Proceedings of the 6th Nordic Signal Processing Symposium - NORSIG 2004
June 9 - 11, 2004
Espoo, Finland

Least Squares Optimization of 2-D IIR Filters

Bogdan Dumitrescu

Tampere University of Technology
Institute of Signal Processing
Tampere International Center for Signal Processing
P.O.Box 553, 33101 Tampere
FINLAND
e-mail: bogdand@cs.tut.fi

ABSTRACT

We present an algorithm for the least squares optimization of 2-D IIR filters with separable or nonseparable denominator. The algorithm is iterative and each iteration consists of solving a semidefinite programming problem. We adapt the Gauss-Newton idea which outcomes to a convex approximation of the optimization criterion. The stability of the 2-D IIR filters is ensured by keeping the denominator inside convex stability domains, defined by linear matrix inequalities. In the experimental section, 2-D IIR filters with separable and nonseparable denominator are designed and compared.

1. INTRODUCTION

The interest in the optimization of IIR filters has been revived in the latest years by several techniques. Among the recent improvements we remark different ways to replace the optimization criterion by a convex approximation and the use of convex stability domains [1, 2, 3]. A benefit of these approaches is the use of efficient convex optimization methods like semidefinite programming (SDP). Some of the ideas can be transferred easily to the 2-D case, especially for filters with separable denominators [2, 4]. In this paper, we present an algorithm for least squares (LS) optimization of 2-D IIR filters which merges in a single framework the separable and the nonseparable cases.

The 2-D IIR filter has the transfer function

$$H(z_1,z_2) = \frac{B(z_1,z_2)}{A(z_1,z_2)} = \frac{\sum_{k_1=0}^{m_1}\sum_{k_2=0}^{m_2} b_{k_1,k_2} z_1^{-k_1} z_2^{-k_2}}{\sum_{k_1=0}^{n_1}\sum_{k_2=0}^{n_2} a_{k_1,k_2} z_1^{-k_1} z_2^{-k_2}}, \qquad (1)$$

with $a_{00} = 1$. If the factorization $A(z_1,z_2) = A_1(z_1)A_2(z_2)$ is possible, then the denominator is separable. We assume that a complex desired frequency response $D(\omega_1,\omega_2)$ is given, with $\omega_1,\omega_2 \in [-\pi,\pi]$, whose values are D_{ℓ_1,ℓ_2} on a grid of frequencies defined by $\omega_1(\ell_1)$, $\omega_2(\ell_2)$, $\ell_1 = 1 : L_1$, $\ell_2 = 1 : L_2$. The optimization problem is to find the coefficients of a filter (1), with given degrees of numerator and denominator, that minimize the (discretized) LS error

$$J(A,B) = \sum_{\ell_1=1}^{L_1}\sum_{\ell_2=1}^{L_2} \lambda_{\ell_1,\ell_2} \left| D_{\ell_1,\ell_2} - \frac{B(\omega_1(\ell_1),\omega_2(\ell_2))}{A(\omega_1(\ell_1),\omega_2(\ell_2))} \right|^2, \qquad (2)$$

where $A(\omega_1,\omega_2)$, $B(\omega_1,\omega_2)$ are the frequency responses of the denominator and numerator of (1). The numbers $\lambda_{\ell_1,\ell_2} \geq 0$ represent weights. We name $J(A,B)/(L_1L_2)$ the normalized criterion. For given coefficients of (1), we denote $A_{\ell_1,\ell_2} = A(\omega_1(\ell_1),\omega_2(\ell_2))$ and $B_{\ell_1,\ell_2} = B(\omega_1(\ell_1),\omega_2(\ell_2))$.

The difficulties in optimizing (2) are the nonconvexity of the criterion and the necessity of ensuring the stability of the IIR filter. Generally, the nonconvexity is treated like in the 1-D case, with the burden of an increased number of parameters in (1). For separable denominator, stability can be enforced using 1-D techniques. In the nonseparable case, stability was ensured by various means, mainly using sufficient conditions. The positive realness condition $\mathrm{Re}A(e^{j\omega_1}, e^{j\omega_2}) > 0$ was used in [5], while in [6] the denominator was expressed as a product of low-order polynomials for which stability conditions are known.

The contents and contributions of this paper are as follows. In section 2, we describe an algorithm for the LS optimization of 2-D IIR filters; it uses a Gauss-Newton convexification of the LS criterion (2); the stability of the filter is ensured by using 2-D convex stability domains built around the current denominator $A(z_1,z_2)$; the description of the stability domain is based on a positive realness condition and has the form of a linear matrix inequality (LMI); each iteration of the method consists of solving an SDP problem. In section 3, we give examples showing that best performance may be given by filters with either separable and nonseparable denominator, depending on the desired response $D(\omega_1,\omega_2)$; we also give a comparison with a recent design from [4].

2. THE NEW ALGORITHM

With $n_1' = n_1 + 1$ and $n_2' = n_2 + 1$, we denote $\mathbf{A} \in \mathbb{R}^{n_1' \times n_2'}$ the matrix of coefficients of the polynomial $A(z_1,z_2)$, such that we have

$$A(z_1,z_2) = [1\ z_1^{-1}\ \ldots\ z_1^{-n_1}] \cdot \mathbf{A} \cdot [1\ z_2^{-1}\ \ldots\ z_2^{-n_2}]^T.$$

Work supported by Academy of Finland, project No. 44876 (Finnish Centre of Excellence Program (2000-2005)). B.Dumitrescu is on leave from the Department of Automatic Control and Computers, "Politehnica" University of Bucharest, Romania.

We denote $\text{vec}(A) \in \mathbb{R}^N$ the vector of coefficients obtained by concatenating the columns of $\mathbf{A}$.

2.1 2-D convex stability domain

Let us assume that $A(z_1,z_2)$ is the denominator of a stable 2-D IIR filter (1), i.e. $A(z_1,z_2) \neq 0$ for any $|z_1| \geq 1$, $|z_2| \geq 1$; we name such $A(z_1,z_2)$ a Schur polynomial; remind that $a_{0,0} = 1$. Let $\Delta(z_1,z_2)$ be a polynomial of degree (n_1,n_2), with $\delta_{0,0} = 0$. Let $\mathcal{D}_A$ be a set of stable polynomials, built around $A(z_1,z_2)$ and containing polynomials of the form $\tilde{A} = A + \Delta$. To build a convex domain $\mathcal{D}_A$, we appeal at the following results presented in [7] and using a parameterization of polynomials that are sum-of-squares and thus nonnegative on the unit bicircle. We consider polynomials

$$R(z_1,z_2) = \sum_{k_1=-n_1}^{n_1} \sum_{k_2=-n_2}^{n_2} r_{k_1,k_2} z_1^{-k_1} z_2^{-k_2}, \tag{3}$$

which are symmetric, i.e. $r_{k_1,k_2} = r_{-k_1,-k_2}$, and may be written as a sum-of-squares, i.e. in the form

$$R(z_1,z_2) = \sum_{\ell=1}^{\nu} F_\ell(z_1^{-1},z_2^{-1}) F_\ell(z_1,z_2), \tag{4}$$

with $\nu \leq N = n_1' n_2'$ and $F_\ell(z_1,z_2)$ polynomials of degree (n_1,n_2).

Theorem 1 *A polynomial (3) is sum-of-squares if and only if there exists a positive semidefinite matrix $\mathbf{Q} \in \mathbb{R}^{N \times N}$ such that*

$$r_{k_1,k_2} = tr[(\mathbf{T}_{k_2} \otimes \mathbf{T}_{k_1}) \cdot \mathbf{Q}], \tag{5}$$

where $\mathbf{T}_{k_1} \in \mathbb{R}^{n_1' \times n_1'}$ and $\mathbf{T}_{k_2} \in \mathbb{R}^{n_2' \times n_2'}$ are elementary Toeplitz matrices with ones only on the k_1-th and k_2-th diagonals, respectively; $tr\mathbf{X}$ is the trace of matrix $\mathbf{X}$ and $\otimes$ is the symbol of Kronecker product.

The total number of coefficients in (3) is $M = (2n_1+1)(2n_2+1)$, but due to the symmetry there are only $M' = (M+1)/2$ distinct coefficients. We denote $\text{vec}(R) \in \mathbb{R}^{M'}$ the vector of the coefficients of $R(z_1,z_2)$ situated in the upper half plane and enumerated along the first axis, i.e. in the order $r_{0,0}, r_{1,0}, \ldots, r_{n_1,0}, r_{-n_1,1}, \ldots, r_{n_1,n_2}$.

Theorem 2 *Let Δ be a 2-D polynomial of degree (n_1,n_2) such that*

$$vec(R) = \mathbf{\Phi} \cdot vec(\Delta) + \mathbf{g}, \tag{6}$$

where the coefficients of the polynomial R are described by (5), with $\mathbf{Q} > 0$, and the matrix $\mathbf{\Phi} \in \mathbb{R}^{M' \times N}$ and the vector $\mathbf{g} \in \mathbb{R}^{M'}$ are constant and their coefficients depend only on A (complete formulas are given in [7]).

The domain

$$\mathcal{D}_A = \{\tilde{A} = A + \Delta \mid \Delta \text{ such that (6) is true}\} \tag{7}$$

contains only Schur polynomials and is convex.

We note that relation (6) is an LMI (which makes obvious the convexity of $\mathcal{D}_A$). The LMI (6) can be modified such that robust stability is ensured, in the sense that $\tilde{A}(z_1,z_2) \neq 0$, for $|z_1| \geq \rho$, $|z_2| \geq \rho$, with $\rho < 1$ (of course, it is assumed that $A(z_1,z_2)$ satisfies the same condition). To this purpose we simply rewrite Theorem 2 for $\tilde{A}_\rho(z_1,z_2) = \tilde{A}(\rho z_1, \rho z_2)$ and obtain a new LMI (6), in which $\mathbf{\Phi}$ and $\mathbf{g}$ depend also on ρ.

2.2 Gauss-Newton LS optimization of 2-D IIR filters

We give here a general description of the Gauss-Newton (GN) method applied to the optimization of the LS criterion (2). Let us suppose that, at iteration i of the method, the denominator and numerator of the filter (1) are $A^{(i)}$ and $B^{(i)}$. We seek $\Delta_A^{(i)}$, $\Delta_B^{(i)}$ such that the new filter with $A^{(i+1)} = A^{(i)} + \Delta_A^{(i)}$, $B^{(i+1)} = B^{(i)} + \Delta_B^{(i)}$ gives a better value of the criterion (2), in the sense that $J(A^{(i+1)},B^{(i+1)}) < J(A^{(i)},B^{(i)})$. The GN method is based on a convexification of the criterion, using a first order approximation of the filter (1) viewed as a function of its coefficients. We denote $\delta^{(i)} = [\text{vec}(\Delta_A^{(i)})^T \ \text{vec}(\Delta_B^{(i)})^T]^T$ the vector of coefficients of the variable polynomials $\Delta_A^{(i)}$, $\Delta_B^{(i)}$. Also, we denote $\nabla H_{\ell_1,\ell_2}^{(i)}$ the gradient of (1) with respect to A, B, computed in $A^{(i)}$, $B^{(i)}$, for the frequencies $\omega_1(\ell_1)$, $\omega_2(\ell_2)$. The main operation in an iteration of the GN method consists of solving

$$\begin{array}{ll} \min\limits_{\delta^{(i)}} & \displaystyle\sum_{\ell_1=1}^{L_1} \sum_{\ell_2=1}^{L_2} \lambda_{\ell_1,\ell_2} \left| D_{\ell_1,\ell_2} - \frac{B_{\ell_1,\ell_2}^{(i)}}{A_{\ell_1,\ell_2}^{(i)}} - \nabla^T H_{\ell_1,\ell_2}^{(i)} \cdot \delta^{(i)} \right|^2 \\ \text{s.t.} & A^{(i)} + \Delta_A^{(i)} \in \mathcal{D}_{A^{(i)}} \end{array} \tag{8}$$

The stability domain $\mathcal{D}_{A^{(i)}}$ is described by the LMI (6). Since the criterion is quadratic and the constraint is an LMI, the optimization problem (8) is convex and may be brought to an SDP form (similar to that given in [3] for the 1-D case); hence, its solution can be computed reliably. The whole GN algorithm is presented in Figure 1. Some comments are in order. We remark that in the LS problem, if the denominator A is given, then the optimal numerator can be found directly by solving (in LS sense) an overdetermined system of linear equations; we denote $B = B_{LS}(A)$ such an optimal numerator, as seen in step 2 of the algorithm. Also, we notice that the polynomials $\Delta_A^{(i)}$, $\Delta_B^{(i)}$ obtained by solving (8) are actually used as maximum steps in the descent direction; an optimal step is computed in (9), by line search. The stopping decision may take different forms; we have shown only a decision based on the relative improvement of the criterion (2); in a practical implementation we should also impose a maximum number of iterations.

If the denominator of the filter (1) is separable, then a convex stability domain may be built easily using two LMIs (one for each factor of the denominator) corresponding to 1-D positive realness conditions [3]. Otherwise, the general form (8) of a GN iteration remains valid; again, the implementation form is SDP.

3. EXPERIMENTAL RESULTS

We have implemented our algorithms in Matlab, using the SDP library SeDuMi [8]. We report results on three 2-D

Algorithm GN_LS

Input: Degrees m_1, m_2, n_1, n_2 of (1). Desired response D and the weights λ from (2), on a grid of frequencies with $L_1 \times L_2$ points. A tolerance ε.

1. Set $A^{(1)}(z_1,z_2) = 1$ and $i = 1$.
2. Compute $B^{(i)} = B_{LS}(A^{(i)})$.
3. Compute $\Delta_A^{(i)}$, $\Delta_B^{(i)}$ by solving the GN optimization problem (8).
4. Compute optimal step $\alpha^\star$ by solving the line search problem

$$\begin{array}{ll} \min\limits_{\alpha} & J(A^{(i)} + \alpha\Delta_A^{(i)}, B^{(i)} + \alpha\Delta_B^{(i)}) \\ \text{s.t.} & 0 \le \alpha \le 1 \end{array} \tag{9}$$

5. Compute new filter $A^{(i+1)} = A^{(i)} + \alpha^\star\Delta_A^{(i)}$, $B^{(i+1)} = B^{(i)} + \alpha^\star\Delta_B^{(i)}$.
6. If

$$\frac{J(A^{(i)},B^{(i)}) - J(A^{(i+1)},B^{(i+1)})}{J(A^{(i)},B^{(i)})} < \varepsilon, \tag{10}$$

stop. Otherwise, put $i = i+1$ and go back to 2.

Output: IIR filter with $A = A^{(i+1)}$, $B = B^{(i+1)}$.

Fig. 1. Gauss-Newton algorithm for LS optimization of 2-D IIR filters.

IIR linear phase filter design problems. In all problems, the desired response $D(\omega_1,\omega_2)$ is lowpass, but the passband has different shapes: circular, rhomboidal and elliptic. In the first two problems, the desired response is quadrantally symmetric (i.e. $|D(\omega_1,\omega_2)| = |D(-\omega_1,\omega_2)| = |D(\omega_1,-\omega_2)|$), but in the third it is not. All desired responses have the general form

$$D(\omega_1,\omega_2) = |D(\omega_1,\omega_2)| \cdot \mathrm{e}^{-j(\tau_1\omega_1+\tau_2\omega_2)}. \tag{11}$$

We consider only equal group delays, i.e. $\tau_1 = \tau_2 = \tau$. The magnitude responses have the following expressions.

Problem 1:

$$|D(\omega_1,\omega_2)| = \begin{cases} 1, & \text{if } \sqrt{\omega_1^2+\omega_2^2} \le \omega_p, \\ 0, & \text{if } \sqrt{\omega_1^2+\omega_2^2} \ge \omega_s, \end{cases} \tag{12}$$

with $\omega_p = 0.5\pi$, $\omega_s = 0.7\pi$.

Problem 2:

$$|D(\omega_1,\omega_2)| = \begin{cases} 1, & \text{if } |\omega_1|+|\omega_2| \le \omega_p, \\ 0, & \text{if } |\omega_1|+|\omega_2| \ge \omega_s, \end{cases} \tag{13}$$

with $\omega_p = 0.55\pi$, $\omega_s = 0.7\pi$.

Problem 3:

$$|D(\omega_1,\omega_2)| = \begin{cases} 1, & \text{if } r_1 \le 1, \\ 0, & \text{if } r_2 \ge 1, \end{cases} \tag{14}$$

where

$$r_i^2 = \frac{(\omega_1\cos\theta + \omega_2\sin\theta)^2}{\alpha_i^2} + \frac{(-\omega_1\sin\theta + \omega_2\cos\theta)^2}{\beta_i^2}, \tag{15}$$

with $\alpha_1 = 1$, $\beta_1 = 0.5$, $\alpha_2 = 1.2$, $\beta_2 = 0.7$, $\theta = \pi/4$.

In the criterion (2), the weights are 1 in the passband and stopband and zero in the transition band. We took $L_1 = L_2 = 80$ and a uniform grid of frequency points. The stopping tolerance is $\varepsilon = 10^{-5}$. The maximum pole radius is $\rho = 0.9$ if not otherwise specified.

We notice that all three responses are symmetric with respect to the first bisector. Filters (1) whose coefficients form symmetric matrices $\mathbf{A}$, $\mathbf{B}$ (i.e. $H(z_1,z_2) = H(z_2,z_1)$) have such symmetric responses. Although we have not enforced explicitly the symmetry of coefficients, all obtained optimal filters obey to this symmetry.

We design filters with separable and nonseparable denominator. A fair comparison between filters designed for the same desired response should take into account the implementation complexity of the filters. Since the magnitude of the desired responses is symmetric, we take $n_1 = n_2 = n$. A separable denominator of degree (n,n) is implemented with $2n$ multiplications and the same number of additions; the symmetry of coefficients does not reduce complexity. The implementation of a nonseparable denominator of same degree requires $(n+1)^2$ multiplications and the same number of additions; symmetry reduces the number of multiplications to $(n+1)(n+2)/2$. With good approximation we can say that a nonseparable denominator with $n = 4$ has the same complexity as a separable denominator with $n = 8$; we will focus our comparisons around these values of the degrees, for which the values of the criterion (2) will be shown with bold characters.

We present in table 1 the values of the normalized LS criterion, obtained by running *GN_LS* for the three design problems, with $m_1 = m_2 = 12$ and various degrees of the denominator. The values of the group delays were chosen such that, for the given degree, the LS criterion is minimum for the FIR filter. In the nonseparable case, some improvements are obtained if the method is run twice, first with a smaller stability radius, e.g. 0.9ρ, then with radius ρ (the results of the first run are initializations to the second). This is a way to alleviate the greediness of the method, in which the advance to the border of the stability domain is sometimes too fast.

We remark that the results are better for separable denominator filters in Problem 1, but better for nonseparable in Problem 2, while in Problem 3 they are rather similar. So, the better filter may depend on the shape of the passband, the width of the transition band and probably other factors; an a priori evaluation seems difficult. We can remark that a quadrantally symmetric desired response does not imply that an IIR filter with separable denominator gives the best performance; the results for Problem 2 are a good example. On the contrary, a quadrantally nonsymmetric specification, as in Problem 3, may be well satisfied by a filter with separable denominator.

The execution times of *GN_LS* for the problems at stake were usually of 3-5 minutes on a PC computer with a Pentium III processor at 1GHz. Although an iteration takes more time in the nonseparable case, the number of iterations is higher in the separable case and so the execution times are more or less similar. For the higher degrees ($n_1 = n_2 = 10$), the separable design may take 10 minutes.

Table 1. Values of the LS normalized criterion for the algorithm *GN_LS*.

$n_1 = n_2$	Problem 1, $\tau = 7.5$		Problem 2, $\tau = 6$		Problem 3, $\tau = 6$	
	nonseparable	separable	nonseparable	separable	nonseparable	separable
2	4.93e-5	6.52e-5	7.89e-4	1.08e-3	1.13e-4	1.14e-4
4	**2.07e-5**	3.35e-5	**5.98e-4**	9.39e-4	**9.15e-5**	1.04e-4
6	1.14e-5	1.49e-5	5.76e-4	8.19e-4	7.85e-5	1.01e-4
8	-	**8.76e-6**	-	**7.05e-4**	-	**9.18e-5**
10	-	7.62e-6	-	6.55e-4	-	7.27e-5

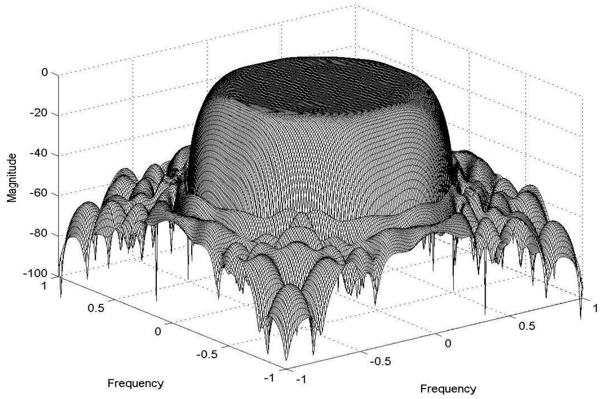

Fig. 2. Magnitude of the frequency response of a separable denominator filter designed with *GN_LS* for the specifications of Problem 1, with $\rho = 0.8238$.

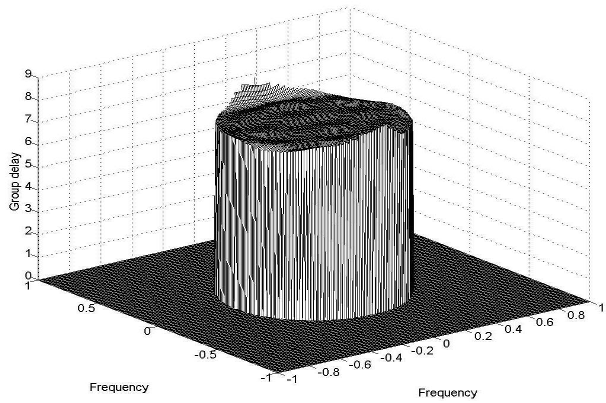

Fig. 3. Group delay in passband, same filter as in Fig. 2.

We give now one example of design for Problem 1, with $m_1 = m_2 = 12$, $n_1 = n_2 = 8$, $\tau = 7.5$ and $\rho = 0.8238$, i.e. the specification used in [4]. The value of the criterion (2) is 1.45e-5 and is smaller than the corresponding value from Table 1, due to the tougher pole radius constraint. The frequency response of the filter is shown in figures 2 (magnitude response) and 3 (group delay). Our filter has better performances than the design reported in [4] (although the latter is optimized using a minimax criterion): 0.0227 vs. 0.0319 stopband error; 0.0197 vs. 0.0315 maximum amplitude deviation in passband. The maximum group delay error in passband is 0.951 while in [4] is 0.665. The execution time of our method is about 3 minutes, which compares favorably with the more than 3 hours reported in [4], on a computer slightly slower than ours.

4. CONCLUSIONS

We have presented an algorithm for the LS optimization of 2-D IIR filters with separable or nonseparable denominator. The algorithm uses the Gauss-Newton convexification of the criterion (2); robust stability is ensured by the use of convex stability domains; each iteration of the algorithm is a semidefinite programming problem. The design of 2-D IIR filters upon several specifications shows that it is difficult to decide a priori which is better between separable and nonseparable denominator.

REFERENCES

[1] M.C. Lang, "Least-Squares Design of IIR Filters with Prescribed Magnitude and Phase Response and a Pole Radius Constraint," *IEEE Trans. Signal Processing*, vol. 48, no. 11, pp. 3109–3121, Nov. 2000.

[2] W.-S. Lu, S.-C. Pei, and C.-C. Tseng, "A Weighted Least-Squares Method for the Design of Stable 1-D and 2-D IIR Digital Filters," *IEEE Trans. Signal Processing*, vol. 46, pp. 1–10, Jan. 1998.

[3] B. Dumitrescu and R. Niemistö, "Multistage IIR Filter Design Using Convex Stability Domains Defined by Positive Realness," *IEEE Trans. Signal Proc.*, vol. 52, no. 4, pp. 962–974, April 2004.

[4] W.S. Lu, "A Unified Approach for the Design of 2D Digital Filters via Semidefinite Programming," *IEEE Trans. Circ. Syst. I*, vol. 49, no. 6, pp. 814–826, June 2002.

[5] A.T. Chottera and G.A. Jullien, "Design of Two-Dimensional Recursive Digital Filters Using Linear Programming," *IEEE Trans. Circ. Syst.*, vol. CAS-29, no. 12, pp. 817–826, Dec. 1982.

[6] T. Hinamoto and S. Maekawa, "Design of Two-Dimensional Recursive Digital Filters using Mirror Image Polynomials," *IEEE Trans. Circ. Syst.*, vol. 33, no. 8, August 1986.

[7] B. Dumitrescu, "2-D Convex Stability Domain," in *Proc. 2nd IEEE Symp. on Signals, Circuits and Systems*, Iaşi, Romania, 2003, vol. 2, pp. 365–368.

[8] J.F. Sturm, "Using SeDuMi, a Matlab Toolbox for Optimization over Symmetric Cones," *Optimization Methods and Software*, vol. 11-12, pp. 625–653, 1999, http://fewcal.kub.nl/sturm/software/sedumi.html.