# Primal-dual first-order methods for conic optimization

Lieven Vandenberghe

Department of Electrical and Computer Engineering, UCLA

Midwest Optimization Meeting

University of Waterloo, November 9, 2024

# Conic linear program

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax - b \in L \\ & x \in K \end{array}$$

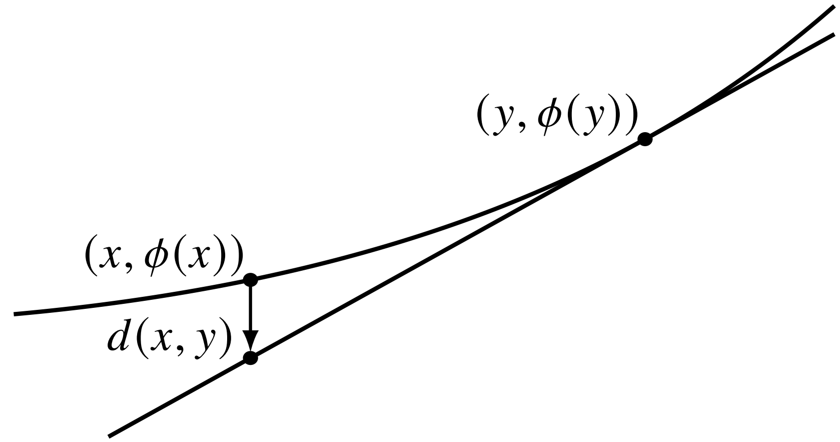
- $K, L$ are closed convex cones
- widely used for convex optimization modeling since 1990s
- often solved via IPMs for second order cone and semidefinite optimization

## Primal-dual proximal methods

- avoid cost of assembling and solving "Schur complement" system in IPMs
- per-iteration complexity dominated by application of $A, A^T$, projections on $K, L^*$
- use of non-Euclidean (Bregman) projections can further reduce complexity

# Bregman distance

$$d(x,y) = \phi(x) - \phi(y) - \nabla\phi(y)^T(x-y)$$




- $\phi$ is the *kernel function*, convex and continuously differentiable on $\mathrm{int}\,(\mathrm{dom}\,\phi)$
- squared Euclidean distance

$$d(x,y) = \frac{1}{2}\|x-y\|_2^2, \qquad \phi(x) = \frac{1}{2}\|x\|_2^2$$

- relative entropy

$$d(x,y) = \sum_{i=1}^n (x_i \log(x_i/y_i) - x_i + y_i), \qquad \phi(x) = \sum_{i=1}^n x_i \log x_i, \qquad \mathrm{dom}\,\phi = \mathbf{R}_+^n$$

# Generalized (Bregman) proximal mapping

- for $d(x,y) = \frac{1}{2}\|x-y\|_2^2$, the proximal mapping is defined as

$$\mathrm{prox}_f(y-a) = \operatorname*{argmin}_x \left(f(x) + \frac{1}{2}\|x-y+a\|_2^2\right)$$

- the proximal mapping for a generalized distance maps $a$ and $y \in \mathrm{int}(\mathrm{dom}\,\phi)$ to

$$\operatorname*{argmin}_x \left(f(x) + a^T x + d(x,y)\right) = \operatorname*{argmin}_x \left(f(x) + (a - \nabla\phi(y))^T x + \phi(x)\right)$$

**Requirements**

- for all $y \in \mathrm{int}(\mathrm{dom}\,\phi)$ and all $a$, unique minimizer exists in $\mathrm{int}(\mathrm{dom}\,\phi)$
- minimizer is inexpensive to compute
- convergence results often assume $\phi$ is strongly convex on $\mathrm{dom}\,f$:

$$d(x,y) \geq \frac{1}{2}\|x-y\|^2 \quad \text{for all } x, y \in \mathrm{dom}\,f$$

# Generalized projection on cone

suppose $K$ is a proper (regular) cone and $S$ is the "simplex"

$$S = \{x \in K \mid \hat{s}^T x \leq 1\}$$

where $\hat{s} \in \operatorname{int}(K^*)$

- generalized projection on $S$ maps $y \in \operatorname{int}(\operatorname{dom} \phi)$ and $a$ to

$$\operatorname*{argmin}_{x \in S} \left(a^T x + d(x, y)\right)$$

  this is the generalized proximal mapping of indicator function $\delta_S$

- conjugate of indicator function $\delta_S$ is Minkowski gauge for $K^*$:

$$\begin{aligned} \delta_S^*(h) &= \sup \{h^T x \mid \hat{s}^T x \leq 1, \; x \in K\} \\ &= \inf \{\beta \geq 0 \mid \beta \hat{s} - h \in K^*\} \end{aligned}$$

- non-Euclidean kernel may be strongly convex on $S$, not necessarily on $K$

# Outline

1. Nonnegative trigonometric polynomials
   [Hsiao-Han Chao, LV 2018]

2. Sparse positive semidefinite completable matrices

3. Proximal methods for self-dual LP

# Nonnegative trigonometric polynomials

$$F_x(\omega) = x_0 + \sum_{k=1}^{n} (x_k e^{-\mathrm{j}k\omega} + \bar{x}_k e^{\mathrm{j}k\omega}) \geq 0 \quad \text{for all } \omega \qquad (\mathrm{j} = \sqrt{-1})$$

- coefficients $x = (x_0, \ldots, x_n)$ form a semidefinite-representable convex cone $K$
- for simplicity, we'll assume $x$ is real ($F_x$ is a cosine polynomial)

**Applications**

- source of many SDP applications in signal processing since 1990s
- 2010s: applications to grid-free compressed sensing
- via transformation $t = \cos\omega$, a nonnegative polynomial in Chebyshev basis
- SDP formulations extend to matrix polynomials, rational (Popov) functions, ...

# Semidefinite representation of $K$ and dual cone $K^*$

$$K = \{D(X) \mid X \in \mathbf{S}^{n+1},\ X \succeq 0\}, \qquad K^* = \{y \in \mathbf{R}^{n+1} \mid T(y) \succeq 0\}$$

- we use the inner product $\langle x, y \rangle = x_0 y_0 + 2x_1 y_1 + \cdots + 2x_n y_n$
- $D : \mathbf{S}^{n+1} \to \mathbf{R}^{n+1}$ maps symmetric matrix $X$ to vector of diagonal sums

$$D(X) = \begin{bmatrix} X_{00} + X_{11} + \cdots + X_{nn} \\ X_{01} + X_{12} + \cdots + X_{n-1,n} \\ \vdots \\ X_{0n} \end{bmatrix}$$

- $T : \mathbf{R}^{n+1} \to \mathbf{S}^{n+1}$ maps vector $(y_0, \ldots, y_n)$ to the symmetric Toeplitz matrix

$$T(y) = \begin{bmatrix} y_0 & y_1 & \cdots & y_n \\ y_1 & y_0 & \cdots & y_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ y_n & y_{n-1} & \cdots & y_0 \end{bmatrix}$$

# Complexity of common algorithms

## Interior-point algorithms

- conic inequality contributes dense term to Schur complement system
- general-purpose interior-point SDP solvers: $O(n^4)$ per iteration
- customized interior-point solvers: $O(n^3)$ per iteration

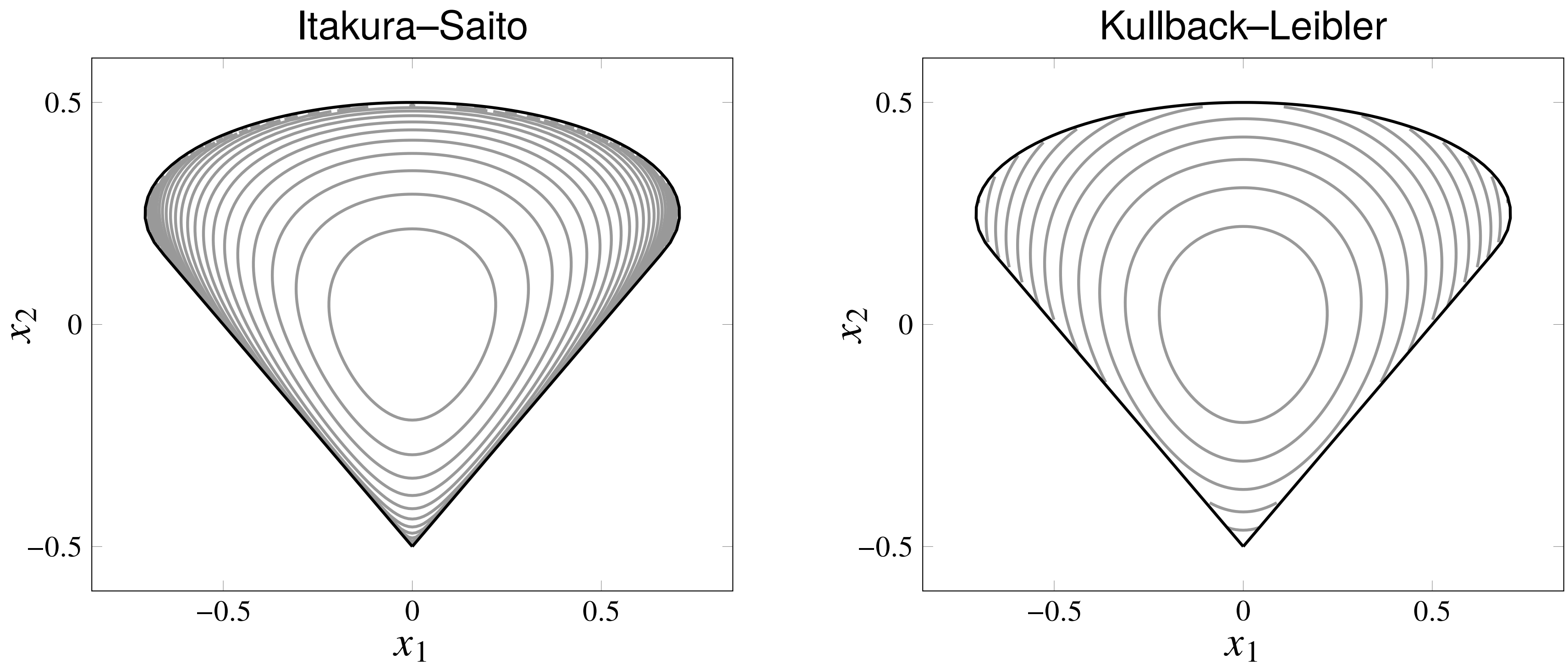
## Proximal algorithms

- for Euclidean norm: $O(n^3)$ per iteration (for projection on p.s.d. cone)
- reduction below $O(n^3)$ requires non-Euclidean distance

# Generalized distances

kernels for Itakura–Saito distance and Kullback–Leibler distance

$$\phi(x) = -\frac{1}{2\pi}\int_0^{2\pi} \log F_x(\omega)\,d\omega \qquad \phi_{\mathrm{kl}}(x) = \frac{1}{2\pi}\int_0^{2\pi} F_x(\omega)\log F_x(\omega)\,d\omega$$




- plots show contour lines of $\phi$ and $\phi_{\mathrm{kl}}$ on section $\{x \in K \mid x_0 = 1\}$
- $\phi$ is essentially smooth; $\phi_{\mathrm{kl}}$ is not

# Semidefinite representation of entropy kernel $\phi$

$$\begin{array}{ll} \text{minimize (over } X\text{)} & -\log X_{00} \\ \text{subject to} & D(X) = x \\ & X \succeq 0 \end{array}$$

- for $x \in K \setminus \{0\}$, optimal value is

$$\phi(x) = -\frac{1}{2\pi} \int_0^{2\pi} \log F_x(\omega) d\omega$$

- optimal $X$ has rank one:

$$X = bb^T, \qquad \phi(x) = -2 \log b_0$$

- $b$ is minimum-phase spectral factor ($b_0 + b_1 z^{-1} + \cdots + b_n z^{-n} \neq 0$ for $|z| > 1$)
- $b$ is efficiently computed by spectral factorization of $x$: solve quadratic equation

$$D(bb^T) = x$$

# Dual of semidefinite representation of $\phi$

$$\text{maximize (over } y) \quad -\psi(y) - \langle x, y \rangle + 1$$

- convex function $\psi$ is defined as

$$\psi(y) = \log(e^T T(y)^{-1} e), \qquad \operatorname{dom} \psi = \{y \mid T(y) \succ 0\} = \operatorname{int}(K^*)$$

  where $e = (1, 0, \ldots, 0)$

- by duality, optimal value is $\phi(x)$
- optimal $y$ is $y = -\nabla \phi(x)$, and related to primal solution $X = bb^T$ as

$$T(y)b = e$$

  $y$ can be computed from spectral factor $b$ by reverse Levinson algorithm

# Itakura–Saito distance and projection

$$d(x,y) = \frac{1}{2\pi}\int_0^{2\pi} \left(\frac{F_x(\omega)}{F_y(\omega)} - \log\frac{F_x(\omega)}{F_y(\omega)} - 1\right) d\omega$$

where $F_x(\omega) = x_0 + 2x_1 \cos\omega + \cdots + 2x_n \cos n\omega$

- proposed in 1970s as spectral distance measure in speech processing
- generalized projection on $H = \{x \in K \mid \langle e, x\rangle = 1\}$ (where $\langle e, x\rangle = x_0$):

$$\operatorname*{argmin}_{x_0=1} \left(\langle a, x\rangle + d(x,y)\right) = \operatorname*{argmin}_{x_0=1} \left(\langle a - \nabla\phi(y), x\rangle + \phi(x)\right)$$

- dual problem (scalar variable $\lambda$ is multiplier for constraint $x_0 = 1$)

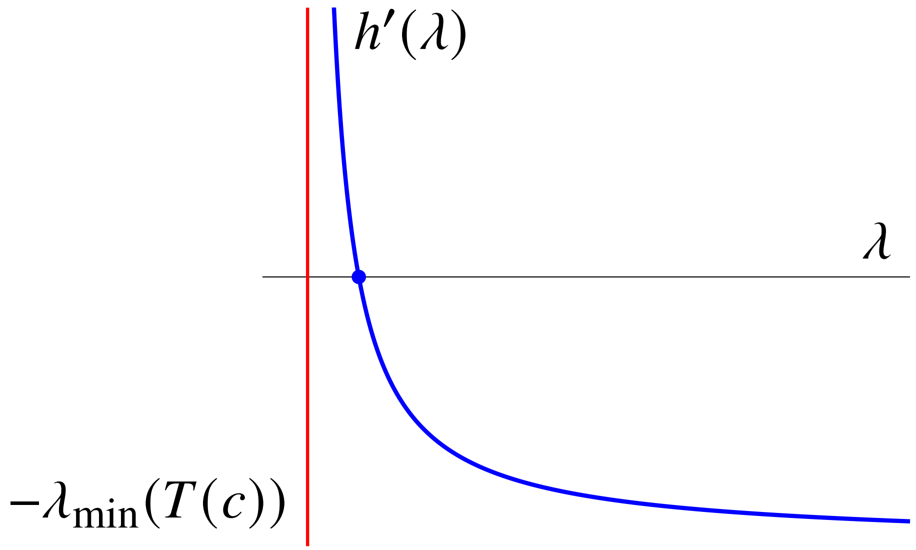
$$\text{maximize} \quad -\log\left(e^T (T(c) + \lambda I)^{-1} e\right) - \lambda \qquad \text{(where } c = a - \nabla\phi(y))$$

$e^T(T(c) + \lambda I)^{-1} e$ is leading element of inverse Toeplitz matrix $T(c + \lambda e)^{-1}$

# Computing Itakura–Saito projection

solve dual problem with scalar variable $\lambda$, for example, by Newton's method

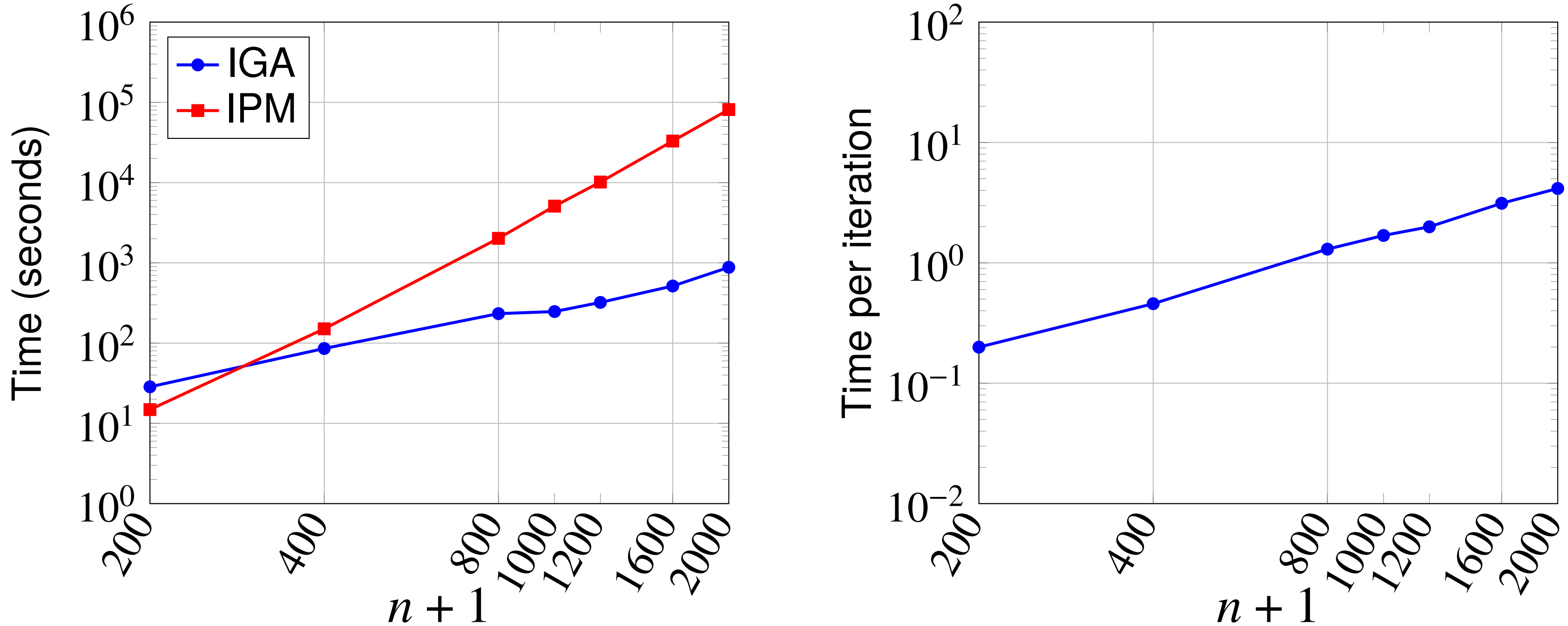
$$\text{maximize} \quad h(\lambda) = -\log(e^T (T(c) + \lambda I)^{-1} e) - \lambda$$




- at each Newton step, factorize positive definite Toeplitz matrix $T(c + \lambda e)$
- complexity: $O(n^2)$ with Levinson algorithm, $O(n(\log n)^2)$ with superfast solvers
- from optimal $\lambda$, compute solution $x = (1/b_0)D(bb^T)$ where $b = T(c + \lambda e)^{-1} e$

# Euclidean projection

$$\begin{array}{ll} \text{minimize} & \sum_{k=0}^{n} (x_k - a_k)^2 \\ \text{subject to} & x \in K, \quad x_0 = 1 \end{array}$$




- IPM is SDPT3/SeDuMi via CVX
- IGA is Auslender–Teboulle proximal gradient algorithm [2006]
- number of IGA iterations is 100–200 to reach relative accuracy $10^{-4}$
- about 10 Newton steps per projection; Toeplitz solver is Levinson algorithm

# Extensions

- projection on $\{x \in K \mid \langle \hat{s}, x \rangle = 1\}$ or $\{x \in K \mid \langle \hat{s}, x \rangle \leq 1\}$ where $\hat{s} \in K^*$
- nonnegative trigonometric matrix polynomials [Cederberg 2023]

# Outline

1. Nonnegative trigonometric polynomials
2. Sparse positive semidefinite completable matrices
   [Xin Jiang, LV 2022]
3. Proximal methods for self-dual LP

# Sparse semidefinite program

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

- $C, A_1, \ldots, A_m$ are sparse with common sparsity pattern $E$
- without loss of generality, we assume $E$ is chordal (a filled Cholesky pattern)
- optimal $X$ is typically dense, even for sparse coefficients $C, A_1, \ldots, A_m$

## Equivalent conic linear program

$$\begin{array}{ll} \text{minimize} & \mathrm{tr}(CX) \\ \text{subject to} & \mathrm{tr}(A_iX) = b_i, \quad i = 1, \ldots, m \\ & X \in K \end{array}$$

- variable $X$ is a *sparse* matrix with sparsity pattern $E$ (notation: $\mathbf{S}^n_E$)
- $K$ is cone of matrices in $\mathbf{S}^n_E$ that have a positive semidefinite completion

# Bregman distance generated by logarithmic barrier

**Logarithmic barrier** (for cone $K$ of p.s.d. completable matrices)

$$\phi(X) = \sup_{S \in \operatorname{int} K^*} (\log\det S - \operatorname{tr}(XS))$$

- dual cone $K^*$ is cone of positive semidefinite matrices in $\mathbf{S}^n_E$
- optimal $\hat{S}_X$ is inverse of maximum determinant pos. definite completion of $X$

$$\phi(X) = \log\det \hat{S}_X - n, \qquad \nabla\phi(X) = -\hat{S}_X$$

- for chordal $E$: efficient algorithms for computing $\hat{S}_X$ given $X$
- complexity is comparable with sparse Cholesky factorization with pattern $E$

**Bregman distance**

$$d(X,Y) = -\log\det(\hat{S}_Y \hat{S}_X^{-1}) + \operatorname{tr}(\hat{S}_Y \hat{S}_X^{-1}) + n$$

the relative entropy (Kullback–Leibler divergence) between completions $\hat{S}_Y$ and $\hat{S}_X$

# Bregman projection

Bregman projection on $H = \{X \mid \operatorname{tr} X = 1\}$

$$\operatorname*{argmin}_{\operatorname{tr} X=1} (\operatorname{tr}(AX) + d(X,Y)) = \operatorname*{argmin}_{\operatorname{tr} X=1} (\operatorname{tr}(BX) + \phi(X))$$

where $B = A - \nabla\phi(Y)$

- solution is projection $\Pi_E((B+\lambda I)^{-1})$ on $\mathbf{S}^n_E$, where $\lambda$ satisfies

$$\operatorname{tr}((B+\lambda I)^{-1}) = 1, \qquad B + \lambda I \succ 0$$

- easily extended to projection on $\{X \mid \operatorname{tr}(NX) \leq 1\}$ with $N \in K^*$
- similar problem for Bregman proximal operator of centering objective

$$f(X) = \operatorname{tr}(CX) + \mu\phi(X) + \delta_H(X)$$

# Evaluating the Bregman projection

$$\begin{array}{ll} \text{minimize} & \operatorname{tr}(BX) + \phi(X) \\ \text{subject to} & \operatorname{tr} X = 1 \end{array}$$

- use Newton's method to find unique solution $\lambda$ of the nonlinear equation

$$\operatorname{tr}((B + \lambda I)^{-1}) = 1 \qquad \text{(with } B + \lambda I \succ 0\text{)}$$

- from $\lambda$, compute solution $\hat{X} = \Pi_E((B + \lambda I)^{-1})$ on $\mathbf{S}^n_E$
- for chordal sparsity patterns $E$, efficient algorithms exist for computing

$$g(\lambda) = \operatorname{tr}((B + \lambda I)^{-1}), \qquad g'(\lambda) = -\operatorname{tr}((B + \lambda I)^{-2}), \qquad \hat{X} = \Pi_E((B + \lambda I)^{-1})$$

  from sparse Cholesky factorization of $B + \lambda I$

complexity $\approx$ # Newton iterations $\times$ cost of sparse Cholesky factorization

# Maximum-cut problem

$$\begin{array}{ll} \text{maximize} & \operatorname{tr}(LX) \\ \text{subject to} & \operatorname{diag}(X) = \mathbf{1}, \;\; X \succeq 0 \end{array}$$

- compute approximate solution on central path (parameter $\mu = 0.001/n$)
- Bregman variant of PDHG
- four problems from SDPLIB, four graphs from SuiteSparse matrix collection

| | $n$ | time per Cholesky factorization | Newton steps per iteration | time per PDHG iteration | PDHG iterations |
|---|---|---|---|---|---|
| maxG51 | 1000 | 0.05 | 2.45 | 0.12 | 267 |
| maxG32 | 2000 | 0.12 | 1.56 | 0.18 | 240 |
| maxG55 | 5000 | 0.29 | 2.10 | 0.58 | 249 |
| maxG60 | 7000 | 0.60 | 2.55 | 1.22 | 279 |
| barth4 | 6019 | 0.42 | 3.57 | 1.55 | 346 |
| tuma2 | 12992 | 0.48 | 4.36 | 1.89 | 375 |
| biplane-9 | 21701 | 0.95 | 2.58 | 2.12 | 287 |
| c-67 | 57975 | 0.76 | 3.58 | 3.56 | 378 |

# Outline

1. Nonnegative trigonometric polynomials
2. Sparse positive semidefinite completable matrices
3. Proximal methods for self-dual LP

# Primal-dual proximal methods

$$\begin{array}{lll} \text{primal:} & \text{minimize} & c^T x + f(x) + g(b - Ax) \\ \text{dual:} & \text{maximize} & b^T z - g^*(z) - f^*(A^T z - c) \end{array}$$

- $f, g$ are closed convex functions
- conic LP is special case with $f = \delta_K$, $g = \delta_{-L}$ (indicators of cones $K, -L$)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax - b \in L \\ & x \in K \end{array}$$

we discuss methods that evaluate prox-operators of $f, g^*$ and products with $A, A^T$

# PDHG and extragradient method

## Primal-dual hybrid gradient (PDHG) method

$$\begin{aligned} x_{k+1} &= \operatorname{prox}_{\tau f}(x_k + \tau(A^T z_k - c)) \\ z_{k+1} &= \operatorname{prox}_{\sigma g^*}(z_k + \sigma(b - A(2x_{k+1} - x_k)) \end{aligned}$$

[Esser, Zhang, Chan 2010, Pock, Cremers, Bischof, Chambolle 2009, Chambolle & Pock 2011, ...]

- dual variant applies this iteration to the dual problem
- recently used for large-scale linear programming [Applegate *et al.* 2021, 2023]

## Proximal generalization of extragradient method

$$\begin{aligned} \bar{x}_k &= \operatorname{prox}_{\tau f}(x_k + \tau(A^T z_k - c)) \\ \bar{z}_k &= \operatorname{prox}_{\sigma g^*}(z_k + \sigma(b - A x_k)) \\ x_{k+1} &= \operatorname{prox}_{\tau f}(x_k + \tau(A^T \bar{z}_k - c)) \\ z_{k+1} &= \operatorname{prox}_{\sigma g^*}(z_k + \sigma(b - A\bar{x}_k)) \end{aligned}$$

# Extensions with generalized distances

use primal and dual distances $d_{\mathrm{p}}, d_{\mathrm{d}}$ generated by kernels $\phi_{\mathrm{p}}, \phi_{\mathrm{d}}$

**PDHG** [Chambolle and Pock 2016]

$$
\begin{aligned}
x_{k+1} &= \operatorname*{argmin}_x \left(c^T x + f(x) - z_k^T A x + \frac{1}{\tau} d_{\mathrm{p}}(x, x_k)\right) \\
z_{k+1} &= \operatorname*{argmin}_z \left(-b^T z + g^*(z) + z^T A(2x_{k+1} - x_k) + \frac{1}{\sigma} d_{\mathrm{d}}(z, z_k)\right)
\end{aligned}
$$

**Extragradient method** [Nemirovski 2004, Auslender & Teboulle 2005, Tseng 2008]

$$
\begin{aligned}
\bar{x}_k &= \operatorname*{argmin}_x \left(c^T x + f(x) - z_k^T A x + d_{\mathrm{p}}(x, x_k)/\tau\right) \\
\bar{z}_k &= \operatorname*{argmin}_z \left(-b^T z + g^*(z) + z^T A x_k + d_{\mathrm{d}}(z, z_k)/\sigma\right) \\
x_{k+1} &= \operatorname*{argmin}_x \left(c^T x + f(x) - \bar{z}_k^T A x + d_{\mathrm{p}}(x, x_k)/\tau\right) \\
z_{k+1} &= \operatorname*{argmin}_z \left(-b^T z + g^*(z) + z^T A \bar{x}_k + d_{\mathrm{d}}(z, z_k)/\sigma\right)
\end{aligned}
$$

# Condat–Vũ algorithm

$$\text{minimize} \quad h(x) + f(x) + g(b - Ax)$$

additional term $h$ is differentiable convex function

## Algorithm

$$\begin{aligned} x_{k+1} &= \operatorname*{argmin}_x \left(\nabla h(x_k)^T x + f(x) - z_k^T A x + \frac{1}{\tau} d_\mathrm{p}(x, x_k)\right) \\ z_{k+1} &= \operatorname*{argmin}_z \left(-b^T z + g^*(z) + z^T A(2x_{k+1} - x_k) + \frac{1}{\sigma} d_\mathrm{d}(z, z_k)\right) \end{aligned}$$

- Bregman generalization of Condat–Vũ algorithm [Condat 2013, Vũ 2013]
- three-term extension of PDHG (the special case with $h(x) = c^T x$)

[Xin Jiang, LV 2022]

# Derivation of Bregman Condat–Vũ from proximal point method

apply Bregman proximal point method to optimality conditions

$$0 \in -A^T z + \partial f(x) + \nabla h(x), \qquad 0 \in Ax - b + \partial g^*(z)$$

with distance generated by

$$\phi_{\mathrm{pd}}(x, z) = \frac{1}{\tau}\phi_{\mathrm{p}}(x) + \frac{1}{\sigma}\phi_{\mathrm{d}}(z) + z^T Ax - h(x)$$

**Stepsize conditions:** $\phi_{\mathrm{pd}}$ is convex if the following assumptions hold

- $\phi_{\mathrm{p}}$, $\phi_{\mathrm{d}}$ are $1$-strongly convex with respect to norms $\|\cdot\|_{\mathrm{p}}$, $\|\cdot\|_{\mathrm{d}}$
- the function $L\phi_{\mathrm{p}} - h$ is convex
- $\sigma\tau\|A\|^2 + \tau L \leq 1$ where $\|A\|$ is the matrix norm induced by $\|\cdot\|_{\mathrm{p}}$, $\|\cdot\|_{\mathrm{d}}$

this extends the PPA interpretation of PDHG [He & Yuan 2012]

# Application to conic LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax - b \in L \\ & x \in K \end{array} \qquad \begin{array}{ll} \text{maximize} & b^T z \\ \text{subject to} & -A^T z + c \in K^* \\ & z \in L^* \end{array}$$

- minimize $c^T x + f(x) + g(b - Ax)$ with $f = \delta_K$, $g = \delta_{-L}$
- primal-dual methods require (generalized) projections on $K$, $L^*$
- iterates satisfy $x_k \in K$, $z_k \in L^*$, not $Ax_k - b \in L$ and $-A^T z_k + c \in K^*$

use of Bregman distances requires additional bounds on $x, z$

- often needed for well-defined generalized projections
- convergence results assume strong convexity of Bregman kernels
- stopping conditions, convergence results simplify if iterates are feasible

# Self-dual problem

$$\begin{array}{lll} \text{primal:} & \text{minimize} & c^T x + f(x) + g(b - Ax) \\ \text{dual:} & \text{maximize} & b^T z - g^*(z) - f^*(A^T z - c) \end{array}$$

- if $b = -c$, $g = f^*$, $A = -A^T$ problem is self-dual and can be written as

$$\text{minimize} \quad -b^T x + f(x) + f^*(b - Ax)$$

- special case with $f = \delta_K$ is self-dual conic LP

$$\begin{array}{ll} \text{minimize} & -b^T x \\ \text{subject to} & Ax - b \in K^* \\ & x \in K \end{array}$$

[Duffin 1956]

- if strictly consistent, optimal value is zero, optimal set is nonempty and bounded

[Rockafellar 1970, Rockafellar and Wets 1998]

# Exact penalty formulation of self-dual LP

$$\begin{array}{ll} \text{minimize} & -b^T x \\ \text{subject to} & Ax - b \in K^* \qquad \text{(with } A = -A^T\text{)} \\ & x \in K \end{array}$$

- we assume (for simplicity) that $K$ is a proper cone
- assume that a point $\hat{x} \in \mathrm{int}(K)$ is known that satisfies $\hat{s} = A\hat{x} - b \in \mathrm{int}(K^*)$
- for example, constructed via Ye–Todd–Mizuno embedding of general LP

**Self-dual exact penalty formulation**

$$\text{minimize} \quad -b^T x + f(x) + f^*(b - Ax)$$

- $f$ is indicator function of $\{x \in K \mid \hat{s}^T x \leq \mu\}$ where $\mu > \hat{s}^T \hat{x}$
- $f^*(y) = \inf \{\beta \geq 0 \mid \beta \hat{s} - \mu y \in K^*\}$ is Minkowski gauge for $K^*$
- $f^*(b - Ax)$ is an exact penalty for constraint $Ax - b \in K^*$

# Primal-dual methods for self-dual problem

$$\text{minimize} \quad -b^T x + f(x) + f^*(b - Ax) \qquad (\text{with } A = -A^T)$$

- if $x_0 = z_0$, $d_\mathrm{p} = d_\mathrm{d}$, $\sigma_k = \tau_k$ the four steps in extragradient algorithm reduce to

$$\begin{aligned}
\bar{x}_k &= \operatorname*{argmin}_x \left( f(x) + (Ax_k - b)^T x + \frac{1}{\tau_k} d(x, x_k) \right) \\
x_{k+1} &= \operatorname*{argmin}_x \left( f(x) + (A\bar{x}_k - b)^T x + \frac{1}{\tau_k} d(x, x_k) \right)
\end{aligned}$$

- for the exact penalty formulation of the self-dual LP

$$\begin{aligned}
\bar{x}_k &= \operatorname*{argmin}_{x \in K,\ \hat{s}^T x \leq \mu} \left( (Ax_k - b)^T x + \frac{1}{\tau_k} d(x, x_k) \right) \\
x_{k+1} &= \operatorname*{argmin}_{x \in K,\ \hat{s}^T x \leq \mu} \left( (A\bar{x}_k - b)^T x + \frac{1}{\tau_k} d(x, x_k) \right)
\end{aligned}$$

- PDHG iterations are similar

# Convergence result

- assume kernel is $1$-strongly convex on $\operatorname{dom} f$ with respect to norm $\|\cdot\|$
- $\tau_k = 1/\|A\|$ where $A$ is matrix norm induced by $\|\cdot\|$, or determined adaptively
- define $\mathcal{L}(y,x) = b^T(x-y) + f(y) - f(x) - y^T Ax$
- define averaged iterate

$$y_k = \frac{1}{\sum_{i=0}^k \tau_i} \sum_{i=0}^k \tau_k \bar{x}_i$$

- adapting results for extragradient method to self-dual problem [Tseng 2008]:

$$\mathcal{L}(y_k, x) \leq \frac{d(x, x_0)}{\sum_{i=0}^k \tau_i} \quad \text{for all } x \in \operatorname{dom} f$$

- maximizing over $x \in \operatorname{dom} f$ gives

$$-b^T y_k + f(y_k) + f^*(b - Ay_k) \leq \frac{1}{\sum_{i=0}^k \tau_i} \sup_{x \in \operatorname{dom} f} d(x, x_0)$$

# Summary

**Bregman projections for two classes of SDP-representable cones**

- Itakura–Saito distance for nonnegative trigonometric polynomials

  cost of generalized projection is roughly $O(n^2)$

- distance generated by logarithmic barrier of p.s.d. completable sparse matrices

  cost roughly on the same order as sparse Cholesky factorization

**Primal-dual proximal methods for conic LP**

- adding bounds to conic constraints is important for several reasons
- self-dual embedding provides useful bounding inequality constraint
- can be interpreted as solving self-dual exact penalty reformulation