Equality constrained minimization

A supplementary note to Chapter 10 of *Convex Optimization* by S. Boyd and L. Vandenberghe

Optimization Lab.

IE department
Seoul National University

25th November 2009

Equality constrained minimization

Consider linear-constrained minimization

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & Ax = b \end{array} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is convex and twice cont. diff'ble (hence domain is open) and $A \in \mathbb{R}^{p \times n}$ is of full row-rank .

Assumption

There exists an optimal solution $x^* \in \mathbf{dom} f$ such that $p^* = f(x^*) = \inf_x \{f(x) : Ax = b\}$.

From KKT conditions, a point $x^* \in \mathbf{dom} f$ is optimal iff $\exists\ \nu^* \in \mathbb{R}^p$ such that

$$\begin{array}{rl} Ax^* = b, & (\text{“Primal feasibility”}) \\ \nabla f(x^*) + A^T \nu^* = 0. & (\text{“Dual feasibility”}) \end{array} \tag{2}$$

Thus, solving (1) is equivalent to solving KKT equations (2) of $n + p$ variables.

Linear-constrained quadratic minimization

Consider linear constrained convex quadratic minimization

$$\begin{array}{ll} \min & \frac{1}{2}x^T P x + q^T x + r \\ \text{s.t.} & Ax = b \end{array} \tag{3}$$

where $P \in \mathbb{S}_+^n$ and $A \in \mathbb{R}^{p \times n}$. Then the optimality conditions are

$$\begin{array}{rl} Ax^* = b, & \text{(primal feasibility)} \\ \nabla P x^* + q + A^T \nu^* = 0. & \text{(dual feasibility)} \end{array} \tag{4}$$

Or,

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}, \tag{5}$$

where coefficient matrix is called *KKT matrix*.

Linear-constrained quadratic minimization(cont'd)

Example

- *If (5) is infeasible, the problem is unbounded below.*
  Infeasibility implies $\exists\ v \in \mathbb{R}^n$ *and* $w \in \mathbb{R}^p$ *such that*

$$Pv + A^T w = 0, \quad Av = 0, \quad -q^T v + b^T w > 0.$$

Let $\hat{x}$ *be any feasible point. The point* $x = \hat{x} + tv$ *is feasible for all* t *and*

$$\begin{aligned} f(\hat{x} + tv) &= f(\hat{x}) + t(v^T P\hat{x} + q^T v) + (1/2)t^2 v^T Pv \\ &= f(\hat{x}) + t(-\hat{x}^T Aw + q^T v) - (1/2)t^2 w^T Av \quad (v^T P = -w^T A) \\ &= f(\hat{x}) + t(-b^T w + q^T v) \quad (Av = 0,\ A\hat{x} = b) \end{aligned}$$

Thus, $f(\hat{x} + tv) \to -\infty$ *as* $t \to \infty$.

Linear-constrained quadratic minimization(cont'd)

The followings are equivalent conditions that KKT matrix is nonsingular.

- $\mathcal{N}(P) \cap \mathcal{N}(A) = \{0\}$.
- $Ax = 0,\ x \neq 0 \Rightarrow x^T P x > 0$, i.e., P is PD on $\mathcal{N}(A)$.
- $F^T P F \succ 0\ \forall\ F$ such that $\mathcal{R}(F) = \mathcal{N}(A)$.

Eliminating linear constraints

- We have already seen that a linear-constrained problem can be posed as an unconstrained problem relying on implicit function.
- Let matrix $F \in \mathbb{R}^{n\times(n-p)}$ be such that $\mathcal{R}(F) = \mathcal{N}(A)$ and $\hat{x} \in \mathbb{R}^n$ a feasible solution. Then

$$\{x|Ax = b\} = \{Fz + \hat{x}| \in \mathbb{R}^{n-p}\}.$$

- Then, (1) is equivalent to unconstrained minimization

$$\min \ \tilde{f}(z) = f(Fz + \hat{x}). \tag{6}$$

- From its opt solution z^*, we can find an opt solution of (1) by $x^* = Fz^* + \hat{x}$.

Eliminating linear constraints (cont'd)

Example (Optimal allocation with resource constraint)

$$\begin{array}{ll} \min & \sum_{i=1}^n f_i(x_i) \\ \textit{sub to} & \sum_{i=1}^n x_i = b, \end{array}$$

where $f_i : \mathbb{R} \to \mathbb{R}$ *are convex and twice differentiable, and* $b \in \mathbb{R}$*.*

- *Eliminate* x_n *using the parametrization:* $x_n = b - x_1 - \cdots - x_{n-1}$.
- $\hat{x} = be_n, \quad F = \begin{bmatrix} I \\ -\mathbf{1}^T \end{bmatrix} \in \mathbb{R}^{n \times (n-1)}$.
- *The reduced problem is*

$$\min \quad f_n(b - x_1 - \cdots - x_{n-1}) + \sum_{i=1}^{n-1} f_i(x_i).$$

Definition (via second-order approximation)

- Newton step $\Delta x_{\rm nt}$ for $\min\{f(x)|Ax=b\}$ at feasible x is defined based on second-order Taylor approximation of f near x:

$$\begin{array}{ll} \min & \hat{f}(x+v) = f(x) + \nabla f(x)^T v + (1/2) v^T \nabla^2 f(x) v \\ \text{sub to} & A(x+v) = b. \end{array} \tag{7}$$

- Newton step at x, $\Delta x_{\rm nt}$ is defined to be an opt solution of (7).
- Then, from (5) we get

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\rm nt} \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}. \tag{8}$$

An interpretation: linearly approximated optimality conditions

- Consider the optimality condition

$$Ax^* = b, \quad \nabla f(x^*) + A^T \nu^* = 0.$$

- By substituting $x + \Delta x_{\mathrm{nt}}$ for x^* and w for ν^* and replacing gradient term by its linearized approximation near x, we obtain

$$A(x+\Delta x_{\mathrm{nt}}) = b, \ \nabla f(x+\Delta x_{\mathrm{nt}})+A^T w \approx \nabla f(x)+\nabla^2 f(x)\Delta x_{\mathrm{nt}}+A^T w = 0.$$

- Using $Ax = b$,

$$A\Delta x_{\mathrm{nt}} = 0, \quad \nabla^2 f(x)\Delta x_{\mathrm{nt}} + A^T w = -\nabla f(x),$$

which is (8).

Algorithm (Newton's method for linear-constrained minimization)

given *initial feasible point* x, *tolerance* $\epsilon > 0$.

repeat

1. *Compute Newton step* Δx *and decrement* $\lambda(x) = (\Delta x_{nt}^T \nabla^2 f(x) \Delta x_{nt})^{1/2}$.
2. *Stopping criterion:* $\lambda(x)^2/2 \leq \epsilon$.
3. *Line search. Choose a step size* $t > 0$ *by backtracking line search.*
4. *Update.* $x := x + t\Delta x_{nt}$.

Newton's method and elimination

We can recapture the Newton step from unconstrained Newton step for unconstrained minimization (6): $\min\ \tilde{f}(z) = f(Fz + \hat{x})$.

- Note that

$$\nabla \tilde{f}(z) = F^T \nabla f(Fz + \hat{x}), \quad \nabla^2 \tilde{f}(z) = F^T \nabla^2 f(Fz + \hat{x}) F.$$

- The Newton step for (6) is

$$\Delta z_{\text{nt}} = -\nabla^2 \tilde{f}(z)^{-1} \nabla \tilde{f}(z) = -(F^T \nabla^2 f(x) F)^{-1} F^T \nabla f(x)$$

  where $x = Fz + \hat{x}$.

- The corresponding direction for the linear constrained optimization is (1) is $F\Delta z_{\text{nt}}$. We will observe that $F\Delta z_{\text{nt}}$ is no other than the Newton step Δx_{nt} for (1).

Newton's method and elimination(*cont'd*)

Theorem

$\Delta x_{nt} = F\Delta z_{nt}$.

Proof. It suffices to show that $\Delta x_{\text{nt}} := F\Delta z_{\text{nt}}$ and $w := -(AA^T)^{-1}A \left(\nabla f(x) + \nabla^2 f(x)\Delta x_{\text{nt}}\right)$ satisfy (8).
But, the second condition holds as

$$A\Delta x_{\text{nt}} = AF\Delta z_{\text{nt}} = 0.$$

Regrading the first condition,

$$\begin{bmatrix} F^T \\ A \end{bmatrix} \left(\nabla^2 f(x)\Delta x_{\text{nt}} + A^T w + \nabla f(x)\right) = \begin{bmatrix} F^T\nabla^2 f(x)F\Delta z_{\text{nt}} + F^T A^T w + F^T\nabla f(x) \\ A\nabla^2 f(x)F\Delta z_{\text{nt}} + AA^T w + A\nabla f(x) \end{bmatrix} = 0.$$

Nonsingularity of $\begin{bmatrix} F^T \\ A \end{bmatrix}$ implies the desired condition. □

We assume

a. $S = \{x | x \in \mathrm{dom} f,\ f(x) \leq f(x^{(0)}), Ax = b\}$ is closed where $Ax^{(0)} = b$,

b. $\nabla^2 f(x) \preceq MI$ and

$$\left\| \begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}^{-1} \right\|_2 \leq K, \text{ on } S, \text{ and} \tag{9}$$

c. $\nabla^2 f$ satisfies the Lipschitz condition with L:

$$\|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \leq L\|x - y\|_2,\ \forall x, y \in S.$$

The assumptions imply that $\tilde{f}$ of (6), together with initial point $z^{(0)}$ satisfy the assumptions required in the convergence analysis of unconstrained Newton's method for (6).
Here, we only show that b implies $\nabla^2 \tilde{f}(x) \succeq mI$ where,

$$m = \frac{\sigma_{\min}(F)^2}{K^2 M}. \tag{10}$$

Suppose on the contrary that $F^T HF \not\succeq mI$ where $H = \nabla^2 f(x)$. Then $\exists$ u with $\|u\| = 1$ such that $u^T F^T HFu < m$, or $\|H^{\frac{1}{2}} Fu\|_2 < m^{\frac{1}{2}}$. Using $AF = 0$ we have,

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} Fu \\ 0 \end{bmatrix} = \begin{bmatrix} HFu \\ 0 \end{bmatrix}, \text{ and hence}$$

$$\left\| \begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}^{-1} \right\|_2 \geq \frac{\left\| \begin{bmatrix} Fu \\ 0 \end{bmatrix} \right\|_2}{\left\| \begin{bmatrix} HFu \\ 0 \end{bmatrix} \right\|_2} = \frac{\|Fu\|_2}{\|HFu\|_2}.$$

Since $\|Fu\|_2 \geq \sigma_{\min}(F)$ and

$$\|HFu\|_2 \leq \|H^{\frac{1}{2}}\|_2\|H^{\frac{1}{2}}Fu\|_2 < M^{\frac{1}{2}}m^{\frac{1}{2}},$$

$$\left\| \begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}^{-1} \right\|_2 \geq \frac{\|Fu\|_2}{\|HFu\|_2} > \frac{\sigma_{\min}(F)}{M^{\frac{1}{2}}m^{\frac{1}{2}}} = K \text{ from (10).}$$

A contradiction. $\square$

Homework

10.1, 10.2, 10.4, 10.6.