Lecture 7
Midterm : 6:10 – 8 pm
same room
Briefly: will cover through conjugate direction method, conjugate gradiant method
Hessian, convex: local $\Rightarrow$ global
Global cvg thm
Steepest descent
$E(X_{n+1}) \leq \left(\frac{r-1}{r+1}\right)^2 E(X_n)$
$E(x) = f(x) - f(x^*)$, f quadratic

F 3-5 HU 1001B extra OH

<u>Main topic for now</u>
minimize f with constraints $h_1(x) = \cdots = h_m(x) = 0$
$g_1, \cdots, g_r \leq 0$
we will write
$\underline{h}(x) = 0$ for $h = (h_1, \cdots, h_m)$
$\underline{g}(x) \leq 0$ for $g = (g_1, \cdots, g_r)$

Last week: equality constraint.
- defined tangent space (plane)
- defined a regular pt. of the surface S defined by constrain

X is a regular point if $\nabla h_i(X)$, $i = 1, \cdots, m$ are Linearly independent.
Remark: for well-formulated problem, most of all pts are regular.

Thm: if $x \in S$ is a regular point, then the tangent space at $x \in S$ is
$\{y : \mathbb{R}^n : \nabla \underline{h}(x) y = 0\}$

i.e. $\nabla h_i(x) y = 0$, for $i = 1, 2, \cdots, m$

$h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$ So $\nabla \underline{h} = \begin{pmatrix} \nabla h_1 \\ \vdots \\ \nabla h_m \end{pmatrix}$

Thm: If x^* is a local min of f with constraint $\underline{h}(x) = 0$, then there is some $\lambda \in \mathbb{R}^m$ s.t.
$\nabla f(x^*) + \lambda \nabla \underline{h}(x^*) = 0$
$\underline{h}(x^*) = 0$

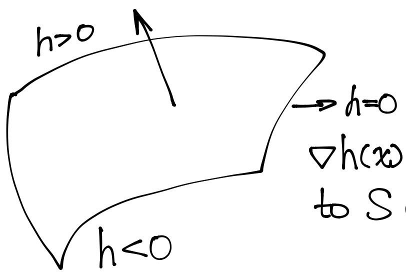
n eqt'n

m eqt'n

Note: $(m+n)$ eqt'n, $(m+n)$ unknown, $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^m$

λ is called "Lagrange multiplier"

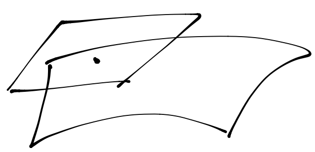
If only 1 constraint easy to . . .



$\nabla h(x)$ orthogonal to S at x

$\nabla f(x) + \lambda \nabla h(x) = 0$
says $\nabla f(x^*)$ is multiple of $\nabla h(x)$

2nd order conditions
Necessary conditions
Suppose x^* is a local min pt of f, (also regular point), s.t. constraint $\underline{h}(x)=0$, $f, h \in C^2$
$\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0$ for some $\lambda \in \mathbb{R}^m$
Let $L = \nabla^2 f + \lambda^T \nabla^2 \underline{h}$ (same λ)
Then $y^T L y \geq 0$ for $\forall y \in M = \{y : \nabla h(x^*) y = 0\}$



Proof: For $y \in M$, know from last week that $\exists$ a curve $x(t)$, $a<t<b$, $a<0, b>0$ s.t
$x(t) \in S$ for all t, $x(0) = x^*$ & $x'(0) = y$
Let $g(t) = f(x(t)) + \lambda^T h(x(t)) = f(x(t))$
since $h(x(t)) = 0$ for all $x \in S$. $g(t)$ has local min. at $t=0$, since f has local min at $x(t) = x^*$

so $g'(0) = 0$, $g''(0) \geq 0$
compute: $g'(t) = \nabla f(x(t)) \cdot x'(t) + \lambda^T \cdot \nabla h(x(t)) \cdot x'(t)$
$g''(t) = x'(t)^T \nabla^2 f(x(t)) \cdot x'(t) + \nabla f(x(t)) \cdot x''(t) + \text{same for } \lambda^T h$?

so when $t=0$, $x(0) = x^*$, $x'(0) = y$ and so
$g''(0) = y^T(\nabla^2 f(x^*) + \lambda^T \nabla^2 \underline{h}(x^*))y + \underbrace{(\nabla f(x^*) + \lambda^T \nabla h(x^*))}_{=0 \text{ by F.O.C.}} x''(0)$

$g''(0) \geq 0 \implies y^T L y \geq 0$

Sufficient condition:
same set-up as above,
$y^T L y > 0$, $\forall y \neq 0$ in $\mathbb{R}^n$, $\implies$ strict local min
$(\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0$, x^*: regular point$)$.

Ex: $\min f(x_1, \cdots, x_4) \equiv (x_1 x_2 x_3 + x_1 x_2 x_4 + x_1 x_3 x_4 + x_2 x_3 x_4)$ s.t $x_1 + x_2 + x_3 + x_4 = 4$

Procedure:
① Identify "candidates by Lagrange multiplier problem: ② investigate, have I found (local) minimizer?

① Lagrange multiplier problem:

$$\begin{cases} \nabla f(x^*)+\lambda \nabla h(x^*)=0 \\ h(x^*)=0 \end{cases}$$

1st constraint:

$\mathcal{L}_{x_1}: -(x_2x_3+x_2x_4+x_3x_4)+\lambda(1)=0$

$\mathcal{L}_{x_2}: -(x_1x_4+x_1x_3+x_3+x_4)+\lambda=0$

$\mathcal{L}_{x_3}: -(x_1x_2+x_1x_4+x_2x_4)+\lambda=0$

$\mathcal{L}_{x_4}: -(x_1x_2+x_1x_3+x_2x_3)+\lambda=0$

$x_1+x_2+x_3+x_4=4$

1 soln: $x=(1,1,1,1)$, $\lambda=3$ (only soln)

$$L=\nabla^2 f(x^*) = -\begin{pmatrix} 0 & 2 & 2 & 2 \\ 2 & 0 & 2 & 2 \\ 2 & 2 & 0 & 2 \\ 2 & 2 & 2 & 0 \end{pmatrix}$$

Q: is it true that $y^T L y > 0$ if $\nabla h(x^*)\cdot y = 0$ i.e. $y_1+y_2+y_3+y_4=0$

Note: a basis for M is $\{(1,-1,0,0),(1,0,-1,0),(1,0,0,-1)\}$ enough to check $y_j^T L y_j \geq 0$ $j=1,2,3$

For y_1, $y_1^T L y_1 = (1,-1,0,0)\begin{pmatrix} & \end{pmatrix}\begin{pmatrix} -1 \\ -1 \\ 0 \\ 0 \end{pmatrix}$

y_2, y_3 basically the same so y=basis should be orthogonal, so "encouraging"

In this case, note that

$$L=2\begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}-2\begin{pmatrix} 1&1&1&1 \\ 1&1&1&1 \\ 1&1&1&1 \\ 1&1&1&1 \end{pmatrix}=2I-2\begin{pmatrix}1\\1\\1\\1\end{pmatrix}(1\ 1\ 1\ 1)$$

$$=2I-2\nabla h^T\cdot\nabla h$$

So if $y\in M$, $\nabla h(x)y=0$.

So $y^T L y = y^T(2I)y - 2y^T\nabla h^T \underbrace{\nabla h y}_{0} = 2y^Ty > 0$ if $y\neq 0$

Last example:

① found x^*, λ

② computed $L=\nabla^2 f(x^*)+\lambda\underbrace{\nabla^2 h(x^*)}_{}$

0, because h linear

③ Is $y^T L y>0$ for non-zero $y\in M$

For ③, give 2 answers

1) choose basis for M and do some computations
2) a trick: usually not available
1) is incomplete, will post online

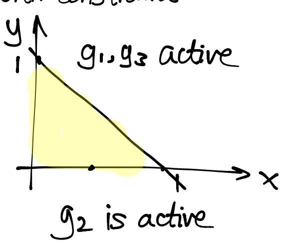
Inequality constraints
minimize f with $\underline{h}(x)$, $\underline{g}(x) \le 0$

Idea: reduce to the case of equality constraints

Def'n: Sps x satisfied
$\underline{h}(x)=0$, $\underline{g}(x) \le 0$
$g_j(x) \le 0$, $j=1,2,\cdots,r$

Def'n: jth constraint is active if $g_j(x)=0$
$T = \{j : g_j(x)=0\}$ = set of active constraints

Say we minimize $f(x,y)$ with constraint
$g_1: -x \le 0$
$g_2: -y \le 0$
$g_3: x+y-1 \le 0$



Idea: relate problem to minimization with equality constraints $\underline{h}(x)=0$, $g_j(x)=0$, j active and mostly forget about inactive constraints

Def'n: Sps x satisfies $\underline{h}(x)=0$, $\underline{g}(x) \le 0$, x is regular if $\nabla h_i(x)$, $\nabla g_j(x)$, $i=1,2,\cdots,m$
j active are L.I.

Like before, most or all pts are regular for well-formulated problems

F.O.C.
(Karush-Kuhn-Tucker conditions)
Let x^* be local min of f, s.t. $\underline{h}(x)=0$, $\underline{g}(x) \le 0$. Then $\exists x \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$ s.t.

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) + \mu^T \nabla g(x^*) = 0$$
$$\mu^T g(x^*) = 0$$
$$\mu \ge 0, \text{ i.e. } \mu_j \ge 0 \text{ for } \forall j, \quad h(x^*)=0$$

Note: $\mu^T g = 0$
says $\mu_j = 0$ if jth constraint is not active

Partial proof:
x^* minimizes f with constraints $h(x)=0$, $g(x)<0$
The def'n of active says x^* minimizes (is a local min) of f with constraints $\underline{h}(x)=0$, $g_j(x)=0$, j active

So previous theorem (Lagrange multipliers) applies with $\underline{h}$ replace by (h_i, g_j),
$i=1,\cdots,m$, j active

So I get a Lagrange multiplier, which I call (λ_i, μ_j) $i=1,\cdots,n$, $j\in J$

If j inactive, set $\mu_j=0$
Then

$$\nabla f(x^*)+\lambda^T\nabla h(x^*)+\mu^T\nabla g(x^*)=0$$
$$\mu^T g(x)=0$$

because of choice of μ_j for inactive & definition of active.
This finishes the proof except for $\mu>0$

Let's give the idea without a complete proof

For simplicity, first consider only 1 inequality constraint.
i.e. minimizes $f(x)$ with constraint $g(x)<0$ ← scalar

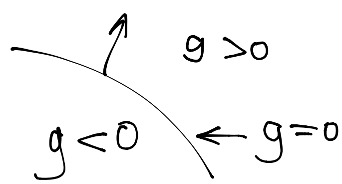
know that at x^*

$$\nabla f(x^*)+\mu\cdot\nabla g(x^*)=0$$
$$\mu\cdot g(x^*)=0$$

i.e. $\mu=0$ if inactive

Claim: $\mu\geqslant 0$
True if constraint is inactive, so consider active constraint

$$\nabla f(x^*)=-\mu\nabla g(x^*)$$



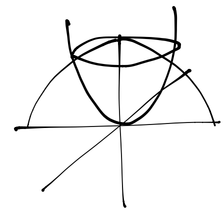
If $\mu<0$, then $\nabla f(x^*)$ and $\nabla g(x^*)$ point in the same direction away from the set where $g(x)\leq 0$

If this was true, can decrease f by moving in $-\nabla f(x^*)$ direction (into constraint set) impossible.

In general case,
$\underline{h}(x)=0$, $\underline{g}(x)\leq 0$
vectors

$\mu\geqslant 0$ for some reason (but slightly more complicated)

Ex: $f=x_1+x_2+x_3$
$h=x_1^2+x_2^2-x_3$
$g=x_1^2+x_2^2+x_3^2-6$
min f with $h=0$, $g\leq 0$



$h=0$ if $x_3=x_1^2+x_2^2$ parabola
$g=0$ if $x_1^2+x_2^2+x_3^2\leq 6$

Unfortunately, don't know if g is active, consider different cases.

Case 1: g inactive

$\nabla f(x^*) + \lambda \nabla h(x^*) = 0$ since $\mu = 0$

$\frac{\partial}{\partial x_1}$ $1 + \lambda \cdot 2x_1 = 0$

$\frac{\partial}{\partial x_2}$ $1 + \lambda \cdot 2x_2 = 0$

$\frac{\partial}{\partial x_3}$ $1 + \lambda(-1) = 0$

$\lambda = 1$, $x_2 = x_1 = -\frac{1}{2}$ $x_3 = x_1^2 + x_2^2 = \frac{1}{2}$

$x^* = (\frac{1}{2}, \frac{1}{2}, -\frac{1}{2})$, $f(x^*) = -\frac{1}{2}$

Case 2: g active

$\nabla f + \lambda \nabla h + \mu \nabla g = 0$

$g(x^*) = 0$

$h(x^*) = 0$

$1 + 2\lambda x_1 + 2\mu x_2 = 0$

$1 + 2\lambda x_2 + 2\mu x_2 = 0$

$1 - \lambda + 2\mu x_3 = 0$

$x_1 = x_2 = \frac{1}{2(\mu + \lambda)}$

Also, $h = g = 0$

$x_1^2 + x_2^2 - x_3 = 0$

$x_1^2 + x_2^2 + x_3^2 = 6$

$\Rightarrow x_3^2 + x_3 - 6 = 0$

$(x_3 - 2)(x_3 + 3) = 0$

$\Rightarrow x_3 = 2$

So $x_1 = x_2$, $x_3 = 2$

$x^* = (1, 1, 2)$, $\lambda = -\frac{2}{10}$, $\mu = -\frac{3}{10}$

$\Downarrow f = 4$

$x^* = (-1, -1, 2) \longrightarrow f = 0$, $\lambda = \frac{6}{10}$, $\mu = -\frac{1}{10}$

not minimizer since $\mu < 0$.