar{g}_1, \ldots, \bar{g}_n\}$ *in* $\mathbb{R}[\mathbf{B}]$ *with respect to the graded reverse lexicographic order* $\beta_n \succ \beta_{n-1} \succ \cdots \succ \beta_0$*, with* β_k^2 *being the leading monomial of* $\bar{\bar{g}}_k$*. (The leading coefficients are 1.)*

The proof is given in the Appendix.

We call the ideal $\langle \mathcal{G} \rangle$ of $\mathbb{R}[\mathbf{B}]$ the *ideal of spectral factorization*. The set of the leading monomials of the elements of $\mathcal{G}$ is $\{\beta_0^2, \beta_1^2, \ldots, \beta_n^2\}$. Therefore,

$$\mathcal{LB} := \left\{\beta_0^{d_0}\beta_1^{d_1}\cdots\beta_n^{d_n} \mid d_k \in \{0,1\}\right\}$$

forms a basis of the residue class ring $\mathbb{R}[\mathbf{B}]/\langle \mathcal{G} \rangle$ as an $\mathbb{R}$-linear space, and $\dim_{\mathbb{R}} \mathbb{R}[\mathbf{B}]/\langle \mathcal{G} \rangle = \#\mathcal{LB} = 2^{n+1}$. These facts lead to the following lemma.

Lemma 6 *The ideal of spectral factorization is* 0 *dimensional and the number of its zeros with multiplicities counted is* 2^{n+1}.

Once this ideal of spectral factorization is established based on the unconventional representation (12) of $g(x)$, the rest of the development follows essentially in the same line as the *continuous-time* case. Here some vital points are presented. Readers are referred to [19, 14] for full details of the continuous-time case. If $f(x)$ does not have multiple roots, there are exactly 2^n distinct possible root combinations and thus 2^{n+1} different polynomials $g(x)$ that satisfy (3). This also implies that there are exactly 2^{n+1} zeros of $\langle \mathcal{G} \rangle$. The 'true' $g(x)$, namely the spectral factor, corresponds to a zero with the largest real β_n. In order to simplify the search for this particular zero, we convert the Gröbner basis $\mathcal{G}$ into the so-called shape basis.

The ideal of spectral factorization for the discrete-time system has two cases, the *generic* case and the *singular* case, just as for the continuous-time system. Almost all $f(x)$ arising from practical applications fall into the *generic* case and we mainly discuss the generic case; some remark on the *singular* case will be made at the end of this subsection.

Definition 7 *Given* $f(x)$*, when distinct* $(\epsilon_1, \epsilon_2, \ldots, \epsilon_n) \in \mathcal{P}$ *give distinct* $C(\epsilon_1, \epsilon_2, \ldots, \epsilon_n)$*, we call the situation a* generic case*. Otherwise it is called a* singular case.

In either case we can show by using 'generic coefficient' and Lemma 11 in the next subsection that $\tilde{S}_f(\tilde{y})$ in (9) is identical to the characteristic polynomial of β_n modulo $\langle \mathcal{G} \rangle$, i.e., the characteristic polynomial of the linear map derived from the multiplication map [28]:

$$\mathbb{R}[\mathbf{B}]/\langle \mathcal{G} \rangle \ni g \to \beta_n g \in \mathbb{R}[\mathbf{B}]/\langle \mathcal{G} \rangle\ .$$

Moreover, in the generic case, $\tilde{S}_f(\tilde{y})$ is square-free and its degree is 2^{n+1}. Noting that the ideal of spectral factorization has at most 2^{n+1} distinct zeros, we can immediately deduce that β_n is a separating element [29]. Due to the facts that $\langle \mathcal{G} \rangle$ is 0 dimensional and radical and that β_n is a separating element, we can obtain a special Gröbner basis called the *shape basis* in the discrete-time case as well.

Theorem 8 *In the generic case the ideal of spectral factorization has a Gröbner basis so-called* shape basis *with respect to any elimination ordering* $\{\beta_0, \beta_1, \ldots, \beta_{n-1}\} \succ\succ \beta_n$*:*

$$\mathcal{F} := \left\{\tilde{S}_f(\beta_n), \beta_{n-1} - \tilde{h}_{n-1}(\beta_n), \ldots, \beta_0 - \tilde{h}_0(\beta_n)\right\} ,$$

where $\tilde{S}_f$ *is a polynomial of degree exactly* 2^{n+1} *and* $\tilde{h}_i$*'s are polynomials of degree strictly less than* 2^{n+1}*.*

The theorem guarantees that, in the discrete-time case, just like the continuous-time case, all the coefficients of the spectral factor can be expressed as polynomials in β_n and therefore that the problem of polynomial spectral factorization can in essence be solved by finding the largest real root of $\tilde{S}_f(\tilde{y})$. Last but not least, since $\mathcal{G}$ is already a Gröbner basis, we can effectively compute a shape basis from $\mathcal{G}$ by way of the basis conversion (change-of-order) technique [29].

Here it should be emphasized that the proposed approach has properties favourable to practical applications. A Gröbner basis with respect to the graded reverse lexicographic order is obtained almost instantly from the problem formulation and the investigation of the properties of the ideal is thus possible. It is then seen that a shape basis is obtainable, and that relatively easily. Lastly we only have to examine the largest real root of $\tilde{S}_f(\tilde{y})$. A typical scenario is that one need to resort to an expensive calculation to derive a shape basis and then one has to investigate all the roots of the characteristic polynomial in search of the true solution. Those advantages are acquired by discovering the representation (12) that yields helpful structural properties.

Before closing this subsection we briefly discusses the the *singular* case, where the characteristic polynomial $\tilde{S}_f(\tilde{y})$ has multiple roots. Even in this case we are still able to derive a shape basis in the same manner as in [19]. The key point is to add the *simple part* $\tilde{T}(\tilde{y})$ of $\tilde{S}_f(\tilde{y})$ to the ideal, where the simple part $\tilde{T}(\tilde{y})$ is the factor of $\tilde{S}_f(\tilde{y})$ obtained as the product of $\tilde{y} - \gamma_i$ for all simple roots γ_i of $\tilde{S}_f(\tilde{y})$. Note that $\tilde{T}(\tilde{y})$ can be computed by GCD of $\tilde{S}_f$ and its derivative. Then we can show the following.

Theorem 9 *The ideal* $\langle \mathcal{G}, \tilde{T}(\beta_n)\rangle$ *has a shape basis with respect to any elimination ordering* $\{\beta_0, \ldots, \beta_{n-1}\} \succ\succ \beta_n$*:*

$$\left\{\tilde{T}(\beta_n), \beta_{n-1} - \bar{h}_{n-1}(\beta_n), \ldots, \beta_0 - \bar{h}_0(\beta_n)\right\} ,$$

where $\bar{h}_i$*'s are polynomials of degree strictly less than that of* $\tilde{T}$*.*

3.3 Parametric Case

This section deals with the case where each coefficient a_k is some polynomial in real parameters $\mathbf{p} = (p_1, p_2, \ldots, p_m)$ over $\mathbb{Q}$. Now the polynomial $f(x)$ is considered as a multivariate one $f(x, \mathbf{p})$ in $\mathbb{Q}[x, \mathbf{p}]$. Even in the parametric case it often happens that the ideal of spectral factorization is generic for almost all combinations of parameter values. In such a case we can apply efficient *parametric basis conversion*, where we can compute the shape basis directly over the rational function field of parameters, as pointed out in [19] (See also Remark 13 below). Nevertheless we need to pay special attention to singular situations so that optimization that follows may be carried out thoroughly. It is shown here that such singularities can also be dealt with systematically. To this end the notion of 'comprehensive Gröbner system' is crucial and we can apply several techniques for its computation.

For each element $\mathbf{c} = (c_1, c_2, \ldots, c_m)$ in $\mathbb{R}^m$, we denote by $\varphi_{\mathbf{c}}$ the ring homomorphism from $\mathbb{Q}[\mathbf{p}][\mathbf{B}]$ to $\mathbb{R}[\mathbf{B}]$ obtained by substitution of $\mathbf{p}$ with $\mathbf{c}$. For simplicity we denote by $f_{\mathbf{c}}(x)$ the polynomial $\varphi_{\mathbf{c}}(f)(= f(x, \mathbf{c}))$. To perform spectral factorization, we consider the following semi-algebraic set.

Definition 10 *A semi-algebraic set* $\mathtt{C} \subset \mathbb{R}^m$ *is called a* regular region *if, for any* $\mathbf{c} \in \mathtt{C}$, $a_n(\mathbf{c}) \neq 0$ *and there exist no roots of unit modulus in* $f_{\mathbf{c}}(x)$.

In the same way as the continuous-time case, the condition that f has no roots of unit modulus can be computed by the *quantifier elimination technique* or *real root counting* methods. Notice that a polynomial in x over $\mathbb{R}$ has a root of unit modulus if and only if it has a factor of the form of $x^2 + cx + 1$, $x - 1$, or $x + 1$ over $\mathbb{R}$. Due to the special structure (2), factors such as $x - 1$ and $x + 1$ in fact appear in $f(x)$ as their squares and thus are represented by $x^2 + cx + 1$. Substitute c with a new variable z. Then the resultant of $x^n f(x)$ and $x^2 + zx + 1$ becomes a polynomial $R(z)$ in z over $\mathbb{Q}[\mathbf{p}]$. It can be shown that, for a parameter value $\mathbf{c} \in \mathbb{R}^m$, $\varphi_{\mathbf{c}}(f)$ has no roots of unit modulus if and only if $\varphi_{\mathbf{c}}(R(z))$ has no real roots.

Now assume that $\mathtt{C}$ is a regular region for $f(x, \mathbf{p})$. We can compute the polynomial set $\mathcal{G}$ as in Lemma 5, where all polynomials are treated as ones over $\mathbb{Q}[\mathbf{p}]$. Also, for each $\mathbf{c} \in \mathtt{C}$, we can compute the polynomial set, say $\mathcal{G}_{\mathbf{c}}$, for $f_{\mathbf{c}}$ as in Lemma 5. Then, $\mathcal{G}_{\mathbf{c}} = \varphi_{\mathbf{c}}(\mathcal{G})$. Lemma 5 now implies that $\mathcal{G}$ and $\mathcal{G}_{\mathbf{c}}$ are still Gröbner bases over $\mathbb{Q}(\mathbf{p})$ and $\mathbb{Q}$ for ideals generated by themselves with respect to the graded reverse lexicographic order $\beta_n \succ \cdots \succ \beta_0$.

We now consider the ideal $\langle \mathcal{G} \rangle$ of spectral factorization in $\mathbb{Q}(\mathbf{p})[\mathbf{B}]$ and the ideal $\langle \mathcal{G}_{\mathbf{c}} \rangle$ of spectral factorization in $\mathbb{R}[\mathbf{B}]$. We note that all arguments in the previous subsections can be applied to the ideals in $\mathbb{Q}(\mathbf{p})[\mathbf{B}]$, as $\mathbb{Q}(\mathbf{p})$ is a field.

An important computational property of the characteristic polynomial of β_n can be derived as follows: $\mathcal{LB} = \left\{\beta_0^{d_0}\beta_1^{d_1} \cdots \beta_n^{d_n} \mid d_k \in \{0, 1\}\right\}$ is still a linear basis for $\mathbb{Q}(\mathbf{p})[\mathbf{B}]/\langle \mathcal{G} \rangle$, and, for the linear map $\mathbb{Q}(\mathbf{p})[\mathbf{B}]/\langle \mathcal{G} \rangle \ni g \to \beta_n g \in \mathbb{Q}(\mathbf{p})[\mathbf{B}]/\langle \mathcal{G} \rangle$, its matrix representation $M_{\mathbf{p}}$ with respect to $\mathcal{LB}$ is a matrix over $\mathbb{Q}[\mathbf{p}]$, since all the leading coefficients of the elements of $\mathcal{G}$ are 1 as polynomials in $\mathbb{Q}[\mathbf{p}, \mathbf{B}]$. Thus the characteristic polynomial $\tilde{S}_f$ is a polynomial in $\tilde{y}$ over $\mathbb{Q}[\mathbf{p}]$.

In the same manner, for each $\mathbf{c} \in \mathtt{C}$, we can compute the characteristic polynomial $\tilde{S}_{f_{\mathbf{c}}}$ as the characteristic polynomial of the matrix $M_{\mathbf{c}}$ derived from the linear map. Then, $M_{\mathbf{c}}$ coincides with the matrix obtained from $M_{\mathbf{p}}$ by substituting $\mathbf{p}$ with $\mathbf{c}$, and thus, $\varphi_{\mathbf{c}}(\tilde{S}_{f(x,\mathbf{p})}) = \tilde{S}_{f_{\mathbf{c}}}$.

Lemma 11 *The characteristic polynomial $\tilde{S}_f$ is a monic polynomial over $\mathbb{Q}[\mathbf{p}]$, and, for each $\mathbf{c} \in \mathtt{C}$, the characteristic polynomial $\tilde{S}_{f_{\mathbf{c}}}$ can be computed by $\tilde{S}_{f_{\mathbf{c}}}(\beta_n) = \varphi_{\mathbf{c}}\big(\tilde{S}_f(\beta_n)\big)$.*

Again there are the *generic case* and the *singular case* for the ideal spectral factorization over $\mathbb{Q}(\mathbf{p})$. In the generic case, $\tilde{S}_f(y)$ is square-free over $\mathbb{Q}(\mathbf{p})$ and the ideal $\langle \mathcal{G} \rangle$ is radical. Then we have the shape basis $\mathcal{F}$ with respect to an elimination ordering $\{\beta_0, \ldots, \beta_{n-1}\} \succ\succ \beta_n$ over $\mathbb{Q}(\mathbf{p})$ as in Theorem 8. For the singular case we compute the simple part $\tilde{T}$ of $\tilde{S}_f$ by using GCD computation, and consider the ideal $\langle \mathcal{G} \cup \{\tilde{T}\} \rangle$ instead of the original one. Then, over $\mathbb{Q}(\mathbf{p})$, we have the shape basis as in Theorem 9. Thus, from now on, we only deal with the generic case.

Now we outline our approach for the computation of the shape basis for the parametric case. Using *comprehensive Gröbner systems* [30, 31], we can compute the Gröbner basis of the ideal of spectral factorization with respect to the elimination ordering, where the region $\mathtt{C}$ may be decomposed to a number of cells:

$$\mathtt{C} = \bigcap_{i \in I} \mathtt{C}_i \text{ and } \#I < \infty \ , \tag{16}$$

where each $\mathtt{C}_i$ is given as a semi-algebraic set and its associated polynomial set $\mathcal{G}_i$ in $\mathbb{Q}(\mathbf{p})[\mathbf{B}]$ is computed such that $\varphi_{\mathbf{c}}(\mathcal{G}_i)$ forms a Gröbner basis of $\langle \mathcal{G}_{\mathbf{c}} \rangle$ for any $\mathbf{c} \in \mathtt{C}_i$. We can make the above comprehensive Gröbner system *reduced*, that is, for each $\mathbf{c} \in \mathtt{C}$, $\varphi_{\mathbf{c}}(\mathcal{G}_i)$ is reduced. Then, from the discussion in the previous subsection, it follows that, for each cell $\mathtt{C}_i$ in (16), $\mathcal{G}_i$ is a shape basis if $\tilde{S}_{f_{\mathbf{c}}}$ is square-free for any $\mathbf{c}$ in $\mathtt{C}_i$.

Definition 12 *In the resulted comprehensive Gröbner system* (16)*, if $\mathcal{G}_i$ is a shape basis, we say that $\mathtt{C}_i$ is* generic.

Thus, examining the forms of $\mathcal{G}_i$'s, we can extract all generic cells $\mathtt{C}_i$'s where $\mathcal{G}_i$'s are shape bases.

Remark 13 *To extract such a generic cell more efficiently, we can use the technique of* parametric basis conversion *proposed in [19]. Over $\mathbb{Q}(\mathbf{p})$, both $\mathcal{G}$ and its shape basis $\mathcal{F}$ are Gröbner bases (with respect to different orderings) and so it can be shown that, for any $\mathbf{c} \in \mathbb{R}^m$, if the denominators of elements of $\mathcal{F}$ do not vanish on $\mathbf{c}$, then $\varphi_{\mathbf{c}}(\mathcal{F})$ is also the shape basis of $\langle \varphi_{\mathbf{c}}(\mathcal{G}) \rangle$.*

For a non-generic cell $\mathtt{C}_i$, we can compute a certain shape basis by Theorem 9. In this case, by introducing new variable γ, we can extract the simple part of $\tilde{S}_f$. Let $\tilde{\mathcal{G}} := \mathcal{G} \cup \{\gamma \frac{d\tilde{S}_f}{d\beta_n} - 1\}$. Since the additional polynomial excludes multiple roots of $\tilde{S}_f$, computing the comprehensive Gröbner system for $\tilde{\mathcal{G}}$, we have a desired shape basis whose first element has β_n as its largest real root. Also we can apply efficient technique in [19].

4 Numerical Example

This section considers a particular control problem to demonstrate the discrete-time polynomial spectral factorization algorithm developed in Section 3 and further to show the applicability of the optimization framework proposed in this report. The problem we deal with is the $\mathcal{H}_2$ regulation problem with input penalty, which is often employed in a research direction in control called the characterization of performance limitations [32]. The problem is formulated as follows. In the feedback configuration in Figure 1, $P(z)$ is a plant, i.e., a system we want to control, and $K(z)$ is a controller that we have to design. The disturbance signal $d(k)$ is taken to be a unit pulse signal and the cost function we employ is

$$E := \sum_{k=0}^{\infty} \left(|y(k)|^2 + |u(t)|^2\right) .$$

The task is then to find a controller that achieves the smallest value of E among all stabilizing controllers and also the smallest value of E:

$$E^\star := \min_{K \text{ stabilizing}} E .$$

Given a fixed $P(z)$, a standard numerical procedure allows us to find the value of $E^\star$ and the controller that achieves this. Our focus is thus on a plant with parameters $P(z; \mathbf{p})$, where $\mathbf{p}$ is a vector of real parameters that can be tuned, and we aim to find the minimum value of $E^\star$:

$$\min_{\mathbf{p} \in \mathcal{Q}} E^\star ,$$

where $\mathcal{Q}$ is the feasible region of $\mathbf{p}$, specified as $\mathcal{F}(\mathbf{p})$ (cf. (1)). Write an n-th order single-input-single-output plant $P(z)$ as

$$P(z) = \frac{P_N(z)}{P_D(z)} ,$$

where $P_N(z)$ and $P_D(z)$ are coprime polynomials and $P_D(z)$ is n-th order and monic. Construct a self-reciprocal polynomial

$$P_N(z)P_N\left(\frac{1}{z}\right) + P_D(z)P_D\left(\frac{1}{z}\right) \tag{17}$$

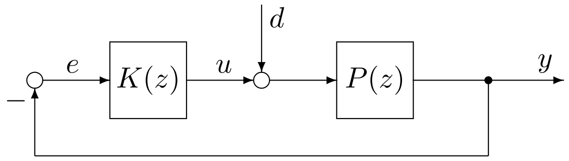

Figure 1: Unity feedback system configuration.

(which corresponds to $f(x)$ in (2)), and write its spectral factor as

$$M_D(z) = m_n(x+1)^n + m_{n-1}(x+1)^{n-1} + \cdots + m_0$$

(which corresponds to $g(x)$ in Definition 1 represented in the form of (12)). Under the assumptions that $P(z)$ is strictly proper and minimum-phase, $E^\star$ can be expressed as [27]

$$E^\star = m_n^2 - 1 \; . \tag{18}$$

Notice that (18) satisfies the condition for $\phi(\mathbf{p})$ in (1) since the development in Section 3 allows us to get a polynomial which has m_n as a real root and whose coefficients are polynomial in $\mathbf{p}$. It is also noted that, using the expressions of m_j's, we can get an expression for the optimal controller. Thus the essential part of the solution is to compute the shape basis for m_j's. It is emphasized that, although it is not immediate to see the relationship between the $\mathcal{H}_2$ regulation problem and the problem formulation (1), Gröbner basis theory can reformulate the control problem into (1) and thus that the formulation (1) is relevant to the control problem. Other important things to be noted here are that m_n (b_n and β_n in Section 3) represents an essential quantity in control and that it is beneficial to develop a polynomial spectral factorization algorithm in the discrete-time domain.

As a numerical example we employ the following:

$$P(z;p_1,p_2) = \frac{z - \left(p_1 - \frac{1}{8}\right)p_1}{z^2 + \left(1 + \frac{1}{100}p_2\right)z + p_2^2 + \frac{1}{4}} \; ,$$
$$(p_1,p_2) \in \mathcal{Q} := \left\{(p_1,p_2) \,\middle|\, p_1 \in [-\tfrac{1}{4}, \tfrac{1}{4}], p_2 \in [-\tfrac{1}{2}, \tfrac{1}{2}]\right\} \; .$$

Firstly it can be confirmed that the region $\mathcal{Q}$ is regular. Comparing the coefficients of (17) and $M_D(z)M_D\left(\frac{1}{z}\right)$, we obtain a set of polynomials. Using Lemma 5, we can get a Gröbner basis with respect to the graded reverse lexicographic order $m_2 \succ m_1 \succ m_0$:

$$\begin{aligned}\big\{m_0^2 - p_1^4 - p_2^4 + \tfrac{1}{4}p_1^3 + \tfrac{1}{50}p_2^3 - \tfrac{129}{64}p_1^2 - \tfrac{5001}{10000}p_2^2 + \tfrac{1}{4}p_1 + \tfrac{1}{200}p_2 - \tfrac{17}{16},& \\ m_1^2 - 2m_2m_0 + m_1m_0 - \tfrac{1}{100}p_2^3 + p_1^2 + 3p_2^2 - \tfrac{1}{8}p_1 - \tfrac{1}{80}p_2 - \tfrac{1}{4},& \\ m_2^2 + m_2m_1 + m_2m_0 - p_2^2 - \tfrac{1}{4}\big\} \; .&\end{aligned}$$

It can further be converted into a shape basis (which is not included due to space limitation) by using the basis conversion computation over the rational function field of parameters. None of the denominators vanishes inside $\mathcal{Q}$ and thus $\mathcal{Q}$ is generic (see Remark 13). Lemma 11 then confirms that the first element is the correct characteristic polynomial of m_2 for all $\mathbf{p} \in \mathcal{Q}$. In fact, since we are focusing on m_2 only in this problem, it suffices to construct the multiplication matrix of m_2 and to compute its characteristic polynomial. From the characteristic polynomial of m_2, we can obtain a polynomial in λ whose largest real root is equal to m_2^2:

$$\begin{aligned}
&\lambda^4 + (-p_1^4 - p_2^4 + \tfrac{1}{4}p_1^3 - \tfrac{1}{64}p_1^2 + \tfrac{14999}{10000}p_2^2 - \tfrac{1}{50}p_2 - \tfrac{41}{16})\lambda^3 \\
&\quad + (-2p_1^4p_2^2 - \tfrac{19999}{10000}p_2^6 + \tfrac{1}{2}p_1^3p_2^2 - \tfrac{1}{50}p_1^2p_2^3 + \tfrac{1}{50}p_2^5 + \tfrac{1}{2}p_1^4 - \tfrac{65}{32}p_1^2p_2^2 \\
&\quad + \tfrac{1}{400}p_1p_2^3 + \tfrac{30001}{20000}p_2^4 - \tfrac{1}{8}p_1^3 - \tfrac{1}{40}p_1^2p_2 + \tfrac{1}{4}p_1p_2^2 + \tfrac{1}{100}p_2^3 \\
&\quad - \tfrac{319}{128}p_1^2 + \tfrac{1}{320}p_1p_2 - \tfrac{459983}{160000}p_2^2 + \tfrac{5}{16}p_1 + \tfrac{17}{800}p_2 + \tfrac{5}{32})\lambda^2 \\
&\quad + (-p_1^4p_2^4 - p_2^8 + \tfrac{1}{4}p_1^3p_2^4 - \tfrac{1}{2}p_1^4p_2^2 - \tfrac{1}{64}p_1^2p_2^4 + \tfrac{9999}{10000}p_2^6 \\
&\quad + \tfrac{1}{8}p_1^3p_2^2 - \tfrac{1}{50}p_2^5 - \tfrac{1}{16}p_1^4 - \tfrac{1}{128}p_1^2p_2^2 - \tfrac{37501}{20000}p_2^4 + \tfrac{1}{64}p_1^3 \\
&\quad - \tfrac{1}{100}p_2^3 - \tfrac{1}{1024}p_1^2 - \tfrac{190001}{160000}p_2^2 - \tfrac{1}{800}p_2 - \tfrac{41}{256})\lambda \\
&\quad + p_2^8 + p_2^6 + \tfrac{3}{8}p_2^4 + \tfrac{1}{16}p_2^2 + \tfrac{1}{256} \; .
\end{aligned}$$

By means of a special QE algorithm [20] and the relationship (18), we can find that the global optimum of $E^\star$ is

$$\min_{\mathbf{p}\in\mathcal{Q}} \; E^\star \simeq 1.508 \; ,$$

and that this is achieved at

$$\mathbf{p}_{\text{opt}} \simeq (0.0625, -0.3457) \; .$$

It is emphasized that the optimal value and $\mathbf{p}_{\text{opt}}$ are the true global optimal values (cf. Figure 2) and moreover can be obtained as algebraic numbers and that they can be computed with arbitrary accuracy. The computation times required for obtaining the result are

Computation of the shape basis :	0.01 [sec]
Optimization of $E^\star$:	403.57 [sec]

by programs implemented in Maple running on a 1.33GHz PC with Intel Core Solo U1500.

5 Conclusions

This report has proposed a new framework of symbolic optimization that may encompass a wide range of problems arising from practical applications. Algebraic approaches play a crucial rôle in the framework: formulation into this framework and solution of the problem formulated in this framework. Also emphasized is the significance of the exploitation of the structural properties inherent (sometimes in an obscure way) in the problem. As an example this report has developed an algebraic approach to discrete-time polynomial spectral factorization, which is believed to deserve attention in its own right. It is expected that the development will find many applications in the area of signal processing.

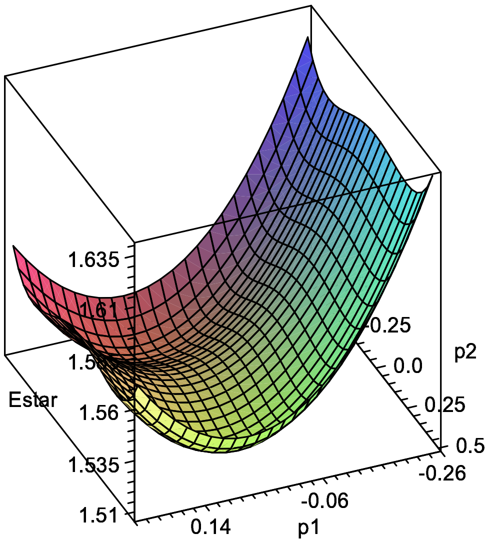

Figure 2: Plot of $E^{\star}$.

6 Acknowledgments

The authors would like to thank Ms. Silvia Gandy and Mr. Hideaki Tanaka for their help in preparing the numerical example.

References

[1] B. Hanzon and J. M. Maciejowski. Constructive algebra methods for the L_2-problem for stable linear systems. *Automatica*, 32(12):1645–1657, December 1996.

[2] H. Park. Optimal design of synthesis filters in multidimensional perfect reconstruction FIR filter banks using Gröbner bases. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 49(6):843–851, June 2002.

[3] G. Collins. Quantifier elimination for real closed fields by cylindrical algebraic decomposition. In *Proceedings Second GI Conference on Automata Theory and Formal Languages*, volume 33 of *Lecture Notes in Computer Science*, pages 134–183, Berlin, 1975. Springer-Verlag.

[4] G. E. Collins and H. Hong. Partial cylindrical algebraic decomposition for quantifier elimination. *Journal of Symbolic Computation*, 12(3):299–328, September 1991.

[5] A. Dolzmann, A. Seidl, and T. Sturm. Efficient projection orders for CAD. In *Proceedings of the ACM SIGSAM International Symposium on Symbolic and Algebraic Computation, ISSAC2004*, pages 111–118. ACM Press, 2004.

[6] V. Weispfenning. A new approach to quantifier elimination for real algebra. In B. F. Caviness and J. R. Johnson, editors, *Quantifier Elimination and Cylindrical Algebraic Decomposition*, Texts and Monographs in Symbolic Computation, pages 376–392. Springer, Wien, 1998.

[7] C. W. Brown. QEPCAD B: A program for computing with semi-algebraic sets using CADs. *ACM SIGSAM Bulletin*, 37(4):97–108, December 2003.

[8] H. Anai, H. Yanami, K. Sakabe, and S. Hara. Fixed-structure robust controller synthesis based on symbolic-numeric computation: Design algorithms with a CACSD toolbox (invited paper). In *Proceedings of CCA/ISIC/CACSD 2004*, pages 1540–1545, Taipei, Taiwan, 2004.

[9] B. D. O. Anderson, N. K. Bose, and E. I. Jury. Output feedback stabilization and related problems—solution via decision methods. *IEEE Transactions on Automatic Control*, AC-20(1):53–66, February 1975.

[10] P. Dorato, W. Yang, and C. Abdallah. Robust multi-objective feedback design by quantifier elimination. *Journal of Symbolic Computation*, 24(2):153–159, August 1997.

[11] I. A. Fotiou, P. Rostalski, P. A. Parrilo, and M. Morari. Parametric optimization and optimal control using algebraic geometry methods. *International Journal of Control*, 79(11):1340–1358, November 2006.

[12] H. Hoon and R. Liska, editors. *Journal of Symbolic Computation: Special Issue on Application of Quantifier Elimination*, volume 24, number 2. Academic Press, August 1997.

[13] S. Basu, R. Pollack, and M.-F. Roy. *Algorithms in Real Algebraic Geometry*, volume 10 of *Algorithms and Computation in Mathematics*. Springer-Verlag, Berlin, 2nd edition, 2006.

[14] M. Kanno, K. Yokoyama, H. Anai, and S. Hara. Parametric optimization in control using the sum of roots for parametric polynomial spectral factorization. In *Proceedings of the International Symposium on Symbolic and Algebraic Computation, ISSAC 2007*, pages 211–218, Waterloo, Ontario, Canada, July-August 2007.

[15] V. Weispfenning. Simulation and optimization by quantifier elimination. *Journal of Symbolic Computation*, 24(2):189–208, August 1997.

[16] A. M. Cohen, editor. *Computer Algebra in Industry: Problem Solving in Practice*. John Wiley & Sons, Chichester, 1993.

[17] D. Lazard and F. Rouillier. Solving parametric polynomial systems. *Journal of Symbolic Computation*, 42(6):636–667, June 2007.

[18] L. González-Vega, T. Recio, H. Lombardi, and M.-F. Roy. Sturm-Habicht sequences determinants and real roots of univariate polynomials. In B. F. Caviness and J. R. Johnson, editors, *Quantifier Elimination and Cylindrical Algebraic Decomposition*, Texts and Monographs in Symbolic Computation, pages 300–316. Springer, Wien, New York, 1998.

[19] H. Anai, S. Hara, M. Kanno, and K. Yokoyama. Parametric polynomial spectral factorization using the sum of roots and its application to a control design problem. Technical Report METR 2008-04, Department of Mathematical Informatics, The University of Tokyo, January 2008. Also submitted to the *Journal of Symbolic Computation*.

[20] M. Kanno, S. Gandy, H. Anai, and K. Yokoyama. Optimizing the maximal real root of a polynomial by a special cylindrical algebraic decomposition. Presented at *Mathematical Aspects of Computer and Information Sciences 2007*, Paris, France, December 2007.

[21] M. Kanno, S. Hara, H. Anai, and K. Yokoyama. Sum of roots, polynomial spectral factorization, and control performance limitations. In *Proceedings of the 46th IEEE Conference on Decision and Control*, pages 2968–2973, New Orleans, Louisiana USA, December 2007.

[22] B. Dumitrescu. *Positive Trigonometric Polynomials and Signal Processing Applications*. Signals and Communication Technology. Springer, Dordrecht, The Netherlands, 2007.

[23] V. Kučera. A tutorial on H_2 control theory: The continuous time case. In M. J. Grimble and V. Kučera, editors, *Polynomial Methods for Control Systems Design*, pages 1–55. Springer, London, 1996.

[24] A. H. Sayed and T. Kailath. A survey of spectral factorization methods. *Numerical Linear Algebra with Applications*, 8(6-7):467–496, September-November 2001.

[25] M. Kanno, H. Anai, and K. Yokoyama. On the relationship between the sum of roots with positive real parts and polynomial spectral factorization. In T. Boyanov et al., editors, *Numerical Methods and Applications — 6th International Conference, NMA 2006, Borovets, Bulgaria, August, 2006, Revised Papers*, volume 4310 of *Lecture Notes in Computer Science*, pages 320–328. Springer-Verlag, Heidelberg, 2007.

[26] K. Ogata. *Discrete-time Control Systems*. Prentice-Hall International, London, 2nd edition, 1995.

[27] H. Tanaka, M. Kanno, and K. Tsumura. Expressions for discrete-time $\mathcal{H}_2$ control performance limitations based on poles and zeros. In *Proceedings of SICE 8th Annual Conference on Control Systems (CD-Rom)*, Kyoto, Japan, March 2008.

[28] K. Yokoyama, M. Noro, and T. Takeshima. Solutions of systems of algebraic equations and linear maps on residue class rings. *Journal of Symbolic Computation*, 14(4):399–417, October 1992.

[29] M. Noro and K. Yokoyama. A modular method to compute the rational univariate representation of zero-dimensional ideals. *Journal of Symbolic Computation*, 28(1-2):243–264, July 1999.

[30] A. Montes. A new algorithm for discussing Gröbner bases with parameters. *Journal of Symbolic Computation*, 33(2):183–208, February 2002.

[31] A. Suzuki and Y. Sato. A simple algorithm to compute comprehensive Gröbner bases using Gröbner bases. In *Proceedings of the ACM SIGSAM International Symposium on Symbolic and Algebraic Computation, ISSAC2006*, pages 326–331. ACM Press, 2006.

[32] J. Chen and R. H. Middleton, editors. *IEEE Transactions on Automatic Control: Special Section on New Developments and Applications in Performance Limitation of Feedback Control*, volume 48, number 8. IEEE Control Systems Society, August 2003.

[33] D. Cox, J. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms*. Springer, New York, NY, 3rd edition, 2007.

Appendix: Proof of Lemma 5

Firstly the conversion between $\bar{g}_i$ and $\bar{\bar{g}}_k$ are linear and also triangular, and $\bar{g}_i$'s can also be expressed as linear combinations of $\bar{\bar{g}}_k$'s. Therefore the ideals $\langle \bar{g}_0, \bar{g}_1, \ldots, \bar{g}_n \rangle$ and $\langle \bar{\bar{g}}_0, \bar{\bar{g}}_1, \ldots, \bar{\bar{g}}_n \rangle$ are identical. So it is sufficient to show that $\mathcal{G}$ is the reduced Gröbner basis of the ideal generated by itself.

We first show that $\mathcal{G}$ is a Gröbner basis. Each $\bar{g}_k$ is quadratic and contains terms β_ℓ^2, $k \leq \ell \leq n$, but not $\beta_{\ell'}^2$, $0 \leq \ell' < k$. Also the coefficient of the term β_k^2 in $\bar{g}_k$ is 1. These facts imply that $\bar{\bar{g}}_k$ is quadratic and contains terms β_ℓ^2, $k \leq \ell \leq n$ (but not $\beta_{\ell'}^2$, $0 \leq \ell' < k$), with the coefficient of β_k^2 being 1. Moreover Claim A below states that, when constructing $\bar{\bar{g}}_k$, the coefficients of term $\beta_i \beta_j$, $k \leq i \leq j \leq n$, become 0 except for $i = j = k$, namely, all the terms bigger than β_k^2 disappear. Therefore the leading monomial of $\bar{\bar{g}}_k$ is β_k^2. (It is also noted that $\bar{\bar{g}}_k$ does not have monomials of the form β_ℓ^2 except β_k^2.) Furthermore the leading monomials of any pair of polynomials in $\mathcal{G}$ are prime to each other. Thus Buchberger's criterion [33, Section 2-2, Theorem 6] is satisfied, and $\mathcal{G}$ is a Gröbner basis of $\langle \mathcal{G} \rangle$.

Now the reducedness of the basis is proven. For any $\bar{\bar{g}}_k \in \mathcal{G}$,

$$\langle \mathrm{LT}(\mathcal{G} - \{\bar{\bar{g}}_k\}) \rangle = \{\beta_0^2, \beta_1^2, \ldots, \beta_n^2\} \setminus \{\beta_k^2\} \ , \tag{19}$$

where $\mathrm{LT}(\cdot)$ denotes the set of leading terms of elements. It is immediate that no monomial of $\bar{\bar{g}}_k$ lies in (19). That is, $\mathcal{G}$ satisfies the definition of the reduced Gröbner basis [33, Section 2-2, Definition 5] and is thus the reduced Gröbner basis.

Claim A *The coefficients of term $\beta_i\beta_j$, $k \le i \le j \le n$, in the polynomial $\bar{\bar{g}}_k$ are* 0 *except for $i = j = k$, and the coefficient of β_k^2 is 1.*

Proof: Let $d_{i,j}^k$ be the coefficient of the term $\beta_i\beta_j$ in $\bar{g}_k$ for $0 \le k \le n$, $0 \le i \le j \le n$. By (11) and (13), it follows that

$$\begin{cases} d_{i,j}^k = \displaystyle\sum_{\ell=0}^{n} \binom{i}{\ell}\binom{j}{\ell+k} + \sum_{\ell=0}^{n} \binom{j}{\ell}\binom{i}{\ell+k} & i < j \ , \\ d_{i,i}^k = \displaystyle\sum_{\ell=0}^{n} \binom{i}{\ell}\binom{i}{\ell+k} \ , \end{cases}$$

where $\binom{s}{t}$ is the binomial coefficient for $s, t \in \mathbb{N}$ and its value is defined 0 when $t < 0$ or $t > s$. Then it is straightforward to show that

$$\sum_{\ell=0}^{n} \binom{i}{\ell}\binom{j}{\ell+k} = \binom{i+j}{j-k} \ ,$$

which can be considered as a generalization of Pascal's arithmetic triangle. We thus have

$$\begin{cases} d_{i,j}^k = \dbinom{i+j}{j-k} + \dbinom{i+j}{i-k} & i < j \ , \\ d_{i,i}^k = \dbinom{2i}{i-k} \ . \end{cases}$$

Now we consider the coefficients of $\bar{\bar{g}}_k$. Let $e_{i,j}^k$ be the coefficient of the term $\beta_i\beta_j$ in $\bar{\bar{g}}_k$ for $0 \le k \le n$, $0 \le i \le j \le n$. By the definition (14) of $\bar{\bar{g}}_k$, we have

$$e_{i,j}^k = \sum_{\ell=k}^{n} c_{k,\ell} d_{i,j}^\ell \ . \tag{20}$$

To show that $e_{i,j}^k = 0$ for $j > i \ge k$ and that $e_{i,i}^k = 0$ for $i > k$, we use some kind of "induction argument on $i+j$". By Pascal's arithmetic triangle, we have the following for $k \le i < j$ and $k \le \ell$:

$$\begin{aligned} d_{i,j}^\ell &= \binom{i+j}{j-\ell} + \binom{i+j}{i-\ell} \\ &= \binom{i+j-1}{j-\ell} + \binom{i+j-1}{j-1-\ell} + \binom{i+j-1}{i-\ell} + \binom{i+j-1}{i-1-\ell} \\ &= \begin{cases} d_{i,j-1}^\ell + d_{i-1,j}^\ell & j-1 \ne i \ , \\ 2d_{i,i}^\ell + d_{i-1,i+1}^\ell & j-1 = i \ . \end{cases} \end{aligned}$$

Also we have

$$d_{i,i}^{\ell} = \binom{2i}{i-\ell} = \binom{2i-1}{i-\ell} + \binom{2i-1}{i-1-\ell} = d_{i-1,i}^{\ell} \ .$$

Hence, by (20), we have

$$\begin{aligned} e_{i,j}^k &= \begin{cases} e_{i,j-1}^k + e_{i-1,j}^k & j-1 \neq i \ , \\ 2e_{i,i}^k + e_{i-1,i+1}^k & j-1 = i \ , \end{cases} \\ e_{i,i}^k &= e_{i-1,i}^k \ . \end{aligned}$$

Thus, if $e_{i-1,j}^k = e_{i,j-1}^k = 0$, then we have $e_{i,j}^k = 0$. Moreover, if $e_{i-1,i}^k = 0$, then $e_{i,i}^k = 0$.

Using this fact, we can apply some kind of “induction argument”. Assume, to the contrary, that $e_{i,j}^k \neq 0$ for some $j \geq i \geq k$ (but not $i = j = k$). Then, at least either $e_{i-1,j}^k \neq 0$ or $e_{i,j-1}^k \neq 0$. Applying this argument recursively, we reach the conclusion that there is some $i > k$ such that $e_{k,i}^k \neq 0$. Conversely, it means that, if $e_{k,i}^k = 0$ for any $i > k$, then $e_{i,j}^k = 0$ for $j \geq i \geq k$ except for the case $i = j = k$.

Finally we show that $e_{k,i}^k = 0$ for any $i > k$, which completes the proof of Claim A. To this end we use an induction argument on i.

For $i = k+1$, direct computation shows that $e_{k,k+1}^k = 0$. Thus we assume that $e_{k,k+s}^k = 0$ for some $s > 0$, and show that $e_{k,k+s+1}^k = 0$. By the definition of $e_{k,k+s+1}^k$, we have

$$\begin{aligned} e_{k,k+s+1}^k &= \sum_{\ell=k}^{n} c_{k,\ell} d_{k,k+s+1}^{\ell} \\ &= \sum_{\ell=k}^{n} c_{k,\ell} \left\{ d_{k,k+s}^{\ell} + \binom{2k+s}{k+s+1-\ell} + \binom{2k+s}{k-1-\ell} \right\} \\ &= e_{k,k+s}^k + \sum_{\ell=k}^{k+s+1} c_{k,\ell} \binom{2k+s}{k+s+1-\ell} \ . \end{aligned}$$

We note that $\binom{2k+s}{k-1-\ell} = 0$, as $\ell \geq k$, and that $d_{k,k+s}^t = 0$ for $t > k+s$.

By the assumption, $e_{k,k+s}^k = 0$ and, thus it is enough to show that the second term vanishes:

$$\sum_{\ell=k}^{k+s+1} c_{k,\ell} \binom{2k+s}{k+s+1-\ell} = 0 \ .$$

By direct calculation, we have the desired conclusion:

$$
\begin{aligned}
&\sum_{\ell=k}^{k+s+1} c_{k,\ell}\binom{2k+s}{k+s+1-\ell}\\
&\quad= \sum_{\ell=k}^{k+s+1}(-1)^{k+\ell}\frac{2}{(2k)!}\frac{(k+\ell-1)!}{(\ell-k)!}\ell\frac{(2k+s)!}{(k+s+1-\ell)!(k+\ell-1)!}\\
&\quad= \sum_{t=0}^{s+1}(-1)^t\frac{2}{(2k)!}\frac{(2k+s)!(k+t)}{t!(s+1-t)!}\\
&\quad= \Delta k\sum_{t=0}^{s+1}(-1)^t\frac{1}{t!(s+1-t)!}+\Delta\sum_{t=1}^{s+1}(-1)^t\frac{1}{(t-1)!(s+1-t)!}\\
&\quad= \Delta\frac{k}{(s+1)!}\sum_{t=0}^{s+1}(-1)^t\binom{s+1}{t}+\Delta\frac{1}{s!}\sum_{t=1}^{s+1}(-1)^t\binom{s}{t-1}\\
&\quad= 0\ ,
\end{aligned}
$$

where $\Delta = \frac{2(2k+s)!}{(2k)!}$. □