5. Theory of constrained Optimization
5.1. Implicit function theorem.

$h: \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^n$.

Assume:

i) $h(x^*, y^*) = 0$

ii) ~~$\nabla_x h(x,y)$~~ $h \in C^1$ in $\mathcal{N}(x^*, y^*)$

iii) $\nabla_x h(x,y)$ is invertible at x^*

Then $\exists\ \varphi: \mathcal{N}(y^*) \to \mathbb{R}^k$ such that

i) $\varphi(y^*) = x^*$

ii) $h(\varphi(y), y) = 0 \ ; \forall\ y \in \mathcal{N}(y^*)$

iii) $\nabla_y \varphi(y) =$

$$\left[\nabla_x h \cdot \nabla_y \varphi + \nabla_y h = 0 \Rightarrow \nabla_y \varphi = -(\nabla_x h)^{-1} \cdot \nabla_y h \right]$$

Corollary: $h \in C^r \Rightarrow \varphi \in C^r$

Corollary: $h(x) = 0$; rank $h: \mathbb{R}^{k+m} \to \mathbb{R}^m$;
$\nabla_x h$ full rank; $\exists\ x = (x_1, x_2)$; $h(\varphi(x_2), x_2) = 0$

Miami University Notes

5.1.2 Corollary

$h: \mathbb{R}^n \to \mathbb{R}^m \quad (n > m) \; ; \; h \in C^1$

$h(x^*) = 0$; $\nabla_x h(x)\big|_{x^*}$ is full rank (m)

$\Rightarrow \exists \; x = (x_1, x_2)$ such that $x_1 = \varphi(x_2)$

$\nabla_{x_2} \varphi = -(\nabla_{x_1} h)^{-1} \nabla_{x_2} h$

Mihai Anitescu Notes

5. CONSTRAINED OPTIMIZATION
5.2 Equality Constraints

- Problem:

$$\min_x f(x) \qquad f: \mathbb{R}^n \to \mathbb{R}$$
$$h(x) = 0 \qquad h: \mathbb{R}^m \to \mathbb{R}^m \; ; \; m < n$$

- Assume: Let x^* be the solution.
  $\nabla_x h|_{x^*}$ has full row rank

- Implication: by Implicit Function theorem.
  $x = (y^*, z^*)$ (or by a permutation)
  such that $y^* = \varphi(z^*)$ in $\mathcal{N}(y^*, z^*) \cap \mathcal{F}$

- Problem is equivalent to:

$$\min_z \psi(z) = \min_z f(\varphi(z), z) \; ; \text{ with solution } z^*$$

- First-order optimality conditions

$\boxtimes$

Minai Anitesou Notes

5.2.2 $\nabla_z \psi(z) = 0$ This section assume row gradients

$$\Rightarrow \nabla_y f \cdot \nabla_z \varphi + \nabla_z f = 0$$

$$\Rightarrow \underbrace{\nabla_y f \cdot (-\nabla_y h)^{-1}} \cdot \nabla_z h + \nabla_z f = 0$$

Define $\lambda^{*T} = \nabla_y f \cdot (\nabla_y h)^{-1}$

$$\Rightarrow -\lambda^{*T} \nabla_z h + \nabla_z f = 0$$

expand

$$\Rightarrow -\lambda^{*T} \nabla_y h + \nabla_y f = 0$$

$$\Rightarrow \boxed{\nabla_x f - \lambda^{*T} \nabla_x h = 0}$$

Define $\mathcal{L}(x, \lambda) \Rightarrow \exists \lambda^*$ such that

$$\nabla_x \mathcal{L}(x, \lambda^*) \Big|_{x = x^*} = 0$$

Mihai Anitescu Notes

5.2.3 We knew this: Lagrange's Theorem.

But what about SOC?

Define $x^d(t) = \begin{pmatrix} \varphi(td) \\ td \end{pmatrix} = \begin{pmatrix} y \\ z \end{pmatrix}$

$\Rightarrow h(x^d(t)) = 0.$

Choose $x^d(0) = 0$

$\Rightarrow f(x^d(t)) \geq f(x^d(0)) \ \forall t$

WRITE S.O. optimality conditions

$$\frac{d^2}{dt^2} f(x^d(t))\Big|_{t=0} \geq 0$$

$$\|$$

$$\dot{x}^d(0)\, \nabla^2_{xx} f\, \dot{x}^d(0) + \nabla_x f \cdot \ddot{x}^d(0)$$

Mihai Anitescu Notes

5.2.4

$\cancel{\nabla_x f \ddot{x}^d(0) =}$

But $\ddot{x}^d(0) = \begin{pmatrix} \ddot{y}^d(0) \\ 0 \end{pmatrix}$

$\Rightarrow \nabla_x f \ddot{x}_d(0) = \nabla_y f \cdot \ddot{y}^d(0) =$

$\underbrace{\nabla_y f \cdot \nabla_y^{-1} h}_{+\lambda^{*T}} \cdot \nabla_y h \cdot \ddot{y}^d(0) = +\lambda^{*T} \cdot \nabla_y h \, \ddot{y}^d(0)$

$= +\lambda^{*T} \cdot \nabla_x h \cdot \ddot{x}^d(0)$

But $h(x^d(t)) = 0 \Rightarrow \cancel{\frac{d^2}{dt^2} h(x^d(t)) = 0}$

$\lambda^{*T} h(x^d(t)) = 0 \Rightarrow \frac{d^2}{dt^2} \lambda^{*T} h(x^d(t)) = 0$

$\Rightarrow \lambda^{*T} \cdot \nabla_x h \cdot \ddot{x}^d + \dot{x}^d \nabla_{xx}^2 (\lambda^{*T} h) \dot{x}^d = 0$

$\Rightarrow \nabla_x f \ddot{x}_d(0) = -\dot{x}^d \nabla_{xx}^2 (\lambda^{*T} h) \dot{x}^d$

Mihai Anitescu Notes

5.2.5 Which implies

$$\dot{x}_{(0)}^{d}\left(\nabla_{xx}^{2}\left(f+\lambda^{kT}h\right)\right)\dot{x}_{(0)}^{d}\geq 0$$

$$\Rightarrow \begin{bmatrix} J^{d} \\ d \end{bmatrix} \nabla_{xx}^{2}\mathcal{L}(x,\lambda^{k})\begin{bmatrix} J^{d} \\ d \end{bmatrix} \geq 0$$

Where $J = \nabla_{z}\rho\Big|_{z=z^{k}}$

$$\Leftarrow \boxed{\begin{array}{l} u^{T}\nabla_{xx}^{2}\mathcal{L}(x,\lambda^{k})\,u \geq 0 \quad \forall\, u \\ \nabla_{x}h\cdot u = 0 \end{array}}$$

$$\boxed{\nabla_{x}\mathcal{L}(x,\lambda^{k})=0} \quad KT$$

h has full rank $\Rightarrow$ ↑

5. Constrained Optimization

5.2.6

Sufficient Conditions: $\nabla_x h$ full rank P

$$\nabla_x \mathcal{L}(x, \lambda^*) = 0$$

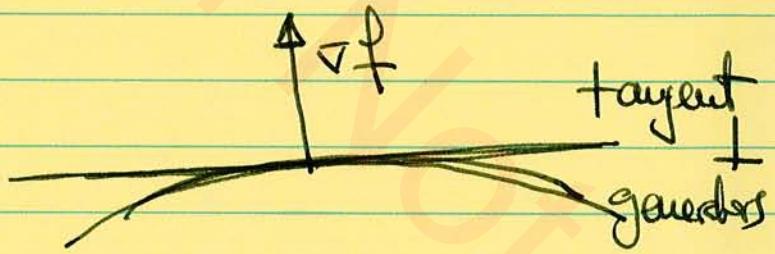
$$u^T \nabla_x^2 \mathcal{L}(x, \lambda^*) u \geq 0 \cdot \|u\|^2$$

$$\forall u \text{ s.t. } \nabla_x h \, u = 0$$

$\Downarrow$

x^* local min ; SOSC

GEOMETRY



Miguel Anitescu Notes

5 Constrained Optimization
5.3 Augmented Lagrangian

$u^T M u \geq \sigma \|u\|^2$; $\forall u$ s.t $Au = 0$

A full rank

$\Rightarrow \exists c$ such that

$M + c A^T A$ is P.D.

"Augmented Lagrangian"

① $\Rightarrow f(x) = \mathcal{L}(x, \lambda) + c \cdot \|h(x)\|^2 = \mathcal{L}_c(x, \lambda)$

Satisfies $\nabla_x \mathcal{L}_c(x, \lambda^*) = 0$; $\nabla_{xx}^2 \mathcal{L}_c(x, \lambda^*) > 0$.

$\Rightarrow x$ is an unconstrained minof $\mathcal{L}_c(x, \lambda^*)$!

② $\begin{pmatrix} M & A^T \\ A & 0\,(-c) \end{pmatrix}$ is not singular $\Rightarrow$ system is locally Newton!

③ Can obtain that these are sufficient by using augmented Lagrangian approach

5 Constrained Optimization
5.4 Optimality Conditions for Problems w. Inequality Constraints

Problem

$$\min_x f(x)$$
$$\text{s.t } c_i(x) = 0 \quad i \in \mathcal{E}$$
$$c_i(x) \geq 0 \quad i \in \mathcal{I}$$

Define $\mathcal{A}(x) = \{ i \in \mathcal{E} \cup \mathcal{I} \mid c^i(x) = 0 \}$

Assume $\nabla_x c_{\mathcal{A}}$ is full row rank at x^*

Problem 2

$$\min_x f(x)$$

$$h(x) = \begin{cases} c_i(x) = 0 & i \in \mathcal{E} \\ c_i(x) - z_i^2 = 0 & i \in \mathcal{I} \end{cases}$$

has $\nabla_{x,z} h(x,z) = \begin{bmatrix} \nabla_x c_{\mathcal{A}} & 0 \\ \nabla_x c_{\mathcal{A}} & 2I \end{bmatrix}$ AND FULL RANK

5.4.2 Apply ~~unconstrained~~ equality constrained optimality conditions

$\Rightarrow \exists\ \lambda^* \in \mathbb{R}^{n_E + n_I}$ ↓ "complementarity condition"

$\nabla_x \mathcal{L}(x, \lambda^*) = 0 \ ; \ c_i(x) \geq 0 \Rightarrow \lambda_i = 0 \ ; \ i \in I$

$c_i(x) = 0 \Rightarrow \lambda_i \geq 0 \ ; \ i \in I$

$u^T \nabla_{xx}^2 \mathcal{L}(x, \lambda^*) u \geq 0 \ ; \ \forall u \ ; \ \nabla_x c_A \cdot u = 0$

Karush (KKT)
Kuhn-Tucker conditions

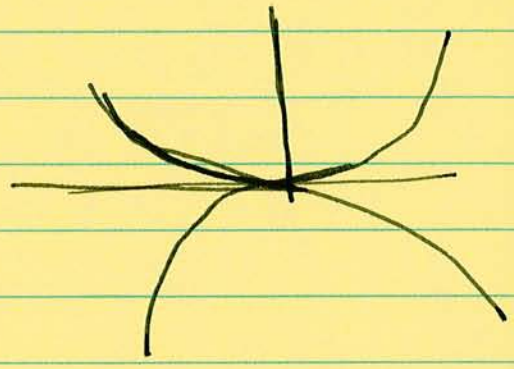
$\nabla c_A(x^*)$ is full rank: constraint qualification

SOSC: $\nabla c_A(x^*)$: full rank ;

$u^T \nabla_{xx}^2 \mathcal{L}(x, \lambda^*) u > \|u\|^2 \ ; \ \forall u \ \nabla_x c_A u = 0$

$c_i(x) = 0 \ ; \ i \in I \Rightarrow \lambda_i > 0$: STRICT COMPLEMENTARITY

5.4.3 Does KKT hold W.O constraint qualification?



$\max_{x,y} x$

$-y \leq 0$

~~$x^3 - y \leq 0$~~

$y \leq x^3$

$$\nabla f = \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \lambda_1 \begin{pmatrix} 0 \\ -1 \end{pmatrix} - \lambda_2 \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0$$ impossible

~~Geometr~~

Why is KKT so important $\Rightarrow$ locally is a nonlinear equation $\begin{cases} \nabla L(x,\lambda) = 0 \\ h = 0 \end{cases}$ $\Rightarrow$ Newton

5.4.4. Why is KKT so important ?

Optimization looks locally like a system of nonlinear equations

$$\begin{cases} \nabla_x \mathcal{L}(x, \lambda) = 0 \\ \nabla_{c_A}(x) = 0 \end{cases} \xrightarrow{J} \begin{bmatrix} \nabla^2_{xx}\mathcal{L} & \nabla_{c_A}^T \\ \nabla_{c_A} & 0 \end{bmatrix}$$

$\Rightarrow$ nonsingular

$\Rightarrow$ Newton ?

Mihai Anitescu Notes

5.5.1 Types of Methods

(1) SQP:

$$\min_{x^k} \begin{cases} \frac{1}{2} p^T \nabla_{xx}^2 \mathcal{L}(x^k, \lambda^k) p + \nabla f_k^T p \\ \text{s.t. } \nabla c_i(x) p = 0 \quad i \in \mathcal{E} \\ \nabla_x c_i(x) p \geq 0 \quad i \in \mathcal{I} \end{cases}$$

[use with TR ; or ; line search for merit function

After ones step $x^{k+1} = x^k + \alpha^k p^k$

$\lambda^{k+1} = (\dots)$

(2) Aug Lag

$$\min_x \mathcal{L}_A \quad \text{s.t. } x \geq 0 \quad \Big\} \Rightarrow \text{update } \lambda, \mu$$

$\Downarrow$

μ does not go to ∞ !!

5.5.2

(3) Interior-point methods (barrier methods)

$$\min_{x,s} f(x) - \mu \sum_{i}^{m} \log(s_i)$$

$$c_i(x) = 0 \; ; \; i \in \mathcal{E}$$

$$c_i(x) = s_i \; ; \; i \in \mathcal{I}$$

Latter is considered best w. Hessian

Mihai Anitescu Notes