15. Barrier functions

- self-concordant functions
- Newton’s method
- normal barriers
Self-concordant functions

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

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

$$\left. \frac{d}{d\alpha} \nabla^2 f(x + \alpha v) \right|_{\alpha=0} \leq 2\|v\|_x \nabla^2 f(x)$$

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

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

two-sided matrix inequality

\[-2\|v\|_x \nabla^2 f(x) \leq \left. \frac{d}{d\alpha} \nabla^2 f(x + \alpha v) \right|_{\alpha=0} \leq 2\|v\|_x \nabla^2 f(x) \quad (1)\]

lower bound follows from the upper bound applied to \(-v\)

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

\[-2g''(\alpha)^{3/2} \leq g'''(\alpha) \leq 2g''(\alpha)^{3/2} \quad (2)\]

• 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 \( \mathbb{R}_{++} \)

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 \( af \) is self-concordant for \( a \geq 1 \)
- 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 $g(\alpha) = f(x + \alpha v)$

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

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

• bounds on Hessian

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

these inequalities hold for $0 \leq \alpha \|v\|_x < 1$
proof of first pair of inequalities

• from (2) on page 15-3,

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

• 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

\[\frac{g''(0)}{(1 + \alpha g''(0)^{1/2})^2} \leq g''(\alpha) \leq \frac{g''(0)}{(1 - \alpha g''(0)^{1/2})^2}\]
**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 15-3

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

- therefore, from (3) on page 15-5,

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

- 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

$$\omega(u) = u - \log(1 + u)$$

$$\omega^*(u) = -u - \log(1 - u)$$

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

**definition:** the ellipsoid

\[ E_x = \{ y \mid \|y - x\|_x \leq 1 \} = \{ y \mid (y - x)^T \nabla^2 f(x)(y - x) \leq 1 \} \]

is called the *Dikin ellipsoid* centered at \( x \in \text{dom } f \)

**Dikin ellipsoid theorem**

\[ \text{int } E_x \subseteq \text{dom } f \]

follows from the upper bound on \( f(y) \) (page 15-8), which is finite for \( \|y - x\|_x < 1 \), and 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**

$$\lambda(x) = \left( (\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)) \right)^{1/2}$$

$$= \|\Delta x\|_x$$

$$= \|\nabla f(x)\|_{x^*}$$

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 $x + \alpha \Delta x \in \text{dom } f$ for $\alpha < 1/\lambda(x)$
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 $x \in \text{dom } f$

repeat:

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

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

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

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

$$f(x^*) \geq f(x) - \omega^*(\lambda(x)) = f(x) + \lambda(x) + \log(1 - \lambda(x))$$

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

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

useful as stopping criterion
proof:

• from the lower bound on p.15-8, in an arbitrary direction $v$,

$$ f(x + \alpha v) \geq f(x) + \alpha \nabla f(x)^T v + \omega(\alpha \|v\|_x) $$

$$ \geq f(x) - \alpha \lambda(x) \|v\|_x + \alpha \|v\|_x - \log(1 + \alpha \|v\|_x) \quad \text{(5)} $$

(Second line from the Cauchy-Schwarz inequality)

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

$$ \inf_{\alpha} f(x + \alpha v) \geq f(x) + \lambda(x) + \log(1 - \lambda(x)) $$

r.h.s. is a lower bound on $\inf_x f(x)$ because $v$ is arbitrary

• if $\lambda(x) < 1$, the r.h.s. 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 \( x^+ \in \text{dom} \, f \) and

\[
\begin{align*}
    f(x^+) & \leq f(x) - \omega(\lambda(x)) \\
    &= f(x) - \lambda(x) + \log(1 + \lambda(x))
\end{align*}
\]

**consequences** (for Newton algorithm on page 15-11)

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

(from p. 15-12, since \( \lambda(x) < 1 \) after a finite number of damped steps)
\textit{proof:} from the upper bound on page 15-8:

\begin{align*}
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^*\left(\frac{\lambda(x)}{1 + \lambda(x)}\right) \\
& = f(x) - \omega(\lambda(x))
\end{align*}
Quadratic convergence

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

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

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

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

• 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 $\lambda^+ = \lambda(x^+)$, $\lambda = \lambda(x)$)

$$
\lambda^+ = \left( \nabla f(x^+)^T \nabla^2 f(x^+)^{-1} \nabla f(x^+) \right)^{1/2} \\
\leq \frac{1}{1 - \lambda} \left\| \nabla f(x^+) \right\|_2 \\
= \frac{1}{1 - \lambda} \left\| \nabla f(x^+) - \nabla f(x) - \Delta x \right\|_2 \\
= \frac{1}{1 - \lambda} \frac{\lambda}{1 - \lambda} \left\| \Delta x \right\|_2 \\
= \frac{\lambda^2}{(1 - \lambda)^2}
$$
Summary: Newton’s method

convergence results for the algorithm of page 15-11

- **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

\[
\frac{f(x^{(0)}) - f(x^\star)}{\omega(\eta)} + \log_2 \log_2 (1/\epsilon)
\]

iterations

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

  after \( k \) iterations in quadratically convergent phase,

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

  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 $\text{int } K$
- logarithmically homogeneous with parameter θ:

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

interpretation

a negative ‘logarithm’ for $K$; generalizes $\phi(x) = -\log x$ for $K = \mathbb{R}_+$
Examples

nonnegative orthant: \( K = \mathbb{R}_+^m \)

\[ \phi(x) = - \sum_{i=1}^{m} \log x_i \quad (\theta = m) \]

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

\[ \phi(x, y) = - \log(y^2 - x^T x) \quad (\theta = 2) \]

semidefinite cone: \( K = S^p = \{x \in \mathbb{R}^{p(p+1)/2} \mid \text{mat}(x) \succeq 0\} \)

\[ \phi(x) = - \log \det \text{mat}(x) \quad (\theta = p) \]
exponential cone: \( K_{\exp} = \text{cl}\{(x, y, z) \in \mathbb{R}^3 \mid ye^{x/y} \leq z, \ y > 0\} \)

\[
\phi(x, y, z) = -\log (y \log(z/y) - x) - \log z - \log y \quad (\theta = 3)
\]

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

\[
\phi(x, y) = -\log \left(x_1^{2\alpha_1} x_2^{2\alpha_2} - y^2\right) - \log x_1 - \log x_2 \quad (\theta = 4)
\]
Consequences of logarithmic homogeneity

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

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

- 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:

  $$x^T \nabla^2 \phi(x)x = \theta, \quad \nabla \phi(x)^T \nabla^2 \phi(x)^{-1} \nabla \phi(x) = \theta$$
Strengthened lower bound from convexity

from convexity and logarithmic homogeneity, if \( x, y \in \text{int} \ K \) and \( t > 0 \),

\[
\phi(y) \geq \phi(tx) + \nabla \phi(tx)^T (y - tx)
\]

\[
= \phi(x) - \theta \log t + \frac{1}{t} \nabla \phi(x)^T y + \theta
\]

• implies \( \nabla \phi(x)^T y < 0 \) (otherwise \( t \to 0 \) gives contradiction)

• maximizing right-hand side over \( t \) gives

\[
\phi(y) \geq \phi(x) - \theta \log \frac{-\nabla \phi(x)^T y}{\theta} \quad \forall x, y \in \text{int} \ K
\]

note: this improves the inequality \( \phi(y) \geq \phi(x) + \nabla \phi(x)^T (y - x) \)
Gradient of normal barrier

\[-\nabla \phi(x) \in \text{int } K^* \quad \forall x \in \text{int } K\]

- from previous page, \(-\nabla \phi(x)^T y > 0 \ \forall y \in \text{int } K\); 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 \(y \in \text{int } K^*\) can be written as

\[y = -\nabla \phi(x)\]

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

**Definition**

\[
\phi_*(y) = \sup_{x \in \text{int } K} (-y^T x - \phi(x))
\]

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

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

- \(\phi_*\) is self-concordant

- \(\text{dom } \phi_* = \{-\nabla \phi(x) \mid x \in \text{int } K\} = \text{int } K^*\)

- logarithmically homogeneous with degree \(\theta\):
  \[
  \phi_*(ty) = \phi_*(y) - \theta \log t
  \]
Gradient and Hessian of dual barrier

\[ \hat{x}(y) = \arg\min_{x} (y^T x + \phi(x)) \]

- the (unique) maximizer in the definition of \( \phi_* \)

- 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 \( y \))

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

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


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