Semidefinite programming and experiment design

Lieven Vandenberghe
Department of Electrical and Computer Engineering, UCLA
Joint work with:
Weng Kee Wong and Yuguang Yue
Department of Biostatistics, UCLA

BIRS Workshop
Latest Advances in the Theory and Applications of
Design and Analysis of Experiments
Banff, August 6–11, 2017

Semidefinite program (SDP)

$$\begin{array}{ll} \text{minimize} & \operatorname{tr}(CX) \\ \text{subject to} & \operatorname{tr}(A_i X) = b_i, \quad i = 1, \ldots, m \\ & X \succeq 0 \end{array}$$

- variable is $n \times n$ symmetric matrix X
- inequality $X \succeq 0$ means X is positive semidefinite
- similar to standard form linear program, but with matrix inequality

Applications

- matrix inequalities arise naturally in many areas (for example, control, statistics)
- relaxations of nonconvex quadratic and polynomial optimization
- used in convex modeling systems (CVX, YALMIP, CVXPY, PICOS, ...)

widely studied since 1990s (following Nesterov & Nemirovski's 1994 book)

Outline

- Semidefinite representations of design criteria
- Discriminating design with polynomial models
- Semidefinite descriptions of moment cones

Conic linear programming

Primal:

$$\begin{array}{ll} \text{minimize} & \langle c, x \rangle \\ \text{subject to} & \langle a_i, x \rangle = b_i, \quad i = 1, \ldots, m \\ & x \in K \end{array}$$

Dual:

$$\begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & \sum_{i=1}^{m} y_i a_i + s = c \\ & s \in K^* \end{array}$$

- K is a proper convex cone (closed, pointed, with nonempty interior)
- $K^* = \{s \mid \langle s, x \rangle \geq 0 \ \forall x \in K\}$ is the dual cone

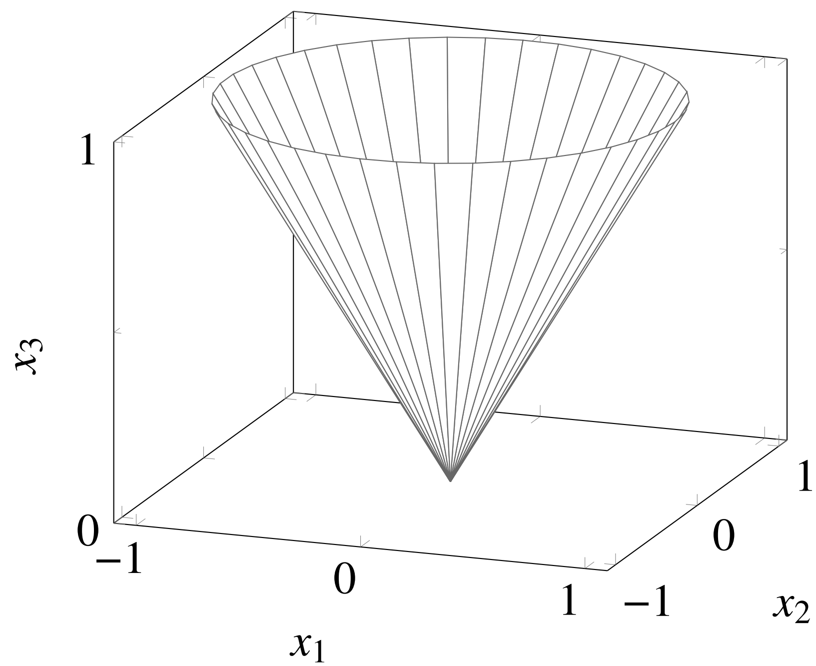
Solvers

- popular solvers include SDPT3, SeDuMi, MOSEK
- implementations of primal-dual interior-point methods
- handle 'symmetric' cones: second order cone and positive semidefinite cone

Symmetric cones

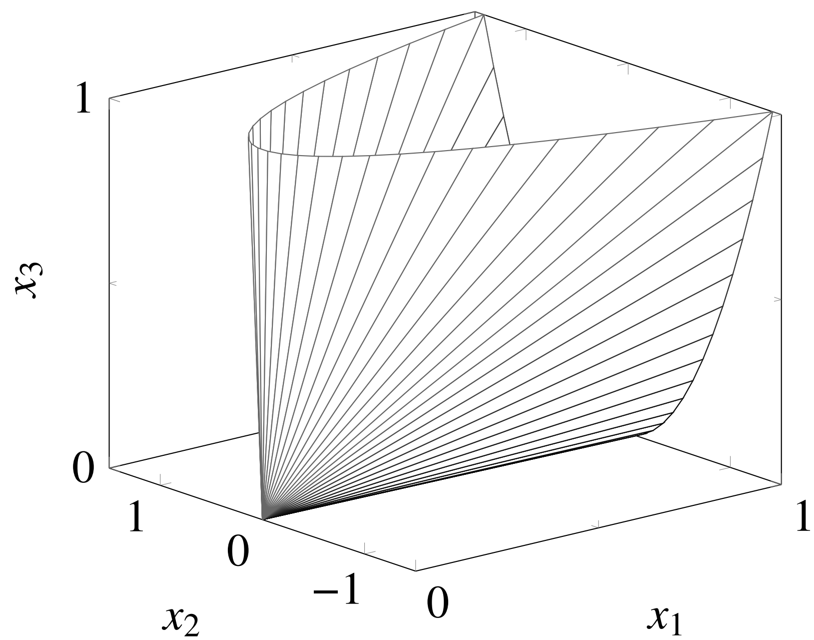
- second order cone (s.o. cone): $\{(x_1, \ldots, x_n) \mid (x_1^2 + \cdots + x_{n-1}^2)^{1/2} \leq x_n\}$
- cone of positive semidefinite symmetric matrices (p.s.d. cone)

s.o. cone in $\mathbf{R}^3$



$$\sqrt{x_1^2 + x_2^2} \leq x_3$$

p.s.d. cone in $\mathbf{R}^3$



$$\begin{bmatrix} x_1 & x_2/\sqrt{2} \\ x_2/\sqrt{2} & x_3 \end{bmatrix} \succeq 0$$

Modeling software

- surprisingly many functions are 's.o.- or p.s.d.-representable'
  [Nesterov & Nemirovski 1994]
- conversion rules implemented in modeling software packages

Modeling packages for convex optimization

- CVX, YALMIP (MATLAB)
- CVXPY, PICOS (Python)
- Convex.jl (Julia)

Optimal experiment design with finite design space

$$\begin{array}{ll} \text{minimize} & f(M) \\ \text{subject to} & M = \sum_{i=1}^{m} w_i f(x_i) f(x_i)^T \\ & w_i \geq 0, \quad i = 1, \ldots, m \\ & \sum_{i=1}^{m} w_i = 1 \end{array}$$

variables: m-vector w and symmetric $p \times p$ matrix M

Design criteria

- c-optimality: $f(M) = c^T M^{-1} c$
- A-optimality: $f(M) = \operatorname{tr} M^{-1}$
- E-optimality: $f(M) = \lambda_{\max}(M^{-1})$
- D-optimality: $f(M) = -(\det M)^{1/n}$
- condition number: $f(M) = \kappa(M)$

these criteria can be minimized using s.o. and p.s.d. conic optimization

Second order cone program formulation of c-optimal design

$$\begin{array}{ll} \text{minimize} & c^T M^{-1} c \\ \text{subject to} & M = \sum_{i=1}^m w_i f(x_i) f(x_i)^T \\ & w_i \geq 0, \quad i = 1, \ldots, m, \quad \sum_{i=1}^m w_i = 1 \end{array}$$

Step 1: equivalent problem with auxiliary variables $y_1, \ldots, y_m$

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^m y_i^2 / w_i \\ \text{subject to} & \sum_{i=1}^m f(x_i) y_i = c \\ & w_i \geq 0, \quad i = 1, \ldots, m, \quad \sum_{i=1}^m w_i = 1 \end{array}$$

equivalence can be shown by optimizing over y

Second order cone program formulation of c-optimal design

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{m} y_i^2/w_i \\ \text{subject to} & \sum_{i=1}^{m} f(x_i) y_i = c \\ & w_i \geq 0, \quad i = 1, \ldots, m, \quad \sum_{i=1}^{m} w_i = 1 \end{array}$$

Step 2: reformulate nonlinear objective as linear objective with s.o. constraints

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{m} t_i \\ \text{subject to} & \left(4y_i^2 + (t_i - w_i)^2\right)^{1/2} \leq t_i + w_i, \quad t_i \geq 0, \quad i = 1, \ldots, m \\ & \sum_{i=1}^{m} f(x_i) y_i = c \\ & w_i \geq 0, \quad i = 1, \ldots, m, \quad \sum_{i=1}^{m} w_i = 1 \end{array}$$

first set of constraints is equivalent to $y_i^2/w_i \leq t_i$ for $i = 1, \ldots, m$

Semidefinite program formulation of E-optimal design

$$
\begin{array}{ll}
\text{minimize} & \lambda_{\max}(M^{-1}) \\
\text{subject to} & M = \sum_{i=1}^{m} w_i f(x_i) f(x_i)^T \\
& w_i \geq 0, \quad i = 1, \ldots, m \\
& \sum_{i=1}^{m} w_i = 1
\end{array}
$$

Equivalent problem: maximize $\lambda_{\min}(M)$ by solving the SDP

$$
\begin{array}{ll}
\text{maximize} & t \\
\text{subject to} & \sum_{i=1}^{m} w_i f(x_i) f(x_i)^T \succeq tI \\
& w_i \geq 0, \quad i = 1, \ldots, m \\
& \sum_{i=1}^{m} w_i = 1
\end{array}
$$

first constraint is equivalent to $\lambda_{\min}(M) \geq t$

Semidefinite program formulation of D-optimal design

$$
\begin{array}{ll}
\text{maximize} & (\det M)^{1/p} \\
\text{subject to} & M = \sum_{i=1}^{m} w_i f(x_i) f(x_i)^T \\
& w_i \geq 0, \quad i = 1, \ldots, m, \quad \sum_{i=1}^{m} w_i = 1
\end{array}
$$

Step 1: introduce Cholesky factor as auxiliary variable

$$
\begin{array}{ll}
\text{maximize} & (\prod_i R_{ii})^{1/p} \\
\text{subject to} & \begin{bmatrix} \sum_{i=1}^{m} w_i f(x_i) f(x_i)^T & R^T \\ R & I \end{bmatrix} \succeq 0 \\
& R \text{ upper triangular}, \quad R_{ii} \geq 0, \quad i = 1, \ldots, p \\
& w_i \geq 0, \quad i = 1, \ldots, m, \quad \sum_{i=1}^{m} w_i = 1
\end{array}
$$

first constraint is equivalent to $M \succeq R^T R$

Semidefinite program formulation of D-optimal design

$$
\begin{array}{ll}
\text{maximize} & t \\
\text{subject to} & t \leq \left(\prod_i R_{ii}\right)^{1/p} \\
& \left[\begin{array}{cc} \sum_{i=1}^{m} w_i f(x_i) f(x_i)^T & R^T \\ R & I \end{array}\right] \succeq 0 \\
& R \text{ upper triangular}, \quad R_{ii} \geq 0, \quad i = 1, \ldots, p \\
& w_i \geq 0, \quad i = 1, \ldots, m, \quad \sum_{i=1}^{m} w_i = 1
\end{array}
$$

Step 2

- first constraint can be expressed as a set of p.s.d. constraints
- reformulation uses repeated application of equivalence

$$
a \leq \sqrt{bc}, \quad a, b, c \geq 0 \qquad \Longleftrightarrow \qquad \left[\begin{array}{cc} b & a \\ a & c \end{array}\right] \succeq 0, \quad a \geq 0
$$

Outline

- Semidefinite representations of design criteria
- Discriminating design with polynomial models
- Semidefinite descriptions of moment cones

Optimal discriminating design

- p-vector $f(x)$ of basis functions, and m models

$$\eta_j(x) = \theta_j^T f(x), \quad \theta_j \in \Theta_j, \quad j = 1, \ldots, m$$

- moment matrix $M = \mathbf{E} f(x) f(x)^T$ depends on distribution of $x \in C$
- $m(m-1)/2$ distance measures

$$\Delta_{ij}(M) = \inf_{\theta_i \in \Theta_i,\, \theta_j \in \Theta_j} \mathbf{E}(\eta_i(x) - \eta_j(x))^2 = \inf_{\theta_i \in \Theta_i,\, \theta_j \in \Theta_j} (\theta_i - \theta_j)^T M (\theta_i - \theta_j)$$

- each $\Delta_{ij}(M)$ is a concave function of M

Design problem: find distribution that makes all $\Delta_{ij}(M)$ large, *e.g.*, by maximizing

$$\min_{j>i} \Delta_{ij}(M) \qquad \text{or} \qquad \sum_{j>i} w_{ij} \Delta_{ij}(M)$$

[Atkinson and Fedorov 1975]

Equivalent expressions for $\Delta_{ij}(M)$

for given M, the function $\Delta_{ij}(M)$ is the optimal value of the optimization problem

$$\begin{array}{ll} \text{minimize} & (\theta_i - \theta_j)^T M (\theta_i - \theta_j) \\ \text{subject to} & \theta_i \in \Theta_i, \ \theta_j \in \Theta_j \end{array}$$

- variables are θ_i, θ_j
- we now assume that Θ_i, Θ_j are convex sets

From convex duality: $\Delta_{ij}(M)$ is the optimal value of the dual problem

$$\begin{array}{ll} \text{maximize} & t - \sigma_i(z) - \sigma_j(-z) \\ \text{subject to} & \begin{bmatrix} M & z/2 \\ z^T/2 & t \end{bmatrix} \succeq 0 \end{array}$$

- variables are t, z
- $\sigma_k(z) = \sup_{\theta \in \Theta_k} z^T \theta$ is support function of Θ_k (a convex function)

Optimal discriminating design

$$\begin{array}{ll} \text{maximize} & \min_{j>i} \Delta_{ij}(M) \\ \text{subject to} & M \in \mathcal{M} \end{array}$$

- $\mathcal{M} = \mathbf{conv}\{f(x)f(x)^T \mid x \in C\}$
- optimization problem is convex in the moment matrix M

Reformulation

$$\begin{array}{ll} \text{maximize} & t \\ \text{subject to} & t \leq t_{ij} - \sigma_i(z_{ij}) - \sigma_j(-z_{ij}), \quad j > i \\ & \left[\begin{array}{cc} M & z_{ij}/2 \\ z_{ij}^T/2 & t_{ij} \end{array} \right] \succeq 0, \quad j > i \\ & M \in \mathcal{M} \end{array}$$

- convex in the variables t, t_{ij}, z_{ij}, M
- requires tractable description or approximation of set $\mathcal{M}$ of moment matrices

Polynomial moments and SDP approximations

- $f(x)$ is vector of $\binom{n+d}{d}$ monomials in $x_1, \ldots, x_n$ of degree d or less
- design space C is a compact set defined by k polynomial inequalities

$$g_1(x) \geq 0, \qquad \ldots, \qquad g_k(x) \geq 0$$

- set of moment matrices is $\mathcal{M} = \mathrm{conv}\{f(x)f(x)^T \mid x \in C\}$

Hierarchy of relaxations: outer approximations $\mathcal{M} \subseteq \mathcal{M}_r$, $r = 0, 1, \ldots$

- $\mathcal{M}_r$ is parameterized by $k+1$ linear matrix inequalities of size up to

$$\binom{n+2d+2r}{n}$$

- approximations are nested and converge to $\mathcal{M}$ as relaxation order r increases

[De Castro, Gamboa, Henrion, Hess, Lasserre 2017] [Lasserre 2010, 2015]

Sums of squares

a polynomial $f : \mathbf{R}^n \to \mathbf{R}$ is a sum of squares (SOS) of degree $2d$ or less if

$$f(x) = \sum_{|\alpha| \leq d} \sum_{|\beta| \leq d} A_{\alpha\beta} x^\alpha x^\beta = v_d(x)^T A v_d(x) \qquad \text{with } A \succeq 0$$

- x^α with $\alpha = (\alpha_1, \ldots, \alpha_n)$ denotes the monomial $x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$
- $|\alpha| = \sum_i \alpha_i$ is degree of monomial x^α
- $v_d(x)$ is vector of $\begin{pmatrix} n+d \\ d \end{pmatrix}$ monomials in x of degree d or less
- SOS property is a semidefinite constraint in coefficients of $f(x)$ and matrix A
- gives a sufficient condition for nonnegativity of $f(x)$

[Parrilo 2000] [Lasserre 2001]

Inner approximation of cone of nonnegative polynomials

- C is a compact set defined by polynomial inequalities $g_1(x) \geq 0, \ldots, g_k(x) \geq 0$
- $\mathcal{P}$ is the cone of polynomials of degree d or less that are nonnegative on C
- sufficient condition for $p \in \mathcal{P}$:

$$p(x) = p_0(x) + \sum_{j=1}^{k} p_j(x) g_j(x)$$

where $p_0(x), \ldots, p_k(x)$ are sums of squares, *i.e.*,

$$p_j(x) = v_{r_j}(x)^T A_j v_{r_j}(x), \quad A_j \succeq 0$$

- defines a p.s.d.-representable inner approximation of $\mathcal{P}$
- increasing the degrees of p_k gives hierarchy of nested inner approximations
- outer approximation of polynomial moment cones follows by duality

surveys in books [Lasserre 2010, 2015]

Example 1

- two models in 7 variables; one exact and one uncertain

$$\begin{aligned} \eta_1(x) &= 1 + x_1 + \cdots + x_7 + x_1^2 + x_1 x_2 + \cdots + x_6 x_7 + x_7^2 \\ \eta_2(x) &= \theta_1 + \theta_2 x_1 + \cdots + \theta_8 x_7 + \theta_9 x_1^2 + \cdots + \theta_{15} x_7^2 \end{aligned}$$

- design space is $C = [-1, 1]^7$
- parameter constraint $\theta \in [0, 4]^{15}$
- relaxation of order 2 gives solution
- optimal design has 72 support points

Example 2

- three models in 3 variables; one exact and two uncertain

$$
\begin{aligned}
\eta_1(x) &= 1 + x_1 + x_2 + x_3 + x_1^2 + x_1x_2 + x_1x_3 + x_2^2 + x_2x_3 + x_3^2 \\
\eta_2(x) &= \theta_{2,1} + \theta_{2,2}x_1 + \theta_{2,3}x_2 + \theta_{2,4}x_3 \\
\eta_3(x) &= \theta_{3,1} + \theta_{3,2}x_1 + \theta_{3,3}x_2 + +\theta_{3,4}x_3 + \theta_{3,5}x_1^2 + \theta_{3,6}x_2^2 + \theta_{3,7}x_3^2
\end{aligned}
$$

- design space is $C = [-1, 1]^3$
- parameter constraints are $\theta_2 = [1, 2]^4$ and $\theta_3 \in [1, 2]^7$
- relaxation of order zero gives solution
- optimal design has 8 support points

Outline

- Semidefinite representations of design criteria
- Discriminating design with polynomial models
- Semidefinite descriptions of moment cones

SDPs in signal processing and system theory

Classical sum-of-squares theorems

- characterize nonnegativity of univariate (trigonometric) polynomials
  [Karlin and Studden 1966] [Krein and Nudelman 1977]
- (generalized) Kalman-Yakubovich-Popov lemma in system theory
- equivalent to sets of linear matrix inequalities
- via convex duality, SDP descriptions of moment cones

Applications

- underlie many of the applications of SDP in control and signal processing
- recent applications to experiment design for system identification
  [Jansson and Hjalmarsson 2005] [Hildebrand, Gevers, Solari 2015]

Positive semidefinite Toeplitz matrices

every $n \times n$ positive semidefinite Toeplitz matrix X can be decomposed as

$$X = \sum_{k=1}^{r} |c_k|^2 \begin{bmatrix} 1 \\ e^{\mathrm{i}\omega_k} \\ e^{\mathrm{i}2\omega_k} \\ \vdots \\ e^{\mathrm{i}(n-1)\omega_k} \end{bmatrix} \begin{bmatrix} 1 \\ e^{\mathrm{i}\omega_k} \\ e^{\mathrm{i}2\omega_k} \\ \vdots \\ e^{\mathrm{i}(n-1)\omega_k} \end{bmatrix}^H$$

- cone of positive semidefinite Toeplitz matrices is convex hull of

$$\{aa^H \mid a = c(1, e^{\mathrm{i}\omega}, \ldots, e^{\mathrm{i}(n-1)\omega})\}$$

- this is also the cone of trigonometric moment matrices
- next: extensions from papers on Kalman-Yakubovich-Popov lemma

Quadratic matrix equation

let U, V be $p \times r$ matrices that satisfy

$$UU^H = VV^H$$

- $U = VS$ with S unitary: follows from singular value decompositions

  $$U = P\Sigma Q_1^H, \qquad V = P\Sigma Q_2^H$$

  and $S = Q_2 Q_1^H$

- take Schur decomposition $S = Q\,\mathrm{diag}(\lambda)Q^H$:

  $$UQ = VQ\,\mathrm{diag}(\lambda)$$

  with Q unitary and $|\lambda_1| = \cdots = |\lambda_r| = 1$

Decomposition of positive semidefinite Toeplitz matrix

- $n \times n$ matrix X is Toeplitz if $FXF^H = GXG^H$ where

$$F = \begin{bmatrix} 0 & I_{n-1} \end{bmatrix}, \qquad G = \begin{bmatrix} I_{n-1} & 0 \end{bmatrix}$$

- factorize $X = YY^H$; the matrix Y satisfies $(FY)(FY)^H = (GY)(GY)^H$:

$$FYQ = GYQ\,\mathrm{diag}(\lambda) \quad \text{with } Q \text{ unitary}, \quad |\lambda_1| = \cdots = |\lambda_r| = 1$$

- columns $a_1, \ldots, a_r$ of YQ give the decomposition

$$X = \sum_{k=1}^{r} a_k a_k^H, \qquad Fa_k = \lambda_k G a_k, \qquad |\lambda_k| = 1$$

  vectors a_k have the form $a_k = c_k(1, \lambda_k, \ldots, \lambda_k^{n-1})$ with $\lambda_k = e^{\mathrm{i}\omega_k}$

Note: this holds for any pair F, G of equal dimension

General quadratic equation

suppose $\Phi \in \mathbf{H}^2$ with $\det \Phi < 0$, and U, V are $p \times r$ matrices with

$$\Phi_{11} U U^H + \Phi_{21} U V^H + \Phi_{12} V U^H + \Phi_{22} V V^H = 0$$

- then there exist unitary Q, vectors μ, ν with

$$U Q \operatorname{diag}(\nu) = V Q \operatorname{diag}(\mu), \qquad \left[\begin{array}{c} \mu_k \\ \nu_k \end{array}\right]^H \Phi \left[\begin{array}{c} \mu_k \\ \nu_k \end{array}\right] = 0, \qquad (\mu_k, \nu_k) \neq 0$$

- second condition restricts $\lambda_k = \mu_k / \nu_k$ to circle or line in complex plane

Φ:	$\left[\begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array}\right]$	$\left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}\right]$	$\left[\begin{array}{cc} 0 & \mathrm{i} \\ -\mathrm{i} & 0 \end{array}\right]$
λ:	unit circle	imaginary axis	real axis

pairs (μ_k, ν_k) with $\nu_k = 0$ correspond to point λ_k at infinity

Quadratic matrix equation and inequality

suppose $\Phi, \Psi \in \mathbf{H}^2$ with $\det \Phi < 0$, and U, V are $p \times r$ matrices with

$$\Phi_{11} U U^H + \Phi_{21} U V^H + \Phi_{12} V U^H + \Phi_{22} V V^H = 0$$
$$\Psi_{11} U U^H + \Psi_{21} U V^H + \Psi_{12} V U^H + \Psi_{22} V V^H \leq 0$$

- then there exist unitary Q, vectors μ, ν with $(\mu_k, \nu_k) \neq 0$, such that

$$UQ \operatorname{diag}(\nu) = VQ \operatorname{diag}(\mu)$$

  and

$$\begin{bmatrix} \mu_k \\ \nu_k \end{bmatrix}^H \Phi \begin{bmatrix} \mu_k \\ \nu_k \end{bmatrix} = 0 \qquad \begin{bmatrix} \mu_k \\ \nu_k \end{bmatrix}^H \Psi \begin{bmatrix} \mu_k \\ \nu_k \end{bmatrix} \leq 0$$

- last two conditions restrict $\lambda_k = \mu_k / \nu_k$ to segment of circle or line
- efficiently computed using standard matrix decompositions (SVD, Schur)

[Iwasaki, Meinsma, Hara 2000] [Iwasaki and Hara 2003]

Generalized Carathéodory decomposition

the following two properties are equivalent:

- X is in the convex hull of $\{aa^H \mid a \in \mathcal{A}\}$

$$\mathcal{A} = \{a \mid \mu G a = \nu F a, \ (\mu, \nu) \in C\}$$

  C is a segment of a line or circle in the complex plane, parameterized by

$$(\mu, \nu) \neq 0, \qquad \begin{bmatrix} \mu \\ \nu \end{bmatrix}^H \Phi \begin{bmatrix} \mu \\ \nu \end{bmatrix} = 0, \qquad \begin{bmatrix} \mu \\ \nu \end{bmatrix}^H \Psi \begin{bmatrix} \mu \\ \nu \end{bmatrix} \leq 0$$

- X is positive semidefinite and satisfies the matrix equation and inequality

$$\Phi_{11} F X F^H + \Phi_{21} F X G^H + \Phi_{12} G X F^H + \Phi_{22} G X G^H = 0$$
$$\Psi_{11} F X F^H + \Psi_{21} F X G^H + \Psi_{12} G X F^H + \Psi_{22} G X G^H \preceq 0$$

decomposition $X = \sum_{k=1}^{r} a_k a_k^H$ with $a_k \in \mathcal{A}$ from efficient matrix algorithms

Other interesting choices of F, G

$$F = \begin{bmatrix} J & \beta e_{n-1} \end{bmatrix}, \qquad G = \begin{bmatrix} I_{n-1} & 0 \end{bmatrix}$$

- J is a tridiagonal (Jacobi) matrix
- J and β define 3-term recurrence for system of orthogonal polynomials

$$p_0(\lambda), \quad p_2(\lambda), \quad \ldots, \quad p_{n-1}(\lambda)$$

- SDP description of convex hull of $\{aa^H \mid a \in \mathcal{A}\}$ where $\mathcal{A}$ contains vectors

$$a = c\left(p_0(\lambda), p_1(\lambda), \ldots, p_{n-1}(\lambda)\right), \qquad \lambda \in C$$

where C is an interval of the real axis

Other interesting choices of F, G

$$F = \begin{bmatrix} A & B \end{bmatrix}, \qquad G = \begin{bmatrix} I & 0 \end{bmatrix} \qquad \text{(size } n_\mathrm{s} \times (n_\mathrm{s} + m))$$

- $\lambda G - F$ is controllability pencil of linear system

$$\lambda G - F = \begin{bmatrix} \lambda I - A & B \end{bmatrix}$$

- SDP description of convex hull of $\{aa^H \mid a \in \mathcal{A}\}$ where $\mathcal{A}$ contains the vectors

$$a = \begin{bmatrix} (\lambda I - A)^{-1} B u \\ u \end{bmatrix}, \qquad u \in \mathbf{C}^m, \qquad \lambda \in C$$

and C is (a segment of) the unit circle or imaginary axis

Summary

- optimal experiment design via second-order cone/semidefinite programming
- SDP relaxations of multivariate polynomial moment cones
- exact SDP description of class of univariate moment cones