Non-linear Optimization (2DME20), lecture 6

Gerhard Woeginger

Technische Universiteit Eindhoven

Fall 2015, Q1

Program for this week

- Some convex programs (second-order cone; semi-definite)
- Cones, dual cones, proper cones
- Generalized inequalities
- Convex functions (generalized)

- Lagrangian duality (generalized)
- Convex Lagrangian duality (generalized)
- Descent methods for unconstrained optimization

Recall, recall, recall (1)

Non-linear (continuous) optimization problem

$$\begin{array}{lll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \leq 0 & i = 1, \ldots, r \\ & h_i(x) = 0 & i = 1, \ldots, s \\ & x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n & \end{array}$$

- necessary optimality condition (KKT conditions)
- no proper duality theorem (duality gap)
- interior point algorithms that find locally optimal solution
  ... in an unknown amount of time
- Virtually any problem can be cast as a nonlinear optimization problem

Recall, recall, recall (2)

Convex (continuous) optimization problem

$f_0, \ldots, f_r$ are convex and $h_1, \ldots, h_s$ are affine

$$
\begin{array}{lll}
\text{minimize} & f_0(x) & \\
\text{subject to} & f_i(x) \leq 0 & i = 1, \ldots, r \\
 & h_i(x) = 0 & i = 1, \ldots, s \\
 & x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n &
\end{array}
$$

- necessary and sufficient optimality condition (KKT conditions)
- duality theorem (no duality gap)
- interior point algorithms that find globally optimal solution

More convex problems (1)

Least-squares

For matrix A and vector b solve

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|^2 \\ \text{subject to} & x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \end{array}$$

- necessary and sufficient optimality condition
- closed-form solution $x^* = (A^T A)^{-1} A^T b$
- finding an optimal solution is no harder than solving linear equations

Applications:

regression analysis, optimal control, parameter estimation, data fitting

More convex problems (2)

Linear programming

For matrix A and vectors b, c solve

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax \leq b \\ & x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \end{aligned}$$

- necessary and sufficient optimality condition
- duality theorem
- simplex algorithm finds optimal solution
- interior point algorithm approximates a solution efficiently

More convex problems (3)

Second-order cone optimization

For matrices $A_1, \ldots, A_r$, vectors $b_1, \ldots, b_r, c_1, \ldots, c_r$, reals $d_1, \ldots, d_r$, for matrix F, and vectors f, g solve

$$
\begin{array}{lll}
\text{minimize} & f^T x & \\
\text{subject to} & \|A_i x + b_i\| \le c_i^T x + d_i & i = 1, \ldots, r \\
 & Fx = g & \\
 & x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n &
\end{array}
$$

- generalizes least-squares
- generalizes linear programming
- solved efficiently by interior point method

Example

Second-order cone constraint: $\|[x,\ y]^T\| \ \le\ 2x$

More convex problems (4)

Semi-definite optimization

For matrices $A_0, \ldots, A_n$ and vector c, solve

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & A_0 + x_1 A_1 + \cdots + x_n A_n \succeq 0 \\ & x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \end{aligned}$$

- generalizes second-order cone optimization
  (and hence least-squares and linear programming)
- solved efficiently by interior point method

Applications:

- stability of dynamical systems,
- fitting a maximum-volume ellipsoid inside a polyhedron,
- minimizing a univariate polynomial

Recall from linear algebra (1)

Definition

The inner product of two matrices $A, B \in S^n$ is

$$\langle A, B \rangle := \sum_i \sum_j A_{ij} B_{ij}.$$

As usual for the inner product

- $\langle A, B \rangle = \langle B, A \rangle$
- $\langle \alpha A + \beta B, C \rangle \;=\; \alpha \langle A, C \rangle + \beta \langle B, C \rangle$

Definition

The trace of a square matrix A is $\mathrm{tr}(A) := \sum_i A_{ii}$.

Note that $\langle A, B \rangle = \mathrm{tr}(AB^T)$. The latter is used throughout in the book.
Hence $\mathrm{tr}((\alpha A + \beta B)C) = \alpha\, \mathrm{tr}(AC) + \beta\, \mathrm{tr}(BC)$, etc.

Recall from linear algebra (2)

As a linear space, S^n is isomorphic to $\mathbb{R}^{n(n+1)/2}$.

An affine subspace L in S^n can be given by linear equations as

$$L \;=\; \{X \in S^n \mid \mathrm{tr}(A_1 X) = b_1, \ldots, \mathrm{tr}(A_m X) = b_m\},$$

or alternatively as

$$L \;=\; \{F_0 + x_1 F_1 + \cdots + x_k F_k \mid x_1, \ldots, x_k \in \mathbb{R}\}.$$

Note that if $A, B \in S^n_+$, then $\mathrm{tr}(AB) \geq 0$. Actually:

Lemma

Let $A \in S^n$. Then $A \in S^n_+$ if and only if $tr(AB) \geq 0$ for all $B \in S^n_+$.

More about cones (1)

A set $K \subseteq \mathbb{R}^n$ is a cone, if $\alpha x + \beta y \in K$ for all $x, y \in K$ and all $\alpha, \beta \geq 0$.

Definition

The dual cone for K is

$$K^* \;:=\; \{y \in \mathbb{R}^n \mid x^T y \geq 0 \text{ for all } x \in K\}$$

So $y \in K^*$ if and only if $\{x \mid y^T x < 0\} \cap K = \emptyset$

Note that

- $K_1 \subseteq K_2$ implies $K_2^* \subseteq K_1^*$
- K^* is closed
- If K is closed then $K^{**} = K$

More about cones (2)

Example

For a subspace $L \subseteq \mathbb{R}^n$, the dual cone L^*
is the orthogonal complement $\{y \mid x^T y = 0 \text{ for all } x \in L\}$.

Example

The following cones happen to be self-dual.

- $(\mathbb{R}^n_+)^* = \mathbb{R}^n_+$
- $(L^n)^* = L^n$
- $(S^n_+)^* = S^n_+$

More about cones (3)

Definition

A cone $K \subseteq \mathbb{R}^n$ is proper if

- K is closed
- K is solid (that is, K has non-empty interior)
- K does not fully contain any line ($x \in K$ and $-x \in K$ imply $x = 0$)

Example

- the nonnegative orthant $\mathbb{R}^n_+$ is a proper cone
- the Lorentz cone L^{n+1} is a proper cone
- the cone S^n_+ of PSD matrices is a proper cone
- the cone P^n of nonnegative polynomials of degree n is proper:

$$P^n = \{(p_0, \dots, p_n) \mid 0 \le p_0 + p_1 x + \cdots + p_n x^n \text{ for all } x \in \mathbb{R}\}$$

Generalized inequalities (1)

Definition

Every proper cone $K \subseteq \mathbb{R}^n$ determines
a generalized inequality $\preceq_K$ and a generalized strict inequality $\prec_K$:

For all $x, y \in \mathbb{R}^n$: $\quad x \preceq_K y \iff y - x \in K$
For all $x, y \in \mathbb{R}^n$: $\quad x \prec_K y \iff y - x \in \operatorname{int} K$

Example

The nonnegative orthant $\mathbb{R}^n_+$ yields componentwise inequality

Example

The cone S^n_+ of PSD matrices yields the usual matrix inequality

$$X \preceq Y \iff Y - X \succeq 0$$
$$X \prec Y \iff Y - X \succ 0$$

Generalized inequalities (2)

Some useful properties:

- Since K is a cone, relation $\preceq_K$ is transitive:
  if $x \preceq_K y$ and $y \preceq_K z$ then $x \preceq_K z$
- Since K does not fully contain a line, relation $\preceq_K$ has antisymmetry:
  if $x \preceq_K y$ and $y \preceq_K x$ then $x = y$
- $\preceq_K$ is reflexive: $x \preceq_K x$

- Relation $\preceq_K$ is preserved under addition:
  if $x \preceq_K y$ and $u \preceq_K v$ then $x + u \preceq_K y + v$
- Relation $\preceq_K$ is preserved under positive scaling:
  if $x \preceq_K y$ and $\alpha \in \mathbb{R}^+$ then $\alpha x \preceq_K \alpha y$

Generalized inequalities (3)

Observation

By definition of the dual cone K^*, we have:

- if $x \succeq_K 0$ and $y \succeq_{K^*} 0$ then $y^T x \geq 0$
- $x \preceq_K y$ if and only if $z^T x < z^T y$ for all $z \succeq_{K^*} 0$

Theorem of alternatives & generalized inequalities (1)

Theorem

Let A be an $m \times n$ matrix, let $b \in \mathbb{R}^m$, and let $K \subseteq \mathbb{R}^m$ be a proper cone. Then exactly one of the following two alternatives holds:

(1) *There exists a vector $x \in \mathbb{R}^n$ such that $Ax \prec_K b$*

(2) *There exists a non-zero $y \succeq_{K^*} 0$ such that $y^T A = 0$ and $y^T b \leq 0$*

- If (1) does not hold, then $\{b - Ax \mid x \in \mathbb{R}^n\}$ and $\text{int}\, K$ disjoint
- Separating hyperplane:
  $y^T(b - Ax) \leq \mu$ for all x, and $y^T z \geq \mu$ for all $z \in \text{int}\, K$
- Then $\mu \leq 0$ and $y \in K^*$, and hence (2)

- If (2) does hold, then (1) cannot hold:
- Otherwise $b - Ax \succ_K 0$ and $y \succeq_{K^*} 0$ imply $y^T(b - Ax) > 0$
  But $y^T b \leq 0$ and $y^T A = 0$ imply $y^T(b - Ax) \leq 0$

Theorem of alternatives & generalized inequalities (2)

Example

Does there exist a number $x \in \mathbb{R}$ so that

$$x \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \preceq \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$$

Consider the matrix $Y = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \succeq 0$ with

- $\mathrm{tr}(Y \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}) = 0$
- $\mathrm{tr}(Y \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}) = -1$

If x were a feasible solution, then

$$0 \;\leq\; \mathrm{tr}(Y(\begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} - x \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix})) = -1 - x \cdot 0$$

Homework 6a

Read the following sections in Boyd & Vandenberghe
1, 2.4–2.6, 4.1–4.4, 4.6

Recommended exercises (Boyd & Vandenberghe):
4.8, 4.10, 4.11; 5.6; 2.30, 2.31, 2.32, 2.33, 2.35;

Attention!
Weeks 6-7 (Oct 6; Oct 9; Oct 13; Oct 16):

- all lectures and instructions in flux 1.06

- tuesday: 1+2 instructions; 3+4 lecture
- friday: 5+6 lecture

Convex functions & generalized inequalities (1)

Let $K \subseteq \mathbb{R}^m$ be a proper cone.

Definition

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is K-convex if

$$f(\alpha x + \beta y) \ \preceq_K \ \alpha f(x) + \beta f(y)$$

for all $x, y \in \mathbb{R}^n$ and $\alpha, \beta \geq 0$ with $\alpha + \beta = 1$.

If $K = \mathbb{R}^m_+$, then f is K-convex
if and only if each component f_i is a convex function

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}^m$. Then f is K-convex if and only if

$$g(x) \ := \ z^T f(x)$$

is convex (in the ordinary one-dimensional sense) for each $z \in K^*$.

Convex functions & generalized inequalities (2)

Example

For $A_0, \ldots, A_n \in S^m$, the function $f : \mathbb{R}^n \to S^m$ such that

$$f : x \mapsto A_0 + x_1 A_1 + \cdots + x_n A_n$$

is affine, and hence is S^m_+-convex.

Example

The function $f : \mathbb{R}^{n \times m} \to S^n$ such that

$$f : X \mapsto XX^T$$

is S^n_+-convex.

Conic optimization (1)

Basic scenario:

- Let $f_0 : \mathbb{R}^n \to \mathbb{R}$
- Let $f_i : \mathbb{R}^n \to \mathbb{R}^{k_i}$, where $i = 1, \ldots, r$
- Let $h_i : \mathbb{R}^n \to \mathbb{R}$, where $i = 1, \ldots, s$
- Let $K_1, \ldots, K_r$ be proper cones, where $K_i \subseteq \mathbb{R}^{k_i}$

Basic optimization problem

$$\begin{array}{lll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \preceq_{K_i} 0 & i = 1, \ldots, r \\ & h_i(x) = 0 & i = 1, \ldots, s \end{array}$$

If every K_i is the nonnegative orthant $\mathbb{R}_+^{k_i}$
then we are back to the scenario discussed in week 3.

Conic optimization (2)

Besides the classical nonnegative orthant cone, the most important cones in conic optimization are:

- the Lorentz cone $L^{n+1} = \{(x,t) \mid x \in \mathbb{R}^n,\ t \in \mathbb{R},\ \|x\| \leq t\}$
- the cone S^n_+ of positive semi-definite matrices

Example

The second-order cone optimization problem

$$\min\ x \quad \text{subject to}\ \ [x-y,\ 1,\ x+y]^T \succeq_{L^3} 0$$

is equivalent to

$$\min\ x \quad \text{subject to}\ \ 4xy \geq 1\ \ \text{and}\ \ x+y \geq 0$$

Compare the material on the following slides
to our discussion of Lagrangian duality in week 3

Lagrangian duality (1)

$$\begin{array}{lll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \preceq_{K_i} 0 & i = 1, \ldots, r \\ & h_i(x) = 0 & i = 1, \ldots, s \end{array}$$

Definition

The Lagrangian $L(x, \lambda, \mu) \; = \; f_0(x) + \sum_{i=1}^{r} \lambda_i^T f_i(x) + \sum_{i=1}^{s} \mu_i h_i(x)$

Lemma

If $\lambda_i \in K_i^*$ for each i, then

$$f_0(x) \geq L(x, \lambda, \mu)$$

for any x such that all $f_i(x) \preceq_{K_i} 0$ and $h_i(x) = 0$.

Lagrangian duality (2)

Let $p^* :=$

$$\begin{array}{lll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \preceq_{K_i} 0 & i = 1, \ldots, r \\ & h_i(x) = 0 & i = 1, \ldots, s \end{array}$$

The Lagrange dual function of this problem is:

$$g(\lambda, \mu) := \min_x L(x, \lambda, \mu)$$

Lemma

If $\lambda_i \in K_i^*$ for each i, then

$$p^* \geq g(\lambda, \mu)$$

Lagrangian duality (3)

Consider the (primal) problem

$$\begin{array}{lll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \preceq_{K_i} 0 & i = 1, \ldots, r \\ & h_i(x) = 0 & i = 1, \ldots, s \end{array}$$

with value p^*.

Its Lagrange dual is the problem

$$\begin{array}{lll} \text{maximize} & g(\lambda, \mu) & \\ \text{subject to} & \lambda_i \in K_i^* & i = 1, \ldots, r \end{array}$$

with value d^*.

Lemma

We have $p^* \geq d^*$.

Lagrangian duality (4a)

Example

The second-order cone optimization problem

$$\min \ y \quad \text{subject to} \ [x,\ y,\ x]^T \succeq_{L^3} 0$$

is equivalent to

$$\min \ y \quad \text{subject to} \ y = 0 \ \text{and} \ x \geq 0$$

The Lagrangian is the function $L : \mathbb{R}^3 \to \mathbb{R}$ given by

$$L(x, y, \lambda) = y + \lambda \left(x - \sqrt{x^2 + y^2} \right)$$

The Lagrange dual function is $g(\lambda) = -\infty$.

Hence $p^* = 0$ and $d^* = -\infty$

Lagrangian duality (4b)

Example

The semi-definite optimization problem

$$\min \; y \quad \text{subject to} \quad \begin{bmatrix} 1+y & 0 & 0 \\ 0 & x & y \\ 0 & y & 0 \end{bmatrix} \succeq_{S^3_+} 0$$

is equivalent to

$$\min \; y \quad \text{subject to} \quad x \geq 0 \;\text{ and }\; y = 0$$

The Lagrangian is the function $L : \mathbb{R}^2 \times S^3 \to \mathbb{R}$ given by

$$L(x, y, Z) = y - \mathrm{tr}(Z \begin{bmatrix} 1+y & 0 & 0 \\ 0 & x & y \\ 0 & y & 0 \end{bmatrix})$$

The dual function is $g(Z) = \begin{cases} -z_{11} & \text{if } z_{22} = 0 \text{ and } z_{11} + 2z_{23} = 1 \\ -\infty & \text{otherwise} \end{cases}$

So the dual is $\max -z_{11}$ subject to $z_{22} = 0,\; z_{11} + 2z_{23} = 1,\; Z \succeq 0$.

Conic convex Lagrangian duality (1)

Let $p^* :=$

$$\begin{array}{lll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \preceq_{K_i} 0 & i = 1, \ldots, r \\ & h_i(x) = 0 & i = 1, \ldots, s \end{array}$$

and $d^* :=$

$$\begin{array}{lll} \text{maximize} & g(\lambda, \mu) & \\ \text{subject to} & \lambda_i \in K_i^* & i = 1, \ldots, r \end{array}$$

Theorem (strong duality for convex optimization)

Suppose (convex program) and (Slater's condition is satisfied):

- f_0 convex; f_i is K_i-convex; and $h_1, \ldots, h_s$ are affine
- $\exists y : \; f_i(y) \prec_{K_i} 0$ for $i = 1, \ldots, r$, and
  $h_i(y) = 0$ for $i = 1, \ldots, s$

Then $p^* = d^*$.

Conic convex Lagrangian duality (2)

Assume $s = 0$.

Sketch of proof (strong duality for convex optimization)

Consider $\mathcal{A} := \{ \begin{bmatrix} u \\ t \end{bmatrix} \mid \exists x : f_i(x) \preceq_{K_i} u_i,\ i = 1, \ldots, r,\ f_0(x) \leq t \}$.

- $\mathcal{A}$ is a convex set
- $p^* = \min\{t \mid \begin{bmatrix} 0 \\ t \end{bmatrix} \in \mathcal{A}\}$
- some hyperplane supports $\mathcal{A}$ at $(0, p^*)$; say

$$\begin{bmatrix} \lambda^* \\ \mu^* \end{bmatrix}^T \begin{bmatrix} u \\ t \end{bmatrix} \geq \alpha \text{ for } \begin{bmatrix} u \\ t \end{bmatrix} \in \mathcal{A}; \qquad \begin{bmatrix} \lambda^* \\ \mu^* \end{bmatrix}^T \begin{bmatrix} 0 \\ p^* \end{bmatrix} = \alpha$$

- $\mu^* \geq 0$, $\lambda_i \in K_i^*$
- If $\mu^* = 0$, then $0 > \sum_i (\lambda_i^*)^T f_i(y) \geq \alpha = 0$; contradiction.
- If $\mu^* > 0$, then $g(\lambda^*/\mu^*) = p^*$.

Duality examples (1)

Semi-definite optimization

Consider the semi-definite optimization problem

$$\min\{c^T x \mid A_0 + x_1 A_1 + \cdots + x_n A_n \preceq 0\}$$

The Lagrangian is the function $L : \mathbb{R}^n \times S^k \to \mathbb{R}$ given by

$$L(x, Z) = c^T x + \mathrm{tr}(Z(A_0 + x_1 A_1 + \cdots + x_n A_n))$$

The Lagrange dual function is

$$g(Z) = \min_x L(x, Z) = \begin{cases} \mathrm{tr}(ZA_0) & \text{if } \mathrm{tr}(ZA_i) + c_i = 0 \text{ for all } i \geq 1 \\ -\infty & \text{otherwise} \end{cases}$$

So the dual is

$$\max\{\mathrm{tr}(ZA_0) \mid Z \succeq 0,\ \mathrm{tr}(ZA_i) + c_i = 0 \text{ for all } i \geq 1\}$$

We have strong duality.

Duality examples (2)

Cone program in standard form

For a proper cone $K \subseteq \mathbb{R}^n$, consider the cone program

$$\min\{c^T x \mid Ax = b,\ x \succeq_K 0\}$$

The Lagrangian is

$$L(x, \lambda, \mu) = c^T x - \lambda^T x + \mu^T (b - Ax)$$

The Lagrange dual function is

$$g(\lambda, \mu) = \min_x L(x, \lambda, \mu) = \begin{cases} \mu^T b & \text{if } c^T = \mu^T A + \lambda^T \\ -\infty & \text{otherwise} \end{cases}$$

So the dual is

$$\max\{\mu^T b \mid A^T \mu + \lambda = c,\ \lambda \succeq_{K^*} 0\}$$

We have strong duality.

Duality examples (3)

Second-order cone optimization

Consider the second-order cone optimization problem

$$\min\{f^T x \mid \ \|A_i x + b_i\| \leq c_i^T x + d_i, \ i = 1, \ldots, r\}$$

The Lagrangian is given by

$$L(x, \lambda, u_1, \ldots, u_r) = f^T x + \sum u_i^T (A_i x + b_i) - \lambda_i (c_i^T x + d_i)$$

The dual is (see Exercise 5.43)

$$\max\{d^T \lambda + \sum u_i^T b_i \mid \sum A_i^T u_i + \lambda_i c_i + f = 0, \ \|u_i\| \leq \lambda_i, \ \text{ for } i \geq 1\}$$

We have strong duality.

Unconstrained optimization (1)

Unconstrained convex minimization problem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex and twice differentiable function,
and consider the problem minimize $f(x)$

As f is convex and differentiable,

$$f(x^*) = \min_x f(x) \iff \nabla f(x^*) = 0$$

Sometimes, the latter equation may be solved analytically:

Example

The quadratic convex function $f(x) = \frac{1}{2}x^T Ax + b^T x + c$
has the minimizer x^* with $Ax^* + b = \nabla f(x^*) = 0$.

Unconstrained optimization (2)

Iterative algorithms

An iterative algorithm for the optimization problem $p^* = \min_x f(x)$ generates a sequence $x^{(0)}, x^{(1)}, \ldots$ of points,
so that $f(x^{(k)}) \to p^*$ as $k \to \infty$.

In the k-th step, a typical iterative algorithm

- chooses a search direction $\Delta x^{(k)} \in \mathbb{R}^n$
- chooses a *step size* $t^{(k)} \in \mathbb{R}$
- puts $x^{(k+1)} = x^{(k)} + t^{(k)} \Delta x^{(k)}$

Definition

The search direction Δx is a descent direction
if $(\nabla f(x))^T \Delta x < 0$.

Then there exists a t so that $f(x + t\Delta x) < f(x)$.

Unconstrained optimization (3)

Example

Examples of descent directions are

- $\Delta x_{gd} = -\nabla f(x)$ (gradient descent)
- $\Delta x_{sd} = \arg\min\{(\nabla f(x))^T v \mid \|v\| \leq 1\}$ (steepest descent)
- $\Delta x_{nt} = -(\nabla^2 f(x))^{-1} \nabla f(x)$ (Newton descent)

The Newton direction is affine-invariant:
If we set $g(y) = f(Ay)$,
then for $x = Ay$ we have $\Delta_{nt} x = A^T \; \Delta_{nt} y$
where $\Delta x_{nt} = -(\nabla^2 f(x))^{-1} \nabla f(x)$
and $\Delta y_{nt} = -(\nabla^2 g(y))^{-1} \nabla g(y)$

Unconstrained optimization (4)

Line search

Given point x and a descent direction Δx,
line search is the problem of choosing the value $t \in \mathbb{R}$
for generating the next point $x^+ = x + t\Delta x$

Goal: choose t so that $f(x^+)$ is small.

Exact line search

Choose $t = \arg\min\{f(x + t\Delta x) \mid t \in \mathbb{R}\}$

Backtracking line search (trades off thoroughness for speed)

Pick α and β with $\alpha < 1/2$ and $0 < \beta < 1$; initialize $t := 1$

While $f(x + t\Delta x) \; > \; f(x) + \alpha t \nabla f(x)\Delta x$
do $t := \beta t$

Recall, recall, recall (1): the logarithmic barrier

- Indicator function $I_-(u) = \begin{cases} 0 & \text{if } u \le 0 \\ \infty & \text{otherwise} \end{cases}$
- Approximate indicator function $\widehat{I}_-(u) = \begin{cases} -\frac{1}{t}\log(-u) & \text{if } u < 0 \\ \infty & \text{otherwise} \end{cases}$
- $\widehat{I}_-(u)$: convex; non-decreasing; differentiable; closed sublevel sets
- The problem $\min\{f_0(x) \mid f_i(x) \le 0, i = 1\ldots, r,\ Ax = b\}$ can be approximated by $\min\{f_0(x) + \frac{1}{t}\phi(x) \mid Ax = b\}$
- Logarithmic barrier $\phi(x) := \sum_{i=1}^{r} -\log(-f_i(x))$
- Central path $x^*(t) := \arg\min\{f_0(x) + \frac{1}{t}\phi(x) \mid Ax = b\}$
- As $t \to \infty$, the central path leads to the optimum

Recall, recall, recall (2): the barrier method

In order to approximate $p^* = \min\{f_0(x) \mid f_i(x) \leq 0, i = 1 \ldots, r,\ Ax = b\}$ within error $\varepsilon > 0$, it suffices to solve the unconstrained problem

$$y^*(t) \;=\; \arg\min\{f_0(Wy+v) + \frac{1}{t}\phi(Wy+v)\}$$

with $t = r/\varepsilon$.

The barrier method

Given a strictly feasible $y = y^{(0)}$ and $t = t^{(0)} > 0$, do

1. compute $y^*(t)$, starting the solution algorithm with y (centering)
2. put $y \leftarrow y^*(t)$ (update)
3. if $r/t < \varepsilon$ then quit; else put $t \leftarrow \mu t$ and repeat (increase)

The centering step uses the Newton method

Recall, recall, recall (3): the barrier method

Example: Linear programming

$\min\{c^T x \mid a_i^T x \leq b_i,\ i = 1, \ldots, r\}$

The logarithmic barrier for this problem is

$$\phi(x) \;=\; \sum_{i=1}^{r} -\log(b_i - a_i^T x)$$

Then

$$tf_0(x) + \phi(x) \;=\; tc^T x + \sum_i -\log(b_i - a_i^T x)$$

Homework 6b

Read the following sections in Boyd & Vandenberghe
3.6.2, 5.1–5.3, 5.9

Recommended exercises (Boyd & Vandenberghe):
3.20, 3.22, 3.60; 5.12, 5.39, 5.42, 5.43

Attention!
Weeks 6-7 (Oct 6; Oct 9; Oct 13; Oct 16):

- all lectures and instructions in flux 1.06
- tuesday: 1+2 instructions; 3+4 lecture
- friday: 5+6 lecture