

Optimization with Equality Constraints

minimize $f(x)$

subject to $h_1(x) = 0$

$$\begin{aligned} h_2(x) &= 0 \\ &\vdots \\ h_m(x) &= 0 \end{aligned}$$

combined into

$$h(x) = 0$$

\uparrow vector-valued

$$f: \mathbb{R}^n \rightarrow \mathbb{R}, h_i: \mathbb{R}^n \rightarrow \mathbb{R}, h: \mathbb{R}^n \rightarrow \mathbb{R}^m$$

- If h is continuous, then $\mathcal{H} = \{x : h(x) = 0\}$ is closed but not necessarily convex
- Assume that h_i 's are consistent so that \mathcal{H} is non-empty

Examples

1) $h(x) = a^T x - b$

$\mathcal{H} = \{x : a^T x = b\}$ — hyper-plane — convex

2) $h(x) = \|x\| - 1$

$\mathcal{H} = \{x : \|x\| = 1\}$ — surface of sphere — not convex

Assume that f, h_i 's are continuously differentiable.
Under appropriate conditions, we will show that if x^* is a local min, then

$\exists \lambda_1, \lambda_2, \dots, \lambda_m \in \mathbb{R}$ s.t.

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) = 0 \quad (*)$$

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) = 0 \quad (*)$$

Two ways to interpret (*):

(1) $\nabla f(x^*) = - \sum_{i=1}^m \lambda_i \nabla h_i(x^*)$, meaning that

$$\nabla f(x^*) \in G(x^*) = \text{Span} \{ \nabla h_i(x^*), i=1,2,\dots,m \}$$

Note that $G(x^*)$ is m -dimensional iff $\nabla h_i(x^*)$ are linearly independent ($i=1,2,\dots,m$)

(2) $\nabla f(x^*)$ is orthogonal to the subspace:

$$V(x^*) = \{ \bar{z} : \nabla h_i(x^*)^T \bar{z} = 0, i=1,2,\dots,m \}$$

$V(x^*)$ is the space of first-order feasible variations at x^*

Why? Feasible x must satisfy $h(x)=0$.

If \bar{z} is a first-order feasible variation at x^* ,

$h(x^* + \alpha \bar{z}) = 0$ for α sufficiently small

$$\Rightarrow 0 = h_i(x^* + \alpha \bar{z}) = h_i(x^*) + \alpha \nabla h_i(x^*)^T \bar{z} + o(\alpha)$$

$$\therefore \nabla h_i(x^*)^T \bar{z} = 0$$

Dividing by α and taking $\alpha \rightarrow 0$

$$\nabla h_i(x^*)^T \bar{z} = 0$$

Example (A problem with no Lagrange multipliers)

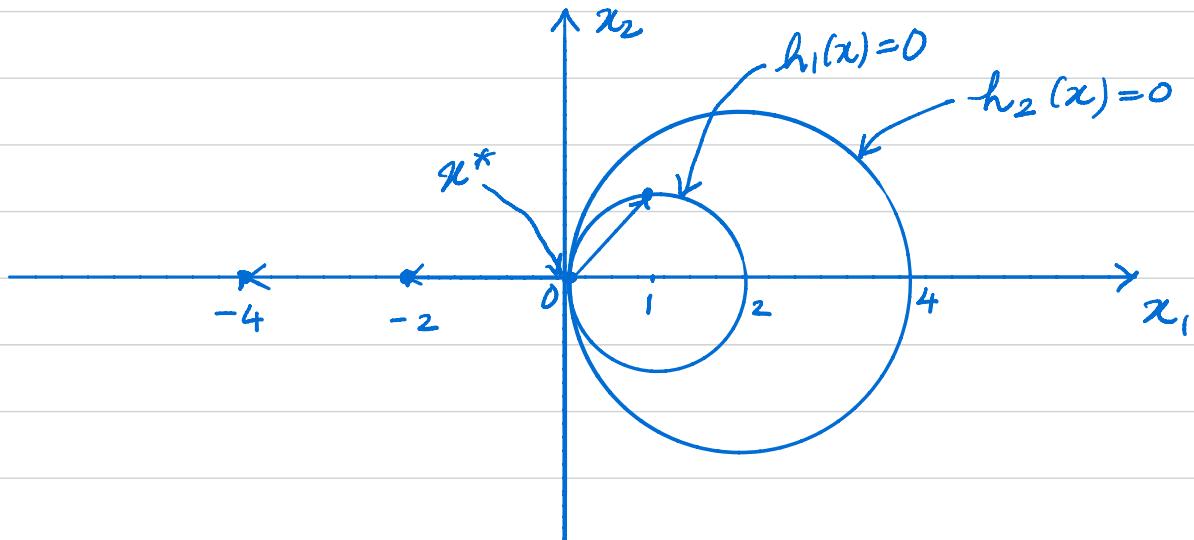
$$\text{minimize } f(x) = x_1 + x_2$$

$$\text{subject to } h_1(x) = (x_1 - 1)^2 + x_2^2 - 1 = 0$$

$$h_2(x) = (x_1 - 2)^2 + x_2^2 - 4 = 0$$

$$\nabla f(x) = (1, 1)$$

$$\nabla h_1(x) = (2(x_1 - 1), 2x_2) \quad \nabla h_2(x) = (2(x_1 - 2), 2x_2)$$



Only feasible point $x = (0, 0) \Rightarrow x^* = (0, 0)$

$$\text{But } \nabla f(x^*) = (1, 1)$$

$$\nabla h_1(x^*) = (-2, 0) \text{ and } \nabla h_2(x^*) = (-4, 0)$$

\Rightarrow There is no choice of λ_1, λ_2 that makes

$$\nabla f(x^*) + \lambda_1 \nabla h_1(x^*) + \lambda_2 \nabla h_2(x^*) = 0$$

Reason: $\nabla h_1(x^*)$ and $\nabla h_2(x^*)$ are linearly dependent

Definition A (feasible) $x \in \mathcal{X}$, i.e., $h(x) = 0$, is called regular if $\nabla h_1(x), \nabla h_2(x), \dots, \nabla h_m(x)$ are linearly independent.

Note For $\nabla h_1(x), \nabla h_2(x), \dots, \nabla h_m(x)$ to be l.i., m has to be $\leq n$.

Lagrange Multiplier Theorem - First-order necessary condition : Let x^* be a local min. of $f(x)$ subject to $h(x) = 0$. Assume that $\nabla h_1(x^*), \nabla h_2(x^*), \dots, \nabla h_m(x^*)$ are linearly independent. Then \exists a unique $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_m)$ s.t.

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) = 0 \quad (*)$$

Remark If $m=n$, and $\nabla h_1(x^*), \dots, \nabla h_m(x^*)$ are l.i., then $(*)$ holds trivially since $\nabla h_1(x^*), \dots, \nabla h_m(x^*)$ form a basis for \mathbb{R}^n . So we need to only consider the case $m < n$ in the proof.

Proof of LMT (Penalty Approach)

Consider sequence of functions:

$$g^{(k)}(x) = f(x) + \frac{\kappa}{2} \|h(x)\|^2 + \underbrace{\frac{\alpha}{2} \|x - x^*\|^2}_{\text{penalty fr violating } h(x) = 0}, \quad k=1, 2, \dots$$

$\nearrow > 0$

\uparrow strictly convex around x^* .

$$g^{(k)}(x) = f(x) + \frac{\kappa}{2} \|h(x)\|^2 + \frac{\alpha}{2} \|x - x^*\|^2, \quad k=1, 2, \dots$$

x^* local min. of $f(x)$ $\Rightarrow \exists \varepsilon > 0$ s.t. $f(x^*) \leq f(x)$
 over $H = \{x : h(x) = 0\}$ for all $x \in H \cap S$,
 $S = \{x : \|x - x^*\| \leq \varepsilon\}$

Let $x^{(k)}$ be an optimal solution to :

$$\begin{aligned} &\text{minimize } g^{(k)}(x) \quad \text{--- (P^(k))} \\ &\text{subject to } x \in S \end{aligned}$$

S is compact \Rightarrow Optimal solution to $(P^{(k)})$
 exists by Weierstrass' Theorem

Lemma $x^{(k)} \rightarrow x^*$ as $k \rightarrow \infty$

$$\text{Proof } g^{(k)}(x^{(k)}) = f(x^{(k)}) + \frac{\kappa}{2} \|h(x^{(k)})\|^2 + \frac{\alpha}{2} \|x^{(k)} - x^*\|^2$$

since $x^{(k)}$
 minimizes $g^{(k)}$ over S and $x^* \in S$ $\Rightarrow g^{(k)}(x^*) = f(x^*) \quad \text{--- (1)}$

Now, since $f(x^{(k)})$ is bounded over S , $\forall k$,

we must have $\lim_{k \rightarrow \infty} \|h(x^{(k)})\| = 0$. Otherwise

LHS of (1) will blow up to ∞ as $k \rightarrow \infty$.

Thus every limit point of $x^{(k)}$, \bar{x} must
 satisfy $h(\bar{x}) = 0$, i.e. $\bar{x} \in H$ (feasible set)

$$(1) \Rightarrow f(x^{(k)}) + \frac{\alpha}{2} \|x^{(k)} - x^*\|^2 \leq f(x^*)$$

$$f(x^{(k)}) + \frac{\alpha}{2} \|x^{(k)} - x^*\|^2 \leq f(x^*)$$

$\Rightarrow f(\bar{x}) + \frac{\alpha}{2} \|\bar{x} - x^*\|^2 \leq f(x^*)$ for every limit point \bar{x} .

Now $\bar{x} \in H$, and $\bar{x} \in S$ (since $x^{(k)} \in S, \forall k$)
Thus $\bar{x} \in H \cap S$, and therefore S is closed

$$f(x^*) \leq f(\bar{x})$$

$$\Rightarrow f(\bar{x}) + \frac{\alpha}{2} \|\bar{x} - x^*\|^2 \leq f(\bar{x})$$

$$\Rightarrow \|\bar{x} - x^*\| = 0 \Rightarrow \bar{x} = x^*.$$

Thus $\lim_{k \rightarrow \infty} x^{(k)}$ exists and $= x^*$. ■

Recall $S = \{x : \|x - x^*\| \leq \varepsilon\}$

$\Rightarrow x^*$ is an interior point of S

$\Rightarrow x^{(k)}$ is an interior point of S for k sufficiently large

$\Rightarrow \nabla g^{(k)}(x^{(k)}) = 0$ for k suff. large

Recall $g^{(k)}(x) = f(x) + \frac{k}{2} \sum_{i=1}^m (h_i(x))^2 + \frac{\alpha}{2} \|x - x^*\|^2$

$$\nabla g^{(k)}(x) = \nabla f(x) + k \sum_{i=1}^m h_i(x) \nabla h_i(x)$$

$$+ \alpha(x - x^*)$$

$$\nabla g^{(k)}(x) = \nabla f(x) + \kappa \sum_{i=1}^m h_i(x) \nabla h_i(x) + \alpha(x - x^*)$$

Define $\underbrace{\nabla h(x)}_{n \times m \text{ matrix}} = [\nabla h_1(x) \ \nabla h_2(x) \ \dots \ \nabla h_m(x)]$

$\nabla h(x^*)$ has rank m (columns are l.i.)

$\Rightarrow \nabla h(x^{(k)})$ has rank m for suff. large κ
by continuity of $\nabla h_i(x)$, $i=1, \dots, m$.

If $A_{n \times m}$ ($n \geq m$) has rank m , then
 $A^T A$ is invertible. $(A^T A)^{-1} A^T$ is called
a pseudo-inverse of A , denoted as A^+

$$A \bar{x} = b \Rightarrow \bar{x} = (A^T A)^{-1} A^T b = A^+ b$$

For κ suff. large

$$\begin{aligned} \nabla g^{(k)}(x) &= \nabla f(x) + \kappa \nabla h(x) h(x) + \alpha(x - x^*) \\ 0 = \nabla g^{(k)}(x^{(k)}) &= \nabla f(x^{(k)}) + \kappa \nabla h(x^{(k)}) h(x^{(k)}) + \alpha(x^{(k)} - x^*) \\ \Rightarrow \kappa \nabla h(x^{(k)}) h(x^{(k)}) &= -(\nabla f(x^{(k)}) + \alpha(x^{(k)} - x^*)) \\ \Rightarrow \kappa h(x^{(k)}) &= -(\nabla h(x^{(k)}))^+ (\nabla f(x^{(k)}) + \alpha(x^{(k)} - x^*)) \\ \Rightarrow \lim_{k \rightarrow \infty} \kappa h(x^{(k)}) &= -(\nabla h(x^*))^+ \nabla f(x^*) \triangleq \lambda \end{aligned}$$

(uniqueness from uniqueness of limit)

$$0 = \nabla g^{(k)}(x^{(k)}) = \nabla f(x^{(k)}) + \kappa \nabla h(x^{(k)}) h(x^{(k)}) + \alpha (x^{(k)} - x^*)$$

$$\Rightarrow \nabla f(x^{(k)}) + \nabla h(x^{(k)}) (\kappa h(x^{(k)})) + \alpha (x^{(k)} - x^*) = 0$$

Taking limits as $k \rightarrow \infty$

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

I.e. $\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) = 0$