5. Primal-dual interior-point methods

- cone programming
- logarithmic barrier function and central path
- symmetrization and Nesterov-Todd scaling
- path-following algorithm
- quadratic cone program
- self-dual embedding

Cone program

primal problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Gx + s = h \\ & Ax = b \\ & s \succeq 0 \end{array}$$

$s \succeq 0$ is inequality with respect to proper convex cone C

dual problem

$$\begin{array}{ll} \text{maximize} & -h^T z - b^T y \\ \text{subject to} & A^T y + G^T z + c = 0 \\ & z \succeq_* 0 \end{array}$$

$z \succeq_* 0$ is generalized inequality with respect to dual cone C^*

Primal-dual path-following methods

closely related to **barrier methods**

- methods follow central path to find approximate optimal points
- steps are computed by linearizing central path equations

incorporate several **modifications** to improve efficiency and robustness

- symmetric treatment of primal and dual iterates
- barrier parameter is updated after each Newton step; no distinction between outer and inner iterations
- more aggressive step sizes
- infeasible starting points
- higher-order approximation of central path
- (some algorithms) detect infeasibility

Three standard cones

we will limit the discussion to three types of cones

nonnegative orthant

$$\{u \in \mathbf{R}^p \mid u_i \geq 0, i = 1, \dots, p\}$$

second-order cone

$$\{(u_0, u_1) \in \mathbf{R} \times \mathbf{R}^{p-1} \mid \|u_1\|_2 \leq u_0\}$$

semidefinite cone

$$\{u \in \mathbf{R}^{p(p+1)/2} \mid \mathbf{mat}(u) \succeq 0\}$$

($\mathbf{mat}(u)$ is symmetric matrix of order p constructed from u; see below)

these cones are self-dual, so we can drop the subscript in $z \succeq_* 0$

Notation for symmetric matrices

symmetric matrix as a vector

if

$$U = \begin{bmatrix} U_{11} & U_{21} & \cdots & U_{p1} \\ U_{21} & U_{22} & \cdots & U_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ U_{p1} & U_{p2} & \cdots & U_{pp} \end{bmatrix} \in \mathbf{S}^p,$$

then

$$\mathbf{vec}(U) = \sqrt{2}\left(\frac{U_{11}}{\sqrt{2}}, U_{21}, \ldots, U_{p1}, \frac{U_{22}}{\sqrt{2}}, U_{32}, \ldots, U_{p2}, \ldots, \frac{U_{pp}}{\sqrt{2}}\right)$$

- $\mathbf{vec}(U) \in \mathbf{R}^{p(p+1)/2}$ are lower triangular entries in column-major order
- off-diagonal entries are scaled by $\sqrt{2}$ so that

$$\mathbf{tr}(UV) = \mathbf{vec}(U)^T \mathbf{vec}(V)$$

vector as symmetric matrix

if $u = \big(u_1, u_2, \ldots, u_{p(p+1)/2}\big) \in \mathbf{R}^{p(p+1)/2}$, then

$$\mathbf{mat}(u) = \frac{1}{\sqrt{2}} \begin{bmatrix} \sqrt{2}u_1 & u_2 & \cdots & u_p \\ u_2 & \sqrt{2}u_{p+1} & \cdots & u_{2p-1} \\ \vdots & \vdots & & \\ u_p & u_{2p-1} & \cdots & \sqrt{2}u_{p(p+1)/2} \end{bmatrix}$$

- $\mathbf{mat}(u) \in \mathbf{S}^p$ is the symmetric matrix of order p constructed from u
- off-diagonal entries are scaled by $\sqrt{2}$ so that

$$u^T v = \mathbf{tr}(\mathbf{mat}(u)\,\mathbf{mat}(v))$$

semidefinite constraints in vector notation

- suppose C is the cone of vectorized positive semidefinite matrices

$$C = \{u \in \mathbf{R}^{p(p+1)/2} \mid \mathbf{mat}(u) \succeq 0\} = \{\mathbf{vec}(U) \mid U \in \mathbf{S}^p_+\}$$

- for $G_i \in \mathbf{S}^p$, $h \in \mathbf{S}^p$, define

$$G = \begin{bmatrix} \mathbf{vec}(G_1) & \mathbf{vec}(G_2) & \cdots & \mathbf{vec}(G_n) \end{bmatrix}, \qquad h = \mathbf{vec}(H)$$

- $Gx \preceq_C h$ is equivalent to a linear matrix inequality

$$\sum_{i=1}^{n} x_i G_i \preceq H$$

- $G^T z + c = 0$, $z \succeq_C 0$ is equivalent to

$$\mathbf{tr}(G_i Z) + c_i = 0, \quad i = 1, \ldots, n, \qquad Z = \mathbf{mat}(z) \succeq 0$$

Optimality conditions

$$\begin{bmatrix} 0 \\ 0 \\ s \end{bmatrix} = \begin{bmatrix} 0 & A^T & G^T \\ -A & 0 & 0 \\ -G & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} c \\ b \\ h \end{bmatrix} \qquad (5.1)$$

$$(s, z) \succeq 0, \qquad s^T z = 0$$

primal feasibility: blocks 2 and 3 of (5.1) and $s \succeq 0$

dual feasibility: block 1 of (5.1) and $z \succeq 0$

zero duality gap: inner product with (x, y, z) on each side of (5.1) gives

$$z^T s = c^T x + b^T y + h^T z$$

Outline

- cone programming
- **logarithmic barrier function and central path**
- symmetrization and Nesterov-Todd scaling
- path-following algorithm
- quadratic cone program
- self-dual embedding

Barrier function for nonnegative orthant

$$C = \{u \in \mathbf{R}^p \mid u_i \geq 0, i = 1, \ldots, p\}$$

barrier function

$$\phi(u) = -\sum_{i=1}^{p} \log u_i$$

gradient

$$\nabla \phi(u) = -\left(\frac{1}{u_1}, \frac{1}{u_2}, \ldots, \frac{1}{u_p}\right)$$

Hessian and inverse Hessian

$$\nabla^2 \phi(u) = \mathbf{diag}(u)^{-2}, \qquad \nabla^2 \phi(u)^{-1} = \mathbf{diag}(u)^2$$

Barrier function for second-order cone

$$C = \left\{ (u_0, u_1) \in \mathbf{R} \times \mathbf{R}^{p-1} \;\middle|\; \|u_1\|_2 \le u_0 \right\}$$

barrier function

$$\phi(u) = -\frac{1}{2} \log(u_0^2 - u_1^T u_1)$$

gradient

$$\nabla \phi(u) = -\frac{1}{u^T J u} J u, \qquad J = \begin{bmatrix} 1 & 0 \\ 0 & -I_{p-1} \end{bmatrix}$$

Hessian and inverse Hessian

$$\begin{aligned} \nabla^2 \phi(u) &= \frac{1}{(u^T J u)^2} \left(2 J u u^T J - (u^T J u) J \right) \\ \nabla^2 \phi(u)^{-1} &= 2 u u^T - (u^T J u) J \end{aligned}$$

Barrier function for positive semidefinite cone

$$C = \left\{ u \in \mathbf{R}^{p(p+1)/2} \;\middle|\; \mathbf{mat}\, u \succeq 0 \right\}$$

barrier function

$$\phi(u) = -\log\det \mathbf{mat}(u)$$

gradient

$$\nabla\phi(u) = -\mathbf{vec}(\mathbf{mat}(u)^{-1})$$

Hessian

$$\nabla^2\phi(u)v = \mathbf{vec}\left(U^{-1}VU^{-1}\right), \qquad \nabla^2\phi(u)^{-1}v = \mathbf{vec}\left(UVU\right)$$

for all v, where $U = \mathbf{mat}(u)$, $V = \mathbf{mat}(v)$

Barrier function for composite cones

$$C = C_1 \times \cdots \times C_N$$

C_k is one of the three standard cones, with barrier ϕ_k

barrier at $u = (u_1, \ldots, u_N)$

$$\phi(u) = \sum_{k=1}^{N} \phi_k(u_k)$$

gradient

$$\nabla \phi(u) = (\nabla \phi_1(u_1), \ldots, \nabla \phi_N(u_N))$$

Hessian

$$\nabla^2 \phi(u) = \begin{bmatrix} \nabla^2 \phi_1(u_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \nabla^2 \phi_N(u_N) \end{bmatrix}$$

Degree of barrier function

the barrier functions satisfy

$$\phi(tu) = \phi(u) - m \log t \quad \text{for } t > 0$$

m is called the logarithmic *degree*

- $m = p$ for nonnegative orthant of dimension p
- $m = 1$ for second-order cone
- $m = p$ for semidefinite cone of order p
- for composite cone, m is sum of degrees of the components

property of gradient

$$\nabla \phi(u) \prec 0, \qquad u^T \nabla \phi(u) = -m$$

Central path

$$\begin{bmatrix} 0 \\ 0 \\ s \end{bmatrix} = \begin{bmatrix} 0 & A^T & G^T \\ -A & 0 & 0 \\ -G & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} c \\ b \\ h \end{bmatrix}$$

$$(s, z) \succ 0, \qquad z = -\mu \nabla \phi(s)$$

- central path is set of solutions s, x, y, z for $\mu > 0$
- points s, z on the central path satisfy

$$\mu = \frac{s^T z}{m}$$

Outline

- cone programming
- logarithmic barrier function and central path
- **symmetrization and Nesterov-Todd scaling**
- path-following algorithm
- quadratic cone program
- self-dual embedding

Symmetrization

we will write the condition $z = -\mu \nabla \phi(s)$ as a product $s \circ z = \mu\, \mathbf{e}$

nonnegative orthant (of dimension p)

$$u \circ v = (u_1 v_1, u_2 v_2, \ldots, u_p v_p), \qquad \mathbf{e} = (1, 1, \ldots, 1)$$

second-order cone (of dimension p): for $u = (u_0, u_1)$, $v = (v_0, v_1)$,

$$u \circ v = \begin{bmatrix} u^T v \\ u_0 v_1 + v_0 u_1 \end{bmatrix}, \qquad \mathbf{e} = (1, 0, \ldots, 0)$$

semidefinite cone (of order p): for $U = \mathbf{mat}(u)$, $V = \mathbf{mat}(v)$,

$$u \circ v = \frac{1}{2} \mathbf{vec}\left(UV + VU\right), \qquad \mathbf{e} = \mathbf{vec}(I_p)$$

composite cone: apply $\circ$-product componentwise

Symmetric parametrization and scaling

$$\begin{bmatrix} 0 \\ 0 \\ s \end{bmatrix} = \begin{bmatrix} 0 & A^T & G^T \\ -A & 0 & 0 \\ -G & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} c \\ b \\ h \end{bmatrix}$$

$$(s, z) \succ 0, \qquad s \circ z = \mu \mathbf{e}$$

can further replace $s \circ z = \mu \mathbf{e}$ with

$$(W^{-T} s) \circ (W z) = \mu \mathbf{e}$$

if multiplications with W and W^{-T} preserve the cone and the central path

W is called a *scaling matrix*

Nesterov-Todd scaling

a specific scaling, used, for example, in Sedumi and SDPT3

- W is associated with a pair $(\hat{s}, \hat{z}) \succ 0$ (and defined on the next pages)
- satisfies $W^T W = \nabla^2 \phi(w)^{-1}$ where w is the unique point for which

$$\nabla^2 \phi(w) \hat{s} = \hat{z}$$

  w is called the *Nesterov-Todd scaling point*

- multiplications by W and W^{-T} map $\hat{s}$ and $\hat{z}$ to the same point:

$$W^{-T} \hat{s} = W \hat{z} = \lambda$$

  this implies that $\hat{s}^T \hat{z} = \|\lambda\|_2^2$

NT scaling for nonnegative orthant

$$W = \begin{bmatrix} \sqrt{\hat{s}_1/\hat{z}_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\hat{s}_2/\hat{z}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\hat{s}_p/\hat{z}_p} \end{bmatrix}$$

- a positive diagonal scaling
- NT scaling point is

$$w = \left(\sqrt{\hat{s}_1/\hat{z}_1}, \sqrt{\hat{s}_2/\hat{z}_2}, \ldots, \sqrt{\hat{s}_p/\hat{z}_p} \right)$$

- scaled $\hat{s}$, $\hat{z}$

$$\lambda = W^{-T}\hat{s} = W\hat{z} = \left(\sqrt{\hat{s}_1\hat{z}_1},\ \sqrt{\hat{s}_2\hat{z}_2},\ \ldots,\ \sqrt{\hat{s}_p\hat{z}_p} \right)$$

NT scaling for second-order cone

$$W = \left(\frac{\hat{s}^T J \hat{s}}{\hat{z}^T J \hat{z}}\right)^{1/4} \begin{bmatrix} \bar{w}_0 & \bar{w}_1^T \\ \bar{w}_1 & I + (\bar{w}_0 + 1)^{-1} \bar{w}_1 \bar{w}_1^T \end{bmatrix}$$

where

$$\bar{w} = \frac{1}{2\gamma}\left(\frac{\hat{s}}{\sqrt{\hat{s}^T J s}} + \frac{J\hat{z}}{\sqrt{\hat{z}^T J z}}\right), \qquad \gamma = \left(\frac{1}{2} + \frac{\hat{z}^T \hat{s}}{2(\hat{s}^T J \hat{s})^{1/2}(\hat{z}^T J \hat{z})^{1/2}}\right)^{1/2}$$

- W can also be expressed as

$$W = \beta(2vv^T - J), \qquad v = \frac{\bar{w} + \mathbf{e}}{(2(\bar{w}_0 + 1))^{1/2}}, \qquad \beta = \left(\frac{\hat{s}^T J \hat{s}}{\hat{z}^T J \hat{z}}\right)^{1/4}$$

  ($2vv^T - J$ is called a *hyperbolic Householder matrix*)

- NT scaling point is $w = \beta \bar{w}$

NT scaling for positive semidefinite cone

$$Wv = \mathbf{vec}\left(R^T \mathbf{mat}(v) R\right)$$

where R simultaneously diagonalizes $\mathbf{mat}(\hat{z})$ and $\mathbf{mat}(\hat{s})^{-1}$

$$R^T \mathbf{mat}(\hat{z}) R = R^{-1} \mathbf{mat}(\hat{s}) R^{-T} = \Lambda$$

- W is a nonsingular congruence transformation
- NT scaling point is $w = \mathbf{vec}(RR^T)$
- R can be computed from two Cholesky factorizations and an SVD: if

$$\mathbf{mat}(\hat{s}) = L_1 L_1^T, \qquad \mathbf{mat}(\hat{z}) = L_2 L_2^T, \qquad L_2^T L_1 = U \Lambda V^T$$

  then $R = L_1 V \Lambda^{-1/2} = L_2^{-T} U \Lambda^{1/2}$

Outline

- cone programming
- logarithmic barrier function and central path
- symmetrization and Nesterov-Todd scaling
- **path-following algorithm**
- quadratic cone program
- self-dual embedding

Basic primal-dual update

suppose we are at $(\hat{s}, \hat{x}, \hat{y}, \hat{z})$ with $\hat{s} \succ 0$, $\hat{z} \succ 0$

- define $\mu = \hat{s}^T \hat{z}/m$ and let W be the NT scaling matrix for $\hat{s}$, $\hat{z}$
- compute $\Delta s, \Delta x, \Delta y, \Delta z$ by linearizing the central path equation

$$\begin{bmatrix} 0 \\ 0 \\ s \end{bmatrix} = \begin{bmatrix} 0 & A^T & G^T \\ -A & 0 & 0 \\ -G & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} c \\ b \\ h \end{bmatrix}$$

$$(W^{-T} s) \circ (W z) = \sigma \mu \, \mathbf{e}$$

around $\hat{s}, \hat{x}, \hat{y}, \hat{z}$, for some $\sigma < 1$

- make an update

$$(\hat{s}, \hat{x}) := (\hat{s}, \hat{x}) + \alpha_{\mathrm{p}}(\Delta x, \Delta s), \qquad (\hat{y}, \hat{z}) := (\hat{y}, \hat{z}) + \alpha_{\mathrm{d}}(\Delta y, \Delta z)$$

that preserves positivity of $\hat{s}$, $\hat{z}$

Linearized central path equation

define $\lambda = W^{-T}\hat{s} = W\hat{z}$ and

$$r = \begin{bmatrix} 0 \\ 0 \\ \hat{s} \end{bmatrix} - \begin{bmatrix} 0 & A^T & G^T \\ -A & 0 & 0 \\ -G & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{bmatrix} - \begin{bmatrix} c \\ b \\ h \end{bmatrix}$$

linearized central path equation

$$\begin{bmatrix} 0 \\ 0 \\ \Delta s \end{bmatrix} - \begin{bmatrix} 0 & A^T & G^T \\ -A & 0 & 0 \\ -G & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = -r$$

$$\lambda \circ (W\Delta z + W^{-T}\Delta s) = \sigma\mu\mathbf{e} - \lambda \circ \lambda$$

second equation is linearization of $(W^{-T}(\hat{s} + \Delta s)) \circ (W(\hat{z} + \Delta z)) = \sigma\mu\mathbf{e}$

Path-following algorithm

choose starting points $\hat{s}$, $\hat{x}$, $\hat{y}$, $\hat{z}$ with $\hat{s} \succ 0$, $\hat{z} \succ 0$

1. **compute residuals and evaluate stopping criteria**

$$r = \begin{bmatrix} 0 \\ 0 \\ \hat{s} \end{bmatrix} - \begin{bmatrix} 0 & A^T & G^T \\ -A & 0 & 0 \\ -G & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{bmatrix} - \begin{bmatrix} c \\ b \\ h \end{bmatrix}$$

   terminate if r and $\hat{s}^T \hat{z}$ are sufficiently small

2. **compute scaling matrix** W associated with $(\hat{s}, \hat{z})$ and set

$$\lambda := W^{-T} \hat{s} = W \hat{z}, \qquad \mu := \frac{\lambda^T \lambda}{m} = \frac{\hat{s}^T \hat{z}}{m}$$

   (m is logarithmic degree of cone)

3. **compute affine scaling direction** by solving the linear equation

$$\begin{bmatrix} 0 \\ 0 \\ \Delta s_{\rm a} \end{bmatrix} - \begin{bmatrix} 0 & A^T & G^T \\ -A & 0 & 0 \\ -G & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\rm a} \\ \Delta y_{\rm a} \\ \Delta z_{\rm a} \end{bmatrix} = -r$$

$$\lambda \circ \left(W \Delta z_{\rm a} + W^{-T} \Delta s_{\rm a} \right) = -\lambda \circ \lambda$$

4. **select barrier parameter**

$$\sigma = \left(\frac{(\hat{s} + \alpha_{\rm p} \Delta s_{\rm a})^T (\hat{z} + \alpha_{\rm d} \Delta z_{\rm a})}{\hat{s}^T \hat{z}} \right)^{\delta}$$

where δ is an algorithm parameter (typical value is $\delta = 3$) and

$$\begin{aligned} \alpha_{\rm p} &= \sup\{\alpha \in [0,1] \mid \hat{s} + \alpha \Delta s_{\rm a} \succeq 0\} \\ \alpha_{\rm d} &= \sup\{\alpha \in [0,1] \mid \hat{z} + \alpha \Delta z_{\rm a} \succeq 0\} \end{aligned}$$

5. **compute search direction** by solving the linear equation

$$\begin{bmatrix} 0 \\ 0 \\ \Delta s \end{bmatrix} - \begin{bmatrix} 0 & A^T & G^T \\ -A & 0 & 0 \\ -G & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = -r$$

$$\lambda \circ (W\Delta z + W^{-T}\Delta s) = \sigma\mu\mathbf{e} - \lambda \circ \lambda$$

6. **update iterates**

$$\begin{aligned} (\hat{x}, \hat{s}) &:= (\hat{x}, \hat{s}) + \min\{1, 0.99\alpha_{\mathrm{p}}\}(\Delta x, \Delta s) \\ (\hat{y}, \hat{z}) &:= (\hat{y}, \hat{z}) + \min\{1, 0.99\alpha_{\mathrm{d}}\}(\Delta y, \Delta z) \end{aligned}$$

where

$$\alpha_{\mathrm{p}} = \sup\{\alpha \geq 0 \mid \hat{s} + \alpha\Delta s \succeq 0\}, \qquad \alpha_{\mathrm{d}} = \sup\{\alpha \geq 0 \mid \hat{z} + \alpha\Delta z \succeq 0\}$$

return to step 1

interpretation and discussion

- step 3: affine scaling direction solves linearized central path equation with $\sigma = 0$, $i.e.$, the linearized optimality conditions

- step 4 is a heuristic for choosing σ based on an estimate of the quality of the affine scaling direction

  σ is small if a step in the affine scaling direction gives a large reduction in $\hat{s}^T \hat{z}$

- step 5: linear equation has same coefficient matrix as equation in step 3

  if a direct method is used, we can reuse the factorization used in step 3, and solve the two equations at the cost of one

Mehrotra correction

in step 5, solve

$$\begin{bmatrix} 0 \\ 0 \\ \Delta s \end{bmatrix} - \begin{bmatrix} 0 & A^T & G^T \\ -A & 0 & 0 \\ -G & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = -r$$

$$\lambda \circ (W\Delta z + W^{-T}\Delta s) = \sigma\mu\mathbf{e} - \lambda \circ \lambda - (W^{-T}\Delta s_{\mathrm{a}}) \circ (W\Delta z_{\mathrm{a}})$$

- extra term on the r.h.s. is approximation of the second-order term in

$$(W^{-T}(\hat{s} + \Delta s)) \circ (W(\hat{z} + \Delta z)) = -\sigma\mu\mathbf{e}$$

- adding the correction typically saves a few iterations

Newton equations

steps 3 and 5 reduce to equations

$$\begin{bmatrix} 0 & A^T & G^T \\ A & 0 & 0 \\ G & 0 & -W^T W \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} d_x \\ d_y \\ d_z \end{bmatrix}$$

usually solved by eliminating Δz:

$$\begin{bmatrix} G^T W^{-1} W^{-T} G & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} d_x + G^T W^{-1} W^{-T} d_z \\ d_y \end{bmatrix}$$

note (from p.5-17) that

$$G^T W^{-1} W^{-T} G = G^T \nabla^2 \phi(w) G,$$

the Hessian of the barrier function $\phi(h - Gx)$ at the scaling point w

Outline

- cone programming
- logarithmic barrier function and central path
- symmetrization and Nesterov-Todd scaling
- path-following algorithm
- **quadratic cone program**
- self-dual embedding

Quadratic cone program

$$\begin{array}{ll} \text{minimize} & (1/2)x^TPx + q^Tx \\ \text{subject to} & Gx + s = h \\ & Ax = b \\ & s \succeq 0 \end{array}$$

optimality conditions

$$\begin{bmatrix} 0 \\ 0 \\ s \end{bmatrix} = \begin{bmatrix} P & A^T & G^T \\ -A & 0 & 0 \\ -G & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} q \\ b \\ h \end{bmatrix}, \quad (s, z) \succeq 0, \quad s^Tz = 0$$

central path

$$\begin{bmatrix} 0 \\ 0 \\ s \end{bmatrix} = \begin{bmatrix} P & A^T & G^T \\ -A & 0 & 0 \\ -G & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} q \\ b \\ h \end{bmatrix}, \quad (s, z) \succ 0, \quad s \circ z = \mu \mathbf{e}$$

Path-following algorithm

algorithm is almost identical to algorithm on page 5-23

- compute search directions from linearized central path equation; for example, step 5 becomes

$$\begin{bmatrix} 0 \\ 0 \\ \Delta s \end{bmatrix} - \begin{bmatrix} P & A^T & G^T \\ -A & 0 & 0 \\ -G & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = -r$$

$$\lambda \circ (W\Delta z + W^{-T}\Delta s) = \sigma\mu\mathbf{e} - \lambda \circ \lambda$$

- use equal primal and dual step sizes

  for example, in step 6,

$$\alpha_{\mathrm{p}} = \alpha_{\mathrm{d}} = \sup\{\alpha \geq 0 \mid \hat{s} + \alpha\Delta s \succeq 0, \hat{z} + \alpha\Delta z \succeq 0\}$$

Outline

- cone programming
- logarithmic barrier function and central path
- symmetrization and Nesterov-Todd scaling
- path-following algorithm
- quadratic cone program
- **self-dual embedding**

Initialization and infeasibility detection

barrier method

- assumes problem is primal and dual feasible
- requires phase I to find initial primal feasible point

primal-dual path-following method

- assumes problem is primal and dual feasible
- allows infeasible starting points

methods based on self-dual embedding

- can detect primal and dual infeasibility
- embed cone program in slightly larger problem that is always feasible
- from solution of embedded problem, extract solution of original problem, or certificates of primal or dual infeasibility

Infeasibility

primal infeasibility: a solution y, z of

$$A^T y + G^T z = 0, \qquad h^T z + b^T y = -1, \qquad z \succeq 0$$

is a certificate of infeasibility of $Gx \preceq h$, $Ax = b$

dual infeasibility: a solution x of

$$Gx \preceq 0, \qquad Ax = 0, \qquad c^T x = -1$$

is a certificate of infeasibility of $A^T y + G^T z + c = 0$, $z \succeq 0$

these are strong alternatives if a constraint qualification holds

Self-dual embedding

$$
\begin{array}{ll}
\text{minimize} & 0 \\
\text{subject to} & \begin{bmatrix} 0 \\ 0 \\ s \\ \kappa \end{bmatrix} = \begin{bmatrix} 0 & A^T & G^T & c \\ -A & 0 & 0 & b \\ -G & 0 & 0 & h \\ -c^T & -b^T & -h^T & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ \tau \end{bmatrix} \\
& (s, \kappa, z, \tau) \succeq 0
\end{array}
$$

- problem has a trivial solution (all variables zero)
- equality constraint implies $s^T z + \kappa\tau = 0$ at feasible points
- problem is not strictly feasible (hence, central path does not exist)

Optimality condition for embedded problem

$$\begin{bmatrix} 0 \\ 0 \\ s \\ \kappa \end{bmatrix} = \begin{bmatrix} 0 & A^T & G^T & c \\ -A & 0 & 0 & b \\ -G & 0 & 0 & h \\ -c^T & -b^T & -h^T & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ \tau \end{bmatrix}$$

$$(s, \kappa, z, \tau) \succeq 0, \qquad z^T s + \tau\kappa = 0$$

- follows from self-dual property
- a (mixed) linear complementarity problem

Classification of nonzero solution

let s, κ, x, y, z, τ be a nonzero solution of the self-dual embedding

case 1: $\tau > 0$, $\kappa = 0$

$$\hat{s} = s/\tau, \qquad \hat{x} = x/\tau, \qquad \hat{y} = y/\tau, \qquad \hat{z} = z/\tau$$

are primal and dual solutions of the cone program, *i.e.*, satisfy

$$\begin{bmatrix} 0 \\ 0 \\ \hat{s} \end{bmatrix} = \begin{bmatrix} 0 & A^T & G^T \\ -A & 0 & 0 \\ -G & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{bmatrix} + \begin{bmatrix} c \\ b \\ h \end{bmatrix}$$

$$(\hat{s}, \hat{z}) \succeq 0, \qquad \hat{s}^T \hat{z} = 0$$

case 2: $\tau = 0$, $\kappa > 0$; this implies $c^T x + b^T y + h^T z < 0$

- if $b^T y + h^T z < 0$, then

$$\hat{z} = \frac{1}{-(h^T z + b^T y)} z, \qquad \hat{y} = \frac{1}{-(h^T z + b^T y)} y$$

  is a proof of primal infeasibility, *i.e.*, satisfies

$$A^T \hat{y} + G^T \hat{z} = 0, \qquad h^T \hat{z} + b^T \hat{y} = -1, \qquad \hat{z} \succeq 0$$

- if $c^T x < 0$, then

$$\hat{x} = \frac{1}{-c^T x} x, \qquad \hat{x} = \frac{1}{-c^T x} s$$

  is a proof of dual infeasibility, *i.e.*, satisfies

$$G\hat{x} \preceq 0, \qquad A\hat{x} = 0, \qquad c^T \hat{x} = -1$$

case 3: $\tau = \kappa = 0$; no conclusion can be made about the original problem

Extended self-dual embedding

$$
\begin{array}{ll}
\text{min.} & (m+1)\theta \\
\text{s.t.} & \begin{bmatrix} 0 \\ 0 \\ s \\ \kappa \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & A^T & G^T & c & q_x \\ -A & 0 & 0 & b & q_y \\ -G & 0 & 0 & h & q_z \\ -c^T & -b^T & -h^T & 0 & q_\tau \\ -q_x^T & -q_y^T & -q_z^T & -q_\tau & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ \tau \\ \theta \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ m+1 \end{bmatrix} \\
 & (s, \kappa, z, \tau) \succeq 0
\end{array}
$$

here, m is the logarithmic degree of the cone and

$$
\begin{bmatrix} q_x \\ q_y \\ q_z \\ q_\tau \end{bmatrix} = \frac{m+1}{s_0^T z_0 + 1} \left(\begin{bmatrix} 0 \\ 0 \\ s_0 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 & A^T & G^T & c \\ -A & 0 & 0 & b \\ -G & 0 & 0 & h \\ -c^T & -b^T & -h^T & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ z_0 \\ 1 \end{bmatrix} \right)
$$

s_0, x_0, y_0, z_0 are arbitrary with $s_0 \succ 0$, $z_0 \succ 0$

Optimality condition

$$\begin{bmatrix} 0 \\ 0 \\ s \\ \kappa \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & A^T & G^T & c & q_x \\ -A & 0 & 0 & b & q_y \\ -G & 0 & 0 & h & q_z \\ -c^T & -b^T & -h^T & 0 & q_\tau \\ -q_x^T & -q_y^T & -q_z^T & -q_\tau & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ \tau \\ \theta \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ m+1 \end{bmatrix}$$

$$(s, \kappa, z, \tau) \succeq 0, \qquad s^T z + \kappa\tau = 0$$

- follows from self-dual property
- a (mixed) linear complementarity problem

Properties of extended self-dual embedding

- problem is strictly feasible; a strictly feasible point is given by

$$(s,\,\kappa,\,x,\,y,\,z,\,\tau,\,\theta) = (s_0,\,1,\,x_0,\,y_0,\,z_0,\,1,\,\frac{s_0^T z_0 + 1}{m+1}) \qquad (5.2)$$

- if $s,\,\kappa,\,x,\,y,\,z,\,\tau,\,\theta$ satisfy equality constraint, then

$$\theta = \frac{s^T z + \kappa\tau}{m+1}$$

  (take inner product with (x, y, z, τ, θ) of each side of the equality)

- at optimum, $\theta = 0$ and problem reduces to the embedding on p.5-33
- classification of p.5-35 also applies to solutions of extended embedding

Central path for extended embedding

$$\begin{bmatrix} 0 \\ 0 \\ s \\ \kappa \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & A^T & G^T & c & q_x \\ -A & 0 & 0 & b & q_y \\ -G & 0 & 0 & h & q_z \\ -c^T & -b^T & -h^T & 0 & q_\tau \\ -q_x^T & -q_y^T & -q_z^T & -q_\tau & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ \tau \\ \theta \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ m+1 \end{bmatrix}$$

$$(s, \kappa, z, \tau) \succeq 0, \qquad s \circ z = \mu \mathbf{e}, \qquad \kappa\tau = \mu$$

- inner product with (x, y, z, τ, θ) shows that on the central path

$$\theta = \frac{z^T s + \kappa\tau}{m+1} = \mu$$

- initial point (5.2) is on the central path with $\mu = (s_0^T z_0 + 1)/(m+1)$

Simplified central path equations

$$\begin{bmatrix} 0 \\ 0 \\ s \\ \kappa \end{bmatrix} = \begin{bmatrix} 0 & A^T & G^T & c \\ -A & 0 & 0 & b \\ -G & 0 & 0 & h \\ -c^T & -b^T & -h^T & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ \tau \end{bmatrix} + \mu \begin{bmatrix} q_x \\ q_y \\ q_z \\ q_\tau \end{bmatrix}$$

$$(s, \kappa, z, \tau) \succeq 0, \qquad s \circ z = \mu \mathbf{e}, \qquad \kappa\tau = \mu$$

- we eliminated variable θ because $\theta = \mu$ on the central path
- we removed the 5th equality, because it is implied by the first four (this follows by taking inner product with (x, y, z, τ))
- can be seen as a 'shifted central path' for the embedding on p.5-33

Path-following algorithm

choose starting points $\hat{s}$, $\hat{x}$, $\hat{y}$, $\hat{z}$, with $\hat{s} \succ 0$, $\hat{z} \succ 0$; set $\hat{\kappa} := 1$, $\hat{\tau} := 1$

1. **compute residuals and evaluate stopping criteria**

$$r = \begin{bmatrix} 0 \\ 0 \\ \hat{s} \\ \hat{\kappa} \end{bmatrix} - \begin{bmatrix} 0 & A^T & G^T & c \\ -A & 0 & 0 & b \\ -G & 0 & 0 & h \\ -c^T & -b^T & -h^T & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \\ \hat{\tau} \end{bmatrix}$$

   terminate if r and $\hat{s}^T\hat{z}/\tau^2$ are sufficiently small, or an approximate certificate of primal or dual infeasibility has been found

2. **compute scaling matrix** W associated with $(\hat{s}, \hat{z})$ and set

$$\lambda := W^{-T}\hat{s} = W\hat{z}, \qquad \mu := \frac{\hat{s}^T\hat{z} + \hat{\kappa}\hat{\tau}}{m+1}$$

3. **compute affine scaling direction** by solving the linear equation

$$\begin{bmatrix} 0 \\ 0 \\ \Delta s_{\rm a} \\ \Delta \kappa_{\rm a} \end{bmatrix} - \begin{bmatrix} 0 & A^T & G^T & c \\ -A & 0 & 0 & b \\ -G & 0 & 0 & h \\ -c^T & -b^T & -h^T & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\rm a} \\ \Delta y_{\rm a} \\ \Delta z_{\rm a} \\ \Delta \tau_{\rm a} \end{bmatrix} = -r$$

$$\lambda \circ \left(W \Delta z_{\rm a} + W^{-T} \Delta s_{\rm a} \right) = -\lambda \circ \lambda, \qquad \hat{\kappa} \Delta \tau_{\rm a} + \hat{\tau} \Delta \kappa_{\rm a} = -\hat{\kappa} \hat{\tau}$$

4. **select barrier parameter**

$$\sigma := (1 - \alpha)^\delta$$

where δ is an algorithm parameter (typical value is $\delta = 3$) and

$$\alpha = \sup \{ \alpha \in [0, 1] \mid (\hat{s}, \hat{\kappa}, \hat{z}, \hat{\tau}) + \alpha (\Delta s_{\rm a}, \Delta \kappa_{\rm a}, \Delta z_{\rm a}, \Delta \tau_{\rm a}) \succeq 0 \}$$

5. **compute search direction** by solving the linear equation

$$\begin{bmatrix} 0 \\ 0 \\ \Delta s \\ \Delta\kappa \end{bmatrix} - \begin{bmatrix} 0 & A^T & G^T & c \\ -A & 0 & 0 & b \\ -G & 0 & 0 & h \\ -c^T & -b^T & -h^T & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \\ \Delta\tau \end{bmatrix} = -(1-\sigma)r$$

$$\lambda \circ (W\Delta z + W^{-T}\Delta s) = \sigma\mu\mathbf{e} - \lambda\circ\lambda - (W^{-T}\Delta s_{\mathrm{a}}) \circ (W\Delta z_{\mathrm{a}})$$

$$\hat{\kappa}\Delta\tau + \hat{\tau}\Delta\kappa = \sigma\mu - \hat{\kappa}\hat{\tau} - \Delta\kappa_{\mathrm{a}}\Delta\tau_{\mathrm{a}}$$

6. **update iterates**

$$(\hat{s}, \hat{\kappa}, \hat{x}, \hat{y}, \hat{z}, \hat{\tau}) := (\hat{s}, \hat{\kappa}, \hat{x}, \hat{y}, \hat{z}, \hat{\tau}) + \min\{1, 0.99\alpha\}\,(\Delta s, \Delta\kappa, \Delta x, \Delta y, \Delta z, \Delta\tau)$$

where $\alpha = \sup\{\alpha \in [0,1] \mid (\hat{s}, \hat{\kappa}, \hat{z}, \hat{\tau}) + \alpha(\Delta s, \Delta\kappa, \Delta z, \Delta\tau) \succeq 0\}$

return to step 1

properties (without proof)

- step 3: affine scaling direction satisfies

$$\hat{s}^T \Delta z_{\mathrm{a}} + \hat{z}^T \Delta s_{\mathrm{a}} = -\hat{s}^T \hat{z}, \qquad \hat{\kappa} \Delta \tau_{\mathrm{a}} + \hat{\tau} \Delta \kappa_{\mathrm{a}} = -\hat{\kappa}\hat{\tau}$$

$$\Delta s_{\mathrm{a}}^T \Delta z_{\mathrm{a}} + \Delta \tau_{\mathrm{a}} \Delta \kappa_{\mathrm{a}} = 0$$

- step 5: search direction satisfies

$$\hat{s}^T \Delta z + \hat{\kappa} \Delta \tau + \hat{z}^T \Delta s + \hat{\tau} \Delta \kappa = -(1-\sigma)(\hat{s}^T \hat{z} + \hat{\kappa}\hat{\tau})$$

$$\Delta s^T \Delta z + \Delta \tau \Delta \kappa = 0$$

discussion

- step 4: expression for σ is based on simplifiying

$$\sigma = \left(\frac{(\hat{s} + \alpha \Delta s_{\rm a})^T (\hat{z} + \alpha \Delta z_{\rm a}) + (\hat{\kappa} + \alpha \Delta \kappa_{\rm a})(\hat{\tau} + \alpha \Delta \tau_{\rm a})}{\hat{s}^T \hat{z} + \hat{\kappa} \hat{\tau}} \right)^{\delta}$$

- steps 5 and 6: gap and residual decrease linearly with α:

$$\mu^+ = (1 - \alpha(1-\sigma))\mu, \qquad r^+ = (1 - \alpha(1-\sigma))r,$$

  if μ^+ and r^+ are the values of μ and r at the next iteration

- $r = \mu q$, with q defined on p.5-37 (a multiple of the initial residual)

- in step 5, $-(1-\sigma)r = -r + \sigma \mu q$: the equation is the linearization of the central path equation of p.5-41 for barrier parameter $\sigma \mu$

Linear algebra complexity

- essentially the same as for the method of lecture 5
- eliminating $\Delta\tau$, $\Delta\kappa$ in steps 3 and 5 requires solution of an extra system

$$\begin{bmatrix} 0 & A^T & G^T \\ A & 0 & 0 \\ G & 0 & -W^TW \end{bmatrix} \begin{bmatrix} \Delta\tilde{x} \\ \Delta\tilde{y} \\ \Delta\tilde{z} \end{bmatrix} = \begin{bmatrix} c \\ b \\ h \end{bmatrix}$$

so number of Newton equations solved per iteration is 3 (as opposed to 2 in the method of page 5-23)

References

- J.F. Sturm, *Implementation of interior-point methods for mixed semidefinite and second order cone optimization problems*, Optimization methods and Software (2002)
- R.H. Tütüncü, K.C. Toh, M.J. Todd, *Solving semidefinite-quadratic-linear programs using SDPT3*, Mathematical Programming (2003)
- E.D. Andersen, C. Roos, T. Terlaky, *On implementing a primal-dual interior-point method for conic quadratic optimization*, Mathematical Programming (2003)
- J. Dahl, L. Vandenberghe, CVXOPT documentation (`abel.ee.ucla.edu/cvxopt`)