CSE203B Convex Optimization:
Chapter 10: Equality Constraint Optimization

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

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 $\rightarrow$ 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, $and\ A \in R^{p\times n}, rank\ A = p, p \le n$

Formula 0 Inequality

$min\ f(x)$

s.t. $Ax \ge b$

$-Ax \ge -b$

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 $\nabla_z f(x_0 + Fz) = F^T \nabla f(x)|_{x=x_0+Fz} = 0.$

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

Formulation 1

Example: min $f(x_1, x_2)$

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

$$x_1 = A_1^{-1}(b - A_2 x_2), \qquad \textit{Suppose the } A_1 \textit{ is nonsingular}.$$

$$f(x_1, x_2) = f(A_1^{-1}(b - A_2 x_2), x_2)$$

Therefore $\nabla_{x_2} f(A_1^{-1}(b - A_2 x_2), x_2) = 0$ *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, $and\ A \in R^{p\times n}, rank\ A = p, p \leq n$

Formula 2 Lagrange Dual Function

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

Formulation 2

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

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

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

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

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

(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. $Ax = b$

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

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

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

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

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

Formulation 3

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

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

Formula 3 KKT condition

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

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

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

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

Formulation 3

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

$s.t. \quad Ax = b, \qquad P \in S_+^n$

KKT Conditions

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ v^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$$

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

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

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 $Q \geq 0$

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

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

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

The amount that the obj. drops

Newton's Method

$Algorithm$.

Given $x \in D, Ax = b, \epsilon > 0$
Repeat

1. Solve NE to find Δx & $\lambda^2 = \Delta x^T \nabla^2 f(x) \Delta x$
2. Quit if $\frac{\lambda^2}{2} \leq \epsilon$
3. Line Search t
4. Update $x := x + t\Delta x$

Newton's Method: Affine Invariant

$$\min f(x)$$
$$Ax = b$$

Theorem: Newton's step is invariant to affine transform.

Proof: Let $x = Ty, T \in R^{nn}, f(x) = f(Ty) = \bar{f}(y)$

For the problem

$$\min \bar{f}(y)$$
$$ATy = b$$

1. We have $\nabla_y \bar{f}(y) = T^T \nabla_x f(Ty)$, $\nabla_y^2 \bar{f}(y) = T^T \nabla^2 f(Ty) T$

2. For Δy_{nt} at y, we have the Newton step,

$$\begin{bmatrix} T^T \nabla^2 f(x) T & T^T A^T \\ AT & 0 \end{bmatrix} \begin{bmatrix} \Delta y_{nt} \\ \bar{v} \end{bmatrix} = \begin{bmatrix} -T^T \nabla f(Ty) \\ 0 \end{bmatrix},$$

which is equivalent to

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

Newton's Method for Reduced Problem

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

pn $n(n-p)$

We have

$$\nabla_z f(Fz + x_o) = F^T \nabla_x f(Fz + x_o)$$
$$\nabla_z^2 f(Fz + x_o) = F^T \nabla_x^2 f(Fz + x_o) F$$

Show this by Taylor's expansion

Thus, the reduced problem has Newton iteration derivation,

$$\Delta z = -(\nabla_z^2 f)^{-1} \nabla_z f = -(F^T \nabla_x^2 f F)^{-1} F^T \nabla_x f$$
$$\Delta x = F \Delta z = -F(F^T \nabla^2 f(x) F)^{-1} F^T \nabla f(x)$$

Theorem: For the reduced operation, the derived $\nabla x, v$ are the same solution as the original NE process.

Proof: Let $\Delta x = F\Delta z, v = -(AA^T)^{-1} A(\nabla f(x) + \nabla^2 f(x) \Delta x)$

We can show that the original NE equations hold, i.e.

$$\nabla^2 f(x) \Delta x + \nabla f(x) + A^T v = 0 \ \ \& \ \ A\Delta x = 0$$

Newton's Method for Reduced Problem

Proof:

1. For the first equation, we multiply the left expression from the left, i.e. $\begin{bmatrix} F^T_{(n-p)n} \\ A_{(pn)} \end{bmatrix} [\nabla^2 f(x)\Delta x + A^T v + \nabla f(x)] =$

$$\begin{bmatrix} F^T \nabla^2 f(x)\Delta x + F^T A^T v + F^T \nabla f(x) \\ A\nabla^2 f(x)\Delta x + AA^T v + A\nabla f(x) \end{bmatrix} = \begin{bmatrix} 0\ (1) \\ 0\ (2) \end{bmatrix}$$

(1) $-F^T \nabla^2 f(x) F(F^T \nabla^2 f(x) F)^{-1} F^T \nabla f(x) + F^T \nabla f(x) + F^T A^T v = 0$

(2) $A\nabla^2 f(x)\Delta x + AA^T\big(-(AA^T)^{-1}A(\nabla f(x) + \nabla^2 f(x)\Delta x)\big) + A\nabla f(x) = 0$

Since $\begin{bmatrix} F^T_{(n-p)n} \\ A_{(pn)} \end{bmatrix}$ is a full ranked matrix, we can conclude that

$$\nabla^2 f(x)\Delta x + A^T v + \nabla f(x)=0$$

2. For the second equation, we have $A\Delta x = AF\Delta z = 0$, since $AF = 0$ by construction.

Infeasible Start Newton's Method

The search of the feasible start point,

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

We can write in incremental derivation,

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta v \end{bmatrix} = - \begin{bmatrix} \nabla f(x) + A^T v \\ Ax - b \end{bmatrix} \qquad \begin{matrix} r_{dual} \\ r_{pri} \end{matrix}$$

Newton Method: Infeasible Start

$Algorithm.$

Given $x \in D, v$, tolerance $\epsilon > 0, \alpha \in \left(0, \frac{1}{2}\right), \beta \in (0,1)$.

Repeat

1. Compute primal and dual Newton steps $\Delta x_{nt}, \Delta v_{nt}$
2. Line search on $\left\|r(x,v)\right\|_2 = ||(r_{dual}(x,v), r_{pri}(x,v)||_2$

   $t := 1$

   while $\left\|r(x + t\Delta x_{nt}, v + t\ \Delta v_{nt})\right\|_2$ >(1-αt)$\left\|r(x,v)\right\|_2$

   $t := \beta t.$
3. Update $x := x + t\Delta x_{nt}, v := v + t\ \Delta v_{nt}$

Until $Ax = b$ and $\left\|r(x,v)\right\|_2 \leq \epsilon$

Summary

KKT Linear Equations:
Quadratic objective function + linear equality constraints

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

Interior Point Method:
Twice differentiable obj function + linear equality + inequality constraints