About the Midterm

(1). minimization problems in E^n (§7.1 through §7.3)

(a). first & second-order necessary conditions for a local minimum. (constrained/unconstrained)

definition of relative (local) minimum, global minimum, strict global minimum

feasible direction d is a vector $\vec{d}$ at $\vec{x}$ if $\exists\ \bar{\alpha} > 0$ s.t. $\vec{x} + \bar{\alpha}\vec{d} \in \Omega\ \forall \alpha,\ 0 \le \alpha \le \bar{\alpha}$

Prop I: First order necessary conditions: (constrained)

Let Ω be a subset of E^n and let $f \in C^1$ be a function on Ω. If $\vec{x}^*$ is a relative minimum point of f over Ω, then for any $\vec{d} \in E^n$ that is a feasible direction at $\vec{x}^*$, we have $\vec{\nabla} f(\vec{x}^*)\vec{d} \ge 0$

Cor: (Unconstrained case).

Let Ω be a subset of E^n, let $f \in C^1$ be a function on Ω. If x^* is a relative minimum ~~of~~ point of f over Ω and if $\vec{x}^*$ is an interior point of Ω, then $\vec{\nabla} f(\vec{x}^*) = \vec{0}$.

Prop: Second-order necessary conditions: (constrained)

Let Ω be a subset of E^n and let $f \in C^2$ be a function on Ω. If $\vec{x}^*$ is a relative minimum point of f over Ω, then for any $\vec{d} \in E^n$ that is a feasible direction at $\vec{x}^*$ we have

i). $\vec{\nabla} f(\vec{x}^*)\vec{d} \ge 0$

ii). if $\vec{\nabla} f(\vec{x}^*)\vec{d} = 0$, then $\vec{d}^T \vec{\nabla}^2 f(\vec{x}^*)\vec{d} \ge 0$

Prop 2. (second-order necessary conditions - unconstrained case)

强化一些 得到

Let $\vec{x}^*$ be an interior point of the set Ω, and suppose $\vec{x}^*$ is a relative minimum point over Ω of the function $f \in C^2$. Then

i). $\vec{\nabla} f(\vec{x}^*) = \vec{0}$

ii). for all $\vec{d}$, $\vec{d}^T \vec{\nabla}^2 f(\vec{x}^*)\vec{d} \ge 0$

[illegible]

(b). sufficient condition for a local minimum

Prop 3: (Second-order sufficient conditions - unconstrained case)

Let $f \in C^2$ be a function defined on a region in which the point $\vec{x}^*$ is an interior point. Sps ~~th~~ in addition that:

(i). $\vec{\nabla} f(\vec{x}^*) = \vec{0}$

(ii). $\vec{F}(\vec{x}^*)$ is positive definite

($\vec{F}(\vec{x}^*) = \vec{\nabla}^2 f(\vec{x})$, the Hessian)

Then $\vec{x}^*$ is a strict relative minimum point of f.

(2). minimization problems in a subset Ω of E^n.

(a). first-order necessary for a local minimum.

(done).

(3). convex functions (§ 7.4–7.5)

(a). definition of convexity

A function f defined on a convex set Ω is said to be convex if $\forall \vec{x}_1, \vec{x}_2 \in \Omega$, and every α, $0 \le \alpha \le 1$, s.t.

$f(\alpha \vec{x}_1 + (1-\alpha)\vec{x}_2) \le \alpha f(\vec{x}_1) + (1-\alpha) f(\vec{x}_2)$

If, $\forall \alpha$, $0 < \alpha < 1$, and $\vec{x}_1 \ne \vec{x}_2$ s.t.

$f(\alpha \vec{x}_1 + (1-\alpha)\vec{x}_2) < \alpha f(\vec{x}_1) + (1-\alpha) f(\vec{x}_2)$

then f is said to be strictly convex.

* Concave: f is concave if $-f$ is convex.

· combinations of convex functions.

Proposition 1: Let f_1 and f_2 be convex functions, on the convex set Ω, then the function $f_1 + f_2$ is convex on Ω.

Proposition 2: f convex on Ω then αf is convex on Ω for any $\alpha \ge 0$.

Proposition 3: Let f be convex on a convex set Ω. The set $\Gamma_c = \{\vec{x} : \vec{x} \in \Omega, f(\vec{x}) \ge c\}$ is convex for every real number c.

all eigenvalues
- positive definite >0
- negative definite <0
- positive semidefinite ≥ 0
- negative semidefinite ≤ 0

Properties of differentiable convex functions

(b) Prop 4: a C^1 function f is convex iff $f(\vec{y}) \geq f(\vec{x}) + \nabla f(\vec{x})(\vec{y}-\vec{x})$ ~~$\forall x, y$~~ $\forall \vec{x}, \vec{y} \in \Omega$

(c) Prop 5: a C^2 function f is convex iff the Hessian matrix $\vec{F}$ is positive ~~definite~~ semidefinite throughout Ω. (at every $\vec{x}$).

(d). §7.5 minimization & maximization of convex functions. (3 results)
(i.e in one word, every critical point of a convex function is a global minimum point).

① Thm 1: f be a convex function defined on the convex set Ω. then the set Γ where f achieves its minimum is convex, and any relative minimum of f is a global minimum.

② Thm 2: f be C^1 convex on convex set Ω, if there's a pt $\vec{x}^* \in \Omega$ s.t. $\forall \vec{y} \in \Omega$, $\vec{\nabla} f(\vec{x}^*)(\vec{y}-\vec{x}^*) \geq 0$, then $\vec{x}^*$ is a global minimum of f over Ω.

③ Thm: f convex defined on the bounded, closed convex set Ω, if f has a maximum ~~on~~ over Ω it is achieved at an extreme point of Ω.

(4). the Global Convergence Theorem for iterative methods for solving minimization problems. (§7.7)

our simpler version:

GCT: Assume X is either E^n or a closed subset Ω of E^n, and that we want to find a global minimum of a function f defined on X.

Let Γ be the set of global minimum points of f ~~on~~ in X

Let A be a point-to-set mapping on X, satisfying

(1). A is closed at x, $\forall x \notin \Gamma$

(2). if $y \in A(x)$, then $f(y) \leq f(x)$, with strict inequality if $x \notin \Gamma$. Let $\{x_k\}_{k=1}^{\infty}$ be ~~a~~ generated by

$$x_{k+1} \in A(x_k)$$

If all points x_k are contained in a compact set $S \subset X$, then any limit of a convergent subsequence is a minimizer of f.

Note that
~~A is set~~

Algorithm $\vec{A}$: is a mapping defined on a space X that assigns to every point $\vec{x} \in X$ a subset of X.
e.g. $\vec{x_{k+1}} \in \vec{A}(\vec{x_k})$.

Γ is called a solution set.

Descent: $\Gamma \subset X$ be a solution set. Let $\vec{A}$ be an algorithm on X, a continuous real-valued function Z on X is said to be a descent function for Γ and $\vec{A}$ if it satisfies

i) if $\vec{x} \notin \Gamma$ and $\vec{y} \in \vec{A}(\vec{x})$, then $Z(\vec{y}) < Z(\vec{x})$

ii) if $\vec{x} \in \Gamma$ and $\vec{y} \in \vec{A}(\vec{x})$, then $Z(\vec{y}) \leq Z(\vec{x})$

Note again that the algorithm A (mapping) is not point-to-point mapping of X, it's point-to-set mapping of X.
$\vec{A}(\vec{x_k})$

For example, $x_0 = 100$, $A(x) = [-\frac{|x|}{2}, \frac{|x|}{2}]$.
might have $100, 50, 25, 12, \cdots$
or $100, -40, 20, -5, \cdots$
or $100, 10, -1, 1/16, \cdots$
...

about "closed mappings":

A point-to-set mapping A from X to Y is said to be closed at $\vec{x} \in X$ if

i). $\vec{x_k} \to \vec{x}$, $\vec{x_k} \in X$

ii). $\vec{y_k} \to \vec{y}$, $\vec{y_k} \in \vec{A}(\vec{x_k})$

i) ii) $\Rightarrow$ iii). $\vec{y} \in \vec{A}(\vec{x})$

The point-to-set map $\vec{A}$ is said to be closed on X if it's closed at each each point of X.

(5). Iterative methods for minimizing functions of a single variable (§8.2)

(a). Newton's method.

Sps f with single variable x to be minimized

Sps a point x_k where a measurement is made to evaluate the 3 numbers $f(x_k)$, $f'(x_k)$, $f''(x_k)$.

Construct a quadratic function q which at x_k agrees with f up to second derivatives.

$$q(x) = f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2} f''(x_k)(x - x_k)^2$$

Setting

$$0 = q'(x) = f'(x_k) + f''(x_k)(x_{k+1} - x_k)$$

then $$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

Let $g(x) = f'(x)$

we get

$$x_{k+1} = x_k - \frac{g(x_k)}{g'(x_k)}$$

- prop (convergence of newton's method) at least order two convergence.

To solve $g(x) = 0$, assume g is C^2, x^* solves $g(x^*) = 0$, $g'(x^*) \neq 0$, then if x_0 is close enough to x^*, the sequence

$$x_{k+1} = x_k - \frac{g(x_k)}{g'(x_k)}$$ converges to x^*.

(b). more general idea of "curve fitting" methods:

Approximate f near x_k by (Gernal minimization plan)

$$q(x) = f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2}(x - x_k)^2 \left(\frac{f'(x_k) - f'(x_{k-1})}{x_k - x_{k-1}} \right)$$

(don't need to know 2nd derivative).

$$q'(x) = f'(x_k) - (x - x_k) \boxed{\frac{f'(x_k) - f'(x_{k-1})}{x_k - x_{k-1}}} \Longrightarrow \text{is } f''(x_k) \text{ if } x_k - x_{k-1} \text{ small}$$

The equation $q'(x) = 0$ implies that

$$x_{k+1} = x_k - \frac{f'(x_k) \cdot (x_k - x_{k-1})}{f'(x_k) - f'(x_{k-1})}$$

(6). The method of steepest descent (§ 8.6 and parts of § 8.4)

(a). definition of the method of steepest descent.

$\vec{\nabla} f(\vec{x})$: n-dim row vector

define $\vec{g}(\vec{x}) = \vec{\nabla} f(\vec{x})^T$ column vector

write $\vec{g}(\vec{x_k}) = \vec{\nabla} f(\vec{x_k})^T = \vec{g_k}$

The method of steepest descent is defined by iterative algorithm:

$$\vec{x_{k+1}} = \vec{x_k} - \alpha_k \vec{g_k}$$

where, α_k is a nonnegative scalar minimizing $f(\vec{x_k} - \alpha \vec{g_k})$
i.e. from point $\vec{x_k}$, we search along the direction of negative gradient $-\vec{g_k}$ to a minimum point on this line, this minimum point is taken to be $\vec{x_{k+1}}$.

in formal terms,
the Algorithm
$\vec{A}: E^n \to E^n$ which gives $\vec{x_{k+1}} \in \vec{A}(\vec{x_k})$
can be decomposed in the form $\vec{A} = \vec{S} \cdot \vec{G}$
where $\vec{G}: E^n \to E^{2n}$ is defined by $\vec{G}(\vec{x}) = (\vec{x}, -\vec{g}(\vec{x}))$
giving the initial pt & direction of a line search.
This is followed by the line search
$\vec{S}: E^{2n} \to E^n$
where $\vec{S}(\vec{x}, \vec{d}) = \{\vec{y} : \vec{y} = \vec{x} + \alpha \vec{d},$ for some $\alpha \geq 0,$
$f(\vec{y}) = \min_{0 \leq \alpha \leq \infty} f(\vec{x} + \alpha \vec{d})\}$

(b). $\left.\begin{array}{l}\vec{S} \text{ ~~異~~ is closed if } \vec{\nabla} f(\vec{x}) \neq \vec{0} \\ \vec{G} \text{ is continuous}\end{array}\right\} \Rightarrow \vec{A}$ is closed

define $\vec{x}$ is solution sets where $\vec{\nabla} f(\vec{x}) = \vec{0}$.

then $Z(\vec{x}) = f(\vec{x})$ is a descent function for $\vec{A}$ since $\vec{\nabla} f(\vec{x}) \neq \vec{0}$

$\lim_{0 \le \alpha < \infty} f(\vec{x} - \alpha \vec{g}(\vec{x})) < f(\vec{x})$

Thus by GCT, if $\{\vec{x_k}\}$ is bdd, then it will have limit pts and each of them is a solution.

(Quadratic case)

Sps ~~Fca~~ $f(\vec{x}) = \frac{1}{2}\vec{x}^T \vec{Q} \vec{x} - \vec{x}^T \vec{b}$ where $\vec{Q}$ is positive definite $n \times n$ symmetric matrix, $\Rightarrow$ all eigenvalues positive, assume $0 < a = \lambda_1 \le \lambda_2 \cdots \le \lambda_n = A$

$\Rightarrow f$ is strictly convex

the unique minimum pt of can be found directly, by setting the gradient to 0, as $\vec{x}^*$ satisfying

$$\vec{Q}\,\vec{x}^* = \vec{b}$$

Moreover, introducing

$$E(\vec{x}) = \frac{1}{2}(\vec{x} - \vec{x}^*)^T Q (\vec{x} - \vec{x}^*)$$

we have

$$E(\vec{x}) = f(\vec{x}) + \frac{1}{2} x^{*T} Q \vec{x}^*$$

(we consider minimizing $E(\vec{x})$ instead of $f(\vec{x})$, b/c it's simpler)

the gradient of (both f and E) is given by

$$\vec{g}(\vec{x}) = \vec{Q}\vec{x} - \vec{b}$$

Thus the steepest descent ~~test~~ can be expressed as

$$\vec{x}_{k+1} = \vec{x_k} - \alpha_k \vec{g_k}$$

where $\vec{g_k} = Q\vec{x_k} - \vec{b}$, α minimizes $f(\vec{x_k} - \alpha \vec{g_k})$

explicitly, $f(\vec{x_k} - \alpha \vec{g_k}) = \frac{1}{2}(\vec{x_k} - \alpha \vec{g_k})^T Q (\vec{x_k} - \alpha \vec{g_k}) - (\vec{x_k} - \alpha \vec{g_k})^T \vec{b}$

$$\alpha_k = \frac{\vec{g_k}^T \vec{g_k}}{\vec{g_k}^T Q \vec{g_k}}$$ (found by differentiating w.r.t. α)

so method of steepest descent is in form of

$$\vec{x}_{k+1} = \vec{x_k} - \left(\frac{\vec{g_k}^T \vec{g_k}}{\vec{g_k}^T Q \vec{g_k}}\right) \cdot \vec{g_k} \quad \text{where } \vec{g_k} = Q\vec{x_k} - \vec{b}$$

Lemma 1: iterative process satisfies

$$E(\vec{x}_{k+1}) = \left\{1 - \frac{(\vec{g_k}^T \vec{g_k})^2}{(\vec{g_k}^T Q \vec{g_k})(\vec{g_k}^T Q^{-1} \vec{g_k})}\right\} E(\vec{x_k})$$

Kantorovich inequality:

Q positive definite symmetric $n \times n$ matrix

For any $\vec{x}$,

we have

$$\frac{(\vec{x}^T \vec{x})^2}{(\vec{x}^T Q \vec{x})(\vec{x}^T Q^{-1} \vec{x})} \geq \frac{4aA}{(a+A)^2}$$

Thm: (Steepest descent – quadratic case)

$\forall a_0 \in E^n$, the method of steepest descent converges to the unique minimum pt $\vec{x}^*$ of f.

Furthermore, with $E(\vec{x}) = \frac{1}{2}(\vec{x} - \vec{x}^*)^T Q (\vec{x} - \vec{x}^*)$

$\forall$ step k,

$$E(\vec{x}_{k+1}) \leq \left(\frac{A-a}{A+a}\right)^2 E(\vec{x_k})$$

we define $r = \frac{A}{a}$ be ~~a~~ "conditional number"

$r \approx 1$ good
$r \gg 1$ bad

s.t. $E(\vec{x}_{k+1}) \leq \left(\frac{r-1}{r+1}\right)^2 E(\vec{x_k})$

minimize $\frac{1}{2}\vec{x}^T Q \vec{x} - \vec{b}^T \vec{x}$

(7). Conjugate directions and conjugate ~~gradients~~ gradient methods (§9.1~9.3)

(a). definition: Given a symmetric matrix Q, 2 vectors $\vec{d_1}$ and $\vec{d_2}$ are said to be Q-orthogonal, or conjugate with respect to Q if

$$\vec{d_1}^T Q \vec{d_2} = 0$$

(if $\vec{Q} = \vec{0}$, 2 vectors are conjugate while if $\vec{Q} = \vec{I}$, conjugacy is equivalent to the usual notion of orthogonality.)

Prop: If Q is positive definite and the set of nonzero vectors $\vec{d_0}, \cdots \vec{d_k}$ are Q-orthogonal, then these vectors are L.I.

(b). Conjugate directions method.

Conjugate directions theorem

Let $\{d_i\}_{i=0}^{n-1}$ be a set of nonzero Q-orthogonal vectors $\forall\ x_0 \in E^n$ the sequence $\{\vec{x_k}\}$ ~~is~~ generated according to

$$\vec{x_{k+1}} = \vec{x_k} + \alpha_k \vec{d_k}, \quad k \geq 0$$

with $\alpha_k = -\dfrac{\vec{g_k}^T \vec{d_k}}{\vec{d_k}^T Q \vec{d_k}}$

and $\vec{g_k} = Q\vec{x_k} - \vec{b}$

converges to the unique solution, $\vec{x}^*$, of $Q\vec{x} = \vec{b}$ after n steps. that is $\vec{x_n} = \vec{x}^*$.

(c). Conjugate gradient method

idea: like conjugate direction method, except that the directions $d_1, d_2, \cdots$, are determined iteratively:

d_{k+1} is found by taking $g_{k+1} = \nabla f(x_{k+1})^T$ and 'correcting it' to make it Q-orthogonal to d_k.

Conjugate gradient algorithm:

Starting at $\forall \vec{x_0} \in E^n$, defined $\vec{d_0} = -\vec{g_0} = \vec{b} - Q\vec{x_0}$

and $\vec{x_{k+1}} = \vec{x_k} + \alpha_k \vec{d_k}$

$$\alpha_k = -\frac{\vec{g_k}^T \vec{d_k}}{\vec{d_k}^T Q \vec{d_k}}$$

$$\vec{d_{k+1}} = -\vec{g_{k+1}} + \beta_k \vec{d_k}$$

$$\beta_k = \frac{\vec{g_{k+1}}^T Q \vec{d_k}}{\vec{d_k}^T Q \vec{d_k}}$$

where $\vec{g} = Q\vec{x_k} - \vec{b}$

(d). Convergence properties of conjugate directions methods (including the conjugate gradient method).

Both the conjugate gradient & conjugate directions methods are guaranteeds to converge to the actual minimizer in at most n steps, for quadratic minimization problems in E^n. (much better than the method of steepest descent).