Nonlinear Optimization

Note Title 11-06-2008

Unconstrained Optimization :

Minimize a non-linear function $F(x) \in \mathbb{R}$ over parameters
$x = [x_1 \cdots x_n]^T$

Denote $x^* = \arg \min\limits_{\underbrace{x \in \mathbb{R}^n}_{\text{unconstrained}}} F(x)$

Types of minima

Strong : $F(x)$ increases locally in all directions from x^*.

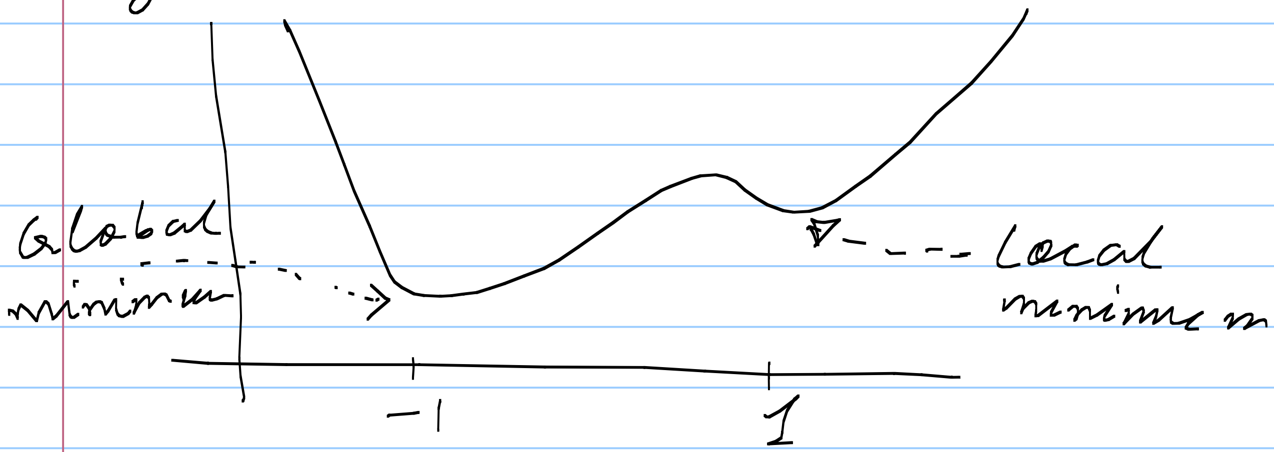
Defn :
A point x^* is a strong minimum of a function $F(x)$ if $\exists \delta > 0$ s.t. $F(x^*) < F(x^* + \Delta x)$ $\forall \Delta x$ s.t. $0 < \|\Delta x\| \leq \delta$

Weak : $F(x)$ remains same or increases locally around x^*

A point x^* is a weak minimum of $F(x)$ if it is not a strong minimum and $\exists \delta > 0$ s.t. $F(x^*) \leq F(x^* + \Delta x)$ $\forall \Delta x$ s.t. $0 < \|\Delta x\| \leq \delta$

Global minimum : A minimum is global if the above definitions hold for $\delta = \infty$. Otherwise there are local minima.

E.g: $F(x) = x^4 - 2x^2 + x + 3$



First Order Conditions

If $F(x)$ has continuous second derivatives, it can be approximated by a Taylor series in the neighborhood of an arbitrary pt.

$$F(x+\Delta x) \approx F(x) + g^T(x)\,\Delta x + \frac{1}{2}\Delta x^T G(x)\,\Delta x + \cdots$$

gradient $g = \left(\frac{\partial F}{\partial x}\right)^T = \begin{bmatrix} \frac{\partial F}{\partial x_1} \\ \frac{\partial F}{\partial x_2} \\ \vdots \\ \frac{\partial F}{\partial x_n} \end{bmatrix}$

$$G = \begin{bmatrix} \frac{\partial^2 F}{\partial x_1^2} & \cdots & \frac{\partial^2 F}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 F}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 F}{\partial x_n^2} \end{bmatrix}$$

Since $g^T(x)\Delta x$ can be both + or −ve, $F(x+\Delta x) > F(x)$ is possible for all Δx if $g(x^*) = 0$

Hence $g(x^*) = 0$ is necessary condition for local minimum.

$g(x^*) = 0$ is necessary and sufficient condition for x^* to be a stationary point.

2nd Order Conditions

Set $g(x^*) = 0$

$$F(x^* + \Delta x) \approx F(x^*) + \frac{1}{2}\Delta x^T G(x^*)\Delta x + \cdots$$

Strong minimum:

$$\Delta x^T G(x^*)\Delta x > 0 \quad \forall \Delta x \neq 0$$

$$\Rightarrow \quad F(x^* + \Delta x) > F(x^*)$$

Hence sufficient condition for a strong local minimum is

$$G(x^*) > 0 \quad \text{(Positive definite)}$$

Summary: $g(x^*) = 0$ & $G(x^*) > 0$

→ Sufficient condition for a strong local minima

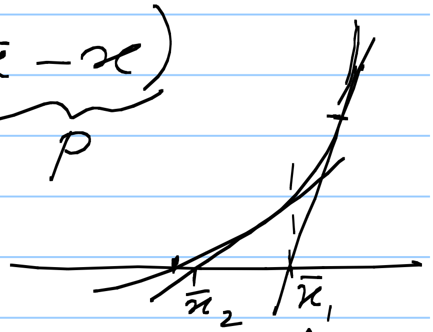
Numerical Solution Methods

Newtons' Method: Solve $g(x^*) = 0$

Step 1: Linearly approximate $g(x)$ about an arbitrary point x using Taylor series. Then for any point $\bar{x}$ in the neighbourhood of x,

$$g(\bar{x}) = g(x) + G(x)\underbrace{(\bar{x} - x)}_{p}$$

$$= g(x) + G(x)p$$



Step 2: Solve for $\bar{x}/p$ s.t. $g(\bar{x}) = 0$

$$p = -G^{-1}(x)\, g(x) \quad \left[\text{Assuming } G(x) \text{ is invertible}\right]$$

and $\bar{x} = x + p$

In general $g(\bar{x}) \neq 0$ but $\bar{x}$ might be a better estimate for x^* than x. Hence $\bar{x}$ is taken as the new x and step 1 & 2 are repeated

Note: 1) This method converges to a stationary pt. (not necessarily to a minimum).

2) computation of G is expensive (Quasi-Newton methods)

Equality Constrained Optimization

$$\min_{u \in \mathbb{R}^m} F(x,u)$$

s.t. $f(x,u)=0 \rightarrow n$ eqns

$x \in \mathbb{R}^n$

$$f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{bmatrix}$$

Basic Requirement: 1) Given u, x should be solvable from $f(x,u)=0$

2) The x, u division is purely for convenience and any suitable partition is fine.

Simple method: direct substitution — works only if $c(x)$ is linear.

Example: $\min F = x_1^2 + x_2^2$
s.t. $x_1 + x_2 + 2 = 0$

$$x_1 = -x_2 - 2$$

Then equivalent problem is

$$\min_{x_2} F_{x_2} = (-2-x_2)^2 + x_2^2$$

Solve $\dfrac{\partial F_{x_2}}{\partial x_2} = 0 \Rightarrow x_2 = -1$

Similarly $x_1 = -1$

Q. What happens if both $F(x, u)$ and $f(x, u)$ are linear in both x and u?

Stationary pt. in the context of the equality constrained problem is one where $dF = 0$ for arbitrary du holding $df = 0$.

Assuming some non-linearity:

$$dF = \frac{\partial F}{\partial x} dx + \frac{\partial F}{\partial u} du$$

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial u} du$$

We require $df = 0$ (since any perturbation must obey the constraint).

Assume: $\frac{\partial f}{\partial x}$ is non-singular

[This is equivalent to the assumption that u determines x through $f(x, u) = 0$. Why?]

Then $dx = -\left[\frac{\partial f}{\partial x}\right]^{-1}\left[\frac{\partial f}{\partial u}\right] du$

Then $dF = \left[\frac{\partial F}{\partial u} - \frac{\partial F}{\partial x}\left[\frac{\partial f}{\partial x}\right]^{-1}\frac{\partial f}{\partial u}\right] du$

Now if $dF = 0$ for arbitrary du
it is necessary that

$$\left\{\begin{array}{l} \dfrac{\partial F}{\partial u} - \dfrac{\partial F}{\partial x}\left[\dfrac{\partial f}{\partial x}\right]^{-1}\dfrac{\partial f}{\partial u} = 0 \Big\} \ m \text{ eqns} \\ + \\ f(x,u) = 0 \quad \rightarrow n \text{ eqns} \end{array}\right.$$

$m+n$ eqns $\longleftrightarrow$ $\underset{\downarrow \atop u}{m} + \underset{\downarrow \atop x}{n}$ unknowns

Better Method: Lagrange Multipliers

Adjoin the constraints to the perf. index by n "undetermined multipliers" $\lambda_1 \cdots \lambda_n$. $\lambda = [\lambda_1 \cdots \lambda_n]^T$

Define the Lagrangian:

$$H(x,u,\lambda) = F(x,u) + \lambda^T f(x,u)$$

$$dH = \frac{\partial H}{\partial x}dx + \frac{\partial H}{\partial u}du$$

Since λ is our choice, we can make $\frac{\partial H}{\partial x} = 0$ by proper choice of λ:

$$\frac{\partial H}{\partial x} = \frac{\partial F}{\partial x} + \lambda^T \frac{\partial f}{\partial x} = 0$$

$$\Rightarrow \quad \lambda^T = -\left[\frac{\partial F}{\partial x}\right]\left[\frac{\partial f}{\partial x}\right]^{-1} \leftarrow \left\{\begin{array}{l}\text{The inv.}\\ \text{assumption}\\ \text{is still req}\end{array}\right.$$

Replacing, $dH = \frac{\partial H}{\partial u} du$. ——— (1)

Now, whatever perturbation we try in u, $f(x,u)=0$ must hold
i.e. $df = 0$

For that we saw, $dx = \left[\frac{\partial f}{\partial x}\right]^{-1} \frac{\partial f}{\partial u} du$

Clearly, for such (dx, du) combination
$dF = dH$

Proof:

$$dF = \frac{\partial F}{\partial x} dx + \frac{\partial F}{\partial u} du$$

$$= \left[\underbrace{\frac{\partial F}{\partial x} \left[\frac{\partial f}{\partial x}\right]^{-1}}_{\lambda^T} \frac{\partial f}{\partial u} + \frac{\partial F}{\partial u} \right] du$$

$$= \left[\frac{\partial F}{\partial u} + \lambda^T \frac{\partial f}{\partial u} \right] du$$

$$= \frac{\partial H}{\partial u} du \rightarrow \text{Same as RHS of (1)}$$

$$= dH$$

Hence for $dF = 0$ $\forall du$ while holding $f(x,u)=0$

$$\frac{\partial H}{\partial u} = 0$$

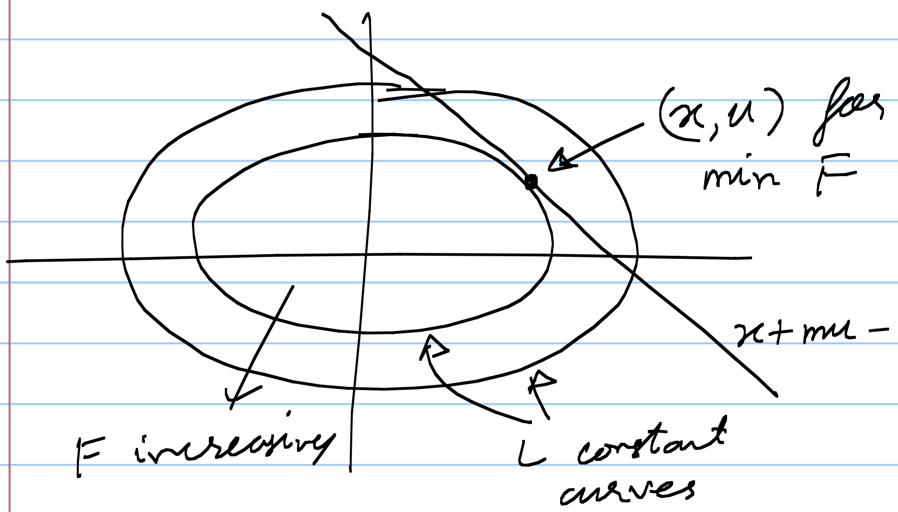
Necessary condition for stationary point:

$$\frac{\partial H}{\partial x} = 0 \; ; \; \frac{\partial H}{\partial u} = 0 \; ; \; \frac{\partial H}{\partial \lambda} = f(x,u) = 0$$

Example: $\min_{u} F(x,u) = \frac{1}{2}\left(\frac{x^2}{a^2} + \frac{u^2}{b^2}\right)$

s.t. $f(x,u) = x + mu - c = 0$

(a, b, m, c) are scalar constants.



$$H = \frac{1}{2}\left(\frac{x^2}{a^2} + \frac{u^2}{b^2}\right) + \lambda(x + mu - c)$$

Necessary conditions:

$$x + mu - c = 0, \quad \frac{\partial H}{\partial x} = \frac{x}{a^2} + \lambda = 0$$

$$\frac{\partial H}{\partial u} = \frac{u}{b^2} + \lambda m = 0$$

Solⁿ:

$$x = \frac{a^2 c}{a^2 + m^2 b^2} \; ; \; u = \frac{b^2 m c}{a^2 + m^2 b^2}, \; \lambda = -\frac{c}{a^2 + m^2 b^2}$$

$$F_{min} = \frac{c^2}{2(a^2 + m^2 b^2)}$$

Q. What if these eqns are not so easily solvable ?

Algebraic Interpretation of Lag. mult

For every allowed perturbations (dx, du) about the stationary pt, the necessary eqns:

$$dF = 0 \quad \text{and} \quad df = 0$$

must be consistent

$$dF = \frac{\partial F}{\partial x} dx + \frac{\partial F}{\partial u} du = 0$$

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial u} du = 0$$

$dF =$

$$\begin{bmatrix} \frac{\partial F}{\partial x_1} & \frac{\partial F}{\partial x_2} & \cdots & \frac{\partial F}{\partial x_n} \end{bmatrix} \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix} + \begin{bmatrix} \frac{\partial F}{\partial u_1} & \cdots & \frac{\partial F}{\partial u_m} \end{bmatrix} \begin{bmatrix} du_1 \\ \vdots \\ du_m \end{bmatrix}$$

$df =$

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix} \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix} + \begin{bmatrix} \frac{\partial f_1}{\partial u_1} & \cdots & \frac{\partial f_1}{\partial u_m} \\ \vdots & & \\ \frac{\partial f_n}{\partial u_1} & \cdots & \frac{\partial f_n}{\partial u_m} \end{bmatrix} \begin{bmatrix} du_1 \\ \vdots \\ du_m \end{bmatrix}$$

$$\begin{array}{cc} & (n+m)\text{ cols} \\ \text{Then}\ (n+1)\text{ rows} & \begin{bmatrix} \frac{\partial F}{\partial x_1} & \cdots & \frac{\partial F}{\partial x_n} & \frac{\partial F}{\partial u_1} & \cdots & \frac{\partial F}{\partial u_m} \\ \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} & \frac{\partial f_1}{\partial u_1} & \cdots & \frac{\partial f_1}{\partial u_m} \\ \vdots & & & \vdots & & \\ \frac{\partial f_n}{\partial x_1} & & \frac{\partial f_n}{\partial x_n} & \frac{\partial f_n}{\partial u_1} & \cdots & \frac{\partial f_n}{\partial u_m} \end{bmatrix} \end{array} \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \\ du_1 \\ \vdots \\ du_m \end{bmatrix} = 0$$

Now whatever dx, du satisfy $df = 0$ should also satisfy $dF = 0$

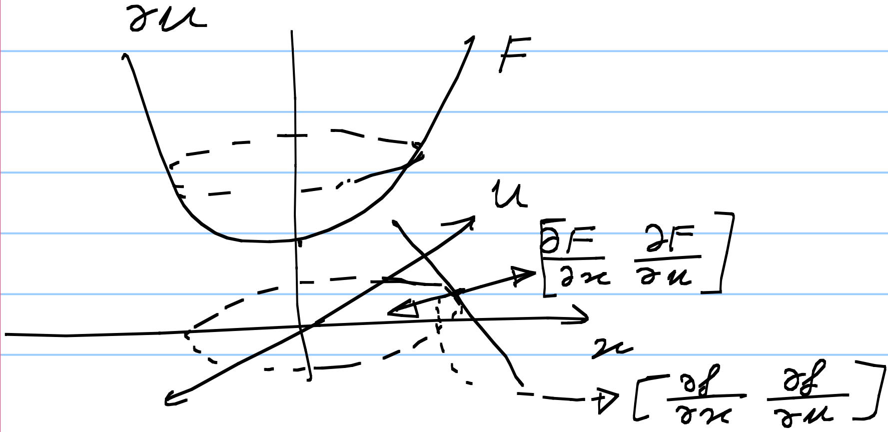
$\Rightarrow$ The top row should be linearly dependent on the bottom n rows

$\Rightarrow \exists$ n constants $\lambda = [\lambda_1 \cdots \lambda_n]^T$ s.t

$$\left.\begin{aligned} \frac{\partial F}{\partial x} + \lambda^T \frac{\partial f}{\partial x} &= 0 \\ \frac{\partial F}{\partial u} + \lambda^T \frac{\partial f}{\partial u} &= 0 \end{aligned}\right\}$$

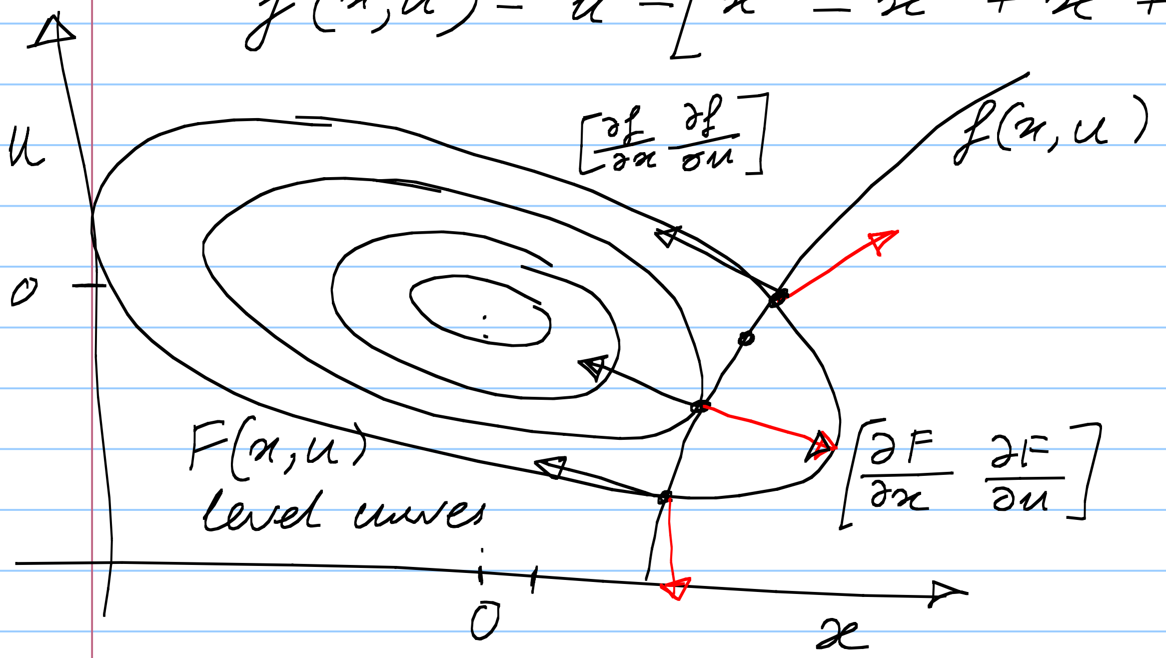
$\Rightarrow$ $\frac{\partial F}{\partial x}$ should be in the span of $\frac{\partial f}{\partial x}$

& $\frac{\partial F}{\partial u}$ " " " $\frac{\partial f}{\partial u}$



Example: $F(x,u) = \frac{1}{2}[x \; u]\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x \\ u \end{bmatrix}$

$$f(x,u) = u - [x^3 - x^2 + x + 2]$$



Note: Soln. to the necessary conditions only imply stationary pt.

Newton's Method for Inequality constrained optimization

Define $y = \begin{bmatrix} x \\ u \end{bmatrix}$

$$H(y) = F(y) + \lambda^T f(y)$$

$$\frac{\partial H}{\partial y} = 0 = \frac{\partial F}{\partial y} + \lambda^T \frac{\partial f}{\partial y} \rightarrow (1)$$

$$\frac{\partial H}{\partial \lambda} = 0 = f(y) \text{ ——— } (2)$$

Linearize ① & ② about (y, λ) using Taylor series: Then at $\bar{y}, \bar{\lambda}$

$$\frac{\partial H}{\partial y}^T(\bar{y}) = \Big\{ \underbrace{\frac{\partial F}{\partial y}^T(y)}_{g} + \underbrace{\frac{\partial f}{\partial y}^T(y)}_{G}\lambda \Big\}$$

$$+ \underbrace{\left[\frac{\partial^2 F}{\partial y^2} + \sum \lambda_i \frac{\partial^2 f_i}{\partial y^2} \right]}_{\text{Name as } Q} (\bar{y} - y)$$

$$+ \underbrace{\frac{\partial f}{\partial y}^T}_{G^T} (\bar{\lambda} - \lambda)$$

Hence, from ①

$$\boxed{g + G^T \lambda + Q(\bar{y} - y) + G^T(\bar{\lambda} - \lambda) = 0}$$

and from ②,

$$f(y) + \frac{\partial f}{\partial y}(\bar{y} - y) = 0$$

or

$$\boxed{f + G(\bar{y} - y) = 0}$$

In matrix form:

$$(KKT) \;---\; \begin{bmatrix} Q & G^T \\ G & 0 \end{bmatrix} \begin{bmatrix} (\bar{y} - y) \\ \bar{\lambda} \end{bmatrix} = \begin{bmatrix} -g \\ -f \end{bmatrix}$$

We can solve for $p = (\bar{y} - y)$ and $\bar{\lambda}$ from KKT and use
$\bar{y} = y + p$ and $\bar{\lambda}$
as the new guesses. Then iterate.

Q. Derive the necessary condition for stationary pt when the objective is quadratic with linear constraints? Is this related to KKT equations?

Problems with Inequality constraints

$$\min \quad F(y)$$
$$\text{s.t.} \quad f(y) \leq 0$$

In general $\dim f \neq \dim y$ but one is not always greater than the other.

One dimensional case

Let y^o be the optimal value:
Then either

Case I: $f(y^o) < 0$

or Case II: $f(y^o) = 0$

For Case I; the constraint is ineffective and can be ignored. $\rightarrow$ Unconstrained optimization

For case II, for small perturbations about y^o,

$$dF = \left.\frac{\partial F}{\partial y}\right|_{y^o} dy \geq 0 \qquad \text{—— (1)}$$

for all admissible values of dy which satisfy

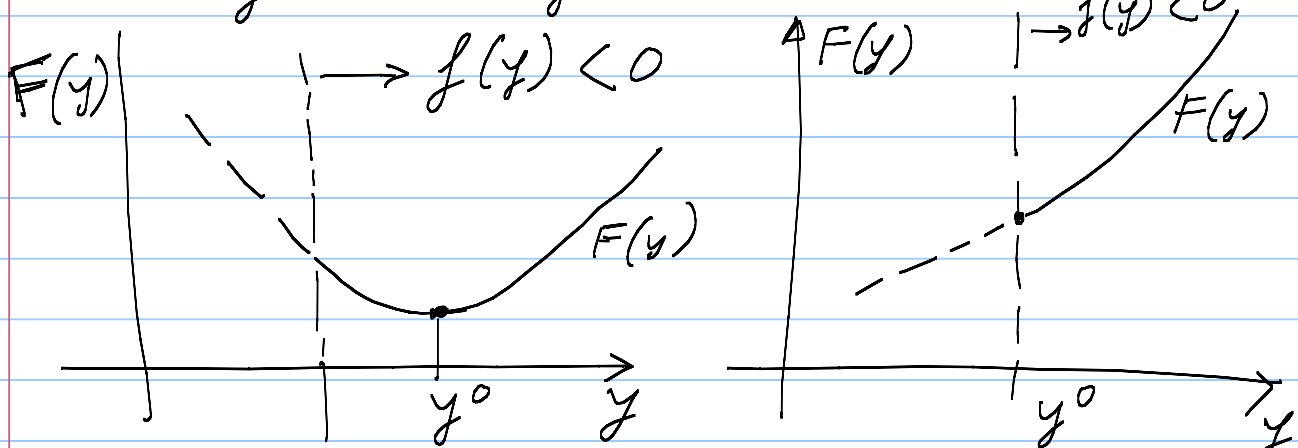
$$df = \left.\frac{\partial f}{\partial y}\right|_{y_o} dy \leq 0 \qquad \text{—— (2)}$$

(1) and (2) are simultaneously possible if

$$\operatorname{sgn} \frac{\partial F}{\partial y} = -\operatorname{sgn} \frac{\partial f}{\partial y} \quad \text{OR} \quad \frac{\partial F}{\partial y} = \frac{\partial f}{\partial y} = 0$$

Expressed together:

$$\frac{\partial F}{\partial y} + \lambda \frac{\partial f}{\partial y} = 0 \qquad \lambda \geq 0 \qquad \text{—— (*)}$$



The problem can be treated by Lag Mult:

$$H(y, \lambda) = F(y) + \lambda f(y)$$

From ⊛, $\frac{\partial H}{\partial y} = 0$ & $f(y) \leq 0$

where $\lambda \geq 0$, $f(y) = 0$
$\lambda = 0$, $f(y) < 0$

General case:

$$\min F(y) \quad \text{s.t.} \quad f(y) \underset{}{\overset{\text{component wise}}{\leq}} 0$$

(Both y and f are vectors)

Still if y° is a minimum, then

$$dF = \left.\frac{\partial F}{\partial y}\right|_{y^{\circ}} dy \geq 0$$

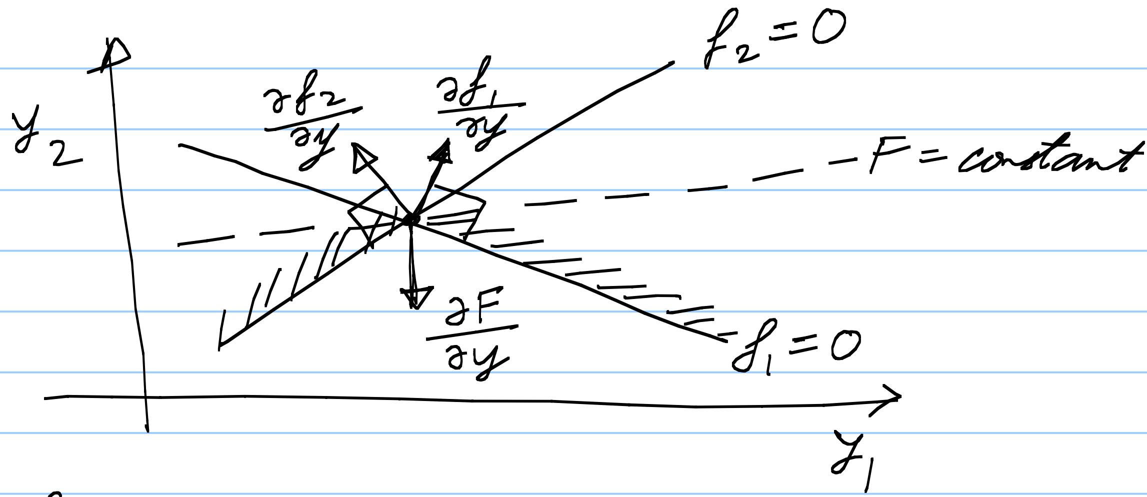
for all dy satisfying

$$\underset{\text{vector}}{df} = \underbrace{\frac{\partial f}{\partial y}}_{\text{matrix}} \underset{\text{vector}}{dy} \underset{\text{component wise}}{\leq} 0$$

Hence:

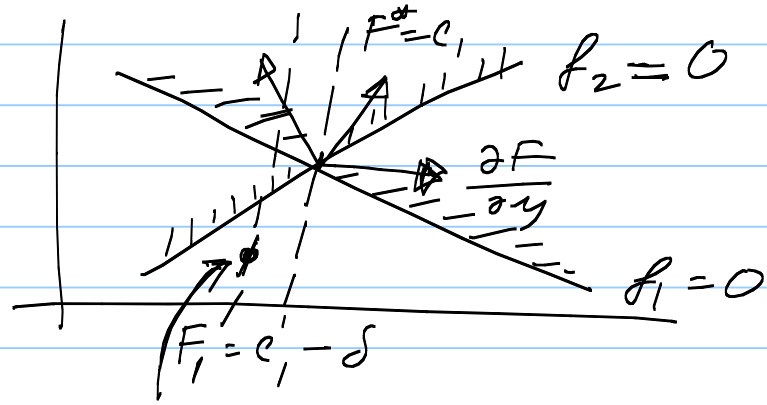
$$\frac{\partial F}{\partial y} + \lambda^{T} \frac{\partial f}{\partial y} = 0, \quad \lambda \geq 0$$

For illustration assume only two components of f, (say f_1 and f_2) are effective



If y^o is a minimizer, $\frac{\partial F}{\partial y}$ must lie between $-\frac{\partial f_1}{\partial y}$ and $-\frac{\partial f_2}{\partial y}$

(Otherwise F could decrease



This pt. for example is satisfying $f_1 < 0$, $f_2 < 0$ and $F_1 < F^*$!

So F^* cannot be a minima.

In general, if q components of f are effective at the optimal pt:

$$\frac{\partial F}{\partial y} + \underset{\text{scalars}}{\lambda_1} \frac{\partial f_1}{\partial y} + \cdots + \lambda_q \frac{\partial f_q}{\partial y} = 0$$

vectors

with $\lambda_1, \ldots, \lambda_q \geq 0$

Suppose y has p component and n components of f are effective.
Then:

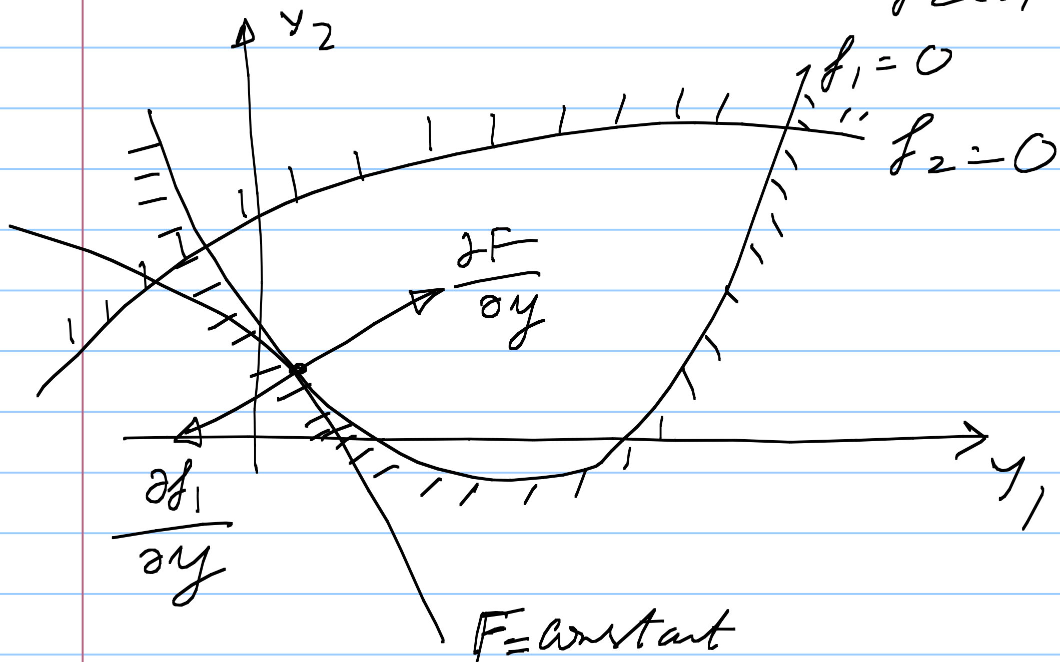
Necessary conditions for minimality

$$\frac{\partial H}{\partial y} = \frac{\partial F}{\partial y} + \lambda^T \frac{\partial f}{\partial y} = 0$$

where $\lambda \geq 0$ for $f_1(y), \dots f_n(y) = 0$
$\lambda = 0$ for $f_{n+1}(y), f_{n+2}(y) < 0$
$- - - -$

<u>Note</u>: For a maximum the sign of λ must change. How?

<u>Example</u>: $F(y_1, y_2)$ with $f_1(y_1, y_2) \leq 0$
$f_2(y_1, y_2) \leq 0$



Numerical Method – Active Set Strategy

$$\min F(y) \quad \text{s.t.} \quad f(y) \leq 0$$

We have seen at y^*,

1> $\quad f_i(y^*) = 0 \quad$ for $i \in A$

2> $\quad f_i(y^*) < 0 \quad$ for $i \in A'$

$A :=$ the active set
$A' :=$ the inactive set

Q. How to identify A / A' ?

A. Use the fact that $\lambda_i^* \geq 0$ for $i \in A$

Example: $F(x) = x_1^2 + x_2^2$

$$f(x) = -(x_1 + x_2 - 2) \leq 0$$

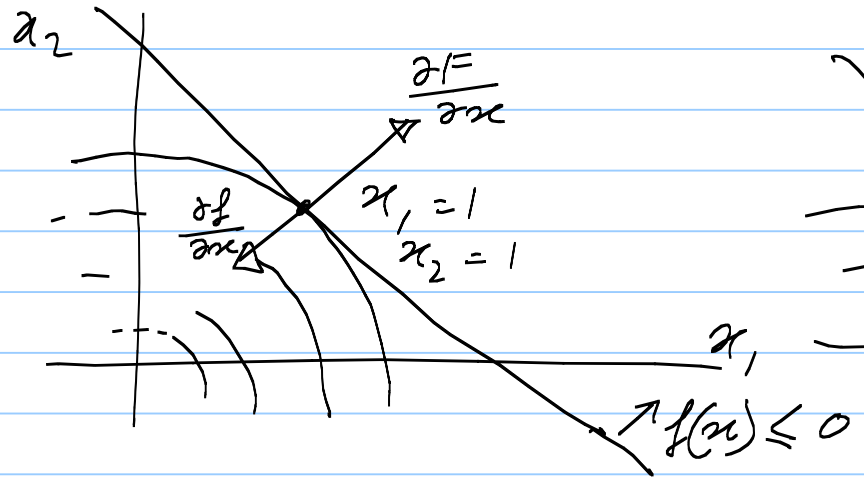
$$\frac{\partial F}{\partial x}^T = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} \qquad \frac{\partial f}{\partial x} = \begin{bmatrix} -1 & -1 \end{bmatrix}$$

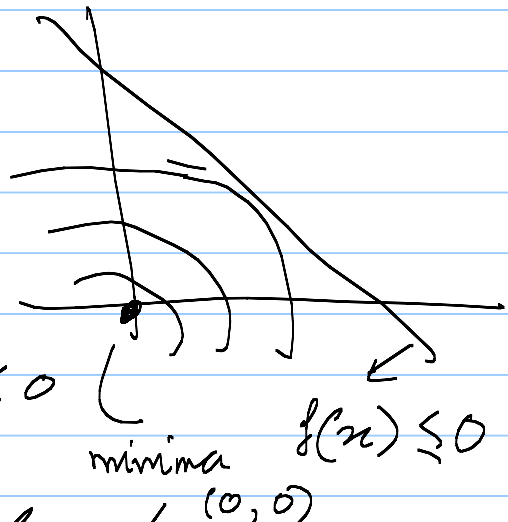
$$H = x_1^2 + x_2^2 + \lambda(-x_1 - x_2 + 2)$$

$$\frac{\partial H}{\partial x}^T = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \end{bmatrix} \lambda = 0$$

At $x_1 = x_2 = 1$, $\quad \lambda = 2 > 0$

So $x_1 = x_2 = 1$ is a minimum





Now change the problem to

$$\min F(x) = x_1^2 + x_2^2 \quad \text{s.t.} \quad x_1 + x_2 - 2 \leq 0$$

$$\frac{\partial H}{\partial x}^T = \left\{ \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \lambda \right\}_{x_1 = x_2 = 1} = 0 \Rightarrow \lambda = -2 \ngtr 0$$

Hence $x_1 = x_2 = 1$ is not a minima
The minima is at $(0,0)$.

Quadratic Programming: Active Set Method

$$\min F(x) = g^T x + \frac{1}{2} x^T H x \quad | \quad H > 0$$

$$\text{s.t.} \quad \begin{matrix} Ax = a \\ Bx \leq b \end{matrix} \quad \left| \quad \begin{matrix} Ax - a = 0 \\ Bx - b \leq 0 \end{matrix} \right\} \text{ compo-nentwise}$$

Assume that an estimate of the active set A^0 + a feasible pt. x^0 is given. Then:

1) Compute min with only active set constraints: Let $\tilde{b}$ be the subset

of b corresponding to active inequality constraint and $\tilde{B}$ be the corresponding Jacobian for the constraints in A_0. Then the KKT system:

$$\begin{bmatrix} Q & G^T \\ G & 0 \end{bmatrix} \begin{bmatrix} (\bar{y} - y) \\ \bar{\lambda} \end{bmatrix} = \begin{bmatrix} -g \\ -f \end{bmatrix}$$

$$= \left[\begin{array}{c|cc} H & A^T & \tilde{B}^T \\ \hline A & 0 & 0 \\ \tilde{B} & 0 & 0 \end{array}\right] \left[\begin{array}{c} \bar{x} - x \\ \hline \bar{\eta} \\ \bar{\lambda} \end{array}\right] = \begin{bmatrix} -g \\ -(Ax - a) \\ -(\tilde{B}x - b) \end{bmatrix}$$

2) Take the largest possible step in the direction of p that does not violate any <u>inactive</u> inequalities

i.e. $\bar{x} = x + \alpha p$

where $0 \leq \alpha \leq 1$ is chosen s.t. $B\bar{x} \leq b$ (only check for inactive inequality const)

3) For restricted step, i.e. $\alpha < 1$

$\rightarrow$ add the limiting inequality to the active set A_0 and return to step 1

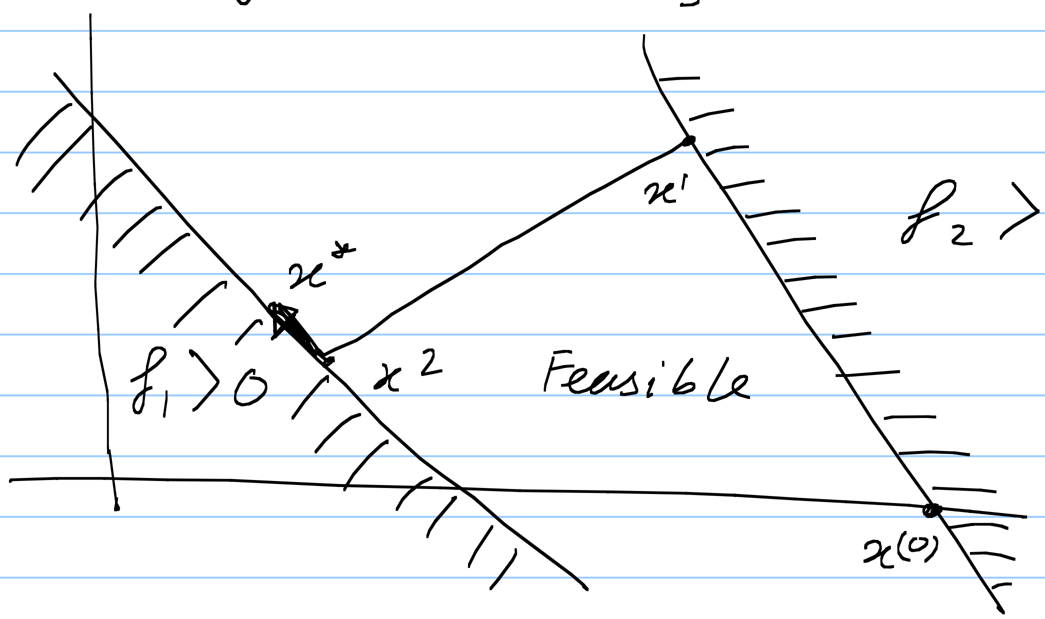
→ otherwise, take full step ($\alpha = 1$) ~~and check~~ sign of Lag. mult.

⇒ if all λ's are +ve, stop.

⇒ otherwise delete the inequality with most −ve λ from active set & return to step 1.

Example: $F(x) = x_1^2 + x_2^2 = \frac{1}{2}[x_1, x_2]\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$f_1(x) = 2 - x_1 - x_2 \leq 0$$
$$f_2(x) = x_1 + \frac{2}{3}x_2 - 4 \leq 0$$



Assume: $x^{(0)} = (4, 0)$
and $A^0 = \{f_2\}$

$$\left[\begin{array}{c|c} H & \frac{\partial f_2}{\partial x}^T \\ \hline \frac{\partial f_2}{\partial x} & 0 \end{array}\right]\begin{bmatrix} \bar{x}_1 - x_1 \\ \bar{x}_2 - x_2 \\ \bar{\lambda}_2 \end{bmatrix} = \begin{bmatrix} -\frac{\partial F}{\partial x}^T \\ -f_2 \end{bmatrix}_{\text{at } (4,0)}$$

$$\equiv \left[\begin{array}{cc|c} 2 & 0 & 1 \\ 0 & 2 & 2/3 \\ \hline 1 & 2/3 & 0 \end{array}\right] \left[\begin{array}{c} p_1 \\ p_2 \\ \hline \bar{\lambda}_2 \end{array}\right] = \left[\begin{array}{c} -8 \\ 0 \\ \hline 0 \end{array}\right]$$

$p_1 = -1.23$; $p_2 = 1.84$; $\bar{\lambda} = -5.5$

$\bar{x}_1 = x_1 + p_1 = 2.76$
$\bar{x}_2 = x_2 + p_2 = 1.84$

Since $\bar{\lambda} < 0$, f_2 is inactive and can be deleted from the active set.

Step 2: 2^{nd} QP: $x^T = (2.76, 1.84)$ with no active constraints, $A^1 = \{\phi\}$

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} -5.53 \\ -3.69 \end{bmatrix}$$

Solving for p_1, p_2, we see that it is not possible to take a full step $\rightarrow$ violates f_1. Instead take
$\bar{x} = x + \alpha p$ with $\alpha = 0.56$

$\bar{x} = \begin{bmatrix} 1.2 \\ 0.8 \end{bmatrix}$ Q. How to compute α?

Hence f_1 must be added $A^2 = \{f_1\}$

Step 3: 3^{rd} QP. $x^T = [1.2 \;\; 0.8]$, $A^2 = \{f_1\}$

$x^* = \begin{bmatrix} 1.2 \\ 0.8 \end{bmatrix} + \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $A^* = \{f_1\}$

Lagrange mult $\rightarrow \lambda^* = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \rightarrow$ so actual minimum.