Linear Matrix Inequalities for Global Optimization of Rational Functions and H_2 Optimal Model Reduction

Dorina Jibetean
Centrum voor Wiskunde en Informatica
P.O. Box 94079, 1090 GB Amsterdam
The Netherlands
d.jibetean@cwi.nl

Bernard Hanzon
Institute of Econometrics, Operations Research and System Theory
Technical University Vienna
Argentinierstrasse 8
A-1040 Vienna
Austria
bhanzon@eos.tuwien.ac.at

Abstract

In this paper we study unconstrained global optimization of rational functions. We give first few theoretical results and study then a relaxation of the initial problem. The relaxation is solved using LMI techniques. Therefore, in general our procedure will produce a lower bound of the infimum of the original problem. However, under no degeneracies, it is possible to check whether the relaxation was in fact exact. The algorithm is then applied to the H_2 optimal model reduction problem.

1 Introduction

Many problems in systems theory can be reformulated as optimization problems where the criterion function is a polynomial or rational function. For example in system identification of linear systems, one may try to estimate from the data the transfer function of a system, which is a (matrix) rational function, by the least squares method. The function to be estimated depends in general on some parameters that need to be identified. This procedure is nothing else than minimization of a multivariate rational function.

Another application area is the model reduction of the order of a system. There one wants to approximate a given stable system by a stable system of a lower order. Finding the *best* approximant (with respect to the H_2 distance) reduces again to optimization of a rational function. We believe however that the applications are much more numerous.

In these problems it is very important to know that the global minimum has been attained but present optimization methods do not give such guarantees. The numerical procedures used in general might return a local minimum. The goal of this paper is to describe an approach for obtaining the global minimum of an arbitrary rational function. For that we follow the approach described in [6] used for (global) minimization of polynomial functions. There, one defines a relaxation of the optimization problem, relaxation which can be solved using linear matrix inequalities (LMI) techniques. It is argued that, under no degeneracies, one can check whether the relaxation was exact. In this paper we give some theoretical results which will allow us to extend the method of [6] to (global) optimization of rational functions. An interesting criterion will be given for a rational function to have the infimum at $-\infty$.

We show then that the algorithm can be applied to the H_2-model reduction problem. This is in itself an interesting problem and has received a lot of attention.

2 Optimization of rational functions

We summarize in this section a general method for finding the global minimum of a multivariate rational function. First the problem is rewritten into an equivalent one. Then a relaxation of the latter using linear matrix inequalities (LMI) is formulated.

2.1 An equivalent formulation

In this section we want to summarize some of the results that will be used in the paper. For the proofs, see [5]. Let us state here a preliminary result which will prove essential in the following.

Lemma 2.1. *Let $a(x)/b(x)$ be a rational multivariate function, with $a(x)$, $b(x)$ relatively prime polynomials. If $a(x)/b(x) \geq 0$, $\forall x \in \{\mathbf{R}^n \mid b(x) \neq 0\}$, then one of the two following statements holds:*

- $a(x) \geq 0, \quad b(x) \geq 0 \quad \forall x \in \mathbf{R}^n$,
- $a(x) \leq 0, \quad b(x) \leq 0 \quad \forall x \in \mathbf{R}^n$.

The result was proved using results from real algebraic geometry. In the sequel, we will say that a polynomial a in n variables changes sign on $\mathbf{R}^n$ if $\exists x_1,\ x_2 \in \mathbf{R}^n$ such that $a(x_1) > 0$ and $a(x_2) < 0$. Otherwise, by a slight abuse of terminology, we say that a has constant sign on $\mathbf{R}^n$.

Consider now the problem

$$\inf_{x\in\mathbf{R}^n} \frac{p(x)}{q(x)}, \quad \text{with} \quad p(x),\ q(x) \in \mathbf{R}[x] \quad \text{relatively prime.} \tag{2.1}$$

In the following we use Lemma 2.1 in order to obtain a criterion for our problem.

Proposition 2.1. *Let $p(x)/q(x)$ be a rational function with $p(x)$, $q(x)$ relatively prime polynomials. If $q(x)$ changes sign on $\mathbf{R}^n$ then $\inf_{x\in\mathbf{R}^n} p(x)/q(x) = -\infty$.*

Proof. We prove it by reduction to absurd. Assume $\exists\ \alpha \in \mathbf{R}$ a lower bound of the function. We have

$$\frac{p(x)}{q(x)} \geq \alpha \quad \forall x \in \{\mathbf{R}^n \mid q(x) \neq 0\} \iff \frac{p(x) - \alpha q(x)}{q(x)} \geq 0 \quad \forall x \in \{\mathbf{R}^n \mid q(x) \neq 0\}.$$

Applying Lemma 2.1, we deduce that both $p(x) - \alpha q(x)$ and $q(x)$ have constant sign on $\mathbf{R}^n$ which contradicts the hypothesis. □

The reciprocal is not true in general. However, we can reformulate now the problem (2.1). Suppose now that q has constant sign on $\mathbf{R}^n$.

Theorem 2.1. *Let q have constant sign on $\mathbf{R}^n$. Assuming, without loss of generality, that $q(x) \geq 0 \quad \forall x \in \mathbf{R}^n$, then the problem* (2.1) *is equivalent to*

$$\begin{array}{ll} \sup & \alpha \\ \text{s.t.} & p(x) - \alpha q(x) \geq 0, \quad \forall x \in \mathbf{R}^n. \end{array} \tag{2.2}$$

In other words, we search for the largest α which is a lower bound (finite or not) of the rational function, as before. Note that if the rational function has no finite lower bound α, then the feasibility domain of (2.2) is the empty set. In this case the supremum will be $-\infty$.

The condition $q(x) \geq 0 \quad \forall x \in \mathbf{R}^n$ can be checked in the following way. Evaluate q at an arbitrary point and suppose that it is indeed positive. Then q is non-negative on $\mathbf{R}^n$ if and only if $\inf_{x \in \mathbf{R}^n} q(x) \geq 0$. Hence we only need to compute the infimum of a polynomial on $\mathbf{R}^n$ and this can be done using for example the algorithm described in [2]. One could also use the algorithm of [6], [7], but we should strongly warn here against numerical errors.

To conclude, in this section we have rewritten the rational optimization problem as a constrained polynomial optimization problem. In order to solve (2.2), we construct an LMI relaxation of it.

2.2 An LMI relaxation

In this section we are going to relax the constraint in the problem (2.2) using the well-known method (see [7], [6]) for checking the non-negativity of a polynomial. The method is based on rewriting a given polynomial as a sum of squares of polynomials. If that succeeds, then we know the original polynomial is non-negative. However not every non-negative polynomial can be written as a sum of squares of polynomials, therefore this is just a relaxation of the original problem.

To be more precise, we construct the matrix N such that

$$p(x) - \alpha q(x) = z^T N z, \quad , z^T = [1, x_1, x_2, \ldots, x_n, x_1 x_2, \ldots, x_n^d].$$

Here $d = \lceil \text{tdeg}(p(x) - \alpha q(x))/2 \rceil$, $\text{tdeg}(p(x) - \alpha q(x))$ is the total degree of $p(x) - \alpha q(x)$ in the variables x (α being considered a parameter), and z contains all monomials of degree less than or equal to d. Note that such a matrix N can always be constructed. Moreover it is symmetric and, in general, not uniquely determined. It is easy to show that the set of all such N forms an affine space, i.e. $N = N_0 + \sum_{l=1}^{k} \lambda_l N_l + \alpha N_{k+1}$ where $N_l,\ l = 0, \ldots, k+1$ are constant matrices (see [5]). Also, if N is positive semi-definite for some value of $\lambda = (\lambda_1, \ldots, \lambda_k)$, then $p(x) - \alpha q(x) \geq 0,\ \forall x \in \mathbf{R}^n$.

Let us denote the generic matrix constructed above by $N(\lambda, \alpha)$, with $\lambda \in \mathbf{R}^k,\ \alpha \in \mathbf{R}$. Consider the LMI problem:

$$\begin{array}{ll} \sup & \alpha \\ \text{s.t.} & N(\alpha, \lambda) \succeq 0. \end{array} \tag{2.3}$$

Indeed, since $N(\alpha, \lambda)$ is symmetric, the matrices $N_l,\ l = 0, \ldots, k+1$ will be symmetric. Moreover, $N(\alpha, \lambda)$ is affine in (α, λ), hence the problem is a standard LMI problem. The relation between the problems (2.3) and (2.2) is studied in the following.

Theorem 2.2. *Let us denote by α_{RAT} the solution of the problem* (2.2)*, and consequently of the rational optimization problem* (2.1)*, and by α_{LMI} the solution of* (2.3)*. Then we have*

$$\alpha_{RAT} \geq \alpha_{LMI}.$$

If $p(x) - \alpha_{RAT} q(x)$ can be written as a sum of squares, then

$$\alpha_{RAT} = \alpha_{LMI}.$$

There are particular cases in which a positive polynomial can always be written as a sum of squares of polynomials. (see [1], Propositions 6.4.3, 6.4.4). Hence, if the polynomial $F(x) = p(x) - \alpha q(x)$ is in one of these cases, the solution of (2.3) will coincide with the solution of (2.1), according to Theorem 2.2. If not, then there is always a polynomial $G(x)$ such that $F(x)G^2(x)$ can be written as a sum of squares of polynomials (see [1], Theorem 6.1.1). It is not clear however how to chose the polynomial $G(x)$. From the practical point of view we are more interested in deciding whether for a particular rational function the infimum was found or just a lower bound of it. A checking procedure is described in [6] which decides whether the LMI relaxation was exact. The same procedure can be applied to our problem.

For the example presented in this paper however, an alternative procedure was used. We compared the lower bound α_{LMI} with different *upper bounds* (actually local minima of the rational function), obtained by running a steepest descent algorithm.

3 Optimal H_2 model reduction

Let us denote by $\Sigma = (A, B, C, D)$ a linear, continuous or discrete time-invariant, stable system of order n. By stable, we mean that all eigenvalues of A are in the open left-half plane in the continuous-time case, respectively in the open unit circle in the discrete-time case. The H_2 model reduction problem is finding the closest (in H_2 distance) system $\hat{\Sigma} = (\hat{A}, \hat{B}, \hat{C}, \hat{D})$

linear, continuous respectively discrete time-invariant, stable, of *given* order $\hat{n}$. Formulated differently, we want to solve:

$$\min_{\hat{\Sigma}-\text{stable}} \|\Sigma - \hat{\Sigma}\|_2^2.$$

That is, for

- continuous-time systems: $\|\Sigma - \hat{\Sigma}\|_2^2 = \text{trace}((D-\hat{D})^T(D-\hat{D})) + \text{trace}(B^T M_1 B + 2B^T M_2 \hat{B} + \hat{B}^T M_3 \hat{B})$,
  with $A^T M_1 + M_1 A = -C^T C,\ A^T M_2 + M_2 \hat{A} = C^T \hat{C},\ \hat{A}^T M_3 + M_3 \hat{A} = -\hat{C}^T \hat{C}$.

- discrete-time systems: $\|\Sigma - \hat{\Sigma}\|_2^2 = \text{trace}((D-\hat{D})^T(D-\hat{D})) + \text{trace}(B^T L_1 B + 2B^T L_2 \hat{B} + \hat{B}^T L_3 \hat{B})$,
  with $L_1 - A^T L_1 A^T = C^T C,\ L_2 - A^T L_2 \hat{A} = -C^T \hat{C},\ L_3 - \hat{A}^T L_3 \hat{A} = \hat{C}^T \hat{C}$.

Obviously, in both cases the criterion is minimized for $D = \hat{D}$ and $\text{trace}((D-\hat{D})^T(D-\hat{D}))$ becomes 0.

Note that one needs to solve the Lyapunov/Sylvester equations in M_1, M_2, M_3 (respectively L_1, L_2, L_3). Actually some canonical forms of (A, B, C) are more advantageous for this problem, as we will see later. It should be remarked however that M_1, M_2, M_3 (respectively L_1, L_2, L_3) will be multivariate rational functions, therefore the criterion to be minimized will be a multivariate rational function as well.

We treat here a particular case of the H_2 optimal model reduction, namely reduction of SISO continuous-time systems. It should be stressed that the procedure is also applicable to the MIMO continuous-time case, using for example canonical forms for stable linear systems constructed in [4].

Let us consider the SISO case. As we have mentioned before, we still have at this point the choice for a parametrization of (A, b, c). In the following we choose the parametrization trying to satisfy two criteria. The first one is the stability requirement, therefore we use canonical forms for *stable* systems. Secondly we want to simplify our calculations (i.e. solving the Lyapunov/Sylvester equations) as much as possible. One way is to choose a so-called output canonical form for (A, c) (respectively $(\hat{A}, \hat{c})$), that is equivalent to saying that M_1, the solution of the Lyapunov equation associated to (A, c), satisfies $M_1 = I_n$ (respectively $M_3 = I_{\hat{n}}$.)

It turns out that in both SISO and MIMO cases there exist parametrizations satisfying both mentioned requirements (see [3], [4]).

In the SISO case, we follow [3]. Let (A, b, c), $(\hat{A}, \hat{b}, \hat{c})$ have Schwarz-like canonical form for *stable* systems:

$$\hat{A} = \begin{pmatrix} -\frac{1}{2}x_1^2 & -x_2 & 0 & \dots & 0 \\ x_2 & 0 & -x_3 & \dots & 0 \\ \vdots & & & & \\ 0 & 0 & \dots & \dots & -x_{\hat{n}} \\ 0 & 0 & \dots & x_{\hat{n}} & 0 \end{pmatrix}, \quad \hat{c} = \begin{pmatrix} x_1 & 0 & 0 & \dots & 0 \end{pmatrix}$$

with $x_i > 0, \ \forall i = 1, \ldots, \hat{n}$, and $b \in \mathbf{R}^{\hat{n}\times 1}$ is a free vector, with the condition that (A, b) is a reachable pair. It is easy to see that $(A, c), \ (\hat{A}, \hat{c})$ are output-normal, i.e. $\hat{A}^T + \hat{A} = -\hat{c}^T\hat{c}, \ A^T + A = -c^T c$. That is equivalent with $M_1 = I_n, \ M_3 = I_{\hat{n}}$.

Moreover, the criterion is quadratic in $\hat{b}$. By optimizing first with respect to $\hat{b}$ one obtains $\hat{b} = -M_2^T b$ and $\|\Sigma - \hat{\Sigma}\|_2^2 = b^T b - b^T M_2 M_2^T b$, where M_2 is the solution of the above mentioned Sylvester equation. The optimization problem becomes

$$\min_{x_i>0, i=1\ldots,\hat{n}} b^T b - b^T M_2 M_2^T b = b^T b - \min_{x_i>0, i=1\ldots,\hat{n}} \frac{p(x_1, \ldots, x_{\hat{n}})}{q(x_1, \ldots, x_{\hat{n}})}.$$

However, the positivity constraints can be dropped due to the following argument. In general, one can show that one property of the Schwarz-like canonical form is that the criterion $p(x)/q(x)$ contains only even powers of $x_1, \ldots, x_{\hat{n}}$. The proof is based on the fact that the H_2 distance between the two systems, and therefore our criterion $p(x)/q(x)$, does not depend on a particularly chosen $(\hat{A}, \hat{b}, \hat{c})$ representation of the approximant system. Moreover, if for a certain $i = 2, \ldots, \hat{n}$ one replaces x_i by $-x_i$ in a given Schwarz-like canonical representation $(\hat{A}, \hat{b}, \hat{c})$, or if one replaces x_1 by $-x_1$ and simultaneously $\hat{b}$ by $-\hat{b}$, then one obtains an *equivalent* Schwarz-like representation. That implies

$$\frac{p(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_{\hat{n}})}{q(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_{\hat{n}})} = \frac{p(x_1, \ldots, x_{i-1}, -x_i, x_{i+1}, \ldots, x_{\hat{n}})}{q(x_1, \ldots, x_{i-1}, -x_i, x_{i+1}, \ldots, x_{\hat{n}})}$$

and therefore the criterion p/q contains only even powers of $x_i, \ i = 1, \ldots, \hat{n}$. Hence, one only needs to solve an unconstrained rational optimization problem. In this case, when minimizing over $(x_1, \ldots, x_{\hat{n}}) \in \mathbf{R}^{\hat{n}}$ we obtain *symmetric* solutions with respect to the origin and the axis, i.e. for any solution $(x_1, \ldots, x_{\hat{n}})$, we know that $(\pm|x_1|, \ldots, \ \pm|x_{\hat{n}})|)$ are solutions as well.

We consider now a concrete example, namely the *Oscillatory system* of [3]. There an optimal approximant was constructed using constructive algebra methods. In the end we intend to compare the results.

Example 3.1. *Find the* best *H_2-approximant of second order ($\hat{n} = 2$), of the system*

$$Tf(s) = \frac{s^2 - s + 2}{s^3 + 0.5s^2 + 2s + 0.5} \qquad (n = 3).$$

Tf corresponds to a Schwarz-like canonical form with parameters $x_1 = x_2 = x_3 = b_1 = b_2 = b_3 = 1, \ d = 0$.

Let us consider a general second order, stable system in Schwarz-like canonical form

$$\hat{A} = \begin{pmatrix} -\frac{1}{2}x_1^2 & -x_2 \\ x_2 & 0 \end{pmatrix}, \quad \hat{c} = \begin{pmatrix} x_1 & 0 \end{pmatrix}, \quad x_1 > 0, \ x_2 > 0$$

We want to find the values of the parameters $x_1, \ x_2$ for which the criterion $b^T b - b^T M_2 M_2^T b$ is minimized. Since b is given, let us now compute the matrix $M_2 \in \mathbf{R}^{3\times 2}$ (as function of $x_1, \ x_2$) from the *linear* system of equations $A^T M_2 + M_2 \hat{A} = c^T \hat{c}$.

We obtain

$$M_2 = \frac{2}{\Delta}\begin{pmatrix} -4x_1 - x_1^5 + 8x_1x_2^2 - 4x_1^3x_2^2 - 4x_1x_2^4 & 16x_1x_2 + 2x_1^5x_2 - 24x_1x_2^3 + 8x_1x_2^5 \\ 2x_1^3 + 16x_1x_2^2 - 8x_1x_2^4 & 4x_1x_2 - 4x_1x_2^3 - 4x_1^3x_2^3 \\ -4x_1 + 4x_1x_2^2 + 4x_1^3x_2^2 & 16x_1x_2 + 2x_1^3x_2 + 2x_1^5x_2 - 8x_1x_2^3 \end{pmatrix}$$

$$\text{with } \Delta = 4 + 8x_1^2 + x_1^4 + x_1^6 + 56x_2^2 - 4x_1^2x_2^2 + 8x_1^4x_2^2 - 60x_2^4 + 4x_1^2x_2^4 + 16x_2^6.$$

The criterion will be

$$b^Tb + \min -b^TM_2M_2^Tb = 3 + \min -\frac{p(x_1,x_2)}{q(x_1,x_2)}$$

where

$p(x_1,x_2) = 4x_1^2(64 - 32x_1^2 + 20x_1^4 - 4x_1^6 + x_1^8 + 848x_2^2 + 256x_1^2x_2^2 + 236x_1^4x_2^2 + 16x_1^6x_2^2 + 16x_1^8x_2^2 - 1616x_2^4 - 480x_1^2x_2^4 - 280x_1^4x_2^4 - 32x_1^6x_2^4 + 1200x_2^6 + 320x_1^2x_2^6 + 80x_1^4x_2^6 - 432x_2^8 - 64x_1^2x_2^8 + 64x_2^{10})$

and

$q(x_1,x_2) = (4 + 8x_1^2 + x_1^4 + x_1^6 + 56x_2^2 - 4x_1^2x_2^2 + 8x_1^4x_2^2 - 60x_2^4 + 4x_1^2x_2^4 + 16x_2^6)^2$.

We apply now the procedure described in Section 2. Note that the denominator of the rational function is the square of a polynomial, hence it is non-negative on R^2. For solving the problem (2.2), we construct the LMI relaxation (2.3), using the vector of monomials z. In order to reduce the size of our problem, and since the polynomials p, q contain only even powers of the variables, we consider only monomials of even power in the vector z as well. We have $\text{tdeg}(p(x) - \alpha q(x)) = 12$, therefore the vector z will contain monomials of degree less or equal its half, that is $m = 6$ and $z^T = \begin{pmatrix} 1 & x_1^2 & x_1^4 & x_1^6 & x_2^2 & x_2^4 & x_2^6 & x_1^2x_2^2 & x_1^4x_2^2 & x_1^2x_2^4 \end{pmatrix}$. In this case, considering this vector z turns out to be sufficient for finding the global minimum. In general however, resticting the number of monomials in z may lead to a strict lower bound of the global minimum.

Let us now construct an arbitrary, symmetric matrix $N \in \mathbf{R}[\alpha,\lambda]^{10\times 10}$. Its dimension is obviously determined by the length of z. We compute $N(\lambda,\alpha)$ by equalizing the coefficients of the polynomials $p(x) - \alpha q(x)$ and z^TNz. It turns out from the computations that $\lambda \in \mathbf{R}^{28}$. The only thing left now it to compute the solution for the LMI relaxation. We obtain, as a lower bound on the infimum of the original problem 1.1117. At this point we still need to decide whether this is a strict lower bound or not. In this case we have run a standard steepest descent algorithm which finds a (local) minimum at $(x_1,x_2) = (1.1916, 0.4183)$ for which the value of the criterion equals our lower bound! This tells us two things, first that the lower bound was exact, secondly that the point $(1.1916, 0.4183)$ is actually a global minimum. Hence, we have found a best approximant in H_2 norm and this is given by

$$\hat{A} = \begin{pmatrix} -0.7099 & -0.4183 \\ 0.4183 & 0 \end{pmatrix}, \quad \hat{c} = \begin{pmatrix} 1.1916 & 0 \end{pmatrix}, \quad \hat{b} = \begin{pmatrix} 0.2080 \\ -1.3118 \end{pmatrix}$$

Under certain conditions, other (more direct) methods than the one presented here can be used to decide whether the obtained lower bound is exact or not. For more details see [6].

For obtaining the lower bound 1.1117 we have run an algorithm which consists out of two parts. The first one, for constructing the LMI relaxation was implemented in Mathematica 4.0 and takes 12 seconds and 16.7 Kb on a Sun Ultra 5 station. Then, for solving the LMI problem we use SeDuMi1.03, a free software package (see [8]) running under Matlab. This takes another 5 seconds (of which 2 are used to read the data obtained with Mathematica). Unfortunately, SeDuMi runs sometimes into numerical problems.

4 Conclusions

In this paper we develop an algorithm for global optimization of rational functions. The approach is based on rewriting a rational optimization problem in $\mathbf{R}^n$ as a constrained polynomial optimization problem of a particular type.

Such equivalent formulation of the problem can in principle be solved using a different algorithm than the one discussed here. We have chosen to apply the algorithm of [7], [6] for its possible relevance in applications. The translation of our problem into this setting was immediate.

We also show how the algorithm can be used for finding the *best* H_2 approximant of a system.

References

[1] J. Bochnak, M. Coste, M-F. Roy, *Géométry algébrique réelle*, Springer-Verlag, 1987

[2] B. Hanzon, D. Jibetean, *Global minimization of a multivariate polynomial using matrix methods,* CWI Report PNA-R0109 July 2001.

[3] B. Hanzon, J. Maciejowski, *Constructive algebra methods for the* L_2*-Problem for stable Linear systems*, Automatica, vol.32 no 12 pp 1645-1657.

[4] B. Hanzon, R. Ober, *Overlapping block-balanced canonical forms for various classes of linear systems*, Linear Algebra and its Applications 281 (1998) 171-225.

[5] D. Jibetean, *Global optimization of rational multivariate functions*, CWI Report PNA-R120 October 2001.

[6] P. Parrilo, *Semidefinite programming relaxations for semialgebraic problems*, submitted

[7] N. Shor, *Class of global minimum bounds of polynomial functions*, Translated from Kibernetica, no. 6,pp. 9-11, November-December, 1987.

[8] J. Sturm, *SeDuMi 1.03* available at http://fewcal.kub.nl/sturm/software/sedumi.html