

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

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

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

Formula 0 Inequality

$$\begin{aligned} & \min f(x) \\ & \text{s. t. } Ax \geq b \\ & \quad -Ax \leq -b \end{aligned}$$

Formulation 1

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

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

Formula 1 Algebraic operation to eliminate the equality constraint

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

Formulation 1

Formula 1: Eliminating equality constraints using algebraic replacement

$$\begin{aligned} & \min f(x) \\ & \text{s.t. } Ax = b, \quad \text{rank } A = p, p \leq n \end{aligned}$$

Let $Ax_0 = b$, nullspace of A is

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

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

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 $\nabla f(x) = -A^T v$, $v \in R^p$

Formulation 1

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

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

$x_1 = A_1^{-1}(b - A_2 x_2)$, *Suppose the A_1 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

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

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

Formula 2 Lagrange Dual Function

$$\begin{aligned} \max_v g(v) &= \max_v \min_x f(x) + v^T Ax - v^T b \\ &= \max_v [-v^T b + \min_x (f(x) + v^T Ax)] \\ &= \max_v [-v^T b - \max_x (-v^T Ax - f(x))] \\ &= \max_v (-v^T b - f^*(-A^T v)) \end{aligned}$$

Formulation 2

Example:
$$\min f(x) = \frac{1}{2}x^T Px + q^T x + r$$
$$\text{s. t. } Ax = b, \quad P \in S_{++}^n$$

(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

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

$$\text{s.t. } Ax = b$$

$$1. L(x, \lambda, v) = -\sum_{i=1}^n \log x_i - \lambda^T x + v^T Ax - v^T b$$

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

$$3. \text{ Let } \min_x g(x, y) = y^T x - \sum_{i=1}^n \log x_i$$

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

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

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

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

Formulation 3

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

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

Formula 3 KKT condition

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

$$f_i(x^*) \leq 0$$

$$h_i(x^*) = 0$$

$$\lambda_i^* \geq 0$$

$$\sum_i \lambda_i^* f_i(x^*) = 0$$

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

$$\begin{aligned} \text{Relation of } v^* \text{ and } x^*: \quad A^T v^* &= -\nabla f(x^*) \\ v^* &= -(AA^T)^{-1} A \nabla f(x^*) \end{aligned}$$

Formulation 3

Example:
$$\min f(x) = \frac{1}{2}x^T P x + q^T x + r$$
$$s. t. Ax = b, \quad P \in S_+^n$$

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

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

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

iv. $P + A^T QA > 0$ 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

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

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

$$\begin{aligned} f(x_0 + tu) &= \frac{1}{2} (x_0 + tu)^T P (x_0 + tu) + q^T (x_0 + tu) + r \\ &= \frac{1}{2} x_0^T P x_0 + tu^T P x_0 + \frac{1}{2} t^2 u^T P u + q^T x_0 + tq^T u + r \end{aligned}$$

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 $f(x)$

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

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

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

$$\begin{aligned} \text{We have } \nabla f(x)^T \Delta x &= -(\nabla^2 f(x) \Delta x + A^T v)^T \Delta x \\ &= -\Delta x^T \nabla^2 f(x) \Delta x - v^T A \Delta x = -\Delta x^T \nabla^2 f(x) \Delta x \end{aligned}$$

$$\begin{aligned} \text{Thus } f(x) + \nabla f(x)^T \Delta x + \frac{1}{2} \Delta x^T \nabla^2 f(x) \Delta x \\ = f(x) - \frac{1}{2} \Delta x^T \nabla^2 f(x) \Delta x \end{aligned}$$

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

$$\begin{aligned} \min f(x) \\ Ax = b \end{aligned}$$

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

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

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

$$\min f(x) = f(Fz + x_o)$$

$$z \in R^{n-p}, Ax_o = b, \text{rank } F = n - p, AF = 0$$

$$pn \quad n(n-p)$$

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.

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

Newton's Method for Reduced Problem

Proof:

1. For the first equation, we multiply the left expression from

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

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

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

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

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 $A \Delta x = A F \Delta z = 0$, since $A F = 0$ 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,

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

Newton Method: Infeasible Start

Algorithm.

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

Repeat

1. Compute primal and dual Newton steps $\Delta x_{nt}, \Delta v_{nt}$
2. Line search on $\|r(x, v)\|_2 = \|(r_{dual}(x, v), r_{pri}(x, v))\|_2$
 $t := 1$
while $\|r(x + t\Delta x_{nt}, v + t\Delta v_{nt})\|_2 > (1-\alpha t)\|r(x, v)\|_2$
 $t := \beta t$.
3. Update $x := x + t\Delta x_{nt}, v := v + t\Delta v_{nt}$

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

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