ON CONVEX STABILITY DOMAIN AND OPTIMIZATION OF IIR FILTERS

*Bogdan Dumitrescu**
Department of Automatic Control and Computers
Politehnica University of Bucharest
313, Spl. Independenţei, 77206 Bucharest, Romania
e-mail: bogdan@lucky.schur.pub.ro

ABSTRACT

We discuss descriptions of convex domains containing Schur polynomials, built around a given Schur polynomial. We show that the domain described by a positive realness constraint always contains the domain characterized by Rouché's theorem. We also show how to handle computationally the positive realness condition, using semidefinite programming, in the context of designing stable IIR filters. Two recent methods of Lang [4] and Lu *et al* [6] for optimizing IIR filters according to a least-squares criterion are modified to incorporate the positive realness condition and shown experimentally to give similar results.

1 INTRODUCTION

Several recent papers [4, 5, 6] propose solutions to handle one of the significant difficulties in designing IIR filters, namely guaranteeing the stability of the filters. The IIR filter

$$H(z) = \frac{b(z)}{a(z)} = \frac{b_0 + b_1 z^{-1} + \ldots + b_m z^{-m}}{1 + a_1 z^{-1} + \ldots + a_n z^{-n}}, \qquad (1)$$

is stable when the denominator is a Schur polynomial, i.e. has all zeros inside the unit circle; for shortness, we say that $a(z)$ is a Schur polynomial, although we actually think at $z^n a(z)$. In [5], the polynomial $a(z)$ is expressed as a product of degree two factors, for which stability is easily imposed; the drawback is the highly nonlinear form of the denominator of (1) in the new parameters. In [6], a *global* positive realness condition is approximated with linear constraints; the fixed convex set thus obtained is a subset of the stability domain, therefore some generality is lost. The most successful approach appears to be the use of Rouché's theorem in [4], in the context of a Gauss-Newton iterative method; a stability domain is built *locally* around the denominator at the previous iteration.

Let us first describe formally the optimization problem to which we confine our study. We assume that the IIR filter (1) has fixed degrees of the numerator and denominator, so the optimization parameters are the coefficients a_i, $i = 1 : n$, and b_k, $k = 0 : m$; we denote $a \in \mathbb{R}^n$, $b \in \mathbb{R}^{m+1}$ the corresponding vectors. We consider a least squares criterion in the complex domain, i.e.

$$J(a, b) = \sum_{k=1}^{N} w_k \left| d_k - H(\omega_k, a, b) \right|^2, \qquad (2)$$

where the complex values d_k, $k = 1 : N$, represent the desired frequency response of the filter in the frequencies ω_k, while $H(\omega_k, a, b)$ is the actual frequency response of $H(z)$ from (1). The numbers $w_k > 0$ represent weights. Such a criterion is very useful when phase properties of the IIR filter are considered in the optimization, e.g. like when linear phase is desired in the passband.

Two difficulties are encountered in the minimization of (2): the non-convexity of the criterion and the non-convexity of the stability domain of Schur polynomials $a(z)$, for $n > 2$. Moreover, we may require some robustness of stability: a small perturbation of the coefficients of $a(z)$ should leave the filter stable. We deal here mainly with the stability issue, in the context of iterative methods. That is, given a Schur polynomial $a(z)$ and a polynomial $b(z)$ (from the previous iteration), we want to find local "steps" $\delta_a(z)$ and $\delta_b(z)$ such that the criterion (2) is improved, i.e. $J(a + \delta_a, b + \delta_b) < J(a, b)$, and the polynomial $\tilde{a}(z) = a(z) + \delta_a(z)$ is (robust) Schur. The contributions of this paper are

- to describe a convex stability domain around a given $a(z)$, based on a positive realness condition which is computationally tractable *exactly*;
- to show that this domain always contains the domain described by the Rouché's theorem and to compare them with the one given by the real stability radius;
- to insert the new description of stability in the methods from [4] and [6] (the latter with modifications) and compare the two resulting methods.

2 CONVEX STABILITY DOMAINS

Let us assume that a Schur polynomial $a(z)$ is given. We desire to characterize—in a computationally advantageous way—a convex vicinity $\mathcal{D}_a \subset \mathbb{R}^n$ of $a(z)$ (of the vector $a \in \mathbb{R}^n$) containing only Schur polynomials. Formally, we express $\tilde{a} \in \mathcal{D}_a$ as $\tilde{a}(z) = a(z) + \delta_a(z)$, with $\delta_a(z) = \delta_1 z^{-1} + \ldots + \delta_n z^{-n}$, i.e. $\tilde{a}_0 = a_0 = 1$, so that only monic polynomials are considered (without loss of generality). In this section we present three stability domains and then compare them.

2.1 Stability radius

We express

$$\mathcal{D}_a^1 = \{\tilde{a} = a + \delta_a \in \mathbb{R}^n \mid \|\delta_a\| < r_a\}, \qquad (3)$$

where r_a is the (real) stability radius of the polynomial $a(z)$, i.e. the minimum 2-norm of a perturbation polynomial that added to $a(z)$ makes the result non-Schur (other norms may

*This work was started when the author was visiting Tampere International Center for Signal Processing, Tampere University of Technology, Finland.

be used as well). The set $\mathcal{D}_a^1$ is a sphere in $\mathbb{R}^n$ and so is obviously convex. The computation of the stability radius is rather complex, as shown e.g. in [7]. On the contrary, after computing r_a, the implementation of the quadratic condition from (3) is immediate. Robustness may be added in a meaningful way by requiring that $\tilde{a}(z)$ have a stability radius greater than a fixed minimum value r_{min}. Accordingly, we define $\mathcal{D}_a^1$ in (3) with $\|\delta_a\| \le r_a - r_{min}$.

2.2 Rouché's Theorem

Rouché's Theorem states that if $|f(z)| < |g(z)|$ on a closed contour in the complex plane, inside which the functions f and g are analytic, then the functions g and $f+g$ have the same number of zeros inside the contour. In our case, if $|\delta_a(z)| < |a(z)|$ on the unit circle, then $\tilde{a}(z)$ and $a(z)$ have the same number of zeros inside the circle, i.e. $\tilde{a}(z)$ is Schur. The set

$$\mathcal{D}_a^2 = \{\tilde{a} = a + \delta_a \in \mathbb{R}^n \mid |\delta_a(\mathrm{e}^{j\omega})| < |a(\mathrm{e}^{j\omega})|,\ \omega \in [0, 2\pi]\} \tag{4}$$

is convex, as an intersection of convex sets: for each ω, the left term of the inequality from (4) is a positive definite quadratic form in the variable δ_a. For implementation, we can consider only a finite number of frequencies in (4), as in [4] (although such an approximated set may contain also non-Schur polynomials, semi-infinite programming or special techniques can be used for working on it). Replacing $\mathrm{e}^{j\omega}$ with $\rho\mathrm{e}^{j\omega}$ in (4), for a given $0 < \rho < 1$, ensures robust stability by forcing the roots of $\tilde{a}(z)$ to stay inside a circle of radius ρ.

2.3 Positive realness

A transfer function $G(z)$ is strictly positive real (SPR) if it is stable and $\mathrm{Re}G(\mathrm{e}^{j\omega}) > 0$, for any $\omega \in [0, 2\pi]$. The following simple result gives a sufficient condition of stability in terms of positive realness (see [10] for the continuous time case).

Proposition 1 *If the transfer function*

$$G(z) = \frac{\tilde{a}(z)}{a(z)} = 1 + \frac{\delta_a(z)}{a(z)} \tag{5}$$

is positive real, then all the convex combinations of $\tilde{a}(z)$ and $a(z)$, i.e. $a_\lambda(z) = \lambda a(z) + (1-\lambda)\tilde{a}(z)$, $\lambda \in [0,1]$, are Schur polynomials.

Proof. Any convex combination of two positive real transfer functions is positive real. Both the numerator and the denominator of a positive real transfer function are Schur polynomials. In our case, the functions are 1 and $G(z)$ and the polynomial $a_\lambda(z)$ is the numerator of $\lambda + (1-\lambda)G(z)$. ∎

Prop. 1 allows us to build the following stability domain "centered" in $a(z)$:

$$\mathcal{D}_a^3 = \{\tilde{a} = a + \delta_a \in \mathbb{R}^n \mid G(z) \text{ from (5) is SPR}\}. \tag{6}$$

Proposition 2 *The set $\mathcal{D}_a^3$ is convex.*

Proof. Any convex combination $\lambda E(z) + (1-\lambda)F(z)$, with $E(z) = c(z)/a(z)$ and $F(z) = d(z)/a(z)$, with $c, d \in \mathcal{D}_a^3$, is positive real, has $a(z)$ as denominator and its numerator is $\lambda c(z) + (1-\lambda)d(z)$, which thus belongs to $\mathcal{D}_a^3$ (and is a Schur polynomial). ∎

The set $\mathcal{D}_a^3$ may be described by a linear matrix inequality (LMI). Therefore, it is appropriate to modern convex optimization tools like semidefinite programming.

Proposition 3 *With $a_0 = 1$, let us define*

$$f = 2[\ \sum_{k=0}^{n} a_k^2\ \dots\ \sum_{k=0}^{n-\ell} a_k a_{k+\ell}\ \dots\ a_0 a_n\ ]^T, \tag{7}$$

and

$$F = \begin{bmatrix} a_1 & a_2 & \dots & a_{n-1} & a_n \\ a_2 & a_3 & \dots & a_n & 0 \\ \vdots & & \vdots & 0 & 0 \\ a_n & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix} + \begin{bmatrix} a_1 & a_2 & \dots & a_n \\ a_0 & a_1 & \dots & a_{n-1} \\ 0 & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & a_1 \\ 0 & 0 & \dots & a_0 \end{bmatrix}. \tag{8}$$

Then, the polynomial $\tilde{a}(z) = a(z) + \delta_a(z)$ belongs to $\mathcal{D}_a^3$ if and only if there exists a positive definite matrix Q such that

$$v = f + F\delta_a, \tag{9}$$

where $v \in \mathbb{R}^{n+1}$, with $v_\ell = trA_\ell Q$; we denote A_ℓ the elementary Toeplitz matrix with all elements on the ℓ'th diagonal equal to 1 and zero elsewhere (note that A_ℓ is not symmetric) and trX is the trace of the matrix X.

Proof. The condition that $G(z)$ defined in (5) be positive real is equivalent to requiring that (since $a(z)$ is Schur)

$$\begin{aligned} G(z) + G(z^{-1}) &= \frac{2a(z)a(z^{-1}) + \delta_a(z)a(z^{-1}) + a(z)\delta_a(z^{-1})}{a(z)a(z^{-1})} \\ &=: \frac{p(z)}{a(z)a(z^{-1})} \end{aligned} \tag{10}$$

is real and positive on the unit circle. Since the denominator of (10) is positive on the unit circle, it follows that the symmetric polynomial $p(z)$ must be positive on the unit circle, which is true if and only if there exists a positive definite matrix Q such that $p_k = \mathrm{tr}A_k Q$, see [2, 1]. The relation (9) results using the definition of $p(z)$ from (10). ∎

For obtaining robust stability, we can enforce the zeros of $\tilde{a}(z)$ to lie in a circle of radius $\rho < 1$, denoted $\mathcal{C}_\rho$. We suppose that the zeros of $a(z)$ are in $\mathcal{C}_\rho$. Let us define $a^\rho(z) = a(\rho z)$; then, the zeros of $a^\rho(z)$ are in $\mathcal{C}_1$ (i.e. $a^\rho(z)$ is Schur). Denoting $R = \mathrm{diag}(\rho, \rho^2, \dots, \rho^n)$, we have $a^\rho = R^{-1}a$. We denote similarly $\delta_a^\rho(z) = \delta(\rho z)$ and $\tilde{a}^\rho(z) = \tilde{a}(\rho z) = a^\rho(z) + \delta_a^\rho(z)$. By virtue of Prop. 1, if $1 + \delta_a^\rho(z)/a^\rho(z)$ is SPR, then $\tilde{a}^\rho(z)$ is Schur, and thus the zeros of $\tilde{a}(z)$ are in $\mathcal{C}_\rho$.

The key point for implementation is that the dependence between δ_a^ρ and δ_a is linear. Instead of (9), we have

$$v = f_\rho + F_\rho R^{-1}\delta_a, \tag{11}$$

where f_ρ and F_ρ are obtained by replacing a with a^ρ in (7) and (8), respectively.

2.4 Comparisons

We are interested now in a comparison between the three above sufficient stability conditions. Let us start with an example. We take $n = 2$, a fixed $a(z)$ and $\delta_a(z) = \delta_1 z^{-1} + \delta_2 z^{-2}$. The stability domain $\mathcal{D}$ of degree-two polynomials is a triangle in the parameter plane $(\tilde{a}_1, \tilde{a}_2)$, as shown in figure 1. The three sets corresponding to the definitions (3), (4) and (6) are drawn in figure 1 for $a(z) = 1$ and in figure 2 for $a(z) = 1 - 0.5z^{-1} + 0.6z^{-2}$. The set $\mathcal{D}_a^1$ given by the stability radius is a circle whose radius is easy to compute (for $n = 2$ only). The set $\mathcal{D}_a^3$, given by the positive realness condition,

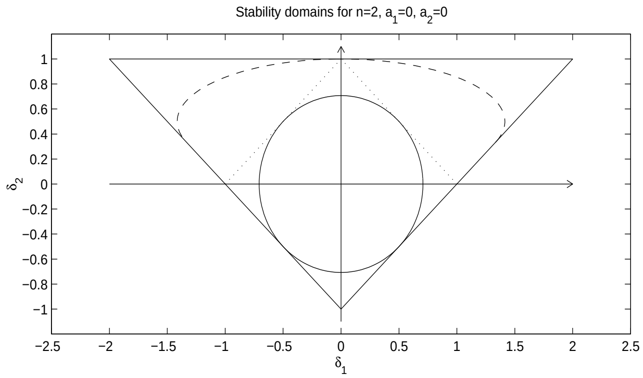

Figure 1: Convex stability domains around $a(z) = 1$, for $n = 2$: real stability radius (solid line), Rouché (dotted), positive realness (dashed).

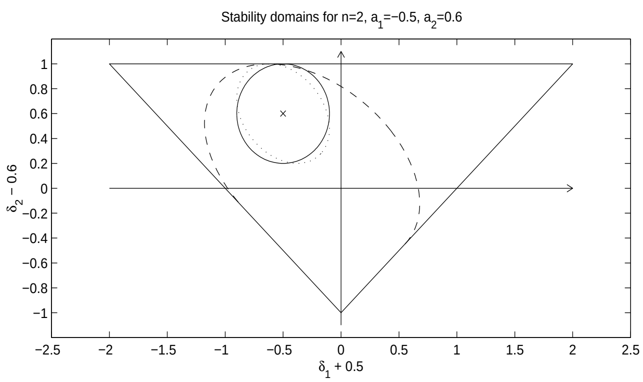

Figure 2: Convex stability domains around $a(z) = 1 - 0.5z^{-1} + 0.6z^{-2}$, for $n = 2$.

has an ellipse as border, completed with the borders of the whole stability domain. We notice that in both examples we have $\mathcal{D}_a^2 \subset \mathcal{D}_a^3$, which is true in general, as shown below. We also notice from figure 2 that $\mathcal{D}_a^1$ contains points that are not in $\mathcal{D}_a^2$ or $\mathcal{D}_a^3$ and conversely. Anyway, from these examples and others not shown here, it appears that the positive realness condition is the most permissive, i.e. the set $\mathcal{D}_a^3$ has the greatest area. The most general result we can prove is the following.

Proposition 4 *For any Schur polynomial $a(z)$, we have $\mathcal{D}_a^2 \subset \mathcal{D}_a^3$.*

Proof. Let us assume that $|\delta_a(\mathrm{e}^{j\omega})| < |a(\mathrm{e}^{j\omega})|$, for any ω. Since $|a(\mathrm{e}^{j\omega})| \neq 0$, we may write

$$\left|\frac{\delta_a(\mathrm{e}^{j\omega})}{a(\mathrm{e}^{j\omega})}\right| < 1 \Rightarrow \left|\mathrm{Re}\frac{\delta_a(\mathrm{e}^{j\omega})}{a(\mathrm{e}^{j\omega})}\right| < 1 \Rightarrow 1 + \mathrm{Re}\frac{\delta_a(\mathrm{e}^{j\omega})}{a(\mathrm{e}^{j\omega})} > 0$$
$$\Rightarrow \mathrm{Re}\frac{\tilde{a}(\mathrm{e}^{j\omega})}{a(\mathrm{e}^{j\omega})} > 0,$$

i.e. $\tilde{a}(z)/a(z)$ is positive real. ■

3 OPTIMIZATION METHODS

For the optimization of the criterion (2), we insert the positive realness description of stability in two iterative methods, obtaining standard convex optimization problems at each iteration.

3.1 Gauss-Newton method

The following method is used in [4]. Let us suppose that we have a, b defining the IIR filter (1) and we seek $a + \delta_a$, $b + \delta_b$ improving (2). To this purpose, we use a first order approximation of $H(\omega, a, b)$ and thus the criterion becomes

$$J(a+\delta_a, b+\delta_b) \approx \sum_{k=1}^{N} w_k |d_k - H(\omega_k, a, b) - \nabla^T H(\omega_k, a, b)\cdot\delta|^2, \tag{12}$$

where $\delta^T = [\delta_a^T \;\; \delta_b^T]$. Simple computation leads to the quadratic form

$$J(a + \delta_a, b + \delta_b) \approx \eta + p^T\delta + \delta^T S\delta, \tag{13}$$

where S is a positive definite matrix, p a vector and η a positive scalar, all known. We minimize the quadratic criterion (13), taking care that $a(z) + \delta_a(z)$ is a (robust) Schur polynomial. To this purpose we use the positive realness description of a convex stability domain around $a(z)$. The resulting optimization problem is

$$\begin{aligned} \min_{\delta} \quad & p^T\delta + \delta^T S\delta \\ \text{s.t.} \quad & 1 + \delta_a(z)/a(z) \text{ is SPR} \end{aligned} \tag{14}$$

This problem may be written as a mixed semidefinite-quadratic-linear program (SQLP) [3], in standard equality form. Let $S = L^T L$ be the Cholesky decomposition of S and denote $y = L\delta + 0.5L^{-T}p$; then, the criterion of (14) is equal to $\|y\|^2$ (minus a negligible positive constant). Let $\tilde{e} \in \mathbb{R}^n$ be a constant vector such that $\tilde{\delta} = \delta + \tilde{e}$ is elementwise positive for any admissible δ (values of order 100-1000 are sufficient for moderately large m, n). With all these notations and taking (9) into account, we obtain the following equivalent of (14):

$$\begin{aligned} \min_{\tilde{\delta},\lambda} \quad & \lambda \\ \text{s.t.} \quad & \begin{bmatrix} \vdots \\ \mathrm{tr} A_\ell Q \\ \vdots \end{bmatrix} - [F\ 0]\tilde{\delta} = f - [F\ 0]\tilde{e} \\ & \qquad - \ L\tilde{\delta} + y = 0.5L^{-T}p - L\tilde{e} \\ & Z \geq 0,\ \|y\| \leq \lambda,\ \tilde{\delta} \geq 0 \end{aligned} \tag{15}$$

We note that other equality constraints may be added immediately to the problem, e.g. $H(0) = 1$, which translate into linear equalities in δ.

The iterative process is initialized with $a(z) = 1$ and a numerator $b(z)$ designed with some fast method for FIR filters. In a stage of the process, after solving the SQLP problem (15)—which is an approximation of the non-convex optimization problem—we find a step length α which ensures $J(a + \alpha\delta_a, b + \alpha\delta_b) < J(a, b)$. This is done by a simple line search, like the golden section (taking a constant α is not always successful). The iterations are stopped when the improvement of the criterion is no more significant.

3.2 Relaxation method

In [6], the Steiglitz-McBride method is revived for optimization of IIR filters. The idea is to write the criterion (2) in the modified form

$$J(\tilde{a}, \tilde{b}) = \sum_{k=1}^{N} \frac{w_k}{|a(\omega_k)|^2} \left| d_k\tilde{a}(\omega_k) - \tilde{b}(\omega_k) \right|^2, \tag{16}$$

where a is kept fixed at the value from the previous iteration, while $\tilde{a}$ and $\tilde{b}$ are the optimization variables in the current iteration. In [6], a global stability condition is put on $\tilde{a}(z)$ using positive realness, i.e. $\tilde{a} \in \mathcal{D}_0^3$. Naturally, here we consider the stability domain $\mathcal{D}_a^3$ from (6), "centered" in $a(z)$, and the optimization variable is $\delta_a(z)$, with $\tilde{a} = a + \delta_a$. (A similar approach using the Rouché condition is proposed in the very recent paper [9], for the particular case of notch filters.) This way, the criterion (16) is quadratic and a problem similar to (14) is obtained; its SQLP formulation is similar to (15).

The iterative process is initialized with $a(z) = 1$. After finding the optimal δ_a in the current iteration, we actually take $\tilde{a} = a + \alpha\delta_a$, where α is a slightly subunitary constant, e.g. $\alpha = 0.99$.

4 Experiments

We implemented the two methods using the SQLP package SeDuMi [8]. Both methods have difficulties in finding the global optimum due to the non-convexity of the criterion (2). The experiments we have done seem to indicate that Gauss-Newton obtains more often better values of the criterion. However, the modified relaxation is faster than Gauss-Newton, requiring usually 2-3 times less iterations (e.g. 6 vs. 15).

We give here only one example, taken from [4], where a lowpass filter is designed, with passband ending at 0.4π and stopband starting at 0.56π. The degrees of the numerator and denominator are $m = 15$ and $n = 4$, respectively. The weights in (2) are taken as 1 in the passband and 10 in the stopband; we use $N = 200$ frequency points, specifically 100 in passband and 100 in the stopband. We seek a linear phase filter and we take the desired response to be $d_k = \mathrm{e}^{-10\omega_k}$ in the passband, i.e. the group delay is 10 (note that the value 12 was used in [4]).

We made 20 designs with each method, forcing the poles of the filter to lie inside circles with radii ρ going from 0.8 to 0.99 with step 0.01. The two methods gave (practically) the same result in 11 cases; the Gauss-Newton method was better in 7 cases, while in 2 cases the modified relaxation was better. We show in figure 3 the magnitude response of three filters designed with the Gauss-Newton method, for $\rho \in \{0.8, 0.9, 1\}$. The corresponding maximum magnitudes of poles of the resulting optimum filters are 0.7998, 0.8994 and 0.9704, respectively. The deviation from 10 of the group delay is 0.469, 0.562 and 0.196 respectively.

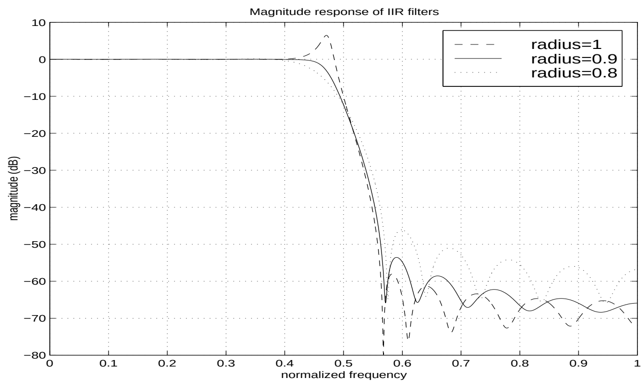

Figure 3: Magnitude response of three filters with different stability constraints.

5 Conclusion

We show in this paper that the stability domain (6) based on positive realness is more meaningful than the Rouché domain (4). In iterative methods for designing IIR filters, it may allow larger steps, thus reducing the number of iterations. We adapt two recent methods for optimizing IIR filters (from [4] and [6]); each stage of the iterative optimization consists of a semidefinite-quadratic-linear programming problem, where the stability constraint (6) is enforced exactly. Future work will be directed towards adapting other optimization techniques to criteria having other form than (2) and towards a thorough comparison of the methods.

Acknowledgement. The author is indebted to Prof. Ezra Zeheb for comments on [10] and to Radu Ştefan for discussions about the real stability radius.

References

[1] B. Dumitrescu, I. Tăbuş, and P. Stoica. On the Parameterization of Positive Real Sequences and MA Parameter Estimation. *IEEE Trans. Signal Processing*, 49(11):2630–2639, Nov. 2001.

[2] Y. Genin, Yu. Nesterov, and P. Van Dooren. Optimization over Positive Polynomial Matrices. In *Proc. Mathematical Theory of Network and Systems*, Perpignan, France, 2000. paper SI27-7.

[3] J.P. Haeberly, M.V. Nayakkankuppam, and M.L. Overton. Mixed Semidefinite-Quadratic-Linear Programs. In L. El Ghaoui and S.I. Niculescu, editors, *Recent advances in LMI methods for control*, pages 41–54. SIAM, 2000.

[4] M.C. Lang. Least-Squares Design of IIR Filters with Prescribed Magnitude and Phase Response and a Pole Radius Constraint. *IEEE Trans. Signal Processing*, 11(48):3109–3121, Nov. 2000.

[5] W.-S. Lu. Design of Recursive Digital Filters with Prescribed Stability Margin: a Parameterization Approach. *IEEE Trans. Circ. Syst. II*, 45:1289–1298, Sept. 1998.

[6] 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*, 46:1–10, Jan. 1998.

[7] L. Qiu, B. Bernhardsson, A. Rantzer, E.J. Davison, P.M. Young, and J.C. Doyle. A Formula for the Computation of the Real Stability Radius. *Automatica*, 31(6):879–890, 1995.

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

[9] C.-C. Tseng and S.-C. Pei. Stable IIR Notch Filter Design with Optimal Pole Placement. *IEEE Trans. Signal Processing*, 49(11):2673–2681, Nov. 2001.

[10] E. Zeheb. How to Derive a Set of Local Convex Directions for Hurwitz Stability. In *Proc. 34th Conf. Decision & Control*, pages 1421–1422, New Orleans, LA, Dec. 1995.