16. Path-following methods

- central path
- short-step barrier method
- predictor-corrector method

Introduction

primal-dual pair of conic LPs

minimize	$c^T x$	maximize	$-b^T z$
subjecct to	$Ax \preceq b$	subjecct to	$A^T z + c = 0$
			$z \succeq_* 0$

- $A \in \mathbf{R}^{m \times n}$ with $\mathbf{rank}(A) = n$
- inequalities are with respect to proper cone K and its dual cone K^*
- we will assume primal and dual problem are strictly feasible

this lecture

- feasible methods that follow the central path to find the solution
- complexity analysis based on theory of self-concordant functions

Outline

- **central path**
- short-step barrier method
- predictor-corrector method

Barrier for the feasible set

definition: as a barrier function for the feasible set we will use

$$\psi(x) = \phi(b - Ax)$$

where ϕ is a θ-normal barrier for K

notation (in this lecture): $\|v\|_{x*} = (v^T \nabla^2 \psi(x)^{-1} v)^{1/2}$

properties

- ψ is self-concordant with domain $\{x \mid Ax \prec b\}$
- Newton decrement of ψ is bounded by $\sqrt{\theta}$, *i.e.*,

$$\|\nabla \psi(x)\|_{x*}^2 = \nabla \psi(x)^T \nabla^2 \psi(x)^{-1} \nabla \psi(x) \leq \theta \quad \forall x \in \mathbf{dom}\, \psi$$

(proof on next page)

proof of bound on Newton decrement

- gradient and Hessian of ψ are (with $s = b - Ax$)

$$\nabla\psi(x) = -A^T\nabla\phi(s), \qquad \nabla^2\psi(x) = A^T\nabla^2\phi(s)A$$

- from page 15-24, $\nabla\phi(s)^T\nabla^2\phi(s)^{-1}\nabla\phi(s) = \theta$; therefore

$$\begin{aligned}
\nabla\psi(x)^T\nabla^2\psi(x)^{-1}\nabla\psi(x) &= \sup_v \left(-v^T\nabla^2\psi(x)v + 2\nabla\psi(x)^T v\right) \\
&= \sup_v \left(-(Av)^T\nabla^2\phi(s)(Av) - 2\nabla\phi(s)^T Av\right) \\
&\leq \sup_w \left(-w^T\nabla^2\phi(s)w + 2\nabla\phi(s)^T w\right) \\
&= \nabla\phi(s)^T\nabla^2\phi(s)^{-1}\nabla\phi(s) \\
&= \theta
\end{aligned}$$

Central path

definition: the set of minimizers $x^\star(t)$, for $t > 0$, of

$$tc^T x + \psi(x) = tc^T x + \phi(b - Ax)$$

optimality conditions

$$A^T \nabla\phi(s) = tc, \qquad s = b - Ax$$

- implies that $z = -(1/t)\nabla\phi(s)$ is strictly dual feasible
- by weak duality,

$$c^T x^\star(t) - p^\star \leq c^T x + b^T z = z^T s = \frac{\theta}{t}$$

hence, $c^T x^\star(t) \to p^\star$ as $t \to \infty$

Existence and uniqueness

centering problem

$$\begin{array}{ll}\text{minimize} & tc^T x + \phi(s) \\ \text{subject to} & Ax + s = b\end{array}$$

Lagrange dual (with dual cone barrier ϕ_* of page 15-27)

$$\begin{array}{ll}\text{maximize} & -tb^T z - \phi_*(z) + \theta \log t \\ \text{subject to} & A^T z + c = 0\end{array}$$

- strictly feasible z for dual conic LP is feasible for dual centering problem
- if dual conic LP is strictly feasible, $tc^T x + \phi(b - Ax)$ is bounded below
- from self-concordance theory (p.15-12), $x^\star(t)$ exists and is unique

Dual points in neighborhood of central path

Newton step Δx for $tc^T x + \psi(x) = tc^T x + \phi(b - Ax)$

- satisfies Newton equation

$$A^T \nabla^2 \phi(s) A \Delta x = -tc + A^T \nabla \phi(s), \qquad s = b - Ax$$

- Newton decrement is $\lambda_t(x) = \left(\Delta x^T \nabla^2 \psi(x) \Delta x\right)^{1/2}$

dual feasible point: define

$$z = -\frac{1}{t}\left(\nabla \phi(s) - \nabla^2 \phi(s) A \Delta x\right)$$

- satisfies $A^T z + c = 0$ by definition
- satisfies $z \succ_* 0$ if $\lambda_t(x) < 1$ (see next page)

proof. $z \succ_* 0$ follows from Dikin ellipsoid theorem

- Newton decrement is

$$\begin{aligned} \lambda_t(x)^2 &= \Delta x^T \nabla^2 \psi(x) \Delta x \\ &= \Delta x^T A^T \nabla^2 \phi(s) A \Delta x \\ &= v^T \nabla^2 \phi(s)^{-1} v \end{aligned}$$

  where $v = \nabla^2 \phi(s) A \Delta x$

- define $u = -\nabla \phi(s)$; then $\nabla^2 \phi_*(u) = \nabla^2 \phi(s)^{-1}$ (see p.15-28) and

$$\lambda_t(x)^2 = v^T \nabla^2 \phi_*(u) v$$

- by Dikin ellipsoid theorem $\lambda_t(x) < 1$ implies

$$u + v = -\nabla \phi(s) + \nabla^2 \phi(s) A \Delta x \succ_* 0$$

Duality gap in neighborhood of central path

$$c^T x - p^\star \leq \left(1 + \frac{\lambda_t(x)}{\sqrt{\theta}}\right)\frac{\theta}{t} \qquad \text{if } \lambda_t(x) < 1$$

- from weak duality, using the dual point z on page 16-7

$$\begin{aligned} s^T z &= \frac{1}{t}\left(\theta - s^T \nabla^2 \phi(s) A \Delta x\right) \\ &\leq \frac{1}{t}\left(\theta + \|\nabla^2\phi(s)^{1/2} s\|_2 \, \|\nabla^2\phi(s)^{1/2} A\Delta x\|_2\right) \\ &= \frac{\theta + \sqrt{\theta}\,\lambda_t(x)}{t} \end{aligned}$$

- implies $c^T x - p^\star \leq 2\theta/t$, since $\theta \geq 1$ holds for any θ-normal barrier ϕ (ϕ is unbounded below, so its Newton decrement $\sqrt{\theta} \geq 1$ everywhere)

Outline

- central path
- **short-step barrier method**
- predictor-corrector method

Short-step methods

general idea: keep the iterates in the region of quadratic convergence for

$$tc^T x + \psi(x),$$

by limiting the rate at which t is increased (hence, 'short-step')

quadratic convergence results (from self-concordance theory)

- if $\lambda_t(x) \leq 1/4$, a full Newton step gives $\lambda_t(x^+) \leq 2\lambda_t(x)^2$
- started at a point with $\lambda_t(x) \leq 1/4$, an accuracy ϵ_{cent} is reached in

$$\log_2 \log_2(1/\epsilon_{\text{cent}}) \text{ iterations}$$

  for practical purposes this is a constant (4–6 for $\epsilon_{\text{cent}} \approx 10^{-5} \ldots 10^{-20}$)

Short-step method with exact centering

simplifying assumptions:

- $x^\star(t)$ is computed exactly
- a central point $x^\star(t_0)$ is given

algorithm: define a tolerance $\epsilon \in (0,1)$ and parameter

$$\mu = 1 + \frac{1}{4\sqrt{\theta}}$$

starting at $t = t_0$, repeat until $\theta/t \leq \epsilon$:

- compute $x^\star(\mu t)$ by Newton's method started at $x^\star(t)$
- set $t := \mu t$

Newton iterations for recentering

Newton decrement at $x = x^\star(t)$ for new value $t^+ = \mu t$ is

$$\begin{aligned}
\lambda_{t^+}(x) &= \|\mu t c + \nabla\psi(x)\|_{x*} \\
&= \|\mu(tc + \nabla\psi(x)) - (\mu - 1)\nabla\psi(x)\|_{x*} \\
&= (\mu - 1)\|\nabla\psi(x)\|_{x*} \\
&\leq (\mu - 1)\sqrt{\theta} \\
&= 1/4
\end{aligned}$$

- line 3 follows because $tc + \nabla\psi(x) = 0$ for $x = x^\star(t)$
- line 4 follows from $\|\nabla\psi(x)\|_{x*} \leq \sqrt{\theta}$ (see page 16-3)

conclusion

#iterations to compute $x^\star(t^+)$ from $x^\star(t)$ is bounded by a small constant

Iteration complexity

number of outer iterations: $t^{(k)} = \mu^k t_0 \geq \theta/\epsilon$ when

$$k \geq \frac{\log(\theta/(\epsilon t_0))}{\log \mu}$$

cumulative number of Newton iterations

$$O\left(\sqrt{\theta} \log\left(\frac{\theta}{\epsilon t_0}\right)\right)$$

(we used $\log \mu \geq (\log 2)/(4\sqrt{\theta})$ by concavity of $\log(1+u)$)

- multiply by flops per iteration to get polynomial worst-case complexity
- $\sqrt{\theta}$ dependence is lowest known complexity for interior-point methods

Short-step method with inexact centering

improvements of short-step method with exact centering

- keep iterates in region of quadratic region, but avoid complete centering
- at each iteration: make small increase in t, followed by *one* Newton step

algorithm: define a tolerance $\epsilon \in (0, 1)$ and parameters

$$\beta = \frac{1}{8}, \qquad \mu = 1 + \frac{1}{1 + 8\sqrt{\theta}}$$

- select x and t with $\lambda_t(x) \leq \beta$
- repeat until $2\theta/t \leq \epsilon$:

$$t := \mu t, \qquad x := x - \nabla^2\psi(x)^{-1}\left(tc + \nabla\psi(x)\right)$$

Newton decrement after update

we first show that $\lambda_t(x) \leq \beta$ at the end of each iteration

- if $\lambda_t(x) \leq \beta$ and $t^+ = \mu t$, then

$$\begin{aligned}
\lambda_{t^+}(x) &= \|t^+ c + \nabla\psi(x))\|_{x*} \\
&= \|\mu(tc + \nabla\psi(x)) - (\mu - 1)\nabla\psi(x)\|_{x*} \\
&\leq \mu\|tc + \nabla\psi(x)\|_{x*} + (\mu - 1)\|\nabla\psi(x)\|_{x*} \\
&\leq \mu\beta + (\mu - 1)\sqrt{\theta} \\
&= \frac{1}{4}
\end{aligned}$$

- from theory of Newton's method for s.c. functions (p.15-16)

$$\lambda_{t^+}(x^+) \leq 2\lambda_{t^+}(x)^2 \leq \frac{1}{8} = \beta$$

Iteration complexity

- from page 16-9, stopping criterion implies $c^T x - p^\star \leq \epsilon$

- stopping criterion is satisified when

$$\frac{t^{(k)}}{t_0} = \mu^k \geq \frac{2\theta}{\epsilon t_0}, \qquad k \geq \frac{\log(2\theta/(\epsilon t_0))}{\log \mu}$$

- taking the logarithm on both sides gives an upper bound of

$$O\left(\sqrt{\theta} \log\left(\frac{\theta}{\epsilon t_0}\right)\right) \quad \text{iterations}$$

(using $\log \mu \geq \log 2/(1 + 8\sqrt{\theta})$)

Outline

- central path
- short-step barrier method
- **predictor-corrector method**

Predictor-corrector methods

short-step methods

- stay in narrow neighborhood of central path (defined by limit on λ_t)
- make small, fixed increases $t^+ = \mu t$

as a result, quite slow in practice

predictor-corrector method

- select new t using a linear approximation to central path ('predictor')
- recenter with new t ('corrector')

allows faster and 'adaptive' increases in t

Global convergence bound for centering problem

$$\text{minimize} \quad f_t(x) = tc^T x + \phi(b - Ax)$$

convergence result (damped Newton algorithm of p.15-11 started at x)

$$\#\text{iterations} \leq \frac{f_t(x) - \inf_u f_t(u)}{\omega(\eta)} + \log_2 \log_2(1/\epsilon_{\text{cent}})$$

- ϵ_{cent} is accuracy in centering; $\eta \in (0, 1/4]$; $\omega(\eta) = \eta - \log(1 + \eta)$
- for practical purposes, second term is a small constant
- first term depends on unknown optimal value $\inf_u f_t(u)$

Bound from duality

dual centering problem (see p.16-6)

$$\begin{array}{ll} \text{maximize} & -tb^Tz - \phi_*(z) + \theta \log t \\ \text{subject to} & A^Tz + c = 0 \end{array}$$

strictly feasible z provides lower bound on $\inf_u f_t(u)$:

$$\inf_u f_t(u) \geq -tb^Tz - \phi_*(z) + \theta \log t$$

bound on centering cost: $f_t(x) - \inf_u f_t(u) \leq V_t(x, s, z)$ where

$$\begin{aligned} V_t(x, s, z) &= t(c^Tx + b^Tz) + \phi(s) + \phi_*(z) - \theta \log t \\ &= ts^Tz + \phi(s) + \phi_*(z) - \theta \log t \end{aligned}$$

Potential function

definition (for strictly feasible x, s, z)

$$\begin{aligned}\Psi(x,s,z) &= \inf_t V_t(x,s,z) \\ &= \theta \log \frac{s^T z}{\theta} + \phi(s) + \phi_*(z) + \theta\end{aligned}$$

(optimal t is $t = \mathrm{argmin}_t\, V_t(x,s,z) = \theta/s^T z$)

properties

- homogeneous of degree zero: $\Psi(\alpha x, \alpha s, \alpha z) = \Psi(x,s,z)$ for $\alpha > 0$
- nonnegative for all strictly feasible x, s, z
- zero only if x, s, z are centered

can be used as a *global* proximity measure

Tangent to central path

central path equation

$$\begin{bmatrix} 0 \\ s^\star(t) \end{bmatrix} = \begin{bmatrix} 0 & A^T \\ -A & 0 \end{bmatrix} \begin{bmatrix} x^\star(t) \\ z^\star(t) \end{bmatrix} + \begin{bmatrix} c \\ b \end{bmatrix}$$

$$z^\star(t) = -\frac{1}{t}\nabla\phi(s^\star(t))$$

derivatives $\dot{x} = dx^\star(t)/dt$, $\dot{s} = ds^\star/dt$, $\dot{z} = dz^\star(t)/dt$ satisfy

$$\begin{bmatrix} 0 \\ \dot{s} \end{bmatrix} = \begin{bmatrix} 0 & A^T \\ -A & 0 \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{z} \end{bmatrix}$$

$$\dot{z} = -\frac{1}{t}z^\star(t) - \frac{1}{t}\nabla^2\phi(s^\star(t))\dot{s}$$

tangent direction: defined as $\Delta x_{\mathrm{t}} = t\dot{x}$, $\Delta s_{\mathrm{t}} = t\dot{s}$, $\Delta z_{\mathrm{t}} = t\dot{z}$

Predictor equations

with $x = x^\star(t)$, $s = s^\star(t)$, $z = z^\star(t)$

$$\begin{bmatrix} (1/t)\nabla^2\phi(s) & 0 & I \\ 0 & 0 & A^T \\ -I & -A & 0 \end{bmatrix} \begin{bmatrix} \Delta s_{\rm t} \\ \Delta x_{\rm t} \\ \Delta z_{\rm t} \end{bmatrix} = \begin{bmatrix} -z \\ 0 \\ 0 \end{bmatrix} \qquad (1)$$

equivalent equations

$$\begin{bmatrix} I & 0 & (1/t)\nabla^2\phi_*(z) \\ 0 & 0 & A^T \\ -I & -A & 0 \end{bmatrix} \begin{bmatrix} \Delta s_{\rm t} \\ \Delta x_{\rm t} \\ \Delta z_{\rm t} \end{bmatrix} = \begin{bmatrix} -s \\ 0 \\ 0 \end{bmatrix} \qquad (2)$$

equivalence follows from primal-dual relations on central path

$$z = -\frac{1}{t}\nabla\phi(s), \qquad s = -\frac{1}{t}\nabla\phi_*(z), \qquad \frac{1}{t}\nabla^2\phi(s) = t\nabla^2\phi_*(z)^{-1}$$

Properties of tangent direction

- from 2nd and 3rd block in (1): $\Delta s_{\mathrm{t}}^T \Delta z_{\mathrm{t}} = 0$
- from first block in (1) and $\nabla^2\phi(s)s = -\nabla\phi(s)$:

$$s^T\Delta z_{\mathrm{t}} + z^T\Delta s_{\mathrm{t}} = -s^T z$$

- hence, gap in tangent direction is

$$(s + \alpha\Delta s_{\mathrm{t}})^T(z + \alpha\Delta z_{\mathrm{t}}) = (1-\alpha)s^T z$$

- from first block in (1)

$$\|\Delta s_{\mathrm{t}}\|_s^2 = \Delta s_{\mathrm{t}}^T \nabla^2\phi(s)\Delta s_{\mathrm{t}} = -tz^T\Delta s_{\mathrm{t}}$$

- similarly, from first block in (2)

$$\|\Delta z_{\mathrm{t}}\|_z^2 = \Delta z_{\mathrm{t}}^T \nabla^2\phi_*(z)\Delta z_{\mathrm{t}} = -ts^T\Delta z_{\mathrm{t}}$$

Predictor-corrector method with exact centering

simplifying assumptions: exact centering, a central point $x^\star(t_0)$ is given

algorithm: define tolerance $\epsilon \in (0,1)$, parameter $\beta > 0$, and set

$$t := t_0, \qquad (x,s,z) := (x^\star(t_0), s^\star(t_0), z^\star(t_0))$$

repeat until $\theta/t \leq \epsilon$:

- compute tangent direction $(\Delta x_{\rm t}, \Delta s_{\rm t}, \Delta z_{\rm t})$ at (x,s,z)
- set $(x,s,z) := (x,s,z) + \alpha(\Delta x_{\rm t}, \Delta s_{\rm t}, \Delta z_{\rm t})$ with α determined from

$$\Psi(x + \alpha\Delta x_{\rm t}, s + \alpha\Delta s_{\rm t}, z + \alpha\Delta z_{\rm t}) = \beta$$

- set $t := \theta/(s^T z)$ and compute $(x,s,z) := (x^\star(t), s^\star(t), z^\star(t))$

Iteration complexity

potential function in tangent direction (proof on next page)

$$\begin{aligned}\Psi(x + \alpha\Delta x_{\mathrm{t}}, s + \alpha\Delta s_{\mathrm{t}}, z + \alpha\Delta s_{\mathrm{t}}) &\leq \omega^*(\alpha\sqrt{\theta}) \\ &= -\alpha\sqrt{\theta} - \log(1 - \alpha\sqrt{\theta})\end{aligned}$$

lower bound on predictor step length: since ω^* is an increasing function

$$\alpha \geq \gamma/\sqrt{\theta} \quad \text{where } \omega^*(\gamma) = \beta$$

reduction in duality gap after one predictor/corrector cycle

$$t/t^+ = 1 - \alpha \leq 1 - \gamma/\sqrt{\theta} \leq \exp(-\gamma/\sqrt{\theta})$$

cumulative Newton iterations: $t^{(k)} \geq \theta/\epsilon$ after

$$O\left(\sqrt{\theta}\log\left(\theta/(t_0\epsilon)\right)\right) \text{ Newton iterations}$$

proof of upper bound on Ψ (with $s^+ = s + \alpha \Delta s_{\mathrm{t}}$, $z^+ = z + \alpha \Delta z_{\mathrm{t}}$)

- bounds on $\phi(s^+)$ and $\phi_*(z^+)$: from the inequality on page 15-8,

$$
\begin{aligned}
\phi(s^+) - \phi(s) &\leq \alpha \nabla \phi(s)^T \Delta s_{\mathrm{t}} + \omega^*(\alpha \|\Delta s_{\mathrm{t}}\|_s) \\
&= -\alpha t z^T \Delta s_{\mathrm{t}} + \omega^*(\alpha \|\Delta s_{\mathrm{t}}\|_s) \\
\phi_*(z^+) - \phi_*(z) &\leq \alpha \nabla \phi(z)^T \Delta z_{\mathrm{t}} + \omega^*(\alpha \|\Delta z_{\mathrm{t}}\|_z) \\
&= -\alpha t s^T \Delta z_{\mathrm{t}} + \omega^*(\alpha \|\Delta z_{\mathrm{t}}\|_z)
\end{aligned}
$$

- add the inequalities and use properties on page 16-23

$$
\begin{aligned}
\phi(s^+) - \phi(s) + \phi_*(z^+) - \phi_*(z) &\leq \alpha\theta + \omega^*(\alpha \|\Delta s_{\mathrm{t}}\|_s) + \omega^*(\alpha \|\Delta z_{\mathrm{t}}\|_z) \\
&\leq \alpha\theta + \omega^*(\alpha \left(\|\Delta s_{\mathrm{t}}\|_s^2 + \|\Delta z_{\mathrm{t}}\|_z^2\right)^{1/2}) \\
&= \alpha\theta + \omega^*(\alpha\sqrt{\theta})
\end{aligned}
$$

- since $(s^+)^T z^+ = (1-\alpha) s^T z$,

$$
\Psi(x^+, s^+, z^+) \leq \theta \log(1-\alpha) + \alpha\theta + \omega^*(\alpha\sqrt{\theta}) \leq \omega^*(\alpha\sqrt{\theta})
$$

References

- Yu. Nesterov, *Introductory Lectures on Convex Optimization. A Basic Course* (2004), chapter 4.
- Yu. Nesterov, *Towards nonsymmetric conic optimization* (2006).