d_1 and d_2 are conjugate if

$$d_1 Q d_2 = 0$$

$Q = I \Rightarrow d_1 \cdot d_2$ are orthogonal

Suppose : $Q = Q' > 0$ $\Rightarrow$ $d_i^T Q d_j = 0$ $i \neq j$

: $d_1, d_2 \ldots d_k$ $\Rightarrow$ $d_i^T Q d_i > 0$

- The directions of $d_1 \ldots d_k$ are linearly independent.

Suppose $d_1 \ldots d_k$ such that

$$\alpha_1 d_1 + \cdots + \alpha_k d_k = 0$$

$(\times Q)$ $\alpha_1 Q d_1 + \alpha_2 Q d_2 + \cdots + \alpha_k Q d_k = 0$

$(\times d_i')$ $\alpha_1 \underbrace{d_i' Q d_1}_{>0} + \alpha_2 \underbrace{d_i' Q d_2}_{=0} + \cdots + \alpha_k \underbrace{d_i' Q d_k}_{=0} = 0$

- There are at most n mutually independent directions

Let

$$f = \frac{1}{2} x' Q x + c' x + d \qquad Q = Q' > 0$$

f has a global min $x^* = -Q^{-1} c$

The conjugate direction method

$$x_0, \underbrace{d_0 \ldots d_{n-1}}_{Q \text{ conjugate}}$$

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

α_k performs a exact line search along d_k

$$\rightarrow \alpha_k = -\frac{\nabla' f(x_k) d_k}{d_k' Q d_k}$$

The sequence $\{x_k\} = \{x_0 \ldots x_n\}$ is such that

$$x_n = x^* = -Q^{-1} c$$

Define a new algorithm such that

$d_0, d_1, \dots d_{n-1}$ are Q-conjugate.

but not a-priori complicated.

$x_0, d_0 = -\nabla f(x_0)$

$x_{k+1} = x_k + \alpha_k d_k \qquad \overbrace{\alpha_k = -\frac{\nabla' f(x_k) d_k}{d_k' Q d_k}}^{\alpha^*}$

$d_{k+1} = -\nabla f(x_{k+1}) + \beta_k d_k$ pick d_k, d_{k+1} Q-conjugate

select β_k such that $\overbrace{d_k' Q d_{k+1} = 0}^{\text{Q-conjugate}}$

$d_k' Q d_{k+1} = -d_k' Q \nabla f(x_{k+1}) + \beta_k d_k' Q d_k$

$$\beta_k = \frac{d_k' Q \nabla f(x_{k+1})}{d_k' Q d_k}$$

$x_0 .\ d_0 = -\nabla f(x_0)$

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

$d_{k+1} = -\nabla f(x_{k+1}) + \beta_k d_k$

exact line search

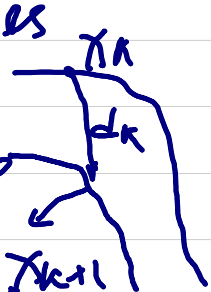
$$\alpha_k = -\frac{\nabla f(x_k)' d_k}{d_k' Q d_k}$$

$$\beta_k = \frac{\nabla f(x_{k+1})' Q d_k}{d_k' Q d_k}$$

Q-conjugacy

This algorithm is such that, for any x_0, the sequence $\{x_k\}$ converges to $x_0 = -Q^{-1}c$ in at most n steps, and the directions $d_0 \dots d_{n-1}$ are mutually Q-conjugate.

Properties



$\underbrace{d_k' \nabla f(x_{k+1}) = 0}_{\text{result of exact line search}}$

$\nabla' f(x_{k+1}) d_k = 0$

$\nabla' f(x_k) \nabla f(x_k) = -\nabla f(x_k)' d_k$

$d_{k+1} = -\nabla f(x_{k+1}) + \beta_k d_k \quad \times (\nabla f'(x_{k+1}))$

$\nabla f(x_{k+1})' d_{k+1} = -\nabla f(_{k+1})$

$$f = \frac{1}{2} x' Q x + c' x + d \qquad Q = Q' > 0$$

$$\nabla f = Qx + c$$

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

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

$$\underbrace{Q x_{k+1} + c}_{\nabla f(x_{k+1})} = \underbrace{Q x_k + c}_{\nabla f(x_k)} + \alpha_k Q d_k$$

$$Q d_k = \frac{\nabla f(x_{k+1}) - \nabla f(x_k)}{\alpha_k}$$

$$\beta_k = \frac{\nabla f(x_{k+1}) \left[\frac{\nabla f(x_{k+1}) - \nabla f(x_k)}{\alpha_k} \right]}{d_k' \left[\frac{\nabla f(x_{k+1}) - \nabla f(x_k)}{\alpha_k} \right]}$$

$x_0 . d_0 = -\nabla f(x_0)$

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

$d_{k+1} = -\nabla f(x_{k+1}) + \beta_k d_k$

$$\beta_k = \nabla f(x_{k+1})' \frac{[\nabla f(x_{k+1}) - \nabla f(x_k)]}{d_k[\nabla f(x_{k+1}) - \nabla f(x_k)]}$$

α_k is a line search parameter obtained using a sufficiently accurate line search algorithm.

$$= \frac{\nabla f(x_{k+1})'[\nabla f(x_{k+1}) - \nabla f(x_k)]}{-d_k' \nabla f(x_k)}$$

Quasi-Newton

$$x_{k+1} = x_k \underbrace{- [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)}_{d_k} \quad \text{Newton}$$

$$x_{k+1} = x_k - \alpha_k \underbrace{[\nabla^2 f(x_k)]^{-1}}_{??} \nabla f(x_k) \quad + \text{line search}$$

$$x_{k+1} = x_k - \alpha_k H_k \nabla f(x_k)$$

$H_{k+1} = ??$ $\quad\quad \{H_k\} \to [\nabla^2 f(x_k)]^{-1}$

an approximate, in some sense, $[\nabla^2 f(x_k)]^{-1}_{x=x_k}$

$H_k > 0$ for all k

$$f = \frac{1}{2} x' Q x + c' x + d$$

$$\nabla f = Qx + c \quad \nabla^2 f = Q$$

$$\nabla f(y) - \nabla f(x) = (Qy + c) - (Qx + c)$$

$$= \underbrace{Q}_{\nabla^2 f}(y - x)$$

$$Q^{-1}[\nabla f(y) - \nabla f(x)] = y - x$$

$$Q^{-1}[\nabla f(x_{k+1}) - \nabla f(x_k)] = (x_{k+1} - x_k)$$

$$\underline{H_{k+1}}[\nabla f(x_{k+1}) \overset{\sigma_k}{-} \nabla f(x_k)] = [x_{k+1} \overset{\delta_k}{-} x_k]$$

$$H_{k+1}\sigma_k = \delta_k$$

n equations in n^2 unknowns
however $H_{k+1} = H_{k+1}'$

The quasi-Newton eqs has several solutions.
Given H_0. we would like an update law for H
like $H_{k+1} = \ldots$

$H_{k+1}\sigma_k = \delta_k$
$H_0 = I$

outer product

$$H_{k+1} = H_k + \frac{\delta_k \delta_k'}{\delta_k' \sigma_k} - \frac{H_k \sigma_k \sigma_k' H_k}{\sigma_k' H_k \sigma_k}$$

inner product

H_0 is symmetric
if H_k is symmetric. then H_{k+1} is also symmetric.

If $H_k > 0$, it is always possible to select α_k such that $H_{k+1} > 0$

$$H_{k+1}\sigma_k = H_k\sigma_k + \frac{\cancel{\delta_k\delta_k'\sigma_k}}{\cancel{\delta_k'\sigma_k}} - \frac{\cancel{H_k\sigma_k\sigma_k' H_k\sigma_k}}{\cancel{\sigma_k' H_k\sigma_k}}$$

$$= H_k\sigma_k + \delta_k - H_k\sigma_k = \delta_k$$

Quasi-Newton (x_0, H_0 are given) $H_0 = H_0' > 0$

$$x_{k+1} = x_k - \alpha_k H_k \nabla f(x_k)$$

$$H_{k+1} = H_k + \frac{\delta_k \delta_k'}{\delta_k'\sigma_k} - \frac{H_k\sigma_k\sigma_k' H_k}{\sigma_k' H_k\sigma_k}$$

If $f = \frac{1}{2}x'Qx + c'x + d$ $\quad Q = Q' > 0$

then $\{x_k\} \to x_* = -Q^{-1}c$ in at most n steps

$\{H_k\} \to Q^{-1}$ in at most n steps

(provided $\alpha_k = \alpha^*$).

For non-quadratic functions under some assumptions. the line-search is suff-accurate f have a global convergence to $x_0 \in \Omega$ with quadratic/superlinear speed.

Moreover if $\nabla^2 f(x_k) > 0$ then $\{H_k\} \to [\nabla^2 f(x_k)]^{-1}$

Methods without derivatives:

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

α_k is selected using parabolic line search.

d_k: $d_0 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, $d_1 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}$, $d_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}$ ... $d_{n-1} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$

Coordinate directions method

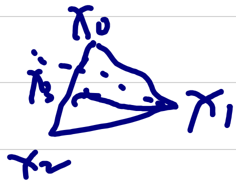
If $L(f(x_0))$ is compact then $\{x_k\}$ is such that $\lim_{k\to\infty}\{x_k\} \in \Omega$

$\{x_k\}$ has a limit if in addition $\lim_{k\to\infty} \|x_{k+1} - x_k\| = 0$

The stopping condition relies on the line search method.

$\min_x f(x) \quad x \in \mathbb{R}^n$

guess/pick 3 points



$f(x_1) > f(x_0) > f(x_2)$

$\{x_0, \cancel{x_1}, x_2\}$

$\downarrow$ projection

$\{x_0, x_2, x_3\}$ $\quad f(x_3) < f(x_1)$?

$\downarrow$

$\{x_2, x_3, x_4\}$

Simplex method: n points in n-dimensional space.
discard the worst.
can cycle between points.

Given P_1 $\min_x f(x)$ $\iff$ $\min_x f(x)$

$g(x)=0$ $\quad$ $g(x)\le 0$

$-g(x)\le 0$

n. decision variables $\iff$ n decision variables

m equation constraints $\quad$ 2m inequation constraints

Given P_2 $\min_x f(x)$ $\iff$ $\min_x f(x)$

$h(x)\le 0$

$$\begin{cases} h_1(x)+y_1^2=0 \\ \vdots \\ h_p(x)+y_p^2=0 \end{cases} \qquad h(x)+Y^2=0$$

n. decision variables $\iff$ n+p decision variables

p inequation constraints $\quad$ p decision variables

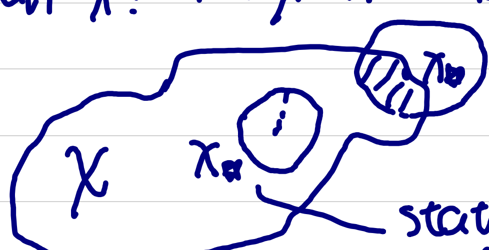
$\min_x f(x)$

$x\in X$

A point x_* is a (strict) local minimiser for the nonlinear programming problem if there exists θ ($\theta>0$) such that

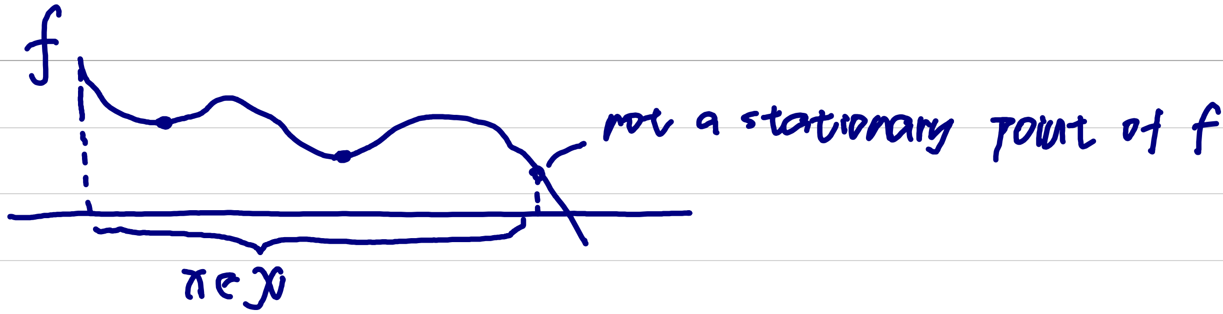
$$f(x_*)\le f(x) \quad (f(x_*)\le f(x))$$

for all x: $x\in X\cap\{\|x-x_*\|<\theta\}$ $\quad$ $(x\ne x_*)$



stationary point

constraints not tight



Problem P_0 (canonical form)

$$\min_x f(x)$$

$$g(x) = 0$$

$$h(x) \le 0 \qquad h(x) = \begin{bmatrix} h_1(x) \\ \vdots \\ h_p(x) \end{bmatrix}$$

The i-th inequality constraint is active at a point $\hat{x}$ if $h_i(\hat{x}) = 0$. If $h_i(\hat{x}) < 0$ then the constraint is not active.

All equality constraints are active at any admissible point $\hat{x}$.

Given P_0 and $\hat{x}$ admissible we define the set $I_a(\hat{x})$ as the set of all active constraints at $\hat{x}$.

$x \in (x_1, x_2)$

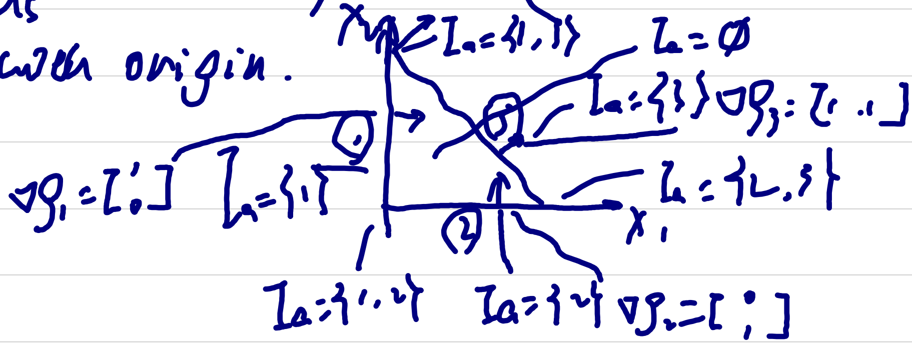
① $x_1 \geq 0$ | $-x_1 \leq 0$

② $x_2 \geq 0$ | $-x_2 \leq 0$

③ $x_1 + x_2 = 1$ | $x_1 + x_2 - 1 \leq 0$

(i) plot equations

(ii) decide side with origin.

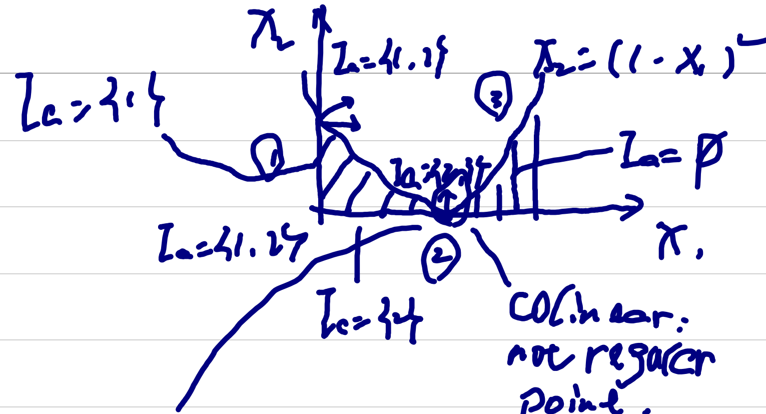


A point $\tilde{x}$ is a regular point for the constraint if the gradients of all active constraint of $\tilde{x}$ are linearly independent.

Given $\tilde{x} \rightarrow \begin{bmatrix} \nabla g_1(\tilde{x}), \nabla g_2(\tilde{x}) \dots \nabla g_m(\tilde{x}) \\ \nabla h_i(\tilde{x}),\ i \in I \end{bmatrix}$

All points in the interior of the admission set are regular points.

$x_1 \geq 0$
$x_2 \geq 0$
$x_2 \leq (1-x_1)^2$



to be discussed one by one.

the admission set around the non-regular points can change drastically with small change of constraints.

$$P_0: \begin{cases} \min f(x) \\ g(x)=0 \quad \times m \\ h(x)\leq 0 \quad \times p \end{cases}$$

Define the Lagrangian of the problem

$$L(x,\lambda,\rho)=f(x)+\underbrace{\lambda' g(x)+\rho' h(x)}_{\text{penalty}}$$

$x\in\mathbb{R}^n,\ \lambda\in\mathbb{R}^m,\ \rho\in\mathbb{R}^p$

First order necessary condition of optimization

Given P_0. Suppose that x_* is a local solution of the problem. Suppose x_* is a regular point. Then there exist (unique) multipliers λ_*, ρ_* such that

$$\nabla_x L(x_*,\lambda_*,\rho_*)=0$$

$$\nabla_\lambda L(x_*,\lambda_*,\rho_*)=g(x_*)=0, \quad h(x_*)\leq 0$$

$$\rho_* \geq 0 \quad \underbrace{(\rho_*)' h(x_*)=0}_{\text{complementary problem}}$$

$\forall i \quad (\rho_*)_i h_i(x_*)=0$

$$(\mu^*)' h(x^*) = \underbrace{(\mu^*)_1}_{\geq 0} \underbrace{h_1(x^*)}_{\leq 0} + \underbrace{(\mu^*)_2}_{\geq 0} \underbrace{h_2(x^*)}_{\leq 0} + \cdots = 0$$

$$\vdots$$

$$\underbrace{(\mu^*)_r h_r(x^*)}_{\leq 0}$$

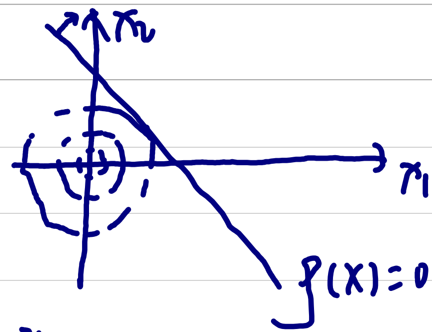
Strict complementary

$$(\mu^*)_i h_i(x^*) = 0 \quad \begin{cases} (\mu^*)_i = 0 \quad h_i(x^*) < 0 \\ (\mu^*)_i > 0 \quad h_i(x^*) = 0 \\ (\mu^*)_i = 0 \quad h_i(x^*) = 0 \end{cases}$$

At x^*, λ^*, μ^* the strict complementary condition hold if $(\mu^*)_i h_i(x^*) = 0$

$$\min \frac{x_1^2 + x_2^2}{2}$$

$$x_1 + x_2 - 1 = 0$$

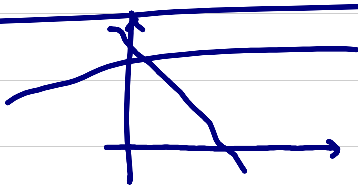


$$L(x, \lambda) = \frac{x_1^2 + x_2^2}{2} + \lambda(x_1 + x_2 - 1)$$

$$0 = \nabla_x L = \begin{bmatrix} x_1 + \lambda \\ x_2 + \lambda \end{bmatrix} \Rightarrow \begin{array}{l} x_1 = -\lambda \\ x_2 = -\lambda \end{array}$$

$$0 = g(x) = [x_1 + x_2 - 1] \quad -2\lambda - 1 = 0 \Rightarrow \lambda = -\frac{1}{2}$$

$x_2 = 1 - x_1$



constraint elimination

constrainted ⇒ unconstrained.

$$\min_{x_1} \frac{x_1^2 + (1 - x_1)^2}{2} \Rightarrow \min_{x_1} \underbrace{\frac{2x_1^2 - 2x_1 + 1}{2}}_{\tilde{f}}$$

$$\nabla \tilde{f} = 4x_1 - 2 = 0 \Rightarrow x_1 = \frac{1}{2}. \quad x_2 = \frac{1}{2}$$

$$\nabla_{xx} L = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

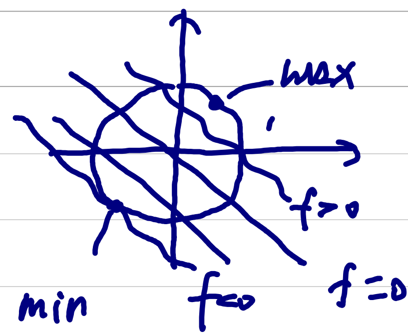
$s' \nabla_{xx} L s > 0$ for $s \neq 0$ and s such that

$$\left[\frac{dg}{dx}\right] s = 0 \quad [1 \ 1] \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = 0 \Rightarrow \begin{bmatrix} \alpha \\ -\alpha \end{bmatrix} \alpha \neq 0.$$

$$[\alpha \ -\alpha] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ -\alpha \end{bmatrix} = 2\alpha^2 > 0$$

$\min x_1 + x_2$

$x_1^2 + x_2^2 - 1 \leq 0$



$$L = (x_1 + x_2) + \mu(x_1^2 + x_2^2 - 1)$$

$$0 = \nabla_x L = \begin{bmatrix} 1 + 2\mu x_1 \\ 1 + 2\mu x_2 \end{bmatrix} \qquad \begin{matrix} \mu \geq 0 \\ x_1^2 + x_2^2 - 1 \leq 0 \end{matrix}$$

complementary condition

$$\mu(x_1^2 + x_2^2 - 1) = 0$$

$\mu = 0 \Rightarrow x_1^2 + x_2^2 - 1 \leq 0$

$\begin{bmatrix} 1+0 \\ 1+0 \end{bmatrix} = 0$ ✗

$\mu > 0 \Rightarrow \begin{bmatrix} 1 + 2\mu x_1 \\ 1 + 2\mu x_2 \end{bmatrix} = 0 \Rightarrow x_1 = x_2 = -\frac{1}{2\mu}$

$x_1^2 + x_2^2 - 1 = 0 \Rightarrow \frac{1}{4\mu^2} \cdot 2 = 1 \Rightarrow \mu_1 = \frac{\sqrt{2}}{2}$

$\mu_2 = -\frac{\sqrt{2}}{2}$ ✗

$x_1 = x_2 = -\frac{\sqrt{2}}{2}$

$$\nabla_{xx} L = \begin{bmatrix} 2\mu & 0 \\ 0 & 2\mu \end{bmatrix} = 2\mu I > 0 \text{ since } \mu > 0$$

$\frac{\partial h}{\partial x} = [2x_1 \;\; 2x_2]\Big|_{x^*} = -\left[\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right]$ $\qquad s' \nabla_{xx} L s = [\alpha \;\; -\alpha]\begin{bmatrix} 2\mu & 0 \\ 0 & 2\mu \end{bmatrix}\begin{bmatrix} \alpha \\ -\alpha \end{bmatrix} = 4\mu\alpha^2 > 0$

$\frac{\partial h}{\partial x} s = -\frac{\sqrt{2}}{2}[1 \;\; 1]\begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = 0 \qquad s = \begin{bmatrix} \alpha \\ -\alpha \end{bmatrix} \; (\alpha \neq 0)$

Sufficient condition

- $L(x, \lambda, \mu) = f(x) + \lambda' g(x) + \mu' h(x)$
- Regulation of constraints
- Strict complementary conditions

Given P_0. Suppose that there exist x_*, λ_*, μ_* such that

- x_* is a regular point
- The strict complementary condition holds
- $0 = \nabla_x L(x_*, \lambda_*, \mu_*)$; $h(x_*) \leq 0$

  $0 = g(x_*)$ $\qquad$ $\mu_* \geq 0$

  $(\mu_*)' h(x_*) = 0$

necessary condition

(+) consider

$$s' \overbrace{\nabla_{xx} L(x_*, \lambda_*, \mu_*)}^{n \times n} s$$

If this is strictly positive for all $s \neq 0$ and

$$\left[\begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial h_*}{\partial x} \end{bmatrix} (x_0) \right] s = 0$$

Then x_* is a local strict optimiser.

Define

$$\phi(x)=\begin{cases}0. & \text{if } x\in\mathcal{X}\\ +\infty. & \text{if } x\notin\mathcal{X}\end{cases}$$

Construct the function

$$F=f+\phi$$

$$\min_{x\in\mathcal{X}} f(x) \Rightarrow \min_{x\in\mathbb{R}} F(x)$$

Approximation of $\phi(x)$

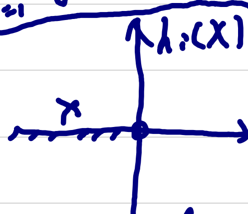
$$\phi(x)=\begin{cases}0, & \text{if } x\in\mathcal{X}\\ >, & \text{if } x\notin\mathcal{X}\end{cases}$$

penalty positive/negative

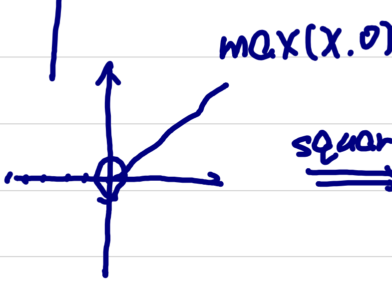
add differentiability

$$\boxed{\phi(x)=\sum_{i=1}^{m}[g_i(x)]^2+\sum_{j=1}^{p}\max(0,h_i(x))^2}$$

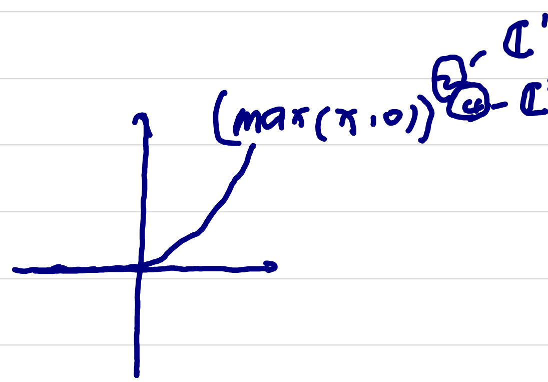
$h_i(x)\le 0$



$\max(h_i(x),0)$



square $\Longrightarrow$



$F_\varepsilon(x) = f(x) + \frac{1}{\varepsilon} P(x)$

$\frac{1}{\varepsilon} P(x) \xrightarrow{\varepsilon \to 0} \phi(x)$

$\frac{1}{\varepsilon} P(x)$ is a good approximation of $\phi(x)$ for $\varepsilon > 0$. sufficiently small.

$P_0 \begin{cases} \min_x f(x) \\ g(x) = 0 \\ h(x) \le 0 \end{cases} \Rightarrow \min_{x \in \mathbb{R}^n} F_\varepsilon(x),\ \varepsilon > 0$

Fix ε_0. solve $\min_{x \in \mathbb{R}^n} F_{\varepsilon_0}(x) \Rightarrow x_0^* \notin \mathcal{X}$

initial guess

Fix $\varepsilon_1 = \frac{\varepsilon_0}{2}$. solve $\min_{x \in \mathbb{R}^n} F_{\varepsilon_1}(x) \Rightarrow x_1^* \notin \mathcal{X}$

$\vdots$

$\varepsilon_{k+1} = \frac{\varepsilon_k}{2}$. solve $\min_{x \in \mathbb{R}^n} F_{\varepsilon_{k+1}}(x) \Rightarrow x_{k+1}^* \notin \mathcal{X}$

We would like to understand what happens as $k \to +\infty$

$\{\varepsilon_k\} \to 0$ $\qquad$ $\underset{\notin \mathcal{X}}{\{x_k^*\}} \to x^* \in \mathcal{X}$

local min of P_0

---

1) $\{\varepsilon_k\} \to 0$ $\quad \frac{1}{\varepsilon_k} P(x) \to \infty$ (?) numerical hard to deal with

2) slow convergence of $\{x_k^*\} \to x^*$.

Solution: $\varepsilon_k \not\to 0$

P_1: $\begin{cases} \min f(x) \\ g(x) = 0 \end{cases} \Rightarrow L(x,\lambda) = f(x) + \lambda^T g(x)$

$0 = \nabla_x L = \nabla f + \frac{\partial g}{\partial x}\lambda$ | necessary condition of optimality

$0 = g(x) = \nabla_\lambda L$

$\min F_\varepsilon$

$F_\varepsilon = f(x) + \frac{1}{\varepsilon}\|f(x)\|^2$

$0 = \nabla_x F_\varepsilon = \nabla f + \frac{2}{\varepsilon}\frac{\partial f}{\partial x} g(x)$

$\tilde{\lambda}^* = \frac{2}{\varepsilon} f(x^*)$

$\min_x F_\varepsilon(x)$

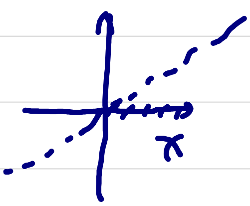
$\tilde{x}^*$ approx of x^*

$\tilde{\lambda}^*$ approx of λ.

$\min_x x$

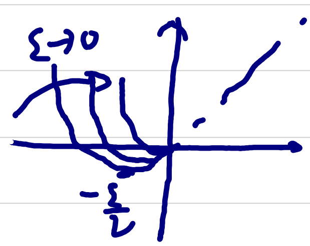
$-x \le 0$

$x^* = 0$



$F_\varepsilon(x) = x + \frac{1}{\varepsilon}\max(0,-x)^2$

$F_\varepsilon(x) = \begin{cases} x, & x \ge 0 \ (-x \le 0) \\ x + \frac{x^2}{\varepsilon}, & x \le 0 \end{cases}$



$0 = \nabla_x F_\varepsilon = \begin{cases} 1, & x \ge 0 \\ 1 + \frac{2x}{\varepsilon}, & x \le 0 \end{cases}$ ∇F_ε is continuous.

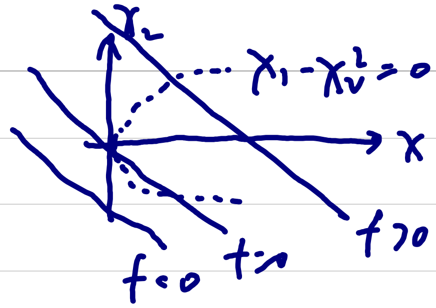
$x_\varepsilon^* = -\frac{\varepsilon}{2} \notin \mathcal{X}$ $\lim_{\varepsilon \to 0} x_\varepsilon^* = 0 = x^* \in \mathcal{X}$

$$\nabla^2 f_\varepsilon = \begin{cases} 0, & \text{if } x \geq 0 \\ \frac{2}{\varepsilon}, & \text{if } x \leq 0 \end{cases}$$

$\nabla^2 f_\varepsilon$ is not continuous.

$\min \; x_1 + x_2$

$x_1 - x_2^2 = 0$



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

$0 = \nabla_x L = \begin{bmatrix} 1+\lambda \\ 1 - 2\lambda x_2 \end{bmatrix}$ $\quad \lambda = -1$

$x_2 = -\frac{1}{2}$

$0 = x_1 - x_2^2$ $\quad x_1 = \frac{1}{4}$

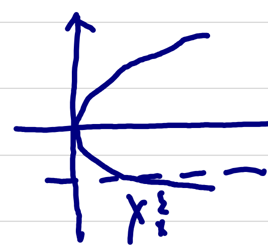
$F_\varepsilon = f(x) + \frac{1}{\varepsilon} g^2(x)$

$= x_1 + x_2 + \frac{1}{\varepsilon}(x_1 - x_2^2)^2$

$0 = \nabla_x F_\varepsilon = \begin{bmatrix} 1 + \frac{2}{\varepsilon}(x_1 - x_2^2) \\ 1 - \frac{4}{\varepsilon}(x_1 - x_2^2)x_2 \end{bmatrix}$ $\quad = -1$

$\Downarrow$

$1 + 2x_2 = 0$ $\quad \begin{cases} x_2 = -\frac{1}{2} \\ x_1 = \frac{1}{4} - \frac{\varepsilon}{2} \end{cases} \rightarrow \lim_{\varepsilon \to 0} x_\varepsilon$



Given P,

$\left.\begin{matrix} \min_x f(x) \\ g(x) = 0 \end{matrix}\right] \rightarrow L(x, \lambda) = f(x) + \lambda^T g(x)$

$0 = \nabla_x L(x, \lambda)$

$0 = g(x)$

if x_* is a candidate optimal point with multiplier λ_0 then

$\nabla_x L(x_0, \lambda_0) = 0$

$g(x_0) = 0$

Suppose λ_0 is given. Then the condition

$$\nabla_x L(x_0, \lambda_0) = 0$$

implies that $L(x, \lambda_0)$ has a stationary point at x_0

$\min_x L(x, \lambda_0) \to x_0 \in X$

$\to x_0$ may not be a min for f_1.

Sequential augmented Lagrangian

$$L_\epsilon(x, \lambda_0) = L(x, \lambda_0) + \frac{1}{\epsilon} \| g(x) \|^2$$

For ϵ small, but strictly positive, there is a one-to-one relation between $\substack{\max \\ \min}$ of P_1 and $\substack{\max \\ \min}$ of P_1

Step 0) $x_0, \lambda_1, \epsilon_1 > 0$

Step 1) Set $k = 1$

Step 2) construct $L_\epsilon(x, \lambda_k)$ and minimise the function with respect to $x \to x_k^*$

Step 3) update $\lambda_k \to \lambda_{k+1}$

Step 4) $\epsilon_{k+1} = \beta \epsilon_k$, $\beta = \begin{cases} 1, & \| g(x_{k+1}) \| \le \frac{1}{4} \| g(x_k) \| \\ < 1 & \end{cases}$

Step 5) $k \to k+1$, go to 2)

$\{x_k^*\}$. $\{\lambda_k\}$. $\{\varepsilon_k\}$

$\downarrow \quad \downarrow \quad \downarrow$

$x^* \in \mathcal{D} \quad \lambda^* \quad \varepsilon^* > 0$

min of P_1

$$\lambda_{k+1} = \lambda_k + \frac{2}{\varepsilon_k} g(x_k)$$

Exact penalty functions

P_1: $\min f(x)$

$g(x) = 0$

$$L(x,\lambda) = f(x) + \lambda' g(x)$$

Necessary condition $\Rightarrow$ $0 = \nabla_x L(x,\lambda) \rightarrow$??

$0 = g(x)$

$$L_Q(x,\lambda) = f(x) + \lambda' g(x) + \frac{1}{\varepsilon} \| g(x) \|^2$$

$$0 = \nabla_x L(x,\lambda)$$

$$= \nabla f + \left(\frac{\partial g}{\partial x}\right)' \lambda$$

tall matrix $n \times m$

if all points are regular points then $\frac{\partial g}{\partial x}$ is full rank.

$$0 = \nabla f + \left(\frac{\partial g}{\partial x}\right)^T \lambda \qquad \times \left(\frac{\partial g}{\partial x}\right) \text{ to the left}$$

$$0 = \frac{\partial g}{\partial x} \nabla f + \underbrace{\left(\frac{\partial g}{\partial x}\right)\left(\frac{\partial g}{\partial x}\right)^T}_{m \times m} \lambda$$

$$\lambda = \lambda(x) = -\left[\left(\frac{\partial g}{\partial x}\right)\left(\frac{\partial g}{\partial x}\right)^T\right]^{-1} \underbrace{\frac{\partial g}{\partial x}}_{m\times n} \underbrace{\nabla f}_{n\times 1} \quad m\times 1$$

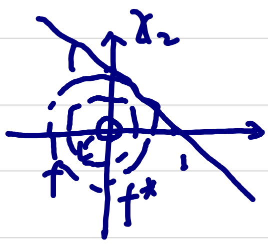
Define the exact penalty function

$$\Sigma_\varepsilon(x) = f(x) + \lambda^T(x)\, g(x) + \frac{1}{\varepsilon}\|g(x)\|^2$$

For $\varepsilon > 0$. and Sufficiently small
there is a one-to-one relationship
between $\underbrace{\text{min of } P_1}_{\text{no } \varepsilon}$ and min of $\Sigma_\varepsilon(x)$

the min of $\Sigma_\varepsilon(x)$ are not function of ε.

$$\min \frac{x_1^2 + x_2^2}{2}$$
$$x_1 + x_2 - 1 = 0$$



$$L = \frac{x_1^2 + x_2^2}{2} + \lambda(x_1 - x_2 - 1)$$

$$0 = \nabla_x L = \begin{bmatrix} x_1 + \lambda \\ x_2 + \lambda \end{bmatrix} = 0 \qquad \lambda_1 = \lambda_2 = -\lambda$$

$$x_1^* = x_2^* = \frac{1}{2}$$

$$0 = x_1 + x_2 - 1 = 0 \qquad -2\lambda - 1 = 0 \Rightarrow \lambda = -\frac{1}{2}$$

$$\Sigma(x) = f(x) + \lambda(x)(x_1 + x_2 - 1) + \frac{1}{2}(x_1 + x_2 - 1)^2$$

$$0 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \lambda \qquad \left(\frac{\partial g}{\partial x}\right)^T$$

$$0 = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \lambda$$

$$0 = x_1 + x_2 + 2\lambda \qquad \lambda(x) = -\frac{1}{2}(x_1 + x_2)$$

$$\xi(x) = \frac{x_1^2 + x_2^2}{2} - \frac{1}{2}(x_1 + x_2)(x_1 + x_2 - 1) + \frac{1}{\xi}(x_1 + x_2 - 1)^2$$

$$\xi(x) = \frac{x_1^2 + x_2^2}{2} - \frac{1}{2}(x_1 + x_2 - 1)^2 - \frac{1}{2}(x_1 + x_2 - 1) + \frac{1}{\xi}(x_1 + x_2 - 1)^2$$

$$\xi(x) = \frac{x_1^2 + x_2^2}{2} - \frac{1}{2}(x_1 + x_2 - 1) + \left(\frac{1}{\xi} - \frac{1}{2}\right)(x_1 + x_2 - 1)^2$$

$$0 = \nabla \xi = \begin{bmatrix} \overset{0}{\overset{\parallel}{x_1 - \frac{1}{2}}} + 2\left(\frac{1}{\xi} - \frac{1}{2}\right)(x_1 + x_2 - 1) \\ \underset{0}{\underset{\parallel}{x_2 - \frac{1}{2}}} + 2\left(\frac{1}{\xi} - \frac{1}{2}\right)(x_1 + x_2 - 1) \end{bmatrix}$$

$$x_1 = x_2 = \frac{1}{2}$$

$$\nabla^2 \xi = \begin{bmatrix} \frac{2}{\xi} - 1 + 1 & \frac{2}{\xi} - 1 \\ \frac{2}{\xi} - 1 & \frac{2}{\xi} - 1 + 1 \end{bmatrix} = I + \overset{>0}{\left(\frac{2}{\xi} - 1\right)} \overset{\geq 0}{\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}} \quad \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix}$$

$\frac{2}{\xi} - 1 > 0$ $\quad$ $\xi < \frac{1}{2}$ then $\nabla^2 \xi > 0$

ξ: shape the Hessian to ensure the stationary point correspond to the optimal solution (regular point).