Semidefinite programming methods for continuous sparse optimization

Lieven Vandenberghe

Department of Electrical Engineering

University of California, Los Angeles

Joint work with Hsiao-Han Chao

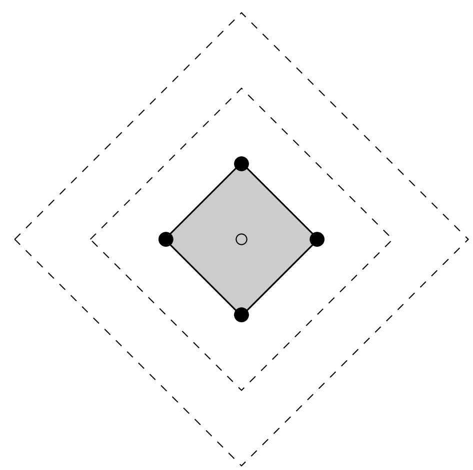
iTWIST 2016

Aalborg, August 26, 2016

Atomic norm

the atomic norm associated with a set C is the **gauge** of its convex hull:

$$g(x) = \inf\{t > 0 \mid x/t \in \operatorname{conv} C\}$$



- a convex, nonnegative, positively homogeneous function
- the largest function with these properties that satisfies $g(x) \leq 1$ for $x \in C$
- not necessarily a norm
- a unified descripton of convex ℓ_1-like penalties

(Chandrasekharan, Recht, Parrilo, Willsky 2012)

Atomic norm

- more explicit expression, obtained by expanding $\mathbf{conv}\, C$ in the definition:

$$g(x) = \inf \{ \sum_{k=1}^{r} \theta_k \mid x = \sum_{k=1}^{r} \theta_k a_k,\ \theta_k \geq 0,\ a_k \in C \}$$

- if C is symmetric ($a \in C$ implies $sa \in C$ for $|s| = 1$):

$$g(x) = \inf \{ \sum_{k=1}^{r} |\theta_k| \mid x = \sum_{k=1}^{r} \theta_k a_k,\ a_k \in C \}$$

Examples

- trace norm $g(X) = \sum_i \sigma_i(X)$: atomic norm of

$$\{ v w^H \mid \|v\| = \|w\| = 1 \}$$

- $g(X) = \mathrm{tr}\, X$ on $\mathbf{dom}\, g = \{ X \mid X \succeq 0 \}$: atomic norm of

$$\{ v v^H \mid \|v\| = 1 \}$$

Regularization with atomic norm

$$\text{minimize} \quad f(x) + g(x)$$

- f a convex function, possibly an indicator of a set
- equivalent problem (assume symmetric C):

$$\begin{array}{ll} \text{minimize} & f(x) + \sum_{k=1}^{r} |\theta_k| \\ \text{subject to} & \sum_{k=1}^{r} \theta_k a_k = x \\ & a_1, \ldots, a_r \in C \end{array}$$

unknowns are variable x, parameters θ_k, a_k, r of the decomposition

- extends LASSO, basis pursuit, noisy basis pursuit, $\ldots$ to non-finite sets C

Complex exponentials

$$C = \left\{ \gamma\left(1, e^{\mathrm{j}\omega}, e^{\mathrm{j}2\omega}, \ldots, e^{\mathrm{j}(n-1)\omega}\right) \mid \omega \in [0, 2\pi),\ |\gamma| = \frac{1}{\sqrt{n}} \right\}$$

- atomic norm $g(x)$ is minimum of $\sum_k |\theta_k|$ subject to

$$x = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{\mathrm{j}\omega_1} & e^{\mathrm{j}\omega_2} & \cdots & e^{\mathrm{j}\omega_r} \\ \vdots & \vdots & & \vdots \\ e^{\mathrm{j}(n-1)\omega_1} & e^{\mathrm{j}(n-1)\omega_2} & \cdots & e^{\mathrm{j}(n-1)\omega_r} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_r \end{bmatrix}$$

- $g(x)$ is optimal value of semidefinite program with variables $V \in \mathbf{H}^n$, $w \in \mathbf{R}$

$$\begin{array}{ll} \text{minimize} & (\operatorname{tr} V + w)/2 \\ \text{subject to} & \begin{bmatrix} V & x \\ x^H & w \end{bmatrix} \succeq 0, \quad V \text{ is Toeplitz} \end{array}$$

(Candès, Fernandez-Granda 2013; Tang, Bhaskar, Shah, Recht 2013; Yang, Xie 2015)

Atomic norm regularization

$$\begin{array}{ll} \text{minimize} & f(x) + \sum_{k=1}^{r} |\theta_k| \\ \text{subject to} & x = \frac{1}{\sqrt{n}} \left[\begin{array}{cccc} 1 & 1 & \cdots & 1 \\ e^{\mathrm{j}\omega_1} & e^{\mathrm{j}\omega_2} & \cdots & e^{\mathrm{j}\omega_r} \\ \vdots & \vdots & & \vdots \\ e^{\mathrm{j}(n-1)\omega_1} & e^{\mathrm{j}(n-1)\omega_2} & \cdots & e^{\mathrm{j}(n-1)\omega_r} \end{array} \right] \left[\begin{array}{c} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_r \end{array} \right] \end{array}$$

variables: x, parameters θ_k, ω_k, r of decomposition

Convex formulation

$$\begin{array}{ll} \text{minimize} & f(x) + (\mathrm{tr}\, V + w)/2 \\ \text{subject to} & \left[\begin{array}{cc} V & x \\ x^H & w \end{array} \right] \succeq 0, \quad V \text{ is Toeplitz} \end{array}$$

applications include superresolution, ‘gridless’ compressed sensing

Matrix extension

$$\begin{array}{ll} \text{minimize} & f(X) + \sum_{k=1}^{r} \|\theta_k\| \\ \text{subject to} & X = \dfrac{1}{\sqrt{n}} \left[\begin{array}{cccc} 1 & 1 & \cdots & 1 \\ e^{\mathrm{j}\omega_1} & e^{\mathrm{j}\omega_2} & \cdots & e^{\mathrm{j}\omega_r} \\ \vdots & \vdots & & \vdots \\ e^{\mathrm{j}(n-1)\omega_1} & e^{\mathrm{j}(n-1)\omega_2} & \cdots & e^{\mathrm{j}(r-1)\omega_m} \end{array} \right] \left[\begin{array}{c} \theta_1^H \\ \theta_2^H \\ \vdots \\ \theta_r^H \end{array} \right] \end{array}$$

variables: matrix X, parameters θ_k, ω_k, r of decomposition

Convex formulation

$$\begin{array}{ll} \text{minimize} & f(X) + (\mathrm{tr}\, V + \mathrm{tr}\, W)/2 \\ \text{subject to} & \left[\begin{array}{cc} V & X \\ X^H & W \end{array} \right] \succeq 0, \quad V \text{ is Toeplitz} \end{array}$$

(Li, Chi 2014; Yang, Xie 2014)

Outline

This talk

- semidefinite representations of a larger class of atomic norms
- applications to low-rank matrix decompositions with structure

Outline

- introduction
- **Carathéodory-type matrix decomposition**
- structured trace norm penalties
- examples
- duality

Decomposition of positive semidefinite Toeplitz matrix

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

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

$$= \sum_{k=1}^{r} |c_k|^2 \begin{bmatrix} 1 & e^{-\mathrm{j}\omega_k} & \cdots & e^{-\mathrm{j}(n-1)\omega_k} \\ e^{\mathrm{j}\omega_k} & 1 & \cdots & e^{-\mathrm{j}(n-2)\omega_k} \\ \vdots & \vdots & \ddots & \vdots \\ e^{\mathrm{j}(n-1)\omega_k} & e^{\mathrm{j}(n-2)\omega_k} & \cdots & 1 \end{bmatrix}$$

- terms in sum are extreme rays of the convex cone of p.s.d. Toeplitz matrices
- next: extensions from papers on Kalman-Yakubovich-Popov lemma
  (starting with Rantzer 1996)

Quadratic matrix equation

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

$$UU^H = VV^H$$

- U and V have singular value decompositions

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

- therefore $U = VS$ with $S = Q_2 Q_1^H$ (a unitary matrix)
- 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 $\lambda_k = e^{\mathrm{j}\omega_k}$

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

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

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

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

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

$$\Phi_{11}UU^H + \Phi_{21}UV^H + \Phi_{12}VU^H + \Phi_{22}VV^H = 0$$
$$\Psi_{11}UU^H + \Psi_{21}UV^H + \Psi_{12}VU^H + \Psi_{22}VV^H \preceq 0$$

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

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

  and

$$\left[\begin{array}{c} \mu_k \\ \nu_k \end{array}\right]^H \Phi \left[\begin{array}{c} \mu_k \\ \nu_k \end{array}\right] = 0 \qquad \left[\begin{array}{c} \mu_k \\ \nu_k \end{array}\right]^H \Psi \left[\begin{array}{c} \mu_k \\ \nu_k \end{array}\right] \leq 0$$

- 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 can be decomposed as

$$X = \sum_{k=1}^{r} a_k a_k^H$$

with vectors a_k taken from the set

$$\mathcal{A} = \{a \in \mathbf{C}^n \mid (\mu G - \nu F)a = 0,\ (\mu, \nu) \in \mathcal{C}_{\Phi\Psi}\}$$

$\mathcal{C}_{\Phi\Psi}$ 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

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

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

Example

$$F = \left[\begin{array}{cc} 0 & I \end{array}\right], \qquad G = \left[\begin{array}{cc} I & 0 \end{array}\right], \qquad \Phi = \left[\begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array}\right], \qquad \Psi = \left[\begin{array}{cc} 0 & -e^{\mathrm{j}\alpha} \\ -e^{-\mathrm{j}\alpha} & 2\cos\beta \end{array}\right]$$

the following two properties are equivalent:

- X can be decomposed as

$$X = \sum_{k=1}^{r} |c_k|^2 \left[\begin{array}{c} 1 \\ e^{\mathrm{j}\omega_k} \\ \vdots \\ e^{\mathrm{j}(n-1)\omega_k} \end{array}\right] \left[\begin{array}{c} 1 \\ e^{\mathrm{j}\omega_k} \\ \vdots \\ e^{\mathrm{j}(n-1)\omega_k} \end{array}\right]^H, \qquad |\omega_k - \alpha| \leq \beta$$

- X is positive semidefinite and satisfies

$$FXF^H - GXG^H = 0 \qquad \text{(X is Toeplitz)}$$

$$-e^{\mathrm{j}\alpha} FXG^H - e^{-\mathrm{j}\alpha} GXF^H + 2(\cos\beta) GXG^H \preceq 0$$

Other interesting choices of F, G

Orthogonal polynomials on the real axis: $\mathcal{C}_{\Phi\Psi}$ defines interval of real axis,

$$\lambda G - F = \left[\begin{array}{cc} \lambda I_{n-1} - J & -\beta e_{n-1} \end{array}\right], \qquad J \text{ a Jacobi matrix}$$

$\mathcal{A}$ contains vectors

$$c\,(p_0(\lambda),\, p_1(\lambda),\, \ldots,\, p_{n-1}(\lambda))$$

for the polynomials p_k defined by 3-term recurrence with coefficients in J, β

State-space linear system model: $\mathcal{C}_{\Phi\Psi}$ is unit circle or imaginary axis,

$$\lambda G - F = \left[\begin{array}{cc} \lambda I - A & B \end{array}\right] \qquad \text{(size } n_{\mathrm{s}} \times (n_{\mathrm{s}} + m)\text{)}$$

$\mathcal{A}$ contains vectors

$$\left[\begin{array}{c} (\lambda I - A)^{-1} B u \\ u \end{array}\right], \qquad u \in \mathbf{C}^m$$

Outline

- Introduction
- Carathéodory-type matrix decompositions
- **Structured trace norm penalties**
- Example
- Duality

Trace penalty for positive semidefinite matrices

define a structured 'trace' penalty function

$$g(X) = \begin{cases} \operatorname{tr} X & \text{if } X = \sum_{k=1}^{r} a_k a_k^H \text{ with } a_1, \ldots, a_r \in \mathcal{A} \\ +\infty & \text{otherwise} \end{cases}$$

- vectors a_k are taken from $\mathcal{A} = \{a \in \mathbf{C}^n \mid (\mu G - \nu F)a = 0,\ (\mu, \nu) \in \mathcal{C}_{\Phi\Psi}\}$
- $g(X)$ is the atomic 'norm' of $C = \{aa^H \in \mathbf{H}^n \mid a \in \mathcal{A},\ \|a\| = 1\}$

Semidefinite representation: $g(X) = \operatorname{tr} X$ if $X \succeq 0$ and

$$\Phi_{11}FXF^H + \Phi_{21}FXG^H + \Phi_{12}GXF^H + \Phi_{22}GXG^H = 0$$
$$\Psi_{11}FXF^H + \Psi_{21}FXG^H + \Psi_{12}GXF^H + \Psi_{22}GXG^H \preceq 0,$$

and $g(X) = +\infty$ otherwise

Regularization with structured trace penalty

$$\begin{array}{ll} \text{minimize} & f(X) + \sum_{k=1}^{r} \|a_k\|^2 \\ \text{subject to} & X = \sum_{k=1}^{r} a_k a_k^H \\ & a_1, \ldots, a_r \in \mathcal{A} \end{array}$$

- variables: $X \in \mathbf{H}^n$, and parameters a_1, …, a_r, r in the decomposition
- regularization term promotes existence of structured low-rank decomposition

Semidefinite formulation

$$\begin{array}{ll} \text{minimize} & f(X) + \mathrm{tr}\, X \\ \text{subject to} & \Phi_{11} F X F^H + \Phi_{21} F X G^H + \Phi_{12} G X F^H + \Phi_{22} G X G^H = 0 \\ & \Psi_{11} F X F^H + \Psi_{21} F X G^H + \Psi_{12} G X F^H + \Psi_{22} G X G^H \preceq 0 \\ & X \succeq 0 \end{array}$$

Example: line spectrum estimation by covariance matrix fitting

$$
\begin{array}{ll}
\text{minimize} & f(R) + \gamma \sum_{k=1}^{r} |c_k|^2 \\
\text{subject to} & R = \sigma^2 I + \sum_{k=1}^{r} |c_k|^2 \begin{bmatrix} 1 \\ e^{\mathrm{j}\omega_k} \\ \vdots \\ e^{\mathrm{j}(n-1)\omega_k} \end{bmatrix} \begin{bmatrix} 1 \\ e^{\mathrm{j}\omega_k} \\ \vdots \\ e^{\mathrm{j}(n-1)\omega_k} \end{bmatrix}^H \\
& \omega_1, \ldots, \omega_r \in [\alpha - \beta, \alpha + \beta]
\end{array}
$$

- for example, $f(R) = \|R - R_{\mathrm{m}}\|_2$, with R_{m} an estimated covariance matrix
- variables are R and parameters σ^2, c_k, ω_k, r in decomposition of R

Semidefinite formulation (variables X, t)

$$
\begin{array}{ll}
\text{minimize} & f(X - tI) + (\gamma/n) \operatorname{tr} X \\
\text{subject to} & t \geq 0, \quad X \succeq 0 \\
& FXF^H = GXG^H \qquad (X \text{ is Toeplitz}) \\
& -e^{\mathrm{j}\alpha} FXG^H - e^{-\mathrm{j}\alpha} GXF^H + 2(\cos\beta) GXG^H \preceq 0
\end{array}
$$

Trace norm (nuclear norm) for nonsymmetric matrices

- $\|Y\|_*$ (sum of singular values) can be expressed in several ways, including

$$\begin{aligned} \|Y\|_* &= \inf \{\sum_{k=1}^{r} \|v_k\|\|w_k\| \mid Y = \sum_{k=1}^{r} v_k w_k^H\} \\ &= \inf \{\frac{1}{2}\sum_{k=1}^{r}(\|v_k\|^2 + \|w_k\|^2) \mid Y = \sum_{k=1}^{r} v_k w_k^H\} \end{aligned}$$

- $\|Y\|_*$ is also the atomic norm of $C = \{vw^H \mid \|v\| = \|w\| = 1\}$

Semidefinite representation: $\|Y\|_*$ is the optimal value of

$$\text{minimize} \quad \frac{1}{2}(\operatorname{tr} V + \operatorname{tr} W) \quad \text{subject to} \quad \begin{bmatrix} V & Y \\ Y^H & W \end{bmatrix} \succeq 0$$

Structured trace norm

- add constraints on v_k, w_k in the definition of trace norm:

$$h(Y) = \inf \{\sum_{k=1}^{r} \|v_k\|\|w_k\| \mid Y = \sum_{k=1}^{r} v_k w_k^H, \ (v_k, w_k) \in \mathcal{A}\}$$

- here $\mathcal{A}$ is defined as before, but with block-diagonal F, G:

$$\mathcal{A} = \{a \mid (\mu G - \nu F)a = 0, \ (\mu, \nu) \in \mathcal{C}_{\Phi\Psi}\}$$

with

$$G = \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix}, \quad F = \begin{bmatrix} F_1 & 0 \\ 0 & F_2 \end{bmatrix}$$

- equivalently,

$$\mathcal{A} = \{(v, w) \mid (\mu G_1 - \nu F_1)v = 0, \ (\mu G_2 - \nu F_2)w = 0, \ (\mu, \nu) \in \mathcal{C}_{\Phi\Psi}\}$$

- row dimension of G_1, F_1 and G_2, F_2 may be zero (*i.e.*, v or w are unrestricted)

Semidefinite representation

$$
\begin{aligned}
h(Y) &= \inf\{\sum_{k=1}^{r} \|v_k\|\|w_k\| \mid Y = \sum_{k=1}^{r} v_k w_k^H,\ (v_k, w_k) \in \mathcal{A}\} \\
&= \inf\{\frac{1}{2}\sum_{k=1}^{r}(\|v_k\|^2 + \|w_k\|^2) \mid Y = \sum_{k=1}^{r} v_k w_k^H,\ (v_k, w_k) \in \mathcal{A}\}
\end{aligned}
$$

Semidefinite representation: $h(Y)$ is the optimal value of the SDP

$$
\begin{array}{ll}
\text{minimize} & (\operatorname{tr} V + \operatorname{tr} W)/2 \\
\text{subject to} & X = \left[\begin{array}{cc} V & Y \\ Y^H & W \end{array}\right] \succeq 0 \\
& \Phi_{11} F X F^H + \Phi_{21} F X G^H + \Phi_{12} G X F^H + \Phi_{22} G X G^H = 0 \\
& \Psi_{11} F X F^H + \Psi_{21} F X G^H + \Psi_{12} G X F^H + \Psi_{22} G X G^H \preceq 0
\end{array}
$$

Regularization with structured trace norm

$$\begin{array}{ll} \text{minimize} & f(Y) + \sum_{k=1}^{r} \|v_k\| \|w_k\| \\ \text{subject to} & Y = \sum_{k=1}^{r} v_k w_k^H \\ & (v_1, w_1), \ldots, (v_r, w_r) \in \mathcal{A} \end{array}$$

variables: $Y \in \mathbf{H}^{m \times n}$, and parameters (v_k, w_k), r in the decomposition

SDP formulation (with variables Y, V, W)

$$\begin{array}{ll} \text{minimize} & f(Y) + (\operatorname{tr} V + \operatorname{tr} W)/2 \\ \text{subject to} & X = \begin{bmatrix} V & Y \\ Y^H & W \end{bmatrix} \succeq 0 \\ & \Phi_{11} F X F^H + \Phi_{21} F X G^H + \Phi_{12} G X F^H + \Phi_{22} G X G^H = 0 \\ & \Psi_{11} F X F^H + \Psi_{21} F X G^H + \Psi_{12} G X F^H + \Psi_{22} G X G^H \preceq 0 \end{array}$$

Special case: column structure

$$\begin{aligned} h(Y) &= \inf\{\sum_{k=1}^{r} \|v_k\|\|w_k\| \mid Y = \sum_{k=1}^{r} v_k w_k^H,\ v_k \in \mathcal{A}_1\} \\ &= \inf\{\sum_{k=1}^{r} \|w_k\| \mid Y = \sum_{k=1}^{r} v_k w_k^H,\ v_k \in \mathcal{A}_1,\ \|v_k\| = 1\} \end{aligned}$$

- colums v_k taken from $\mathcal{A} = \{v \mid (\mu G_1 - \nu F_1)v = 0,\ (\mu, \nu) \in \mathcal{C}_{\Phi\Psi}\}$
- row vectors w_k^H are unconstrained

Semidefinite representation: $h(Y)$ is the optimal value of the SDP

minimize $\quad (\mathrm{tr}\, V + \mathrm{tr}\, W)/2$

subject to
$$\begin{bmatrix} V & Y \\ Y^H & W \end{bmatrix} \succeq 0$$
$$\Phi_{11} F_1 V F_1^H + \Phi_{21} F_1 V G_1^H + \Phi_{12} G_1 V F_1^H + \Phi_{22} G_1 V G_1^H = 0$$
$$\Psi_{11} F_1 V F_1^H + \Psi_{21} F_1 V G_1^H + \Psi_{12} G_1 V F_1^H + \Psi_{22} G_1 V G_1^H \preceq 0$$

Example: fitting sinusoids to noisy data

Penalized least squares formulation

$$\begin{array}{ll} \text{minimize} & f(x) + \gamma\sqrt{n} \sum_{k=1}^{r} |c_k| \\ \text{subject to} & x = \sum_{k=1}^{r} c_k \begin{bmatrix} 1 \\ e^{\mathrm{j}\omega_k} \\ \vdots \\ e^{\mathrm{j}(n-1)\omega_k} \end{bmatrix} \end{array}$$

- for example, $f(x) = \|x - x_{\mathrm{m}}\|^2$, with x_{m} a noisy measurement
- optimization variables: x and signal model parameters c_k, ω_k, r

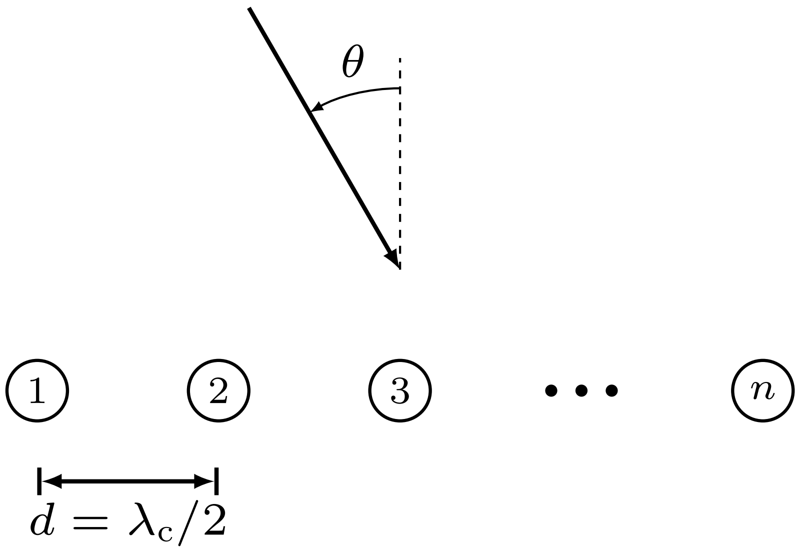
Semidefinite formulation

$$\begin{array}{ll} \text{minimize} & f(x) + \gamma\left(\mathrm{tr}\, V + w\right)/2 \\ \text{subject to} & \begin{bmatrix} V & x \\ x^H & w \end{bmatrix} \succeq 0 \\ & V \text{ is Toeplitz} \end{array}$$

Outline

- Introduction
- Carathéodory-type matrix decompositions
- Structured trace norm penalties
- **Example**
- Duality

Linear sensor array



- r signals s_k arriving from angles θ_k
- linear array of n sensors
- p randomly chosen sensors are used

- output of sensor i is

$$y_i = \sum_{k=1}^{r} d_i(\omega_k) s_k e^{-\mathrm{j}(i-1)\omega_k} \qquad \text{where } \omega_k = \pi \sin\theta_k$$

- two types of sensors, detecting signals in $[-\pi/2, \pi/6]$ or $[-\pi/6, \pi/2]$:

$$d_i(\omega) = \begin{cases} 1 & \text{for } \theta \in [-\pi/2, \pi/6] \text{ or } [-\pi/6, \pi/2] \text{, respectively} \\ 0 & \text{otherwise} \end{cases}$$

Atomic norm formulation

$$\begin{array}{ll} \text{minimize} & \sum_{j=1}^{3} \sum_{k=1}^{r_j} |w_{jk}| \\ \text{subject to} & y_j = \sum_{k=1}^{r_j} v_{jk} w_{jk}, \quad v_{jk} \in \mathcal{A}_j, \quad j = 1,2,3 \\ & (y_1 + y_2)_{I_1} = b_1, \quad (y_2 + y_3)_{I_2} = b_2 \end{array}$$

- three sets $\mathcal{A}_j$, for three sectors $\theta \in [-\frac{\pi}{2}, -\frac{\pi}{6}]$, $[-\frac{\pi}{6}, \frac{\pi}{6}]$, $[\frac{\pi}{6}, \frac{\pi}{2}]$:

$$\mathcal{A}_j = \{(1, e^{\mathrm{j}\omega}, e^{\mathrm{j}2\omega}, \ldots, e^{\mathrm{j}(n-1)\omega}) \mid |\omega - \alpha_j| \leq \beta_j\}, \quad j = 1,2,3$$

- variables y_1, y_2, y_3 are n-vectors (signals at n sensors from the 3 sectors)
- index sets I_1 and I_2 contain indices of used sensor outputs of type 1, 2
- b_1 and b_2 are measurements (assumed exact for simplicity)

Semidefinite formulation

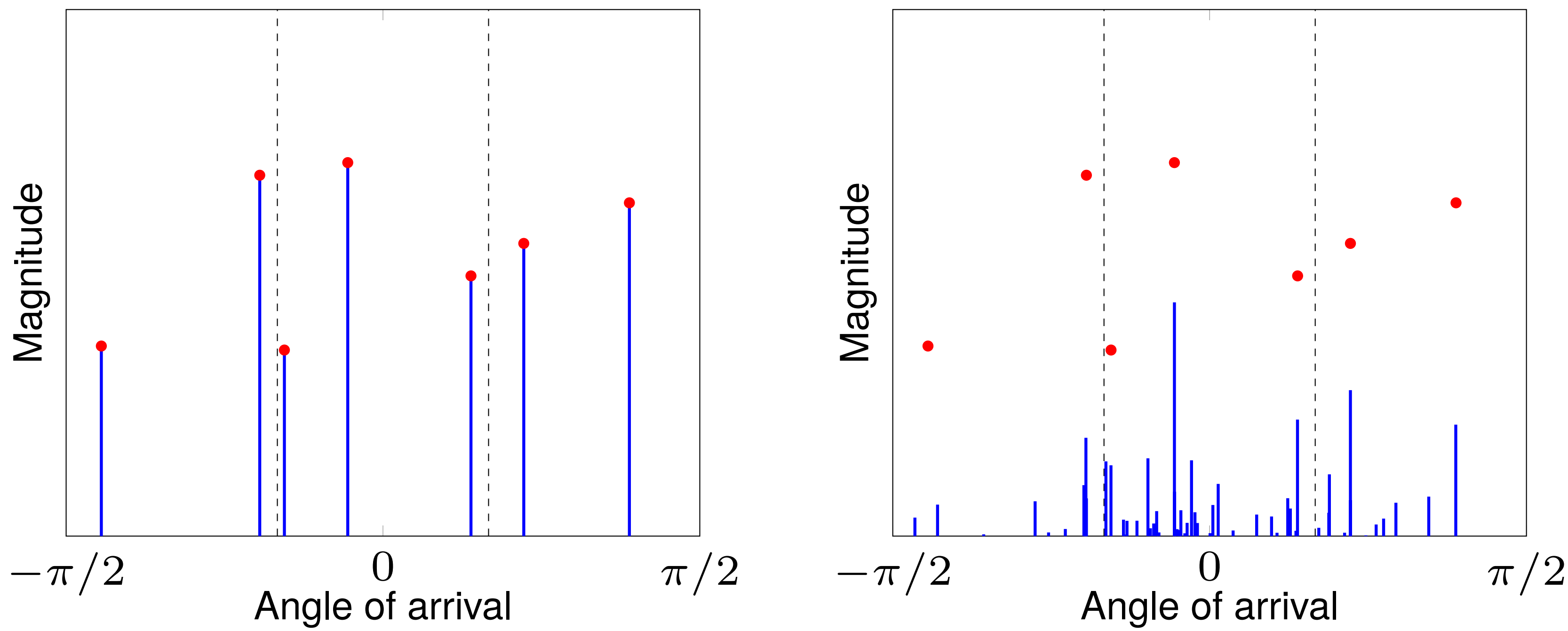
$$\begin{array}{ll} \text{minimize} & \sum_{j=1}^{3} \sum_{k=1}^{r_j} |w_{jk}| \\ \text{subject to} & y_j = \sum_{k=1}^{r_j} v_{jk} w_{jk}, \quad v_{jk} \in \mathcal{A}_j, \quad j = 1,2,3 \\ & (y_1 + y_2)_{I_1} = b_1, \quad (y_2 + y_3)_{I_2} = b_2 \end{array}$$

Equivalent SDP

$$\begin{array}{ll} \text{minimize} & \sum_{j=1}^{3} (\mathrm{tr}\, V_j + w_j)/2 \\ \text{subject to} & \left[\begin{array}{cc} V_j & y_j \\ y_j^H & w_j \end{array}\right] \succeq 0 \\ & FV_jF^H = GV_jG^H, \quad j = 1,2,3 \quad (V_j \text{ is Toeplitz}) \\ & -e^{-\mathrm{j}\alpha_j} FV_jG^H - e^{\mathrm{j}\alpha_j} GV_jF^H + 2\cos\beta_j GV_jG^H \preceq 0, \quad j = 1,2,3 \\ & (y_1 + y_2)_{I_1} = b_1, \quad (y_2 + y_3)_{I_2} = b_2 \end{array}$$

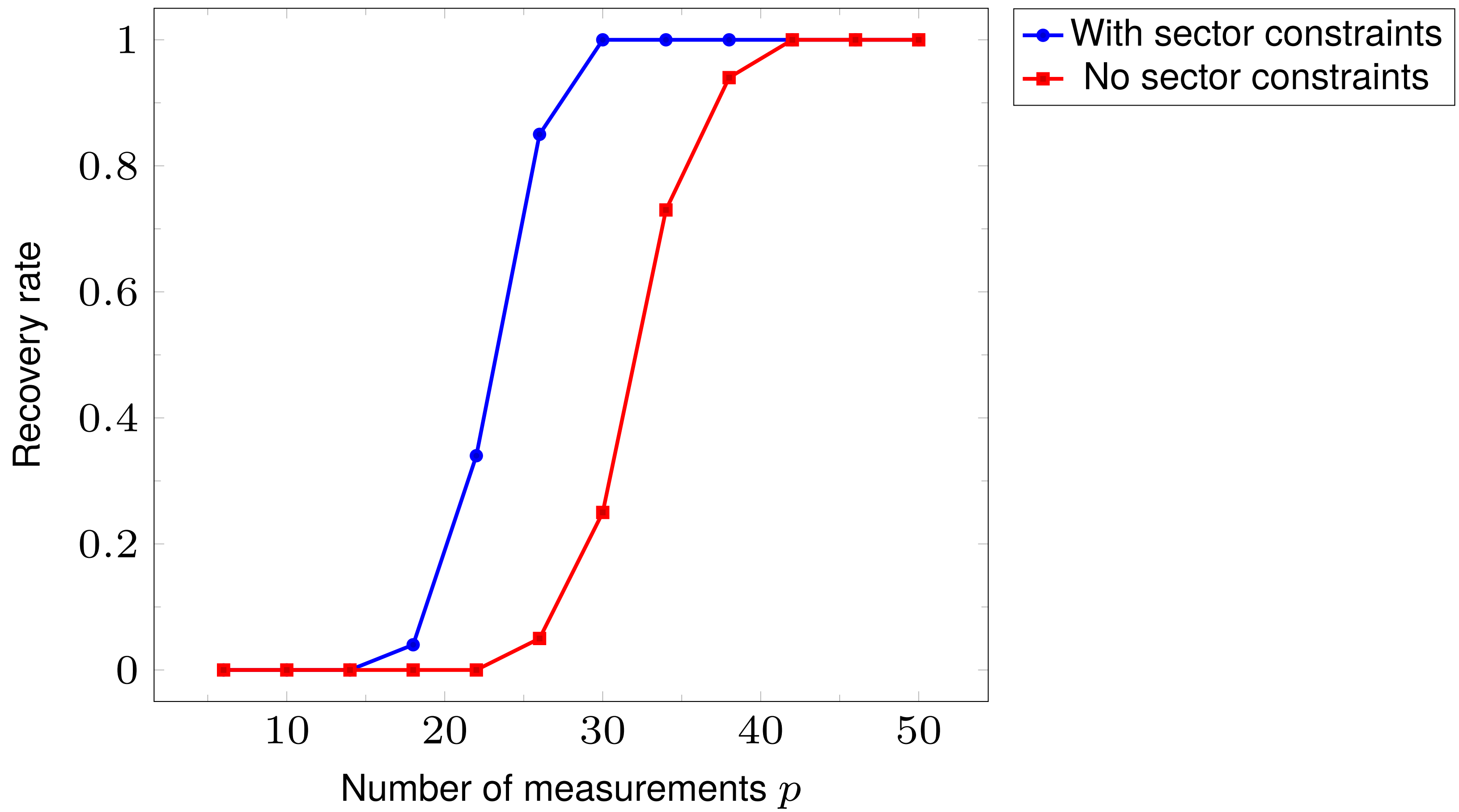
Example

$n = 500$ sensors, 20 sensors used of each type, 7 sources



- red: exact solution
- blue (left): solution from SDP with sector information
- blue (right): solution from SDP omitting sector constraints

Exact recovery



$n = 50$ sensors; 7 sources; p sensor measurements used

Outline

- Introduction
- Carathéodory-type matrix decompositions
- Structured trace norm penalties
- Examples
- **Duality**

Regularization with structured trace penalty

$$
\begin{array}{ll}
\text{minimize} & f(X) + \sum_{k=1}^{r} \|a_k\|^2 \\
\text{subject to} & X = \sum_{k=1}^{r} a_k a_k^H \\
& a_k \in \mathcal{A}
\end{array}
$$

- f a convex function of a Hermitian matrix variable X
- $\mathcal{A} = \{a \in \mathbf{C}^n \mid (\mu G - \nu F)a = 0,\ (\mu, \nu) \in \mathcal{C}_{\Phi\Psi}\}$

Semidefinite formulation

$$
\begin{array}{ll}
\text{minimize} & f(X) + \operatorname{tr} X \\
\text{subject to} & \Phi_{11} F X F^H + \Phi_{21} F X G^H + \Phi_{12} G X F^H + \Phi_{22} G X G^H = 0 \\
& \Psi_{11} F X F^H + \Psi_{21} F X G^H + \Psi_{12} G X F^H + \Psi_{22} G X G^H \preceq 0 \\
& X \succeq 0
\end{array}
$$

Dual problem

Dual of semidefinite formulation

$$
\begin{array}{ll}
\text{maximize} & -f^*(-Z) \\
\text{subject to} & Z - \left[\begin{array}{c} F \\ G \end{array}\right]^H (\Phi \otimes P + \Psi \otimes Q) \left[\begin{array}{c} F \\ G \end{array}\right] \preceq I \\
& Q \succeq 0
\end{array}
$$

- variables: $Z \in \mathbf{H}^n$, P, $Q \in \mathbf{H}^p$
- f^* is conjugate function of f

Interpretation: a problem with infinitely many constraints

$$
\begin{array}{ll}
\text{maximize} & -f^*(-Z) \\
\text{subject to} & a^H Z a \leq 1 \quad \text{for all } a \in \mathcal{A}, \quad \|a\| = 1
\end{array}
$$

equivalence follows from generalized Kalman-Yakubovich-Popov lemma

(Iwasaki and Hara 2005)

Example

Primal problem

$$\begin{array}{ll}
\text{minimize} & f(X) + \sum_{k=1}^{r} |c_k|^2 \\
\text{subject to} & X = \sum_{k=1}^{r} |c_k|^2 \begin{bmatrix} 1 \\ e^{\mathrm{j}\omega_k} \\ \vdots \\ e^{\mathrm{j}(n-1)\omega_k} \end{bmatrix} \begin{bmatrix} 1 \\ e^{\mathrm{j}\omega_k} \\ \vdots \\ e^{\mathrm{j}(n-1)\omega_k} \end{bmatrix}^H \quad \text{with } |\omega_k - \alpha| \leq \beta
\end{array}$$

Dual problem

$$\begin{array}{ll}
\text{maximize} & -f^*(-Z) \\
\text{subject to} & \frac{1}{n} \begin{bmatrix} 1 \\ e^{\mathrm{j}\omega} \\ \vdots \\ e^{\mathrm{j}(n-1)\omega} \end{bmatrix}^H Z \begin{bmatrix} 1 \\ e^{\mathrm{j}\omega} \\ \vdots \\ e^{\mathrm{j}(n-1)\omega} \end{bmatrix} \leq 1 \quad \text{for } |\omega - \alpha| \leq \beta
\end{array}$$

Summary

- atomic norm of sets of matrices with rows/columns chosen from

$$\mathcal{A} = \{a \mid (\lambda G - F)a = 0,\ \lambda \in \mathcal{C}_{\Phi\Psi},\ \|a\|_2 = 1\}$$

  $\mathcal{C}_{\Phi\Psi}$ is segment (interval) of circle or line in the complex plane

- SDP representations based on results for KYP lemma, *i.e.*, for matrix pencil

$$\lambda G - F = \left[\begin{array}{cc} \lambda I - A & B \end{array}\right]$$

- customized interior-point algorithms handle these types of constraints (typically, with complexity $O(n^3)$ instead of $O(n^4)$)

Reference: H.-H. Chao, L. Vandenberghe, *Semidefinite representations of gauge functions for structured low-rank matrix decomposition*, arXiv:1604:02500