Inequality constrained optimization

$$\min_{x \in \mathbb{R}^n} \{ V(x) : h(x) \leq 0 \} \qquad \text{(ICP)}$$

$h : \mathbb{R}^n \to \mathbb{R}^m$

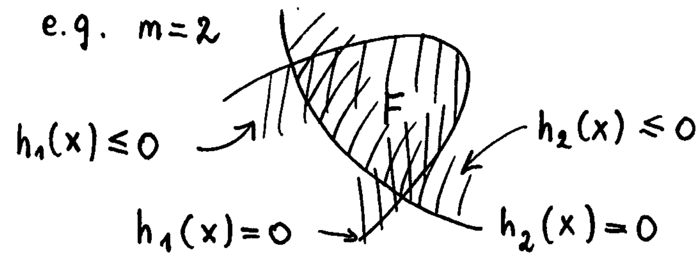
i.e. $h_i(x) \leq 0, \; i = 1, .., m$

Feasible set $F = \{ x \in \mathbb{R}^n : h(x) \leq 0 \}$

$= \{ x \in \mathbb{R}^n : h_i(x) \leq 0, \; i = 1, .., m \}$

(ICP) is called convex iff $V, h_i, i = 1, .., m,$ are convex

e.g. $m = 2$



(17.1) Th $\{ (ICP) = \text{convex} \} \Rightarrow \{ F = \text{convex} \}$

Optimality on convex F

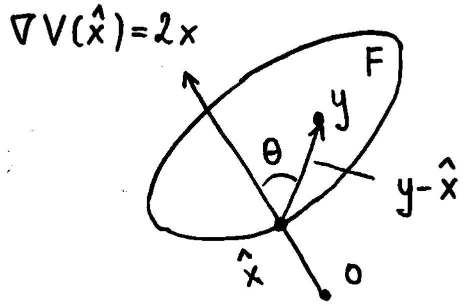
Consider an arbitrary convex set F.

(17.2) Th $\hat{x} \in F$ minimizes a convex, C^1 function V on F, iff:

$$\underbrace{\nabla V(\hat{x})^T (y - \hat{x}) \geq 0}, \quad \forall y \in F$$

i.e. $\nabla V(\hat{x})$ and $(y - \hat{x})$ make an angle $\theta \leq 90^\circ$ between each other

e.g. $\min_{x \in F} V(x)$ where $V(x) = \|x\|^2$



$\hat{x}$ = point in F nearest to the origin

Remark: (17.2) is a useful optimality condition

Proof of Th. (17.2):

Have to show that:

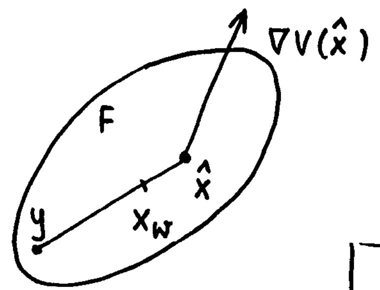
(a) $\hat{x}$ = optimal if $\underbrace{\nabla V(\hat{x})^T (y-\hat{x}) \geq 0,\ \forall y \in F}_{(\alpha)}$

(b) $\hat{x} \neq$ optimal if (α) is false, i.e. if:

$\exists y \in F$ s.t. $\nabla V(\hat{x})^T (y-\hat{x}) < 0$

$\boxed{\text{here we show only (b) :}}$

So, consider such a y and x_ω:



$$x_\omega \triangleq \hat{x} + \omega (y - \hat{x}) \underbrace{\in F}_{\boxed{\text{by convexity of } F}},\ \forall \omega \in [0,1]$$

For small $\omega > 0$:

$$V(x_\omega) = V(\hat{x} + \omega (y-\hat{x}))$$
$$\cong V(\hat{x}) + \nabla V(\hat{x})^T [\omega (y-\hat{x})]$$

Hence, for small $\omega > 0$:

$$V(x_\omega) \cong V(\hat{x}) + \omega \underbrace{\nabla V(\hat{x})^T (y-\hat{x})}_{<0} < V(\hat{x})$$

$\Rightarrow$ for small enough $\omega > 0$: $\begin{cases} x_\omega \in F \\ V(x_\omega) < V(\hat{x}) \end{cases}$

$\Rightarrow$ $\hat{x} \neq$ optimal on F

Duality for inequality-constrained optimization

... similar to the equality-constrained case but with a different perturbation function:

$$\pi(y) \triangleq \min_{x \in \mathbb{R}^n} \{ V(x) : h(x) \leq y \}$$

(17.3) **Th** $\{ (ICP) = \text{convex} \} \Rightarrow$

$\Rightarrow \{ \pi = \text{convex on } \mathbb{R}^m \}$

An important property of π for inequalities, NOT shared by π for equalities, is:

(17.4) **Th** $\quad \pi(\bar{y}) \leq \pi(\tilde{y})$ if $\tilde{y} \leq \bar{y}$

meaning $\tilde{y}_i \leq \bar{y}_i, \ \forall i$

Proof: Let $F_y \triangleq \{x \in \mathbb{R}^n : h(x) \leq y\}$

Then: if $\tilde{y} \leq \bar{y} \Rightarrow F_{\tilde{y}} \subset F_{\bar{y}}$

because:

$$\{x \in F_{\tilde{y}}\} \Rightarrow \{h(x) \leq \tilde{y} \leq \bar{y}\} \Rightarrow \{x \in F_{\bar{y}}\}$$

Now:

$$\pi(\tilde{y}) = \min_{x \in \mathbb{R}^n} \{V(x) : h(x) \leq \tilde{y}\}$$

$$= \min_{x \in F_{\tilde{y}}} V(x) = V(\tilde{x})$$

for some minimizer $\tilde{x} \in F_{\tilde{y}}$

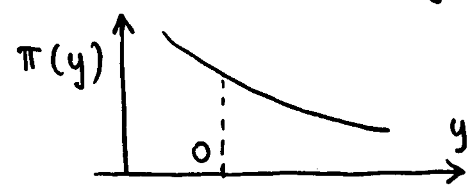
Then if $\tilde{y} \leq \bar{y}$

$$\pi(\bar{y}) = \min_{x \in F_{\bar{y}}} V(x) \leq V(\tilde{x}) = \pi(\tilde{y})$$

since $\tilde{x} \in F_{\tilde{y}} \subset F_{\bar{y}}$
so $\tilde{x}$ may not be optimal for $\pi(\bar{y})$

i.e. $\pi(\bar{y}) \leq \pi(\tilde{y})$ if $\tilde{y} \leq \bar{y}$ QED.

Hence $\pi(y)$ decreases as y increases:



(17.5) **Th** If π is differentiable at $y=0$

then: $\nabla \pi(0) \leq 0$

meaning $\frac{\partial}{\partial y_i} \pi(0) \leq 0, \ i = 1, \ldots, m$

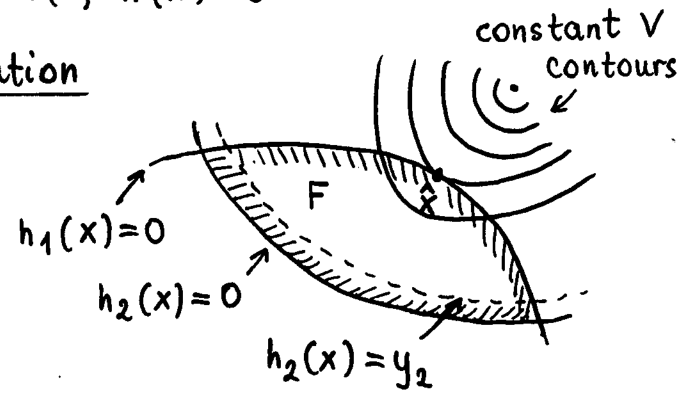
(17.6) Th Suppose π is differentiable at $y=0$ and $\hat{x}$ solves (ICP).

Then: $\nabla\pi(0)_i = 0$ if $h_i(\hat{x})<0$, $\forall i$

and $\nabla\pi(0)^T h(\hat{x}) = 0$

<u>Plausibilification</u>

case $m=2$



Here:

$h_1(\hat{x})=0$

$h_2(\hat{x})<0$ and

$$\underbrace{\min_{x\in\mathbb{R}^2}\left\{V(x) : \begin{array}{l} h_1(x)\le 0 \\ h_2(x)\le y_2 \end{array}\right\}}_{\boxed{\text{i.e. } \pi\left(\begin{bmatrix}0\\ y_2\end{bmatrix}\right) = V(\hat{x}),\ \forall \text{ small } y_2}} = V(\hat{x}), \quad \forall \text{ small } y_2$$

i.e. $\frac{\partial}{\partial y_2}\pi(0) = 0$ if $h_2(\hat{x})<0$

as claimed.

Further:

$$\nabla\pi(0)^T h(\hat{x}) = \frac{\partial}{\partial y_1}\pi(0)\underbrace{h_1(\hat{x})}_{\boxed{=0}} + \underbrace{\frac{\partial\pi}{\partial y_2}(0)}_{\boxed{=0}}\cdot h_2(\hat{x})$$

$= 0$ as claimed

$\boxed{\text{this always happens in the inequality constrained case}}$

For the inequality constrained case, the equality constrained duality theory applies <u>provided</u> the constraint $\lambda \geqslant 0$ is applied.

Exactly as before, define:

(17.7) $$d(\lambda) \triangleq \min_{y\in\mathbb{R}^m}\{\pi(y) + \lambda^T y\}$$

Then we have the <u>same</u> easily-evaluatable formula for $d(\lambda)$ <u>when $\lambda \geqslant 0$.</u>

(17.8) **Th** If $\lambda \geq 0$: $d(\lambda)$ is also given by:

$$d(\lambda) = \min_{x \in \mathbb{R}^n} \{V(x) + \lambda^T h(x)\}$$

Proof: (some of the proof is a little different).

Let
$$J(x,y) = \begin{cases} V(x) & \text{if } h(x) \leq y \\ +\infty & \text{if } h(x) \not\leq y \end{cases}$$

Consider
$$\theta = \min_{\substack{x \in \mathbb{R}^n \\ y \in \mathbb{R}^m}} \{J(x,y) + \lambda^T y\}$$

Then:

$$\theta = \min_{x \in \mathbb{R}^n} \underbrace{\min_{y \in \mathbb{R}^m} \{J(x,y) + \lambda^T y\}}$$

since $J(x,y) = \infty$ if $h(x) \not\leq y$

$$= \min_{y \in \{y \in \mathbb{R}^m : h(x) \leq y\}} \{J(x,y) + \lambda^T y\}$$

Hence:

$$\theta = \min_{x \in \mathbb{R}^n} \min_{y \in \{y \in \mathbb{R}^m : h(x) \leq y\}} \{V(x) + \lambda^T y\}$$

because $J(x,y) = V(x)$ if $h(x) \leq y$

$$= \min_{x \in \mathbb{R}^n} \min_{y \in \{y \in \mathbb{R}^m : h_i(x) \leq y_i,\ i = 1,..,m\}} \{V(x) + \sum_{i=1}^{m} \underbrace{\lambda_i}_{\boxed{\geq 0}} y_i\}$$

$$= \min_{x \in \mathbb{R}^n} \{V(x) + \sum_{i=1}^{m} \lambda_i h_i(x)\}$$

since: $\hat{y}_i = h_i(x)$ as $\lambda_i \geq 0$

$$\equiv \min_{x \in \mathbb{R}^n} \{V(x) + \lambda^T h(x)\}$$

$$\equiv d(\lambda) \text{ of Th (17.8).}$$

But also:

$$\theta = \min_{y \in \mathbb{R}^m} \min_{x \in \mathbb{R}^n} \{J(x,y) + \lambda^T y\}$$

So:

$$\theta = \min_{y \in \mathbb{R}^m} \min_{x \in \{x \in \mathbb{R}^n :\, h(x) \leq y\}} \{ \mathcal{J}(x,y) + \lambda^T y \}$$

as $\mathcal{J}(x,y) = +\infty$ if $h(x) \nleq y$

$$= \min_{y \in \mathbb{R}^m} \min_{x \in \{x \in \mathbb{R}^n :\, h(x) \leq y\}} \{ V(x) + \lambda^T y \}$$

because $\mathcal{J}(x,y) = V(x)$ if $h(x) \leq y$

$$= \min_{y \in \mathbb{R}^m} \left\{ \min_{x \in \{x \in \mathbb{R}^n :\, h(x) \leq y\}} V(x) \right\} + \lambda^T y$$

$$= \min_{y \in \mathbb{R}^m} \pi(y) + \lambda^T y = d(\lambda)$$

of Def^n (17.7)

Hence: $d(\lambda)$ of Def^n (17.7) $= \theta$

$= d(\lambda)$ of Th. (17.8)

QED