The Lagrangian and its role in the necessary conditions for optimality

The Lagrangian, namely:

$$L(x,\lambda) \triangleq V(x) + \sum_{i=1}^{m} \lambda_i h_i(x) \quad , \forall x$$

Lagrange multipliers

$$= V(x) + \lambda^T h(x) \quad , \forall x$$

Lagrange multiplier vector

some use "$-$" here instead of a "$+$"; so $\lambda_{theirs} = -\lambda_{mine}$

plays an important role in constrained opt.

Aside: notation (L is a f^n of x & λ):

$$\nabla_x L \triangleq \left(\frac{\partial L}{\partial x_1}, \dots, \frac{\partial L}{\partial x_n} \right)^T \triangleq L_x^T$$

$$\nabla_\lambda L \triangleq \left(\frac{\partial L}{\partial \lambda_1} \dots \frac{\partial L}{\partial \lambda_n} \right)^T = L_\lambda^T$$

Using L Th. (12.1) can be rewritten as:

(13.0) **Th** $\hat{x}$ is a local minimizer for (E)

Then $\nabla_x L(\hat{x},\lambda) = 0$ for some $\lambda \in \mathbb{R}^m$

i.e. $L(\hat{x},\lambda)_x = 0$ for some $\lambda \in \mathbb{R}^m$

neat notation

Proof: $L(x,\lambda) = V(x) + \sum_{i=1}^{m} \lambda_i h_i(x)$

$\Rightarrow L(\hat{x},\lambda)_x = V(\hat{x})_x + \sum_{i=1}^{m} \lambda_i h_i(\hat{x})_x$

$\Rightarrow L(\hat{x},\lambda)_x^T = V(\hat{x})_x^T + \sum_{i=1}^{m} \lambda_i h_i(\hat{x})_x^T$

$\Rightarrow \nabla_x L(\hat{x},\lambda) = \nabla_x V(\hat{x}) + \sum_{i=1}^{m} \lambda_i \nabla_x h_i(\hat{x})$

$\underset{\text{Th. (12.1)}}{=} 0$ for some $\lambda \in \mathbb{R}^m$ if $\hat{x}$ = a local min. for (E)

QED

(13.1) **Th** (A property of L)

$$V(x) = L(x,\lambda) \ , \ \forall \lambda \in \mathbb{R}^m$$

if x is feasible ($x \in F$)

Proof: $L(x,\lambda) = V(x) + \underbrace{\lambda^T h(x)}_{=0} = V(x) \ , \ \forall \lambda$

if x is feasible

QED

This property useful in proving the following sufficient condition for optimality of $\hat{x}$ for (E):

(13.2) **Th** $\hat{x}$ is a local minimizer for (E) if:

(i) $h(\hat{x}) = 0$

(ii) $L(\hat{x},\hat{\lambda})_x = 0$ for some $\hat{\lambda} \in \mathbb{R}^m$

and either:

(iii) $L(\hat{x},\hat{\lambda})_{xx} > 0$

or

(iv) $H^T L(\hat{x},\hat{\lambda})_{xx} H > 0$

less restrictive than (iii)

where H has linearly indep. columns

and $\mathcal{R}[H] = \mathcal{N}[h(\hat{x})_x]$

Notes on Th. (13.2):

- linear independence of the $\nabla h_i(\hat{x})$ is not needed

- V, h_i, $\forall i$, are assumed to be C^2 functions

  meaning all $\dfrac{\partial^2}{\partial x_i \partial x_j}$ exist and are continuous on $\mathbb{R}^n$, so a 2-nd order expansion is valid in the proof

- $L(\hat{x},\hat{\lambda})_x = V(\hat{x})_x + \sum_{i=1}^m \hat{\lambda}_i h_i(\hat{x})_x$

  $= \left[\nabla V(\hat{x}) + \sum_{i=1}^m \hat{\lambda}_i \nabla h_i(\hat{x})\right]^T$

- $L(\hat{x},\hat{\lambda})_{xx} = \left[V(\hat{x}) + \sum_{i=1}^m \hat{\lambda}_i h_i(\hat{x})\right]_{xx}$

  $= V_{xx}(\hat{x}) + \sum_{i=1}^m \hat{\lambda}_i \underbrace{h_i(\hat{x})_{xx}}_{\equiv\ \boxed{h_{i_{xx}}(\hat{x})}}$

How can one find H ?

- one way to find a suitable H is to apply to $h(\hat{x})_x$ our orthogonal decomposition alg. giving:

$$h(\hat{x})_x = \underset{\perp}{P} \begin{bmatrix} \hat{A}_{r \times r} & 0 \\ 0 & 0 \end{bmatrix} \underset{\perp}{Q^T}$$

from which : $H = [\underbrace{q_{*(r+1)} \cdots q_{*n}}_{\boxed{\text{orthonormal and hence linearly independent}}}]$

and $\mathcal{R}[H] = \mathcal{N}[h(\hat{x})_x]$

<u>Proof of Th (13.2) :</u>

$\boxed{\text{case } L(\hat{x},\hat{\lambda})_{xx} > 0}$

For small δx such that $h(\hat{x}+\delta x) = 0$:

$$V(\hat{x}+\delta x) \underset{\boxed{\text{Th } 13.1}}{=} L(\hat{x}+\delta x, \hat{\lambda})$$

$$\triangleq \underbrace{L(\hat{x},\hat{\lambda})}_{= V(\hat{x})} + \underbrace{L(\hat{x},\hat{\lambda})_x}_{= 0 \;\; \boxed{(ii)}} \delta x + \frac{1}{2}\delta x^T L(\hat{x},\hat{\lambda})_{xx} \delta x$$

$\boxed{(i) \wedge \text{Th } (13.1)}$

$$= V(\hat{x}) + \underbrace{\frac{1}{2}\delta x^T \underbrace{L(\hat{x},\hat{\lambda})_{xx}}_{>0} \delta x}_{>0 \;,\; \forall \delta x \neq 0}$$

$$> V(\hat{x}) \quad , \forall \; \delta x \neq 0 \text{ s.t. } h(\hat{x}+\delta x) = 0$$

$\Rightarrow$ $\hat{x}$ = a local minimizer for (E)

$$\boxed{\text{Case } \; L(\hat{x},\hat{\lambda})_{xx} \not> 0 \;\text{ but with }\; H^T L(\hat{x},\hat{\lambda})_{xx} H > 0}$$

Since $\hat{x}$ = feasible

$\Rightarrow$ $\forall$ sufficiently small $\delta x \neq 0$ such that $h(\hat{x}+\delta x) = 0$;

$$0 = h(\hat{x}+\delta x) \cong \underset{\substack{\| \\ 0}}{h(\hat{x})} + h_x(\hat{x})\delta x$$

$\Rightarrow$ so, for all such $\delta x \neq 0$: $h_x(\hat{x})\delta x = 0$

$\Rightarrow$ $\delta x \in \mathcal{N}[h_x(\hat{x})] = \mathcal{R}[H]$

$\Rightarrow$ $\delta x = H\theta$ for some θ $\underset{0}{\not\|}$ — $\boxed{\text{since } \delta x \neq 0}$

Therefore,

$$V(\hat{x}+\delta x) \cong V(\hat{x}) + \tfrac{1}{2}\delta x^T L(\hat{x},\hat{\lambda})_{xx}\,\delta x$$

$$= V(\hat{x}) + \underbrace{\tfrac{1}{2}\,\theta^T H^T L(\hat{x},\hat{\lambda})_{xx} H\theta}_{>0 \text{ since } \theta \neq 0}$$

so

$V(\hat{x}+\delta x) > V(\hat{x})$, $\forall$ small $\delta x \neq 0$ s.t. $h(\hat{x}+\delta x) = 0$

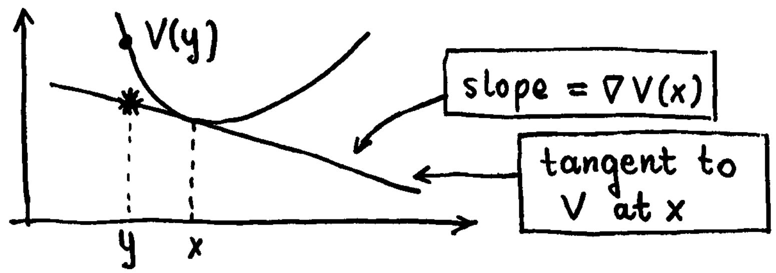
i.e. $\hat{x}$ = a l.m. of (E)

QED

More about convex functions

(13.3) **Th** For a convex C^1 function $V : \mathbb{R}^n \to \mathbb{R}$:

$$V(y) \geqslant \underbrace{V(x) + \nabla V(x)^T (y-x)}_{*} \quad , \forall\, x, y \in \mathbb{R}^n$$



i.e. $V(y)$ lies above its tangent at x, $\forall y$, $\forall x$

e.g. V = standard quadratic on $\mathbb{R}^n$

$$V(y) = V(x) + \nabla V(x)^T (y-x) + \underbrace{\tfrac{1}{2}(y-x)^T C (y-x)}_{\geqslant 0}$$

$$\geqslant V(x) + \nabla V(x)^T (y-x) \quad , \forall\, x, y \in \mathbb{R}^n$$

<u>Special case of sufficient condition</u>
for optimality

(13.4) Th If $h(x) = Dx - f$ — linear equality constraints

V = a convex C^1 function

Then: conditions (i), (ii), (iii) of Th. (13.2) guarantee that:

$\hat{x}$ = a <u>global</u> minimizer for (E)

NOT just the local minimizer predicted by Th (13.2)

<u>Methodology for using Th. (13.2) to find a local minimizer for (E)</u>

Th. (13.2) says (in simplest case):

$\hat{x}$ = a local minimizer for (E)

if $L(\hat{x},\hat{\lambda})_x = 0$ for some $\hat{\lambda} \in \mathbb{R}^m$

$h(\hat{x}) = 0$; $L(\hat{x},\hat{\lambda})_{xx} > 0$

<u>Methodology is:</u>

1. choose $x(\lambda)$ so $L(x(\lambda),\lambda)_x = 0$
2. find $\Lambda = \{\lambda : h(x(\lambda)) = 0\}$
3. find $\tilde{\Lambda} = \{\lambda \in \Lambda : L(x(\lambda),\lambda)_{xx} > 0\}$

   then $\tilde{X} = \{x(\lambda) : \lambda \in \tilde{\Lambda}\}$

   = a set of local minimizers,

   so choose and use an $x \in \tilde{X}$ giving least cost.