16. Barrier functions

- self-concordant functions
- Newton's method
- normal barriers

Self-concordant functions

a function $f : \mathbf{R}^m \to \mathbf{R}$ is self-concordant if

- $\operatorname{dom} f$ is an open convex set
- f is three times continuously differentiable and $\nabla^2 f(x) \succ 0$ on $\operatorname{dom} f$
- f is closed, *i.e.*, $f(x) \to \infty$ as $x \to \operatorname{bd} \operatorname{dom} f$
- the Hessian of f satisfies the inequality

$$\frac{d}{d\alpha} \nabla^2 f(x + \alpha v) \Big|_{\alpha=0} \preceq 2\|v\|_x \nabla^2 f(x)$$

for all $x \in \operatorname{dom} f$ and all $v \in \mathbf{R}^m$, where

$$\|v\|_x = \left(v^T \nabla^2 f(x) v\right)^{1/2}$$

Equivalent definitions

Two-sided matrix inequality

$$-2\|v\|_x \nabla^2 f(x) \preceq \frac{d}{d\alpha}\nabla^2 f(x+\alpha v)\Big|_{\alpha=0} \preceq 2\|v\|_x \nabla^2 f(x) \qquad (1)$$

lower bound follows from the upper bound applied to $-v$

Restriction to a line: $g(\alpha) = f(x+\alpha v)$ satisfies

$$-2g''(\alpha)^{3/2} \le g'''(\alpha) \le 2g''(\alpha)^{3/2} \qquad (2)$$

- at $\alpha = 0$, this follows from (1) and $g''(\alpha) = v^T \nabla^2 f(x+\alpha v)v$
- can be used as equivalent definition of self-concordance

Examples and basic properties

Examples

- $f(x) = -\sum_{i=1}^m \log(b_i - a_i^T x)$ on $\{x \mid a_i^T x < b_i,\ i = 1, \ldots, m\}$
- $f(x) = -\log(x^T P x + q^T x + r)$ on $\{x \mid x^T P x + q^T x + r > 0\}$ if $P \prec 0$
- $f(x) = x \log x - \log x$ on $\mathbf{R}_{++}$

Properties

- f is self-concordant if and only if its restriction to an arbitrary line is s.c.
- if f_1, f_2 are self-concordant, then $f_1 + f_2$ is self-concordant
- if f is self-concordant, then βf is self-concordant for $\beta \geq 1$
- if f is self-concordant, then $f(Ax + b)$ is self-concordant

Bounds on second derivatives

- bounds on second derivative of restriction to a line $g(\alpha) = f(x + \alpha v)$

$$\frac{\|v\|_x^2}{(1 + \alpha\|v\|_x)^2} \leq g''(\alpha) \leq \frac{\|v\|_x^2}{(1 - \alpha\|v\|_x)^2} \tag{3}$$

(note that $\|v\|_x^2 = g''(0)$)

- bounds on Hessian

$$(1 - \alpha\|v\|_x)^2 \, \nabla^2 f(x) \preceq \nabla^2 f(x + \alpha v) \preceq \frac{1}{(1 - \alpha\|v\|_x)^2} \nabla^2 f(x) \tag{4}$$

these inequalities hold for $0 \leq \alpha\|v\|_x < 1$

Proof of first pair of inequalities:

- from (2) on page 16-3,

$$-1 \leq \frac{d}{d\alpha}\left(\frac{1}{\sqrt{g''(\alpha)}}\right) \leq 1$$

- integrate to get

$$\frac{1}{\sqrt{g''(0)}} - \alpha \leq \frac{1}{\sqrt{g''(\alpha)}} \leq \frac{1}{\sqrt{g''(0)}} + \alpha$$

- if $1 - \alpha\sqrt{g''(0)} > 0$ this can be written as

$$\frac{g''(0)}{\left(1 + \alpha g''(0)^{1/2}\right)^2} \leq g''(\alpha) \leq \frac{g''(0)}{\left(1 - \alpha g''(0)^{1/2}\right)^2}$$

Proof of second pair of inequalities:

define $h(\alpha) = w^T \nabla^2 f(x + \alpha v) w$, with arbitrary $w \neq 0$

- from (1) on page 16-3

$$\left| \frac{d}{d\alpha} \log h(\alpha) \right| = \left| \frac{h'(\alpha)}{h(\alpha)} \right| \leq 2\|v\|_{x+\alpha v} = 2\sqrt{g''(\alpha)}$$

- therefore, from (3) on page 16-5,

$$\left| \frac{d}{d\alpha} \log h(\alpha) \right| \leq \frac{2\|v\|_x}{1 - \alpha\|v\|_x}$$

- integrate to get

$$2 \log(1 - \alpha\|v\|_x) \leq \log(h(\alpha)/h(0)) \leq -2 \log(1 - \alpha\|v\|_x)$$

$$(1 - \alpha\|v\|_x)^2 h(0) \leq h(\alpha) \leq (1 - \alpha\|v\|_x)^{-2} h(0)$$

since w is arbitrary, this proves (4)

Bounds on function value

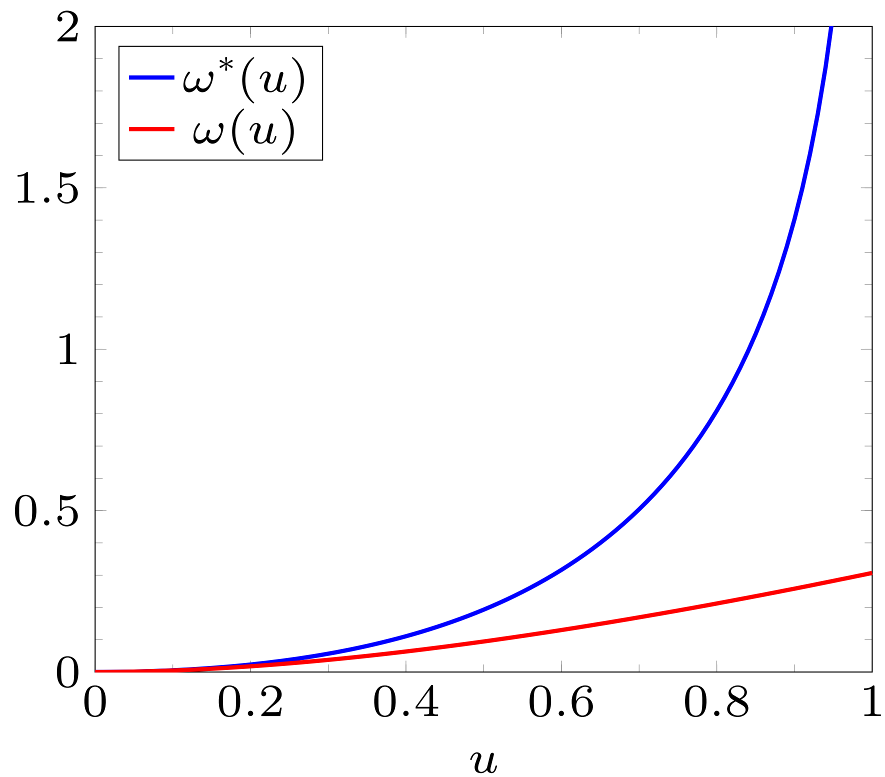
if $\|y - x\|_x < 1$, then

$$\omega(\|y-x\|_x) \;\leq\; f(y) - f(x) - \nabla f(x)^T(y-x) \;\leq\; \omega^*(\|y-x\|_x)$$

- $\omega(u)$ and $\omega^*(u)$ are defined as

$$\begin{aligned} \omega(u) &= u - \log(1+u) \\ \omega^*(u) &= -u - \log(1-u) \end{aligned}$$

- ω and ω^* are conjugates



inequalities follow from integration of (3) with $v = y - x$

Dikin ellipsoid

Definition: the ellipsoid

$$\begin{aligned} \mathcal{E}_x &= \{y \mid \|y - x\|_x \le 1\} \\ &= \left\{y \mid (y-x)^T \nabla^2 f(x)(y-x) \le 1\right\} \end{aligned}$$

is called the *Dikin ellipsoid* centered at $x \in \operatorname{dom} f$

Dikin ellipsoid theorem

$$\operatorname{int} \mathcal{E}_x \subseteq \operatorname{dom} f$$

follows from:

- the upper bound on $f(y)$ (page 16-8), which is finite for $\|y - x\|_x < 1$
- the fact that f is a closed function

Outline

- self-concordant functions
- **Newton's method**
- normal barriers

Newton decrement

Newton step at x:

$$\Delta x = -\nabla^2 f(x)^{-1} \nabla f(x)$$

Newton decrement:

$$\begin{aligned} \lambda(x) &= \left(\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)\right)^{1/2} \\ &= \|\Delta x\|_x \\ &= \|\nabla f(x)\|_{x*} \end{aligned}$$

where $\|v\|_{x*} = (v^T \nabla^2 f(x)^{-1} v)^{1/2}$ is the dual of the local norm $\|\cdot\|_x$

Feasible step size:

Dikin ellipsoid theorem implies that $x + \alpha \Delta x \in \operatorname{dom} f$ for $\alpha < 1/\lambda(x)$

Newton method

we will study the following version of Newton's method

Algorithm

select $\epsilon \in (0, 1/2)$, $\eta \in (0, 1/4]$, and a starting point $x \in \operatorname{dom} f$

repeat:

1. compute Newton step Δx and Newton decrement $\lambda(x)$
2. if $\lambda(x)^2 \leq \epsilon$, return x
3. otherwise, set $x := x + \alpha \Delta x$ with

$$\alpha = \frac{1}{1+\lambda(x)} \quad \text{if } \lambda(x) \geq \eta, \qquad \alpha = 1 \quad \text{otherwise}$$

- stopping criterion guarantees $f(x) - f(x^\star) \leq \epsilon$ (see next page)
- alternatively, can use backtracking line search from EE236B

Bound on suboptimality

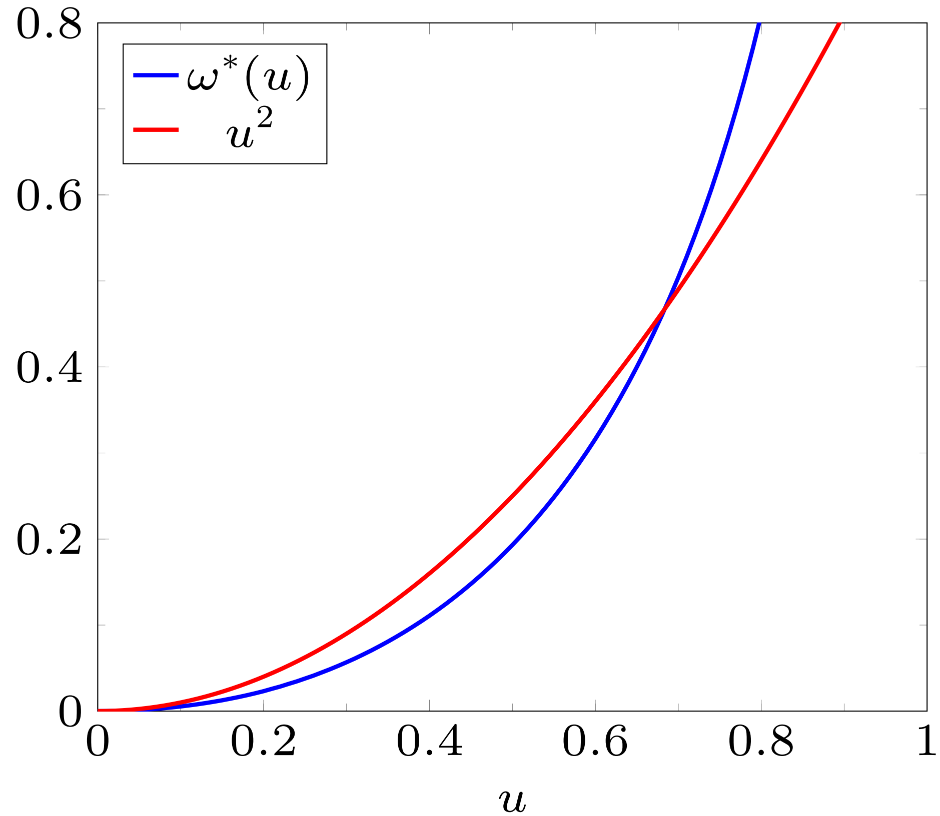
if $\lambda(x) < 1$ then f has a unique minimizer $x^\star$ and

$$\begin{aligned} f(x^\star) &\geq f(x) - \omega^*(\lambda(x)) \\ &= f(x) + \lambda(x) + \log(1 - \lambda(x)) \end{aligned}$$

in particular, if $\lambda(x) \leq 0.68$,

$$f(x) - f(x^\star) \leq \lambda(x)^2$$

useful as stopping criterion



Proof:

- from the lower bound on page 16-8, in an arbitrary direction v,

$$\begin{aligned} f(x+\alpha v) &\geq f(x) + \alpha \nabla f(x)^T v + \omega(\alpha \|v\|_x) \\ &\geq f(x) - \alpha\lambda(x)\|v\|_x + \omega(\alpha\|v\|_x) \\ &\geq f(x) - \alpha\lambda(x)\|v\|_x + \alpha\|v\|_x - \log(1+\alpha\|v\|_x) \qquad (5) \end{aligned}$$

  (second line from the Cauchy-Schwarz inequality)

- if $\lambda(x) < 1$ the r.h.s. of (5) is minimized at $\alpha\|v\|_x = \lambda(x)/(1-\lambda(x))$:

$$\begin{aligned} \inf_\alpha f(x+\alpha v) &\geq f(x) - \omega^*(\lambda(x)) \\ &= f(x) + \lambda(x) + \log(1-\lambda(x)) \end{aligned}$$

  right-hand side is a lower bound on $\inf_x f(x)$ because v is arbitrary

- if $\lambda(x) < 1$, the right-hand side of (5) grows to infinity as $\alpha\|v\|_x \to \infty$

  therefore the sublevel sets of f are bounded and f attains its minimum

- since $\nabla^2 f(x) \succ 0$ the minimizer is unique

Damped Newton step

$$x^+ = x + \frac{1}{1+\lambda(x)}\,\Delta x$$

guarantees $x^+ \in \mathrm{dom}\, f$ and

$$\begin{aligned} f(x^+) &\leq f(x) - \omega(\lambda(x)) \\ &= f(x) - \lambda(x) + \log(1+\lambda(x)) \end{aligned}$$

Consequences (for Newton algorithm on page 16-11)

- each damped Newton step decreases $f(x)$ by at least $\omega(\eta)$
- if f is bounded below, number of damped Newton iterations is finite
- if f is bounded below, its minimum is attained

  (from page 16-12, since $\lambda(x) < 1$ after a finite number of damped steps)

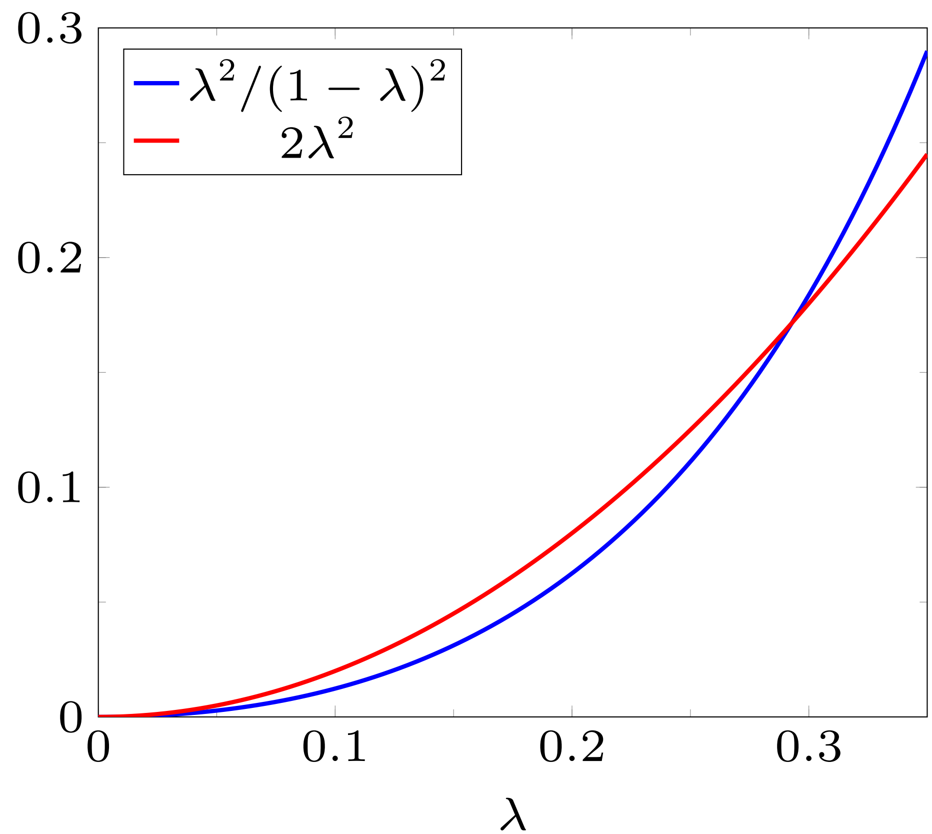
Proof: from the upper bound on page 16-8:

$$
\begin{aligned}
f(x^+) &\leq f(x) + \nabla f(x)^T(x^+ - x) + \omega^*(\|x^+ - x\|_x) \\
&= f(x) - \frac{\lambda(x)^2}{1+\lambda(x)} + \omega^*(\frac{\lambda(x)}{1+\lambda(x)}) \\
&= f(x) - \omega\left(\lambda(x)\right)
\end{aligned}
$$

Quadratic convergence

if $\lambda(x) < 1$ then $x^+ = x + \Delta x \in \mathrm{dom}\, f$ and

$$\lambda(x^+) \leq \left(\frac{\lambda(x)}{1-\lambda(x)}\right)^2$$



in particular, if $\lambda(x) \leq 0.29$,

$$\lambda(x^+) \leq 2\lambda(x)^2$$

Proof: since $\lambda(x)$ and Newton's method are affine invariant, we can assume

$$\nabla^2 f(x) = I, \qquad \Delta x = -\nabla f(x), \qquad \lambda(x) = \|\Delta x\|_2 = \|\nabla f(x)\|_2$$

- from the Hessian bounds (4), with $\nabla^2 f(x) = I$

$$(1-\lambda(x))^2 I \preceq \nabla^2 f(x^+) \preceq \frac{1}{(1-\lambda(x))^2} I$$

- by integrating the Hessian bounds (4),

$$\int_0^1 \nabla^2 f(x + \alpha\Delta x)\, d\alpha - I \preceq \frac{\lambda(x)}{1-\lambda(x)} I$$

and

$$\int_0^1 \nabla^2 f(x + \alpha\Delta x)\, d\alpha - I \succeq -\left(\lambda(x) - \frac{\lambda(x)^2}{3}\right) I \succeq -\frac{\lambda(x)}{1-\lambda(x)} I$$

therefore (with $\lambda^+ = \lambda(x^+)$, $\lambda = \lambda(x)$)

$$\begin{aligned}
\lambda^+ &= \left(\nabla f(x^+)^T \nabla^2 f(x^+)^{-1} \nabla f(x^+)\right)^{1/2} \\
&\leq \frac{1}{1-\lambda} \left\| \nabla f(x^+) \right\|_2 \\
&= \frac{1}{1-\lambda} \left\| \nabla f(x^+) - \nabla f(x) - \Delta x \right\|_2 \\
&= \frac{1}{1-\lambda} \left\| \left(\int_0^1 \nabla^2 f(x + \alpha \Delta x) d\alpha - I \right) \Delta x \right\|_2 \\
&\leq \frac{1}{1-\lambda} \frac{\lambda}{1-\lambda} \|\Delta x\|_2 \\
&= \frac{\lambda^2}{(1-\lambda)^2}
\end{aligned}$$

Summary: Newton's method

convergence results for the algorithm of page 16-11

- **damped Newton phase:** if $\lambda(x) \geq \eta$,

$$f(x^+) - f(x) \leq -\omega(\eta)$$

  function value decreases by at least a positive constant $\omega(\eta)$

- **quadratically convergent phase:** if $\lambda(x) < \eta$,

$$2\lambda(x^+) \leq (2\lambda(x))^2$$

  implies $\lambda(x^+) \leq 2\eta^2 < \eta$, and Newton decrement decreases to zero

Iteration complexity

if f is bounded below, Newton's algorithm terminates after at most

$$\frac{f(x^{(0)}) - f(x^\star)}{\omega(\eta)} + \log_2 \log_2(1/\epsilon) \text{ iterations}$$

- 1st term bounds number of iterations in damped Newton phase
- 2nd term bounds number of iterations in quadratically convergent phase:

  after k iterations in quadratically convergent phase,

$$2\lambda(x) \leq (2\eta)^{2^k} \leq \left(\frac{1}{2}\right)^{2^k}, \qquad f(x) - f(x^\star) \leq \lambda(x)^2 \leq \left(\frac{1}{2}\right)^{2^{k+1}}$$

  so $f(x) - f(x^\star) \leq \epsilon$ if $k \geq \log_2 \log_2(1/\epsilon) - 1$

Outline

- self-concordant functions
- Newton's method
- **normal barriers**

Normal barrier

Definition

ϕ is a θ-normal barrier for the proper cone K if it is

- self-concordant with domain $\mathrm{int}\, K$
- logarithmically homogeneous with parameter θ:

$$\phi(tx) = \phi(x) - \theta \log t \quad \forall x \in \mathrm{int}\, K,\ t > 0$$

Interpretation

a negative 'logarithm' for K; generalizes $\phi(x) = -\log x$ for $K = \mathbf{R}_+$

Examples

Nonnegative orthant: $K = \mathbf{R}^m_+$

$$\phi(x) = -\sum_{i=1}^{m} \log x_i \qquad (\theta = m)$$

Second-order cone: $K = \mathcal{Q}^m = \{(x, y) \in \mathbf{R}^{m-1} \times \mathbf{R} \mid \|x\|_2 \leq y\}$

$$\phi(x, y) = -\log(y^2 - x^T x) \qquad (\theta = 2)$$

Semidefinite cone: $K = \mathcal{S}^p = \{x \in \mathbf{R}^{p(p+1)/2} \mid \mathrm{mat}(x) \succeq 0\}$

$$\phi(x) = -\log\det\mathrm{mat}(x) \qquad (\theta = p)$$

Examples

Exponential cone $K_{\exp} = \mathrm{cl}\{(x,y,z) \in \mathbf{R}^3 \mid ye^{x/y} \leq z,\ y > 0\}$

$$\phi(x,y,z) = -\log\left(y\log(z/y) - x\right) - \log z - \log y \qquad (\theta = 3)$$

Power cone: $K = \{(x_1, x_2, y) \in \mathbf{R}_+ \times \mathbf{R}_+ \times \mathbf{R} \mid |y| \leq x_1^{\alpha_1} x_2^{\alpha_2}\}$

$$\phi(x,y) = -\log\left(x_1^{2\alpha_1} x_2^{2\alpha_2} - y^2\right) - \log x_1 - \log x_2 \qquad (\theta = 4)$$

Consequences of logarithmic homogeneity

- differentiate $\phi(tx) = \phi(x) - \theta \log t$ with respect to x:

$$\nabla \phi(tx) = \frac{1}{t} \nabla \phi(x), \qquad \nabla^2 \phi(tx) = \frac{1}{t^2} \nabla^2 \phi(x)$$

- differentiate $\nabla \phi(tx) = (1/t) \nabla \phi(x)$ with respect to t at $t = 1$:

$$\nabla^2 \phi(x) x = -\nabla \phi(x)$$

- differentiate $\phi(tx) = \phi(x) - \theta \log t$ with respect to t at $t = 1$:

$$\nabla \phi(x)^T x = -\theta$$

- combine the previous two properties:

$$x^T \nabla^2 \phi(x) x = \theta, \qquad \nabla \phi(x)^T \nabla^2 \phi(x)^{-1} \nabla \phi(x) = \theta$$

Strengthened lower bound from convexity

from convexity and logarithmic homogeneity, if $x, y \in \operatorname{int} K$ and $t > 0$,

$$\begin{aligned} \phi(y) &\geq \phi(tx) + \nabla\phi(tx)^T(y - tx) \\ &= \phi(x) - \theta \log t + \frac{1}{t}\nabla\phi(x)^T y + \theta \end{aligned}$$

- implies $\nabla\phi(x)^T y < 0$ (otherwise $t \to 0$ gives contradiction)
- maximizing right-hand side over t gives

$$\phi(y) \geq \phi(x) - \theta \log \frac{-\nabla\phi(x)^T y}{\theta} \qquad \forall x, y \in \operatorname{int} K$$

note: this improves the inequality $\phi(y) \geq \phi(x) + \nabla\phi(x)^T(y - x)$

Gradient of normal barrier

$$-\nabla\phi(x) \in \mathbf{int}\, K^* \qquad \forall x \in \mathbf{int}\, K$$

- from previous page, $-\nabla\phi(x)^T y > 0$ for all $y \in \mathbf{int}\, K$; hence $-\nabla\phi(x) \in K^*$
- $-\nabla\phi(x)$ cannot be in the boundary of K^* because $\nabla^2\phi(x) \succ 0$

  (otherwise $\nabla\phi(x+u) \approx \nabla\phi(x) + \nabla^2\phi(x)u \notin K^*$ for some small u)

conversely, every $y \in \mathbf{int}\, K^*$ can be written as

$$y = -\nabla\phi(x)$$

for some (unique) $x \in \mathbf{int}\, K$ (namely, the minimizer of $y^T x + \phi(x)$)

Dual barrier

Definition

$$\phi_*(y) = \sup_{x \in \mathrm{int}\, K} (-y^T x - \phi(x))$$

(we use a subscript in ϕ_* to distinguish from conjugate $\phi^*(y) = \phi_*(-y)$)

it can be shown that this is a normal barrier for K^*

- ϕ_* is self-concordant
- $\mathrm{dom}\, \phi_* = \{-\nabla\phi(x) \mid x \in \mathrm{int}\, K\} = \mathrm{int}\, K^*$
- logarithmically homogeneous with degree θ: $\phi_*(ty) = \phi_*(y) - \theta \log t$

Gradient and Hessian of dual barrier

define

$$\hat{x}(y) = \underset{x}{\operatorname{argmin}} \, (y^T x + \phi(x))$$

- the (unique) maximizer in the definition of ϕ_*
- satisfies $\nabla \phi(\hat{x}(y)) = -y$

Gradient (from properties of conjugate)

$$\nabla \phi_*(y) = -\hat{x}(y)$$

Hessian (by differentiating $\nabla \phi(\hat{x}(y)) = -y$ with respect to y)

$$\nabla^2 \phi_*(y) = \nabla^2 \phi(\hat{x}(y))^{-1}$$

References

- Yu. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming* (1994).

  introduced a more general definition of self-concordance; the s.c. functions in this lecture correspond to nondegenerate ($\nabla^2 f(x) \succ 0$), standard ($a = 1$), strongly (f closed) self-concordant functions in the book

- Yu. Nesterov, *Introductory Lectures on Convex Optimization. A Basic Course* (2004), chapter 4.

- S. Boyd, L. Vandenberghe, *Convex Optimization* (2004), §9.6.

  explains why the results of Newton's method extend to backtracking line search