(Evector)

Z1. spline function

origin ϑ, parallelogram.

1. vector space $(V, +, \cdot)$ over a field F.

$\Leftrightarrow (V, +, \cdot)$ is left R-module, where F is the ring.

$\Leftrightarrow$ $(V, +)$ is abelian group,

and. $F \times V \to V$:

$\forall\ r \in, s \in F,\ x, y \in V$.

$r \cdot (x+y) = r\cdot x + r \cdot y$

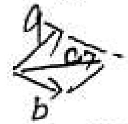
$(r+s)\cdot x = r\cdot x + s \cdot x$

$(r \cdot s)\cdot x = r\cdot(s \cdot x)$

$1_F \cdot x = x$.

Z2:

parallelogram rule.



$c = a+b$.

2 linear combination

$w = a_1 v_1 + a_2 v_2 + \cdots + a_m v_m$.

2.2. affine combination

$\Leftrightarrow$ ① linear combination

② $\sum a_i = 1,\ a_i \in F$.

~~e.x.~~

2.3 convex combination

$\Leftrightarrow$ ① affine combination

② $a_i \geq 0,\ \forall i$.

~~2.3~~ ~~convex combination~~

~~proof~~.

3. set A is convex

$\Leftrightarrow$ ~~convex combination~~

$B = \{ w \mid w = (1-\lambda)u + \lambda v,\ 0 \leq \lambda \leq 1 \}$

$\subset A$, $\forall\ u, v \in A$.

3.3. interval $[a,b] \subset R$

is convex set.

3.4 set $A = \{x \mid \|x\| \leq k, k \in R\}$

is convex.

proof: $\forall u, v \in A$.

$\|(1-\lambda)u + \lambda v\|$

$\leq (1-\lambda)\|u\| + \lambda\|v\|$

$\leq k$

$\therefore A$ is convex.

3.5.

$A = \{ x \in R^n \mid \sum_{i=1}^{n} k_i x_i = c \}$

is convex.

proof: $\forall u, v \in A$.

let $w = (1-\lambda)u + \lambda v$

$\because \sum^n k_i w_i$

$= \sum^n k_i [(1-\lambda)u_i + \lambda v_i]$

$= (1-\lambda)(\sum^n k_i u_i) + (\lambda)(\sum^n k_i v_i)$

$= c$

$\therefore w \in A$. $\therefore A$ convex.

3.6 $S = \{x \in R^n \mid Ax = b\}$

$\Rightarrow S$ convex

proof: $\forall\ u, v \in S$

$A((1-\lambda)u + \lambda v) = b$

3.5. method 2: $\because$ 3.6 $\therefore$ 3.5 proved.

3.7. vector space V is convex

3.8 set S convex $\Rightarrow \bar{S}$ convex

proof: $\forall u, v \in \bar{S}$

$\exists\ x_i \to u,\ x_i \in S$

$y_i \to v,\ y_i \in S$.

$\because$ S convex

$\therefore (1-\lambda)x_i + \lambda y_i \in S$ $\forall i$.

$\because \lim_{i\to\infty} (1-\lambda)x_i + \lambda y_i$

$= (1-\lambda)u + \lambda v$

$\therefore (1-\lambda)u + \lambda v \in \bar{S}$

$\therefore \bar{S}$ convex.

~~3.9. A, B convex $\Rightarrow A \cap B$ convex.~~

~~A_i convex $\Rightarrow \bigcap_{i=1} A_i$ convex~~

3.9. A_α convex, $\forall \alpha$.

$\Rightarrow \bigcap A_\alpha$ convex

3.10. C, C_1, C_2 convex in vector space V,

$k \in F$.

$\Rightarrow$ ① kC convex

② $C_1 + C_2$ convex

proof: ① $u, v \in kC$.

let $w = (1-\lambda)u + \lambda v$

$\because u = ku_k,\ v = kv_k,$

$u_k, v_k \in C$

$\therefore w = k((1-\lambda)u_k + \lambda v_k)$

let $w_k = (1-\lambda)u_k + \lambda v_k$.

$\because C$ convex

$\therefore w_k \in C$

$\therefore w \in kC$. $\therefore kC$ convex.

3.12. A_1, A_2 convex

in vector space V_1, V_2. respectively.

$\Rightarrow A_1 \times A_2$ convex in $V_1 \times V_2$

proof: $\forall u, v \in A_1 \times A_2$

$\therefore u = (u_1, u_2),\ v = (v_1, v_2)$

let $w = (1-\lambda)u + \lambda v$

$\therefore w = ((1-\lambda)u_1, (1-\lambda)u_2)$

$+ (\lambda v_1, \lambda v_2)$

$= ((1-\lambda)u_1 + \lambda v_1,$

$(1-\lambda)u_2 + \lambda v_2)$

$\in A_1 \times A_2$

$\therefore A_1 \times A_2$ convex.

②

$\forall u, v \in C_1 + C_2$

$\therefore u = u_1 + u_2,\ v = v_1 + v_2,$

$u_i \in C_i$, v_i, $i = 1, 2$.

let $w = (1-\lambda)u + \lambda v$

$\therefore w = ((1-\lambda)u_1 + \lambda v_1)$

$+ ((1-\lambda)u_2 + \lambda v_2)$

$= w_1 + w_2$

$\because C_i$ convex $\therefore w_i \in C_i$

$\therefore w \in C_1 + C_2$

$\therefore C_1 + C_2$ convex.

3.13 $\because A_i$ convex in $V_i, \forall i$,

$\Rightarrow A_1 \times A_2 \times \cdots A_n$ convex in $\prod_{i=1}^{n} V_i$

3.11. $\because C_i$ convex $\forall i$. $\Rightarrow \sum_{i=1}^{n} k_i C_i$ convex

3.14. A convex,

$\Rightarrow \sum_{i=1}^{n} k_i x_i \in A$,

$\sum k_i = 1$, $k_i \geq 0$, $x_i \in A$, $\forall i$.

proof: we prove it by induction.

for $n=1$, the result is true.

for $n=2$, true.

If for $n=m$, it is true. then,

$\sum^{m+1} k_i x_i$

$= \sum^{m} k_i x_i + k_{m+1} x_{m+1}$

$= K_m \left(\sum^{m} \frac{k_i}{K_m} x_i \right) + k_{m+1} x_{m+1}$

$\in A$, here $K_m = \sum^{m} k_i$, $K_m \neq 0$.

3.15 A convex $\Leftrightarrow$ $\forall$ convex combination of points in A is still in A. ∵

3.16.

① convex hull of set C is co(C)

② $\Leftrightarrow \cap \{\text{convex set } A \mid A \supset C\}$

③ $\Leftrightarrow \{\text{convex combinations of points in } C\}$ ∵

proof: 3.16.1 ①$\Leftrightarrow$② definition.

16.2 now we need to prove $co(C) = B$, where $B = \{\text{convex combinations of points in } C\}$

16.2.1 $B \subset co(C)$

∵ $\forall x \in B$, ∵ $co(C) \supset C$

co(C) convex

∴ $co(C) \ni x = \sum k_i x_i$, $x_i \in C$.

∴ $B \subset co(C)$

16.2.2. $co(C) \subset B$.

we only need to prove B convex

$\forall x, y \in B$,

∴ $x = \sum k_i x_i$, $x_i \in C$.

$y = \sum h_i y_i$, $y_i \in C$.

$\sum k_i = 1$, $\sum h_i = 1$, $k_i \geq 0$, $h_i \geq 0$.

let $w = (1-\lambda)x + \lambda y$

∴ $w = (1-\lambda)\sum k_i x_i + \lambda \sum h_i y_i$

∵ $(1-\lambda)\sum k_i + \lambda \sum h_i = 1$.

$(1-\lambda) k_i \geq 0$, $\lambda h_i \geq 0$.

$x_i, y_i \in C$.

∴ $w \in B$ ∴ B convex.

17.

17. Carathéodory's theorem. (∵

$S \subset V$, $\dim(V) = n$, V is vectorspace,

$\forall x \in co(S)$.

$\Rightarrow$ x can be represented as a convex combination of no more than $n+1$ elements of S.

proof: ① let $x \in co(S)$.

$\therefore x = \sum^{m} \alpha_i x_i$, $x_i \in S$, $\alpha_i \geq 0$
$\sum \alpha_i = 1$.

let m is the minimal number of vectors for which such a representation of x is possible.

② Suppose $m > n+1$. (We will get a contradictory).

for $\{x_i - x_1, i = 2, \cdots, m\}$,

$\exists \lambda_2, \cdots, \lambda_n$, some $\lambda_j > 0$, s.t.

$$\sum_{i=2}^{m} \lambda_i (x_i - x_1) = 0$$

let $\lambda_1 = -\sum_{i=2}^{m} \lambda_i$

$\therefore \sum_{i=1}^{m} \lambda_i x_i = 0$, $\sum_{i=1}^{m} \lambda_i = 0$,

with some $\lambda_j > 0$.

③ $\because \sum^{m} (\alpha_i - \gamma\lambda_i) x_i$

$= \sum^{m} \alpha_i x_i - \gamma \sum^{m} \lambda_i x_i$

$= x$, $\forall \gamma \in F$.

$\because \sum (\alpha_i - \gamma\lambda_i)$

$= \sum \alpha_i - \gamma \sum \lambda_i$

$= 1$

let $\alpha_i - \gamma\lambda_i \geq 0$

let $\gamma = \min\{\frac{\alpha_i}{\lambda_i} \mid \lambda_i > 0\}$

$\therefore \alpha_i - \gamma\lambda_i \geq 0$

and $\exists j$, s.t. $\alpha_j - \gamma\lambda_j = 0$.

$\therefore \sum^{m} (\alpha_i - \gamma\lambda_i) x_i$ is a convex representation of x in terms of fewer than m points in S.

This contradicts the assumption of minimality of m.

④ $\therefore m \leq n+1$.

proved.

1. $X \subset \mathbb{R}^n$ open set,

$f: X \to \mathbb{R}$ continuously differentiable,

$x^* \in X$ is local minimizer of f,

$\Rightarrow \nabla f(x^*) = 0$

proof:

if $\nabla f(x^*) \neq 0$,

$\Rightarrow \exists d \in \mathbb{R}^n,\ \nabla f(x^*)^T d < 0.$

② $h'(\alpha)\big|_{\alpha=0} = f'(x^*; d)$ (directional derivative)

$= \lim_{\alpha \to 0} \frac{f(x^* + \alpha d) - f(x^*)}{\alpha}$

$= (\nabla f(x^*))^T d < 0$

③ $\therefore \exists \bar{\alpha} > 0,\ \forall\, 0 < \alpha \le \bar{\alpha}$,

$x^* + \alpha d \in X$ (X open)

and $\frac{f(x^* + \alpha d) - f(x^*)}{\alpha} < 0$

(f continuous) (ε) ~~(f continuous)~~

This contradicts the assumption that x^* is local minimizer of f in X.

proved.

1.3 f continuously differentiable ①

$\Rightarrow$ directional derivative $f'(x; d)$ of f exists.

(ε)

2. $X \subset \mathbb{R}^n$ open, $f: X \to \mathbb{R}$, f'' continuous,

x^* local minimizer of f on X,

$\Rightarrow \nabla f(x^*) = 0$,

$\nabla^2 f(x^*)$ positive semi-definite

3. $f(x) = x_1^2 - x_2^4$

$\Rightarrow f'(x) = \begin{pmatrix} 2x_1 \\ -4x_2^3 \end{pmatrix}$

$f''(x) = \begin{pmatrix} 2 & 0 \\ 0 & -12x_2^2 \end{pmatrix}$

$\therefore f'(0,0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $f''(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$

but $x^* = (0,0)^T$ is not local minimizer. [redacted] $\therefore f(x^* + \alpha d) < f(x^*)$ $\forall\, 0 < \alpha \le \bar{\alpha}$

4. $X \subset \mathbb{R}^n$ open, $f: X \to \mathbb{R}$, f'' continuous.

$\nabla f(x^*) = 0$, $\nabla^2 f(x^*)$ positive definite ~~$f(x^*)$~~

$\Rightarrow x^*$ a strict local minimizer.

5. $X \subset \mathbb{R}^n$ convex,

$f: X \to \mathbb{R}$ is called convex

$\Leftrightarrow f(\lambda x + (1-\lambda) y) \le \lambda f(x) + (1-\lambda) f(y)$

$\forall x, y \in X, \forall \lambda \in (0,1)$.

6. $X \subset \mathbb{R}^n$ convex.

$f: X \to \mathbb{R}$ is called uniformly convex

$\Leftrightarrow \exists \mu > 0$,

$f(\lambda x + (1-\lambda) y) + \mu \lambda (1-\lambda) \|x - y\|^2 \le \lambda f(x) + (1-\lambda) f(y)$

$\forall x, y \in X, \forall \lambda \in (0,1)$.

$\Leftrightarrow f$ is called uniformly convex with module μ.

7. $f: \mathbb{R}^n \to \mathbb{R}$,

$f(x) = \frac{1}{2} x^T A x + b^T x + c$

$A \in S_n$, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$.

$\Rightarrow$ ① f convex $\Leftrightarrow A$ positive semi-definite.

② f strictly convex

$\Leftrightarrow f$ uniformly convex

$\Leftrightarrow A$ is positive definite.

8. X open $\subset \mathbb{R}^n$, convex,

$f: X \to \mathbb{R}$, f' continuous

$\Rightarrow$ (a) f convex

$\Leftrightarrow f(x) \ge f(y) + \nabla f(y)^T (x - y)$ $\forall x, y \in X$.

(b) f strictly convex

$\Leftrightarrow f(x) > f(y) + \nabla f(y)^T (x - y)$ $\forall x, y \in X, x \neq y$

(c) f uniformly convex

$\Leftrightarrow \exists \mu > 0$, $f(x) \ge f(y) + \nabla f(y)^T (x - y) + \mu \|x - y\|^2$ $\forall x, y \in X$,

13. $f: R^n \to R$, f' continuous,
f convex, $f'(x^*)=0$
$\Rightarrow x^*$ is a global minimizer of f.

11. $f: R^n \to R$, f' continuous,
$X \subset R^n$ convex.
$\min f(x)$, s.t. $x \in X$, (1)
$\Rightarrow$ (a) f convex on X
$\Rightarrow$ solution set of (1) is convex (possibly empty).
(b) f strictly convex on X
$\Rightarrow$ (1) has at most one solution.
(c) f uniformly convex (on X), X is non-empty, closed,
$\Rightarrow$ (1) has exactly one solution.

12. $f: R^n \to R$, f' continuous, $x^0 \in R^n$, level set $L(x^0)$ convex, f uniformly convex on $L(x^0)$, $x^* \in R^n$ is unique global minimizer of f.
$\Rightarrow \exists \mu > 0$, $\mu \|x - x^*\|^2 \le f(x) - f(x^*)$, $\forall x \in L(x^0)$

9. $X \subset R^n$ open, convex,
$f: X \to R$, f' continuous,
$\Rightarrow$ (a) f convex $\Leftrightarrow \nabla^2 f(x)$ positive semi-definite, $\forall x \in X$
(b) $\nabla^2 f(x)$ p.d. $\forall x \in X$
$\Rightarrow f$ strictly convex.
(c) f uniformly convex
$\Leftrightarrow \nabla^2 f(x)$ uniformly positive definite on X, i.e., $\exists \mu > 0$, s.t.
$d^T \nabla^2 f(x) d \ge \mu \|d\|^2$
$\forall x \in X$, $\forall d \in R^n$

10. $f: R^n \to R$, f' continuous,
$\forall x^0 \in R^n$,
if $L(x^0) := \{x \in R^n \mid f(x) \le f(x^0)\}$ convex, and f uniformly convex in $L(x^0)$
$\Rightarrow L(x^0)$ compact.

descent method, step size strategy

1. $\min f(x)$, $x \in \mathbb{R}^n$

where $f: \mathbb{R}^n \to \mathbb{R}$, f' continuous.

① at $x \in \mathbb{R}^n$, chose a direction $d \in \mathbb{R}^n$ in which f decreases.

② Starting at x, one proceeds along d as long as f reduces sufficiently.

① ~~is descent~~ descent method.

②: step size strategy.

2. $f: \mathbb{R}^n \to \mathbb{R}$, $x \in \mathbb{R}^n$.

$d \in \mathbb{R}^n$ is a descent direction of f at $x \Leftrightarrow \exists \bar{\alpha} > 0$,

$f(x+\alpha d) < f(x)$

$\forall \alpha \in (0, \bar{\alpha}]$

3. f' continuous, $\nabla f(x)^T d < 0$

$\Rightarrow d$ is descent direction.

proof: let $\varphi(\alpha) = f(x+\alpha d)$

$\because f'$ continuous

$\therefore \varphi'(\alpha)$ continuous

$\therefore \varphi(\alpha) = \varphi(0) + \alpha\varphi'(0) + r(\alpha)$ (ε)

$\frac{r(\alpha)}{\alpha} \to 0$, $\alpha \to 0^+$

$\therefore \lim_{\alpha \to 0^+} \frac{\varphi(\alpha) - \varphi(0)}{\alpha} = \varphi'(0) = \nabla f(x)^T d < 0$

$\therefore \exists \bar{\alpha} > 0$,

$\forall \alpha \leq \bar{\alpha}$,

$\frac{\varphi(\alpha) - \varphi(0)}{\alpha} < 0$

$\therefore \varphi(\alpha) < \varphi(0)$

$\therefore f(x+\alpha d) < f(x)$

$\therefore d$ is descent direction.

4. x is a strict local maximizer

$\Rightarrow \forall d \in \mathbb{R}^n$, d is a descent direction of f at x.

but $\nabla f(x)^T d < 0$ doesn't hold.

5. $d_1 = -\nabla f(x)$

$d_2 = -M\nabla f(x)$, $M \in S^n$, M p.d.

$\Rightarrow d_1$ and d_2 are descent directions.

proof: $\nabla f(x)^T d_2$

$= -\nabla f(x)^T M \nabla f(x)$

< 0, $\nabla f(x) \neq 0$

$\therefore d_2$ is descent direction.

6. general descent method.

input: $f: \mathbb{R}^n \to \mathbb{R}$, x^0,

begin

$k \leftarrow 1$

while convergence criterion is not fulfilled

① d^k of f at x^k

② get α_k, s.t. $f(x^k + \alpha_k d^k) < f(x^k)$

③ $x^{k+1} \leftarrow x^k + \alpha_k d^k$, $k \leftarrow k+1$

end while

end.

7. $f: \mathbb{R}^n \to \mathbb{R}$, f' continuous, $\{x^k\}$ a sequence generated by (6), such that, $\exists \theta_1 > 0, \theta_2 > 0$, (independent of $\{x^k\}$, $\{d^k\}$) such that

(a) $-\nabla f(x^k)^T d^k \geq \theta_1 \|\nabla f(x^k)\| \|d^k\|$, $\forall k \in N$, (angle condition).

(b) $f(x^k + \alpha_k d^k) \leq f(x^k) - \theta_2 \left(\frac{\nabla f(x^k)^T d^k}{\|d^k\|}\right)^2$ (sufficient decrease)

with $\alpha_k > 0$, $\forall k \in N$.

$\Rightarrow$ every accumulation point of $\{x^k\}$ is a stationary point of f.

proof: $\because f(x^{k+1}) \leq f(x^k) - \theta_2\theta_1^2 \|\nabla f(x^k)\|^2$

$\therefore f(x^{k+1}) \downarrow$; let x^* be an accumulation point of x^k. $\therefore \exists \{x^{k_i}\} \to x^*$. $\because f$ continuous.

$\therefore f(x^{k_i}) \to f(x^*)$, $i \to \infty$. $\because f(x^k) \downarrow$, $\therefore f(x^k) \to f(x^*)$

$\therefore f(x^k) - f(x^{k+1}) \to 0$. $\therefore \|\nabla f(x^k)\| \to 0$

$\because \nabla f(x)$ continuous. $\therefore \nabla f(x^{k_i}) \to \nabla f(x^*)$

$\therefore \|\nabla f(x^*)\| = 0$, $\nabla f(x^*) = 0$. proved.

Armijo rule

1. $\sigma \in (0,1)$,

$\varphi'(\alpha)$ continuous,

$\varphi'(0) < 0$,

$\Rightarrow \exists \delta > 0$, s.t.

$\varphi(\alpha) \le \varphi(0) + \sigma\alpha\varphi'(0)$

$\forall \alpha \in [0, \delta)$ (3)

proof: $\because \varphi'(0) < \sigma\varphi'(0) < 0$

$\therefore \exists \delta$, $\forall \alpha \in [0,\delta)$,

$\varphi(\alpha) \le \varphi(0) + \sigma\varphi'(0)\alpha$

2. Armijo rule.

$f(x+\alpha d) \le f(x) + \sigma\alpha\nabla f(x)^T d$

where $\sigma \in (0,1)$ fixed, $\nabla f(x)^T d <$

2.2 Armijo rule is a condition which ensures a sufficient descent.

2.3. $\exists a > 0$, $\forall \alpha \in [0, a]$, Armijo rule holds.

2.4. actual calculation of α:

$\alpha = \beta^L$, $L = 0, 1, 2, \cdots$,

where $\beta \in (0,1)$ fixed.

~~$\therefore \alpha = \max_L \{\beta^L\}$, st. Armijo rule holds.~~

~~choose $L \in \mathbb{N}$, st. for $\alpha = \beta^L$, Armijo rule holds.~~

choose L incrementally, until for $\alpha = \beta^L$, Armijo rule holds.

2.5. generalization of 2.4

if $\alpha^{(l)}$ does not fulfill Armijo condition,

then $\alpha^{(l+1)}$ is chosen,

s.t. $\alpha^{(l+1)} \in [\underline{v}\alpha^{(l)}, \bar{v}\alpha^{(l)}]$,

with $0 < \underline{v} \le \bar{v} < 1$.

proof: $\because$

$$\alpha^{(l+1)} \le \bar{v}\alpha^{(l)} \le \cdots \le \bar{v}^{l}\alpha^{(1)}$$

$\therefore \lim_{l\to\infty} \alpha^{(l)} = 0$

$\therefore \exists L$, st. $\alpha^{(L)}$ fulfill Armijo condition.

3. Armijo step size strategy.

input: descent direction d.

begin

$\quad l \leftarrow 0$

$\quad \alpha^{(0)} \leftarrow 1$

$\quad$ while Armijo rule not holds,

$\quad\quad$ choose $\alpha^{(l+1)} \in [\underline{v}\alpha^{(l)}, \bar{v}\alpha^{(l)}]$

$\quad\quad l \leftarrow l+1$

$\quad$ end while

$\quad \alpha_k \leftarrow \alpha^{(l)}$

end

4.

4. There is error in 1.

5. $f: R^n \to R$,

f' Lipschitz-continuous,

L is the Lipschitz constant.

$\sigma \in (0,1)$,

~~If $f(x+\alpha d) \leq f(x) + \sigma\alpha \nabla f(x)^T d$ holds,~~ $d = -M\nabla f(x)$,

$M \in S^n$ positive definite,

let $\lambda_s = \min\lambda(M^{-1})$, $\lambda_g = \max\lambda(M^{-1})$,

If $f'(x) \neq 0$,

$\Rightarrow \forall \alpha \in \left(0, \frac{2\lambda_s(1-\sigma)}{L\,\kappa(M^{-1})}\right]$

with $\kappa(M^{-1}) = \lambda_g/\lambda_s$ the condition number of M^{-1}.

we have

$f(x+\alpha d) \leq f(x) + \sigma\alpha \nabla f(x)^T d$.

proof:

① let $h(\tau) = f(x+\tau\alpha d)$

$\because h'(\tau) = \alpha f'(x+\tau\alpha d)^T d$

$h(1) = f(x+\alpha d)$

$h(0) = f(x)$

$\because h(1) - h(0) = \int_0^1 h'(\tau)\,d\tau$

$\therefore f(x+\alpha d) - f(x)$

$= \alpha\int_0^1 \nabla f(x+\tau\alpha d)^T d\, d\tau$

②

$\therefore f(x+\alpha d)$

$= f(x) + \alpha\nabla f(x)^T d$

$+ \alpha\int_0^1 (\nabla f(x+\tau\alpha d) - \nabla f(x))^T d\, d\tau$

③ $\because (\nabla f(x+\tau\alpha d) - \nabla f(x))^T d$

$\leq \|\nabla f(x+\tau\alpha d) - \nabla f(x)\|\,\|d\|$

$\leq L\|\tau\alpha d\|\,\|d\|$

$= |\tau|\alpha L\|d\|^2$

④ $\therefore f(x+\alpha d)$

$\leq f(x) + \alpha\nabla f(x)^T d$

$+ \frac{\alpha^2}{2}L\|d\|^2$

⑤

$\lambda_g^{-1}\|z\|^2 \leq z^T M z \leq \lambda_s^{-1}\|z\|^2$

$\forall z$.

⑥ $\|d\|^2$

$= \|M\nabla f(x)\|^2$

$= \nabla f(x)^T M^2 \nabla f(x)$

$\leq \lambda_s^{-2}\|\nabla f(x)\|^2$

$\leq \lambda_g\lambda_s^{-2}\nabla f(x)^T M \nabla f(x)$

$= -\lambda_g\lambda_s^{-2}\nabla f(x)^T d$

$= -\kappa(M^{-1})\lambda_s^{-1}\nabla f(x)^T d$

⑦ $\because$ ④, ⑥,

$\therefore f(x+\alpha d)$

$\leq f(x) + \alpha\nabla f(x)^T d$

$+ \frac{\alpha^2}{2}L(-\kappa(M^{-1})\lambda_s^{-1}\nabla f(x)^T d)$

$= f(x)$

$+ \alpha(1 - \kappa(M^{-1})\lambda_s^{-1}\frac{\alpha}{2}L)\nabla f(x)^T d$

⑦ $\therefore$ If $\sigma \leq 1 - \kappa(M^{-1})\lambda_s^{-1}\frac{\alpha L}{2}$

$\Rightarrow f(x+\alpha d)$

$\leq f(x) + \alpha\sigma\nabla f(x)^T d$

$\therefore \alpha \leq \frac{2\lambda_s(1-\sigma)}{L}\frac{1}{\kappa(M^{-1})}$

proved.

∴ the actual step size cannot be smaller than $\frac{2\underline{\lambda}(1-\sigma)}{L\bar{\kappa}} \cdot \underline{\nu}$ ∵.

This proves ①.

②

∵ In Armijo step size strategy (3), we chooses $\alpha^{(0)}=1$, $\alpha^{(l+1)} \le \bar{\nu}\alpha^{(l)}$,

let $\bar{\nu}^m < \frac{2\underline{\lambda}(1-\sigma)}{L\bar{\kappa}}$

then α_k will be found after at most m reductions. ∵

∴ $m \le \log\left(\frac{2\underline{\lambda}(1-\sigma)}{L\bar{\kappa}}\right)/\log(\bar{\nu})$

$m \in N$.

16. $f: \mathbb{R}^n \to \mathbb{R}$, f' Lipschitz-continuous,

L is the Lipschitz-constant.

$\{x^k\}$ is the iteration sequence generated by general descent (6) method with a step size choice according to Armijo step size strategy (3), let $\{M^k\}$ be a sequence of symmetric p.d. matrices, s.t. $\exists\, 0 < \underline{\lambda} \le \bar{\lambda} < +\infty$

with $\underline{\lambda} \le \lambda_s^{(k)} \le \lambda_g^{(k)} \le \bar{\lambda}$, $\forall k \in N$.

where $\lambda_s^{(k)}$ and $\lambda_g^{(k)}$ are the smallest and the largest eigenvalue of $(M^k)^{-1}$ resp.

⇒ ① $\alpha_k \ge \underline{\alpha} = \frac{2\nu\underline{\lambda}(1-\sigma)}{L\bar{\kappa}}$, $\forall k \in N$.

with $\bar{\kappa} = \bar{\lambda}/\underline{\lambda}$.

② In every iteration of general descent method (6), there will be at most

$m \le \log\left(\frac{2\underline{\lambda}(1-\sigma)}{L\bar{\kappa}}\right)/\log(\bar{\nu})$,

$m \in N$

step size reductions necessary.

(here $d^k = -M^k\nabla f(x^k)$)

proof:

① ∵ Armijo step size strategy (3) terminates if

$\alpha \le \frac{2\lambda_s^k(1-\sigma)}{L\kappa((M^k)^{-1})}$

or even before that. ∵.

∵ $\lambda_s^k \ge \underline{\lambda} > 0$,

$\kappa((M^k)^{-1}) = \lambda_g^k/\lambda_s^k \le \bar{\lambda}/\underline{\lambda} = \bar{\kappa}$

∴ $\frac{2\lambda_s^k(1-\sigma)}{L\kappa((M^k)^{-1})} \ge \frac{2\underline{\lambda}(1-\sigma)}{L\bar{\kappa}} > 0$, $\forall k \in N$.