CSE203B Convex Optimization:
Chapter 10: Equality Constraint Optimization

CK Cheng
Dept. of Computer Science and Engineering
University of California, San Diego

 W9B

Chapter 10 Equality Constrained Optimization

- Introduction
- Formulations
  - Eliminating Equality Constraints Using Algebraic Replacement
  - Dual Formulation
  - KKT Condition
- Newton's Method
- Infeasible Start Newton's Method
- Summary

Introduction

Objective Function without Constraints: (Chapter 9)
Gradient descent, Newton's methods

KKT Linear Equations:
Quadratic obj function + linear equality constraints

Newton's Method:
Twice differentiable obj function + linear equality constraints

Interior Point Method: (Chapter 11)
Twice differentiable obj function + linear equality + inequality constraints

Introduction

Formulation 0:
Equality → Inequality

Formulation 1:
Algebraic operation to eliminate the equality constraint

Formulation 2:
Dual formulation

Formulation 3:
KKT conditions

Formulation 0

$$\min\ f(x)$$
$$\text{s.t. } Ax = b$$

where $f: R^n \to R$, convex, twice continuously differentiable, $and\ A \in R^{p\times n}, rank\ A = p, p \leq n$

Formula 0 Inequality

$min\ f(x)$

$\text{s.t. } Ax \geq b$ $-Ax \geq -b$

$-Ax \leq -b$.

Formulation 1

$$\min\ f(x)$$
$$\text{s.t. } Ax = b$$

$f: R^n \to R$, convex, twice continuously differentiable, $and\ A \in R^{p\times n}, rank\ A = p, p \leq n$

Formula 1 Algebraic operation to eliminate the equality constraint

$$\min f(x) = f(Fz + x_o)$$
$$z \in R^{n-p}, Ax_o = b, rank\ F = n - p, AF = 0$$

Formulation 1

Formula 1: Eliminating equality constraints using algebraic replacement

$$\min\ f(x)$$
$$\text{s.t. } Ax = b\,, \qquad rank\ A = p, p \le n$$

Let $Ax_0 = b$, nullspace of A is

$$F \in R^{n\times(n-p)}, \qquad i.e. AF = 0$$

We can write $x = x_0 + Fz, \quad z \in R^{n-p}$

Thus $f(x) = f(x_0 + Fz)$

To minimize $f(x) = f(x_0 + Fz)$

we need $\nabla_z f(x_0 + Fz) = F^T \nabla f(x)|_{x=x_0+Fz} = 0.$ the null of A forms the basis of F.

Remark: This is equivalent to $\nabla f(x) = -A^T v, \quad v \in R^p$

$\nabla f(x)$ is in the range of A^T

$\Rightarrow \nabla f(x)$ is orthogonal to F $(F^T \nabla f(x) = 0)$

because $F^T A^T = 0$. $(AF = 0)$

7

Formulation 1

Example: $\min f(x_1, x_2)$

$[A_1 \quad A_2]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = b, \qquad A_1x_1 + A_2x_2 = b$

$x_1 = A_1^{-1}(b - A_2x_2), \qquad$ *Suppose the A_1 is nonsingular.*

$f(x_1, x_2) = f(A_1^{-1}(b - A_2x_2), x_2) \quad \Rightarrow f(Fz + x_0)$

Therefore $\nabla_{x_2} f(A_1^{-1}(b - A_2x_2), x_2) = 0$ *derive the optimal solution.*

Remark: The equality constraint elimination, e.g. A_1^{-1} operation, may destroy the sparsity of the system.

Formulation 2

$$\min\ f(x)$$
$$\text{s.t. } Ax = b$$

$f: R^n \to R$, convex, twice continuously differentiable, *and* $A \in R^{p\times n}$, *rank* $A = p, p \le n$

Formula 2 Lagrange Dual Function

$$\max_v g(v) = \max_v \min_x f(x) + v^T Ax - v^T b$$
$$= \max_v \left[-v^T b + \min_x \left(f(x) + v^T Ax\right)\right]$$
$$= \max_v \left[-v^T b - \max_x \left(-v^T Ax - f(x)\right)\right]$$
$$= \max_v \left(-v^T b - f^*(-A^T v)\right)$$

$f^*(y) = \max_x y^T x - f(x)$

Formulation 2

Example: $\min f(x) = \frac{1}{2}x^T P x + q^T x + r \Rightarrow -\frac{a^2}{2M}$ $\quad q^T q$

$\quad s.t. \quad Ax = b, \quad P \in S^n_{++}$

(1) *Lagrangian*: $L(x, v) = \frac{1}{2}x^T P x + q^T x + r + v^T(Ax - b)$ $\quad - \frac{q^T P^{-1} q}{2} + r$

(2) Min L vs. x, we have $\nabla_x L(x, v) = Px + q + A^T v = 0$

(3) Thus, $x = -P^{-1}(q + A^T v)$

(4) Therefore, $G(v) = L(x = -P^{-1}(q + A^T v), v)$

$= -\frac{1}{2}v^T A P^{-1} A^T v - (b^T + q^T P^{-1} A^T)v - \frac{1}{2}q^T P^{-1} q + r$

(5) Min G vs. v, we have $\nabla G(v) = -AP^{-1}A^T v - (b + AP^{-1}q) = 0$

(6) Thus, $v = -(AP^{-1}A^T)^{-1}(b + AP^{-1}q)$

(7) Therefore, $\max_v G(v) =$

$$\frac{1}{2}(AP^{-1}q + b)^T (AP^{-1}A^T)^{-1}(AP^{-1}q + b) - \frac{1}{2}q^T P^{-1} q + r$$

Formulation 2

Ex: $\min f(x) = -\sum_{i=1}^{n} \log x_i, \quad x_i > 0$

s.t. $Ax = b$ $\rightarrow$ $x_0 + Fz$

1. $L(x, \lambda, v) = -\sum_{i=1}^{n} \log x_i - \lambda^T x + v^T Ax - v^T b$
2. $G(\lambda, v) = \min\limits_{x} -\lambda^T x + vAx - v^T b - \sum_{i=1}^{n} \log x_i$
3. Let $\min\limits_{x} g(x, y) = y^T x - \sum_{i=1}^{n} \log x_i$

$$\frac{\partial g(x,y)}{\partial x} = y - \begin{bmatrix} \frac{1}{x_1} \\ \cdots \\ \frac{1}{x_n} \end{bmatrix} = 0, \quad x_i = \frac{1}{y_i}$$

We have $\min\limits_{x} g(x, y) = n - \sum \log\left(\frac{1}{y_o}\right) = n + \sum_{i=1}^{n} \log y_i$

4. Thus, we have $\min\limits_{x} g(x, A^T v) = n + \sum \log(A^T v)_i$

Dual $\max\limits_{v} L(v) = -b^T v + n + \sum \log(A^T v)_i, A^T v > 0$

Conjugate function.

Formulation 3

$$\min \ f(x)$$
$$\text{s.t. } Ax = b$$

$f: R^n \rightarrow R$, convex, twice continuously differentiable,

and $A \in R^{p \times n}$, *rank* $A = p, p \leq n$

Formula 3 KKT condition

$$\nabla f(x^*) + \sum_{i=1}^{m} \nabla f_i(x^*)\lambda_i^* + \sum_{i=1}^{p} \nabla h_i(x^*) v_i^* = 0$$
$$f_i(x^*) \leq 0$$
$$h_i(x^*) = 0$$
$$\lambda_i^* \geq 0$$
$$\sum_i \lambda_i^* f_i(x^*) = 0$$

A^T

KKT condition: $\begin{cases} \nabla f(x^*) + A^T v^* = 0 \\ Ax^* = b \end{cases}$

$v \rightarrow p$ variables

$x \rightarrow n$ variables

Relation of v^* *and* x^*: $A^T v^* = -\nabla f(x^*)$

$$v^* = -\underline{(AA^T)^{-1}} A \nabla f(x^*)$$

Formulation 3

Example: $\quad min\, f(x) = \frac{1}{2}x^T P x + q^T x + r$

$s.t. \quad Ax = b, \quad P \in S^n_+$

$\nabla f(x) = Px + q$

KKT Conditions

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} x^* \\ v^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix} \Leftrightarrow \begin{cases} \nabla f(x) + A^T v = 0 \\ Ax = b \end{cases}$$

(1) We know that $Ax = b$ has feasible solution because $p \le n$.

(2) We have $n + p$ equations for $n + p$ variables.

(3) If $\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}$ is nonsingular, then the problem has a unique optimal solution.

(4) If $\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}$ is singular then the problem is unbounded.

Remark: $\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} x^* \\ v^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$ relate to one iteration of Newton's method for a nonlinear function $f(x)$.

Where $P = \nabla^2 f(x), q = \nabla f(x), r = f(0)$

Formulation 3

(3). Nonsingularity

i. $N(P) \cap N(A) = \{0\}$

ii. $Ax = 0, x \neq 0 \rightarrow x^T P x > 0$

iii. $F^T P F > 0$ for $F \in R^{n\times(n-p)}, R(F) = N(A)$

iv. $P + A^T Q A > 0$ for some $Q \ge 0$

Property ii:

If $\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}$ is singular, we can find $\begin{bmatrix} x \\ v \end{bmatrix}$ so that

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} x \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow Ax = 0$$

Therefore, we have

$$\begin{bmatrix} x^T & v^T \end{bmatrix}\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} x \\ v \end{bmatrix} = x^T P x + 2x^T A v = x^T P x = 0$$

Formulation 3

Proof (4): Let $\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} u \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow Pu = -A^T w,\ Au = 0$

Given $Ax_0 = b$*, we have*

$$f(x_0 + tu) = \frac{1}{2}(x_0 + tu)^T P(x_0 + tu) + q^T(x_0 + tu) + r$$

$$= \frac{1}{2}x_0^T P x_0 + tu^T P x_0 + \frac{1}{2}t^2 u^T P u + q^T x_0 + tq^T u + r$$

1. $\frac{1}{2}t^2 u^T P u = \frac{1}{2}t^2 u^T(-A^T w) = 0$
2. $u^T P x_0 = x_0^T P u = x_0^T(-A^T w) = -w^T A x_0 = -w^T b$

Thus, $f(x_0 + tu) = \frac{1}{2}x_0^T P x_0 + t(\underline{-w^T b + q^T u}) + q^T x_0 + r$

Therefore, when $-w^T b + q^T u \neq 0, f(x)$ is unbounded

Newton's Method

$\min f(x)$

$s.t. Ax = b$

(1) Taylor's expansion to approximate $f(x)$

$$f(x + \Delta x) \approx f(x) + \nabla f(x)^T \Delta x + \frac{1}{2}\Delta x^T \nabla^2 f(x) \Delta x$$

$$Ax = b,\ A\Delta x = 0\ \ (A(x + \Delta x) = b)$$

(2) KKT conditions for the quadratic obj.

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} \Delta x \\ v \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}$$

(3) From (2), $\left(\nabla^2 f(x)\Delta x + A^T v = -\nabla f(x)\right)$

We have $\nabla f(x)^T \Delta x = -(\nabla^2 f(x)\Delta x + A^T v)^T \Delta x$

$$= -\Delta x^T \nabla^2 f(x)\Delta x - v^T A \Delta x = -\Delta x^T \nabla^2 f(x) \Delta x$$

Thus $\quad f(x) + \nabla f(x)^T \Delta x + \frac{1}{2}\Delta x^T \nabla^2 f(x) \Delta x$

$$= f(x) - \frac{1}{2}\Delta x^T \nabla^2 f(x) \Delta x \qquad -\frac{a^2}{2m}$$

The amount that the obj. drops