Chapter 20 Problems with Equality Constraints

Constrained Optimization Problem

$\min f(\boldsymbol{x}), \quad \text{s.t. } \boldsymbol{x} \in \Omega$

- $f : \mathbb{R}^n \to \mathbb{R}$ is called the objective/cost function.
- $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^\top \in \mathbb{R}^n$ is called as optimization/decision variables.
- $\Omega \subseteq \mathbb{R}^n$ is called as constraint/feasible set.

★ If f has many minimizers, finding one of them will suffice.

★ $\boxed{\max f(\boldsymbol{x}), \quad \text{s.t. } \boldsymbol{x} \in \Omega} \iff \boxed{-\min -f(\boldsymbol{x}), \quad \text{s.t. } \boldsymbol{x} \in \Omega}$

Definition (representation of Ω)

$$\begin{aligned}\Omega =& \{\boldsymbol{x} \in \mathbb{R}^n \mid h_i(\boldsymbol{x}) = 0, i = 1, \cdots, m;\ g_j(\boldsymbol{x}) \le 0, j = 1, \cdots, p\} \\ =& \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{h}(\boldsymbol{x}) = 0, \ \ \boldsymbol{g}(\boldsymbol{x}) \le 0\},\end{aligned}$$

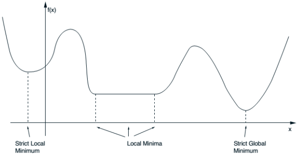
where $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$ (resp. $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^p$) is composed by all h_i (resp. g_i).

- $h_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, are the equality constraint functions.
- $g_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \ldots, p$, are the inequality constraint functions.

Definition (minimizer of optimization $\min f(\boldsymbol{x})$, s.t. $\boldsymbol{x} \in \Omega$)

- $\boldsymbol{x}^* \in \mathbb{R}^n$ is a feasible point if: $\boldsymbol{x}^* \in \Omega$.
- $\boldsymbol{x}^* \in \Omega$ is a local minimizer if: $\exists \varepsilon > 0$ such that $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$ for all $\boldsymbol{x} \in \Omega \backslash \{\boldsymbol{x}^*\}$ and $\|\boldsymbol{x} - \boldsymbol{x}^*\| < \varepsilon$.
- $\boldsymbol{x}^* \in \Omega$ is a global minimizer if: $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$ for all $\boldsymbol{x} \in \Omega \backslash \{\boldsymbol{x}^*\}$.

Remark: If "$\geq$" in the above definitions are "$>$", then it is called strict local/global minimizer.



notations of minimizer

$$\boldsymbol{x}^* \in \Omega \text{ is a global minimizer} \Longrightarrow \begin{cases} f(\boldsymbol{x}^*) = \min_{\boldsymbol{x}\in\Omega} f(\boldsymbol{x}), \\ \boldsymbol{x}^* = \text{argmin}_{\boldsymbol{x}\in\Omega} f(\boldsymbol{x}). \end{cases}$$

If $\Omega = \mathbb{R}^n$, then $f(\boldsymbol{x}^*) = \min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x})$ or $\boldsymbol{x}^* = \text{argmin}_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x})$.

Optimization Problem

$f(x) = (x+1)^2 + 3$

then $\text{argmin}_{x\in\mathbb{R}} f(x) = -1$ and $\text{argmin}_{x\geq 0} f(x) = 0$.

some curious examples

if $f(x) = \frac{1}{x}$ and $\Omega = \mathbb{R}_{++}$, then $f^* = 0$, but no minimizer, $x^* = \infty$

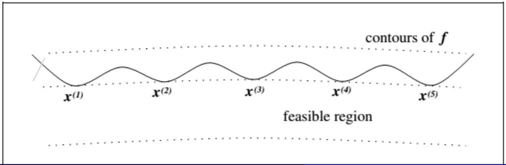
if $f(x) = -\log x$ and $\Omega = \mathbb{R}_{++}$, then $f^* = -\infty$, no minimizer, $x^* = \infty$

if $f(x) = x \log x$ and $\Omega = \mathbb{R}_{++}$, then $f^* = -\frac{1}{e}$, $x^* = \frac{1}{e}$

if $f(x) = x^3 - 3x$, then $f^* = -\infty$, $x^* = 1$ is local minimizer

number of minimizer with/without constraint

$\min\ f(x) = 0.01x_1^2 + (x_2+100)^2$, s.t. $\boldsymbol{x} \in \Omega = \{\boldsymbol{x} \in \mathbb{R}^2 \mid \cos(x_1) - x_2 \leq 0\}$



- without Ω, it has the unique solution $[0, -100]$;
- with Ω: it has infinite number of local solutions near the points: $\boldsymbol{x} = [k\pi, -1]$, $\forall k = \pm 1, \pm 3, \cdots$

Optimization Problem

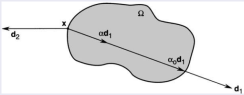
Theorem (Weierstrass Theorem)

The set of minimizers of f over Ω is nonempty and compact if: Ω is closed, f is lower semicontinuous over Ω, and one of the following conditions holds

1. *Ω is bounded.*
2. *some level set $\{\boldsymbol{x} \in \Omega \mid f(\boldsymbol{x}) \leq c\}$ is nonempty and bounded.*
3. *f is coercivity, i.e., for every sequence $\{\boldsymbol{x}^{(k)}\} \subset \Omega$ such that $\|\boldsymbol{x}^{(k)}\| \to \infty$, we have $\lim_{k\to\infty} f(\boldsymbol{x}^{(k)}) = \infty$.*

Definition (feasible direction)

A vector $0 \neq \boldsymbol{d} \in \mathbb{R}^n$ is a feasible direction at $\boldsymbol{x} \in \Omega$ if: $\exists \alpha_0 > 0$ such that $\boldsymbol{x} + \alpha \boldsymbol{d} \in \Omega$ for all $\alpha \in [0, \alpha_0]$.

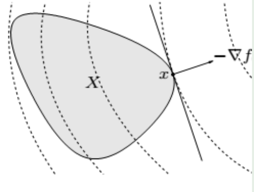


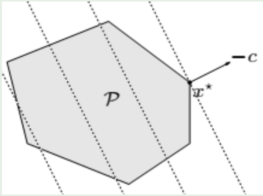
★ minimizer may lie in either int(Ω) or bd(Ω).

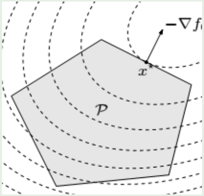
- if $\boldsymbol{x} \in \text{int}(\Omega)$, all feasible directions at $\boldsymbol{x}$ is $\mathbb{R}^n$.
- if $\boldsymbol{x} \in \text{bd}(\Omega)$, the feasible direction at $\boldsymbol{x}$ may be cone, hyperplane, empty...

Conditions for Local Minimizers

Example (where is the minimizer with contour and constraint?)







Definition (linearized feasible direction of Ω)

If $\boldsymbol{x} \in \text{bd}(\Omega)$, the linearized feasible direction of Ω at $\boldsymbol{x}$ is

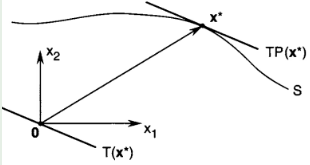
$$0 \neq \boldsymbol{d} = \lim_{\boldsymbol{x}' \to \boldsymbol{x}, \boldsymbol{x}' \in \Omega} (\boldsymbol{x}' - \boldsymbol{x}).$$

All linearized feasible direction of Ω at $\boldsymbol{x}$ is denoted by $T(\boldsymbol{x})$.

★ $T(\boldsymbol{x})$ is usually a cone in $\mathbb{R}^n$.

Example (equality constraint)

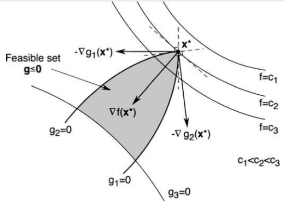
If $S = \{\boldsymbol{x} \in \mathbb{R}^n \mid h_i(\boldsymbol{x}) = 0, i = 1, \ldots, m\}$
with all $h_i : \mathbb{R}^n \to \mathbb{R}$ being differentiable, then
$T(\boldsymbol{x}) = \{\boldsymbol{d} \in \mathbb{R}^n \mid \boldsymbol{D}\boldsymbol{h}_i(\boldsymbol{x})\boldsymbol{d} = 0, i = 1, \ldots, m\}$
is a subspace.



Question: if all $h_i(\boldsymbol{x}) = \boldsymbol{a}_i^\top \boldsymbol{x} - b_i$, $i = 1, \cdots, m$, what is $T(\boldsymbol{x}) =?$

Example (inequality constraint)

If $S = \{\boldsymbol{x} \in \mathbb{R}^n \mid g_j(\boldsymbol{x}) \leq 0, j = 1, \ldots, p\}$
with $g_j : \mathbb{R}^n \to \mathbb{R}$ being differentiable, then
$T(\boldsymbol{x}) = \{\boldsymbol{d} \in \mathbb{R}^n \mid \boldsymbol{D}\boldsymbol{g}_j(\boldsymbol{x})\boldsymbol{d} \leq 0, j \in J(\boldsymbol{x})\}$
is a cone, where
$J(\boldsymbol{x}) = \{j \mid g_j(\boldsymbol{x}) = 0, j = 1, \ldots, p\}$.



Definition (directional derivative)

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $\boldsymbol{d}$ be a feasible direction at $\boldsymbol{x} \in \Omega$. The directional derivative of f w.r.t. $\boldsymbol{d}$ is defined by $\frac{\partial f}{\partial \boldsymbol{d}} = \lim_{\alpha \to 0} \frac{f(\boldsymbol{x}+\alpha\boldsymbol{d})-f(\boldsymbol{x})}{\alpha}$.

- ★ if f is differentiable, then $\frac{\partial f}{\partial \boldsymbol{d}} = \nabla f(\boldsymbol{x})^\top \boldsymbol{d}$.
- ★ if $\boldsymbol{d}$ is a unit vector, then $\frac{\partial f}{\partial \boldsymbol{d}}$ is the rate of increase of f at the point $\boldsymbol{x}$ in the direction $\boldsymbol{d}$.

Conditions for Local Minimizers

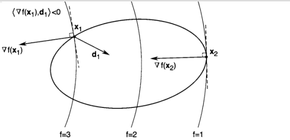
Theorem (first-order necessary condition, FONC)

Let $\Omega \subseteq \mathbb{R}^n$ be a set and $f \in \mathcal{C}^1(\Omega)$. If $\boldsymbol{x}^*$ is a local minimizer of f over Ω, then $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) \geq 0$ for any feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}^*$.

proof. $f(\boldsymbol{x}^* + \alpha \boldsymbol{d}) = f(\boldsymbol{x}^*) + \alpha \nabla f(\boldsymbol{x}^*)^\top \boldsymbol{d} + o(\alpha)$.

Theorem (first-order sufficient condition, FOSC)

Let $\Omega \subseteq \mathbb{R}^n$ be a set and $f \in \mathcal{C}^1(\Omega)$. If $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) > 0$ for any feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}^*$, then $\boldsymbol{x}^*$ is a strict local minimizer of f over Ω.



Corollary ($\Omega = \mathbb{R}^n$ or $\boldsymbol{x}^*$ is an interior point)

Let $f \in \mathcal{C}^1(\Omega)$. If $\boldsymbol{x}^* \in \operatorname{int}(\Omega)$ is a local minimizer, then $\nabla f(\boldsymbol{x}^*) = 0$.
(proof: $T(\boldsymbol{x}^*) = \mathbb{R}^n$, i.e., $\boldsymbol{d}$ is all vectors in $\mathbb{R}^n$.)

Conditions for Local Minimizers

Theorem (second-order necessary condition, SONC)

Let $\Omega \subseteq \mathbb{R}^n$ be a set and $f \in \mathcal{C}^2(\Omega)$. $\boldsymbol{x}^*$ is a local minimizer of f over Ω and $\boldsymbol{d}$ is a feasible direction. If $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) = 0$, then $\boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} \geq 0$.

proof. $f(\boldsymbol{x}^* + \alpha \boldsymbol{d}) = f(\boldsymbol{x}^*) + \alpha \underbrace{\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{d}}_{=0} + \frac{\alpha^2}{2} \boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} + o(\alpha^2)$.

Corollary ($\Omega = \mathbb{R}^n$ or $\boldsymbol{x}^*$ is an interior point)

Let $f \in \mathcal{C}^2(\Omega)$. $\boldsymbol{x}^* \in \operatorname{int}(\Omega)$ is a local minimizer. If $\nabla f(\boldsymbol{x}^*) = 0$, then $\boldsymbol{F}(\boldsymbol{x}^*) \succeq 0$.

- ★ The necessary conditions are not sufficient.
- ★ There may exist points that satisfy FONC and SONC, but are not local minima.

Conditions for Local Minimizers

Example ($f(x) = x^3$)

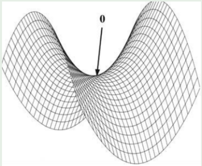
$f'(0) = 0$ and $f''(0) = 0$, $x^* = 0$ satisfies both FONC and SONC. However, $x^* = 0$ is not a minimizer.

Example ($f(\boldsymbol{x}) = x_1^2 - x_2^2$)

$\because \nabla f(\boldsymbol{x}) = [2x_1, -2x_2]^\top = 0$,
$\therefore \boldsymbol{x}^* = [0, 0]^\top$ satisfies FONC.

However, its Hessian $\boldsymbol{F}(\boldsymbol{x}^*) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} \not\geq 0$.

Thus, $x^* = 0$ is not a minimizer.



Conditions for Local Minimizers

Theorem (second-order sufficient condition, SOSC)

Let $\Omega \subseteq \mathbb{R}^n$ be a set and $f \in \mathcal{C}^2(\Omega)$. If $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) = 0$ and $\boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} > 0$, then $\boldsymbol{x}^*$ is a strict local minimizer.

proof. $f(\boldsymbol{x}^* + \alpha \boldsymbol{d}) = f(\boldsymbol{x}^*) + \alpha \underbrace{\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{d}}_{=0} + \frac{\alpha^2}{2}\boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} + o(\alpha^2)$.

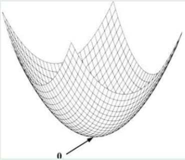
Corollary (SOSC for $\Omega = \mathbb{R}^n$ or $\boldsymbol{x}^*$ is an interior point)

Let $f \in \mathcal{C}^2(\Omega)$ and $\boldsymbol{x}^* \in \operatorname{int}(\Omega)$. If $\nabla f(\boldsymbol{x}^*) = 0$ and $\boldsymbol{F}(\boldsymbol{x}^*) \succ 0$, then $\boldsymbol{x}^*$ is a strict local minimizer.

Example ($f(\boldsymbol{x}) = x_1^2 + x_2^2$)

$\because \nabla f(\boldsymbol{x}) = [2x_1, 2x_2]^\top = 0$, $\therefore \boldsymbol{x}^* = [0,0]^\top$ satisfies FONC.

$\because$ Hessian matrix $\boldsymbol{F}(\boldsymbol{x}^*) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \succ 0$. $\therefore \boldsymbol{x}^* = [0,0]^\top$ satisfies SOSC. $\therefore \boldsymbol{x}^* = 0$ is a strict local minimizer.



Constrained Optimization Problems

standard form of constrained optimization

$\min f(\boldsymbol{x}), \quad$ s.t. $h_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, m;\ \ g_j(\boldsymbol{x}) \le 0,\ j = 1, \ldots, p,$

- $h_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, are the equality constraints (usually $m \le n$)
- $g_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \ldots, p$, are the inequality constraints.

standard form of convex constrained optimization

$\min f(\boldsymbol{x}), \quad$ s.t. $\boldsymbol{a}_i^\top \boldsymbol{x} - b_i = 0,\ i = 1, \ldots, m;\ \ g_j(\boldsymbol{x}) \le 0,\ j = 1, \ldots, p,$

- $h_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, are affine (usually $m \le n$)
- f and $g_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \ldots, p$, are convex functions.

motivation of numerical optimization

Develop convergent methods for seeking the global minimizer $\boldsymbol{x}^*$ throughout the entire feasible domain.

Problems with Equality Constraints

equality constraints optimization

$\min \ f(\boldsymbol{x}) \quad$ s.t. $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}$,
where $\boldsymbol{x} \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{h} = [h_1, \dots, h_m]^\top$, $m \le n$.

Assumption: all $h_i : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable, i.e., $h_i \in \mathcal{C}^1$.
Notation: $S := \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{h}(\boldsymbol{x}) = 0\} = \{\boldsymbol{x} \in \mathbb{R}^n \mid h_1(\boldsymbol{x}) = 0, \dots, h_m(\boldsymbol{x}) = 0\}$.

Definition (regular point of S)

For $\boldsymbol{x}^* \in S$, if vectors $\{\nabla h_1(\boldsymbol{x}^*), \dots, \nabla h_m(\boldsymbol{x}^*)\}$ are linearly independent, then $\boldsymbol{x}^*$ is said to be a regular point.

Theorem

$$D\boldsymbol{h}(\boldsymbol{x}^*) = \begin{bmatrix} \nabla h_1(\boldsymbol{x}^*)^\top \\ \nabla h_2(\boldsymbol{x}^*)^\top \\ \vdots \\ \nabla h_m(\boldsymbol{x}^*)^\top \end{bmatrix}$$

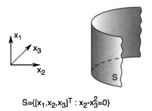
Let $D\boldsymbol{h}(\boldsymbol{x}^*)$ be the Jacobian matrix of $\boldsymbol{h} = [h_1, \dots, h_m]^\top$ at $\boldsymbol{x}^*$. Then, $\boldsymbol{x}^*$ is regular $\iff \operatorname{rank}(D\boldsymbol{h}(\boldsymbol{x}^*)) = m$ (i.e., the Jacobian matrix is of full row rank).

Dimension of Surface

surface and dimension

$S := \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{h}(\boldsymbol{x}) = 0\} = \{\boldsymbol{x} \in \mathbb{R}^n \mid h_1(\boldsymbol{x}) = 0, \ldots, h_m(\boldsymbol{x}) = 0\}$ describes a surface. If all $\boldsymbol{x} \in S$ are regular, then the dimension of S is $n - m$.

Example ($\boldsymbol{x} \in \mathbb{R}^3$, $h_1(\boldsymbol{x}) = x_2 - x_3^2 = 0$)

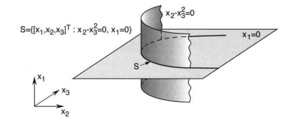


$\because\ \nabla h_1(\boldsymbol{x}) = [0, 1, -2x_3]^\top,$

$\therefore\ \forall \boldsymbol{x} \in \mathbb{R}^3, \nabla h_1(\boldsymbol{x}) \neq \boldsymbol{0}.$

$\therefore\ \dim S = \dim\{\boldsymbol{x} \mid h_1(\boldsymbol{x}) = 0\} = n - m = 2.$

Example ($\boldsymbol{x} \in \mathbb{R}^3$, $h_1(\boldsymbol{x}) = x_1$, $h_2(\boldsymbol{x}) = x_2 - x_3^2$)



$\because\ \nabla h_1(\boldsymbol{x}) = [1, 0, 0]^\top,\ \ \nabla h_2(\boldsymbol{x}) = [0, 1, -2x_3]^\top.$

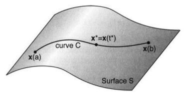
$\therefore\ \{\nabla h_1(\boldsymbol{x}), \nabla h_2(\boldsymbol{x})\}$ are linearly independent.

$\therefore\ \dim S = n - m = 1.$

Curves

Definition (curve on surface S)

$C := \{\boldsymbol{x}(t) \in S \mid t \in (a,b)\}$, which is continuously parameterized by $t \in (a,b)$, i.e., $\boldsymbol{x} : (a,b) \to S$ is continuous.



★ all the points on the curve satisfy the equation of surface. The curve C passes through a point x^* if there exists $t^* \in (a,b)$ such that $\boldsymbol{x}(t^*) = \boldsymbol{x}^*$.

Definition (differentiable curve)

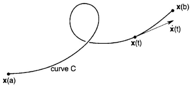
A curve $C = \{\boldsymbol{x}(t) \mid t \in (a,b)\}$ is differentiable if $\dot{\boldsymbol{x}}(t) = \frac{d\boldsymbol{x}}{dt}(t) = [\dot{x}_1(t), \ldots, \dot{x}_n(t)]^\top$ exists for all $t \in (a,b)$.

Definition (twice differentiable curve)

A curve $C = \{\boldsymbol{x}(t) \mid t \in (a,b)\}$ is twice differentiable if $\ddot{\boldsymbol{x}}(t) = \frac{d^2\boldsymbol{x}}{dt^2}(t) = [\ddot{x}_1(t), \ldots, \ddot{x}_n(t)]^\top$ exists for all $t \in (a,b)$.

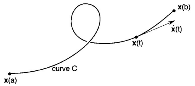
Tangent Spaces

★ $\dot{\boldsymbol{x}}(t)$ and $\ddot{\boldsymbol{x}}(t)$ are n-dimensional vectors and they can be understood as the velocity and acceleration of C at each coordinate.

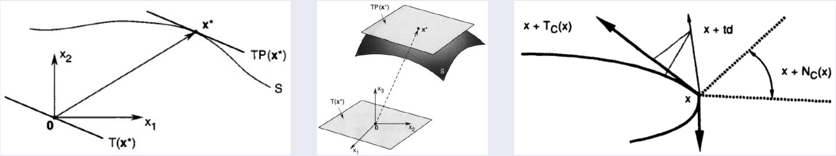


Tangent Spaces

★ $\dot{\boldsymbol{x}}(t)$ and $\ddot{\boldsymbol{x}}(t)$ are n-dimensional vectors and they can be understood as the velocity and acceleration of C at each coordinate.



tangent plane and tangent space



Definition (let $S = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$)

- Tangent space at $\boldsymbol{x}^* \in S$ is $T(\boldsymbol{x}^*) := \{\boldsymbol{y} \in \mathbb{R}^n \mid D\boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{y} = \boldsymbol{0}\}$.
- Tangent plane at $\boldsymbol{x}^* \in S$ is $TP(\boldsymbol{x}^*) := T(\boldsymbol{x}^*) + \boldsymbol{x}^* = \{\boldsymbol{x} + \boldsymbol{x}^* \mid \boldsymbol{x} \in T(\boldsymbol{x}^*)\}$.

★ Tangent space is the nullspace of matrix $D\boldsymbol{h}(\boldsymbol{x}^*)$: $T(\boldsymbol{x}^*) = \mathcal{N}(D\boldsymbol{h}(\boldsymbol{x}^*))$.

★ if $\boldsymbol{x}^*$ is regular, then $\dim(T(\boldsymbol{x}^*)) = n - m$, where m is the number of equality constraints.

Tangent Spaces

$S = \{\boldsymbol{x} \in \mathbb{R}^3 \mid h_1(\boldsymbol{x}) = x_1 = 0, h_2(\boldsymbol{x}) = x_1 - x_2 = 0\}$

$\because\ D\boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} \nabla h_1(\boldsymbol{x})^\top \\ \nabla h_2(\boldsymbol{x})^\top \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix}.$

$\therefore\ \{\nabla h_1, \nabla h_2\}$ are linearly independent.

$\therefore\ \forall \boldsymbol{x} \in S$ is regular point.

$\therefore$ The tangent space at $\boldsymbol{x} \in S$ is

$$T(\boldsymbol{x}) = \left\{ \boldsymbol{y} \;\middle|\; \begin{matrix} \nabla h_1(\boldsymbol{x})^\top \boldsymbol{y} = 0 \\ \nabla h_2(\boldsymbol{x})^\top \boldsymbol{y} = 0 \end{matrix} \right\}$$

$$= \left\{ \boldsymbol{y} \;\middle|\; \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \boldsymbol{0} \right\} = \left\{ \begin{bmatrix} 0 \\ 0 \\ a \end{bmatrix} \;\middle|\; a \in \mathbb{R} \right\}.$$

It is essentially the x_3-axis in $\mathbb{R}^3$.

Theorem

If $\boldsymbol{x}^* \in S$ is regular point, then $\boldsymbol{y} \in T(\boldsymbol{x}^*) \iff \exists$ a differentiable curve $\{\boldsymbol{x}(t) \mid t \in (a,b)\} \subset S$ such that $\exists t^* \in (a,b)$, $\boldsymbol{x}(t^*) = \boldsymbol{x}^*$, $\dot{\boldsymbol{x}}(t^*) = \boldsymbol{y}$.

proof. $(\Leftarrow)$ $\because \ \forall t \in (a,b)$, $\boldsymbol{h}(\boldsymbol{x}(t)) = \boldsymbol{0}$.

$\therefore \ \frac{d}{dt}\boldsymbol{h}(\boldsymbol{x}(t)) = D\boldsymbol{h}(\boldsymbol{x}(t))\dot{\boldsymbol{x}}(t) = \boldsymbol{0}$.

$\therefore \ D\boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{y} = D\boldsymbol{h}(\boldsymbol{x}(t^*))\dot{\boldsymbol{x}}(t^*) = \boldsymbol{0}$.

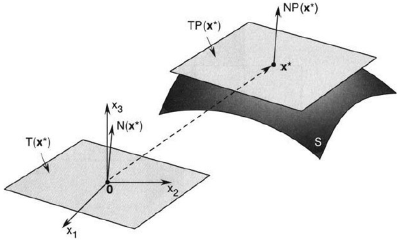
$\therefore \ y \in T(\boldsymbol{x}^*)$.

$(\Rightarrow)$ The implicit function theorem.

Normal Spaces

Definition (let $S = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$)

Normal space at $\boldsymbol{x}^* \in S$ is $N(\boldsymbol{x}^*) := \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} = D\boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{z}, \boldsymbol{z} \in \mathbb{R}^m\}$.



★ Property of normal space:

$$N(\boldsymbol{x}^*) = \mathcal{R}(D\boldsymbol{h}(\boldsymbol{x}^*)^\top)$$
$$= \operatorname{span}[\nabla h_1(\boldsymbol{x}^*), \ldots, \nabla h_m(\boldsymbol{x}^*)].$$

★ if x^* is regular point, the dimension of $N(\boldsymbol{x}^*)$ is m.

Definition (normal plane)

$$NP(\boldsymbol{x}^*) = N(\boldsymbol{x}^*) + \boldsymbol{x}^* = \{\boldsymbol{x} + \boldsymbol{x}^* \in \mathbb{R}^n \mid \boldsymbol{x} \in N(\boldsymbol{x}^*)\}.$$

Lemma (relation of $T(\boldsymbol{x}^*)$ and $N(\boldsymbol{x}^*)$)

$$T(\boldsymbol{x}^*) = N(\boldsymbol{x}^*)^\perp, \qquad T(\boldsymbol{x}^*)^\perp = N(\boldsymbol{x}^*), \qquad \mathbb{R}^n = N(\boldsymbol{x}^*) \oplus T(\boldsymbol{x}^*).$$

Lagrange Condition (2D)

Let $h : \mathbb{R}^2 \to \mathbb{R}$ be the constraint function. Let $\boldsymbol{x}^* = [x_1^*, x_2^*]^\top$ satisfy $h(\boldsymbol{x}^*) = 0$ and $\nabla h(\boldsymbol{x}^*) \neq 0$. The level set through $\boldsymbol{x}^*$ is $\{\boldsymbol{x} \in \mathbb{R}^2 \mid h(\boldsymbol{x}) = 0\}$. Let

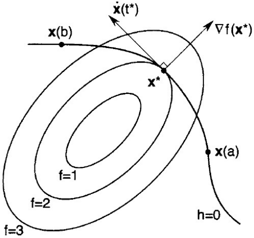
$$\boldsymbol{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \quad t \in (a,b), \quad \boldsymbol{x}^* = \boldsymbol{x}(t^*), \quad \dot{\boldsymbol{x}}(t^*) \neq \mathbf{0}, \quad t^* \in (a,b).$$

$\because$ h is constant on $\{\boldsymbol{x}(t) \mid t \in (a,b)\}$, i.e., $\forall t \in (a,b)$, $h(\boldsymbol{x}(t)) = 0$.
$\therefore$ $\frac{\mathrm{d}}{\mathrm{d}t} h(\boldsymbol{x}(t)) = \nabla h(\boldsymbol{x}(t))^\top \dot{\boldsymbol{x}}(t) = 0.$
$\therefore$ $\nabla h(\boldsymbol{x}^*)$ is orthogonal to $\dot{\boldsymbol{x}}(t^*)$.

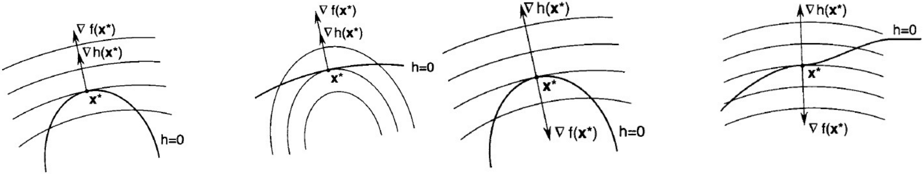
Let $\boldsymbol{x}^*$ be a minimizer of $\boxed{\min_{\boldsymbol{x} \in \mathbb{R}^2} f(\boldsymbol{x}), \text{ s.t. } h(\boldsymbol{x}) = 0\}.}$
Let $\phi(t) = f(\boldsymbol{x}(t))$ achieves a minimizer at $t = t^*$.

$$\therefore\ 0 = \frac{\mathrm{d}}{\mathrm{d}t} \phi(t^*) = \nabla f(\boldsymbol{x}(t^*))^\top \dot{\boldsymbol{x}}(t^*) = \nabla f(\boldsymbol{x}^*)^\top \dot{\boldsymbol{x}}(t^*).$$

$\therefore$ $\nabla f(\boldsymbol{x}^*)$ is orthogonal to $\dot{\boldsymbol{x}}(t^*)$.



Lagrange Condition (2D)



(a) (b) (c) (d)

Lagrange's Theorem: $\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x})$, s.t. $h(\boldsymbol{x}) = 0$.

If $\boldsymbol{x}^*$ is a minimizer and $\boldsymbol{x}^*$ is regular point, then $\exists \boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that

$$\begin{cases} \nabla f(\boldsymbol{x}^*) + D\boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{\lambda}^* = \boldsymbol{0}, \\ \boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}, \end{cases} \quad \text{or} \quad \begin{cases} \nabla f(\boldsymbol{x}^*) + \sum_{i=1}^m \lambda_i^* \nabla h(x^*) = 0, \\ h_i(\boldsymbol{x}^*) = 0, \quad i = 1, \dots, m. \end{cases}$$

proof. It suffices to prove $\nabla f(\boldsymbol{x}^*) \in \mathcal{R}(D\boldsymbol{h}(\boldsymbol{x}^*)^\top) = N(\boldsymbol{x}^*) = T(\boldsymbol{x}^*)^\perp$.

Let $\boldsymbol{y} \in T(\boldsymbol{x}^*)$.

$\therefore$ $\exists \{\boldsymbol{x}(t) \mid t \in (a,b)\}$ such that $\boldsymbol{h}(\boldsymbol{x}(t)) = \boldsymbol{0}$ and $\phi(t) = f(\boldsymbol{x}(t))$ for all t.

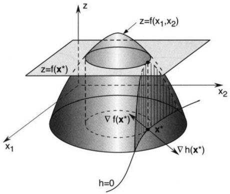
$\therefore$ $\exists t^* \in (a,b)$ s.t. $\boldsymbol{x}(t^*) = \boldsymbol{x}^*$, $\dot{\boldsymbol{x}}(t^*) = \boldsymbol{y}$.

$\therefore$ t^* is local minimizer of ϕ.

$\therefore$ $\frac{d\phi}{dt}(t^*) = Df(\boldsymbol{x}^*)\dot{\boldsymbol{x}}(t^*) = Df(\boldsymbol{x}^*)\boldsymbol{y} = \nabla f(\boldsymbol{x}^*)^\top \boldsymbol{y} = 0$,

i.e., $\nabla f(\boldsymbol{x}^*) \in T(\boldsymbol{x}^*)^\perp$.

- ★ $\lambda^* \in \mathbb{R}$ is called the Lagrange multiplier.
- ★ if $\boldsymbol{x}^*$ is a optimum, then $\nabla f(\boldsymbol{x}^*)$ can be expressed as a linear combination of the gradients of constraints.
- ★ Lagrange theorem is the FONC for local minimizer $\boldsymbol{x}^*$.



Lagrangian Function

equality constrained optimization: $\min\ f(\boldsymbol{x})$ s.t. $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$

Lagrangian function, denoted by $l : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, is defined as

$l(\boldsymbol{x}, \boldsymbol{\lambda}) := f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top \boldsymbol{h}(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i^\top h_i(\boldsymbol{x})$

If $\boldsymbol{x}^*$ is minimizer, then there exists $\boldsymbol{\lambda}^*$ such that $\nabla l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{0}$.

Lagrange's theorem amounts to FONC for unconstrained optimization applied to the Lagrangian function, i.e.,

$$\boldsymbol{0} = \nabla l(\boldsymbol{x}, \boldsymbol{\lambda}) = \begin{bmatrix} \nabla_{\boldsymbol{x}} l(\boldsymbol{x}, \boldsymbol{\lambda}) \\ \nabla_{\boldsymbol{\lambda}} l(\boldsymbol{x}, \boldsymbol{\lambda}) \end{bmatrix} \iff \begin{cases} \nabla f(\boldsymbol{x}) + D\boldsymbol{h}(\boldsymbol{x})^\top \boldsymbol{\lambda} = \boldsymbol{0}, \\ \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}. \end{cases} \tag{1}$$

- ★ Nonlinear equations (1) has $n + m$ equations with $n + m$ unknowns (possibly no solution).
- ★ Lagrange condition is necessary but not sufficient, an $\boldsymbol{x}^*$ satisfying (1) may be not a minimizer.

$\min f(\boldsymbol{x}) = x_1^2 + x_2^2$, s.t. $\boldsymbol{h}(\boldsymbol{x}) = x_1^2 + 2x_2^2 - 1 = 0$

Lagrangian function is $l(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda \boldsymbol{h}(\boldsymbol{x})$.

$$\therefore \begin{cases} \nabla_{\boldsymbol{x}} l(\boldsymbol{x}, \lambda) = \nabla_{\boldsymbol{x}}[f(\boldsymbol{x}) + \lambda \boldsymbol{h}(\boldsymbol{x})] = \begin{bmatrix} 2x_1 + 2\lambda x_1 \\ 2x_2 + 4\lambda x_2 \end{bmatrix}, \\ \nabla_{\lambda} l(\boldsymbol{x}, \lambda) = \boldsymbol{h}(\boldsymbol{x}) = x_1^2 + 2x_2^2 - 1. \end{cases}$$

By the Lagrangian Theorem, $\nabla l(\boldsymbol{x}, \lambda) = \mathbf{0}$.

$$\therefore \begin{cases} 2x_1 + 2\lambda x_1 = 0 \\ 2x_2 + 4\lambda x_2 = 0 \\ x_1^2 + 2x_2^2 = 1 \end{cases} \Longrightarrow \begin{cases} \boldsymbol{x}^{(1)} = [0, \frac{1}{\sqrt{2}}]^\top \\ \boldsymbol{x}^{(2)} = [0, \frac{-1}{\sqrt{2}}]^\top \\ \boldsymbol{x}^{(3)} = [1, 0]^\top \\ \boldsymbol{x}^{(4)} = [-1, 0]^\top \end{cases}.$$

Because $f(\boldsymbol{x}^{(1)}) = f(\boldsymbol{x}^{(2)}) = \frac{1}{2}$, $f(\boldsymbol{x}^{(3)}) = f(\boldsymbol{x}^{(4)}) = 1$.

$\therefore$ if there are local minimizers, then they are $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$;
if there are local maximizers, then they are $\boldsymbol{x}^{(3)}$ and $\boldsymbol{x}^{(4)}$.

Lagrangian Function

exercises

$\min\ x_1 + x_2$, s.t., $x_1^2 + x_2^2 - 2 = 0$, (Ans: $\boldsymbol{x}^* = [-1, -1]^\top$)
$\min\ 2(x_1^2 + x_2^2 - 1) - x_1$, s.t. $x_1^2 + x_2^2 = 1$, (Ans: $\boldsymbol{x}^* = [1, 0]^\top$)
$\min\ x_1 x_2$ s.t. $x_1 + x_2 = 2$, (Ans: xxxx)

Example ($\max\ \frac{\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x}}$, where $\boldsymbol{Q} = \boldsymbol{Q}^\top \succeq 0$ and $\boldsymbol{P} = \boldsymbol{P}^\top \succ 0$)

$\because$ if $\boldsymbol{x}$ is a solution to the problem, then so is $t\boldsymbol{x}$ for all $t \neq 0$. (why?)

$\therefore$ $\max\ \frac{\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x}} \iff \max\{\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} \mid \boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} = 1\}$.

$\because$ all feasible point is regular (why?).

$\therefore$ The Lagrangian function is $l(\boldsymbol{x}, \lambda) = \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \lambda(1 - \boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x})$.

$$\therefore \begin{cases} \nabla_{\boldsymbol{x}} l(\boldsymbol{x}, \lambda) = 2\boldsymbol{Q}\boldsymbol{x} - 2\lambda \boldsymbol{P}\boldsymbol{x} = \boldsymbol{0} \\ \nabla_\lambda l(\boldsymbol{x}, \lambda) = 1 - \boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} = 0 \end{cases} \Longrightarrow \begin{cases} \boldsymbol{Q}\boldsymbol{x} = \lambda \boldsymbol{P}\boldsymbol{x} \\ \boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} = 1 \end{cases}$$

$\therefore$ $\boldsymbol{x}^*$ is an eigenvector of $\boldsymbol{P}^{-1}\boldsymbol{Q}$ and λ^* is the corresponding eigenvalue.

$\because$ $(\boldsymbol{x}^*)^\top \boldsymbol{P}\boldsymbol{x}^* = 1$ and $\boldsymbol{Q}\boldsymbol{x}^* = \lambda^* \boldsymbol{P}\boldsymbol{x}^* \Rightarrow$ obj $= (\boldsymbol{x}^*)^\top \boldsymbol{Q}\boldsymbol{x}^* = \lambda^*(\boldsymbol{x}^*)^\top \boldsymbol{P}\boldsymbol{x}^* = \lambda^*$.

$\therefore$ λ^* is the maximum, more precisely, λ^* is the maximal eigenvalue of $\boldsymbol{P}^{-1}\boldsymbol{Q}$.

★ regularity plays an essential role.

counterexample: $\min\{f(x) \mid h(x) = 0\}$ is lack of regularity

$f(x) = x$, $h(x) = \begin{cases} x^2 & \text{if } x < 0 \\ 0 & \text{if } 0 \le x \le 1 \\ (x-1)^2 & \text{if } x > 1 \end{cases}$. The feasible set is $[0, 1]$.

Clearly, $x^* = 0$ is a local minimizer. However, $f'(x^*) = 1$ and $h'(x^*) = 0$. Therefore, x^* does not satisfy the necessary condition in Lagrange's theorem.

Second-Order Conditions

★ Lagrange condition is merely necessary. Therefore, points satisfying Lagrange condition are only candidates. To classify such points as minimizers, we need the second-order condition.

★ Let f: $\mathbb{R}^n \to \mathbb{R}$ and $\boldsymbol{h}$: $\mathbb{R}^n \to \mathbb{R}^m$ be twice continuously differentiable.

Let $\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ be the Hessian of Lagrangian function $l(\boldsymbol{x}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{x}$:

$$\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{F}(\boldsymbol{x}) + \underbrace{\lambda_1 \boldsymbol{H}_1(\boldsymbol{x}) + \ldots + \lambda_m \boldsymbol{H}_m(\boldsymbol{x})}_{[\boldsymbol{\lambda}\boldsymbol{H}(\boldsymbol{x})]} = \boldsymbol{F}(\boldsymbol{x}) + [\boldsymbol{\lambda}\boldsymbol{H}(\boldsymbol{x})].$$

Second-Order Conditions

Theorem (second-order necessary conditions, SONC)

Let $\boldsymbol{x}^*$ *be a local minimizer of* $\min\{f(\boldsymbol{x}) \mid \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$. *If* $\boldsymbol{x}^*$ *is regular, then there exists* $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ *such that:* ❶ $\nabla f(\boldsymbol{x}^*) + D\boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{\lambda}^* = \boldsymbol{0}$; ❷ $\boldsymbol{h}(\boldsymbol{x}) = 0$;
❸ $\boldsymbol{y}^\top \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)\boldsymbol{y} \geq 0$ *for all* $\boldsymbol{y} \in T(\boldsymbol{x}^*)$.

proof. ❶❷ are the Lagrange's theorem. We now prove ❸.

Let $S = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$ and $\boldsymbol{y} \in T(\boldsymbol{x}^*)$.

$\because$ $\boldsymbol{h} \in \mathcal{C}^2 \implies \exists$ a twice-differentiable curve $\{\boldsymbol{x}(t) \mid t \in (a,b)\} \subset S$ such that $\exists t^* \in (a,b)$, $\quad \boldsymbol{x}(t^*) = \boldsymbol{x}^*$, $\quad \dot{\boldsymbol{x}}(t^*) = \boldsymbol{y}$.

$\because$ $\boldsymbol{x}^*$ is a minimizer of $f \implies t^*$ is a minimizer of $\phi(t) = f(\boldsymbol{x}(t))$.

By the SONC for unconstrained optimization, we obtain $\phi''(t^*) \geq 0$.

$$\begin{aligned} \therefore \quad \phi''(t^*) = \frac{d}{dt}[Df(\boldsymbol{x}(t^*))\dot{\boldsymbol{x}}(t^*)] &= \dot{\boldsymbol{x}}(t^*)^\top \boldsymbol{F}(\boldsymbol{x}^*)\dot{\boldsymbol{x}}(t^*) + Df(\boldsymbol{x}^*)\ddot{\boldsymbol{x}}(t^*) \\ &= \boldsymbol{y}^\top \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{y} + Df(\boldsymbol{x}^*)\ddot{\boldsymbol{x}}(t^*) \geq 0. \qquad (2) \end{aligned}$$

$\because$ $\boldsymbol{h}(\boldsymbol{x}(t)) = \boldsymbol{0}$ for all $t \in (a,b)$, we have $\frac{d^2}{dt^2}\boldsymbol{\lambda}^{*\top}\boldsymbol{h}(\boldsymbol{x}(t)) = 0$.

Second-Order Conditions

$$\therefore \frac{d^2}{dt^2}\boldsymbol{\lambda}^{*\top}\boldsymbol{h}(\boldsymbol{x}(t)) = \frac{d}{dt}\left[\boldsymbol{\lambda}^{*\top}\frac{d}{dt}\boldsymbol{h}(\boldsymbol{x}(t))\right] = \frac{d}{dt}\left[\sum_{k=1}^{m}\lambda_k^*\frac{d}{dt}h_k(\boldsymbol{x}(t))\right]$$

$$= \frac{d}{dt}\left[\sum_{k=1}^{m}\lambda_k^* Dh_k(\boldsymbol{x}(t))\dot{\boldsymbol{x}}(t)\right] = \sum_{k=1}^{m}\lambda_k^*\frac{d}{dt}(Dh_k(\boldsymbol{x}(t))\dot{\boldsymbol{x}}(t))$$

$$= \sum_{k=1}^{m}\lambda_k^*\left[\dot{\boldsymbol{x}}(t)^\top \boldsymbol{H}_k(\boldsymbol{x}(t))\dot{\boldsymbol{x}}(t) + Dh_k(\boldsymbol{x}(t))\ddot{\boldsymbol{x}}(t)\right]$$

$$= \dot{\boldsymbol{x}}^\top(t)\left[\boldsymbol{\lambda}^*\boldsymbol{H}(\boldsymbol{x}(t))\right]\dot{\boldsymbol{x}}(t) + \boldsymbol{\lambda}^{*\top}D\boldsymbol{h}(\boldsymbol{x}(t))\ddot{\boldsymbol{x}}(t) = 0.$$

Substituting $t = t^*$ into the above equality and combining with (2), we have

$$\begin{cases} \boldsymbol{y}^\top[\boldsymbol{\lambda}^*\boldsymbol{H}(\boldsymbol{x}^*)]\boldsymbol{y} + \boldsymbol{\lambda}^{*\top}D\boldsymbol{h}(\boldsymbol{x}^*)\ddot{\boldsymbol{x}}(t^*) = 0, \\ \boldsymbol{y}^\top\boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{y} + Df(\boldsymbol{x}^*)\ddot{\boldsymbol{x}}(\boldsymbol{t}^*) \geq 0. \end{cases}$$

By adding the above two equalities to obtain

$$\boldsymbol{y}^\top(\boldsymbol{F}(\boldsymbol{x}^*) + [\boldsymbol{\lambda}^*\boldsymbol{H}(\boldsymbol{x}^*)])\boldsymbol{y} + (Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*\top}D\boldsymbol{h}(\boldsymbol{x}^*))\ddot{\boldsymbol{x}}(t^*) \geq 0.$$

But, by Lagrange's theorem, $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*\top}D\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}^\top$.

$$\therefore \boldsymbol{y}^\top(\boldsymbol{F}(\boldsymbol{x}^*) + [\boldsymbol{\lambda}^*\boldsymbol{H}(\boldsymbol{x}^*)])\boldsymbol{y} = \boldsymbol{y}^\top\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)\boldsymbol{y} \geq 0.$$

★ $\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ plays a similar role as the Hessian $\boldsymbol{F}(\boldsymbol{x})$ of the objective function f in the unconstrained minimization case. However, we now require that $\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \succeq 0$ only on $T(\boldsymbol{x}^*)$ rather than on $\mathbb{R}^n$.

Theorem (Second-Order Sufficient Conditions)

Let $f, \boldsymbol{h} \in \mathcal{C}^2$. If there exists $\boldsymbol{x}^* \in \mathbb{R}^n$, $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that:
❶ $\nabla f(\boldsymbol{x}^*) + D\boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{\lambda}^* = \boldsymbol{0}$; ❷ $\boldsymbol{h}(\boldsymbol{x}) = 0$; ❸ $\boldsymbol{y}^\top \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)\boldsymbol{y} > 0$ for all $\boldsymbol{y} \in T(\boldsymbol{x}^*)$. Then, $\boldsymbol{x}^*$ is a strict local minimizer of $\min\{f(\boldsymbol{x}) \mid \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$.
(proof. The interested reader can consult [88, p. 334] for a proof)

- ★ if $\boldsymbol{x}^*$ satisfies the Lagrange condition and $\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \succ 0$ on $T(\boldsymbol{x}^*)$, then $\boldsymbol{x}^*$ is a strict local minimizer.
- ★ On the contrary, if $\boldsymbol{x}^*$ satisfies the Lagrange condition and $\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \prec 0$ on $T(\boldsymbol{x}^*)$, then $\boldsymbol{x}^*$ is a strict local maximizer.

Second-Order Conditions

$\min\ f(\boldsymbol{x}) = \frac{1}{2}(x_1^2 + x_2^2 + x_3^2)$, s.t. $h(\boldsymbol{x}) = x_1 + x_2 + x_3 = 3$

$\because$ Lagrangian function is $l(\boldsymbol{x}, \lambda) = \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) + \lambda(x_1 + x_2 + x_3 - 3)$.

$\because$ By the Lagrangian Theorem, $Dl(\boldsymbol{x}, \lambda) = [\nabla_{\boldsymbol{x}} l(\boldsymbol{x}, \lambda), \nabla_\lambda l(\boldsymbol{x}, \lambda)] = \mathbf{0}$.

$$\therefore \begin{cases} x_1 + \lambda = 0, & x_2 + \lambda = 0 \\ x_3 + \lambda = 0, & x_1 + x_2 + x_3 = 3 \end{cases}$$

$\therefore\ \boldsymbol{x}^* = [1, 1, 1]^\top,\ \lambda^* = -1$

We now check that whether $\boldsymbol{x}^*$ is strict local optima.

$\because$ Hessian of $l(\boldsymbol{x}, \lambda)$: $\boldsymbol{L}(\boldsymbol{x}^*, \lambda^*) = \boldsymbol{I} \succ 0$.

By the SOSC, $\boldsymbol{x}^* = [1, 1, 1]^\top$ is a strict local minimizer.

$\min\ f(\boldsymbol{x}) = -(x_1x_2 + x_2x_3 + x_3x_1)$, s.t. $h(\boldsymbol{x}) = x_1 + x_2 + x_3 = 3$

$\because$ Lagrangian function: $l(\boldsymbol{x}, \lambda) = -(x_1x_2 + x_2x_3 + x_3x_1) + \lambda(x_1 + x_2 + x_3 - 3)$.

$\because$ By the Lagrangian Theorem, $Dl(\boldsymbol{x}, \lambda) = [\nabla_{\boldsymbol{x}} l(\boldsymbol{x}, \lambda), \nabla_\lambda l(\boldsymbol{x}, \lambda)] = \mathbf{0}$.

$$\therefore \begin{cases} -x_2 - x_3 + \lambda = 0, & -x_1 - x_3 + \lambda = 0 \\ -x_1 - x_2 + \lambda = 0, & x_1 + x_2 + x_3 = 3 \end{cases} \Longrightarrow \boldsymbol{x}^* = [1, 1, 1]^\top,\ \lambda^* = 2$$

We now check that whether $\boldsymbol{x}^*$ is strict local optima.

$$\because \text{ Hessian of } l(\boldsymbol{x}, \lambda): \quad \boldsymbol{L}(\boldsymbol{x}^*, \lambda^*) = \begin{bmatrix} 0 & -1 & -1 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{bmatrix} \nsucc 0.$$

But, tangent space of constraint $\{\boldsymbol{x} \in \mathbb{R}^3 \mid x_1 + x_2 + x_3 - 3 = 0\}$ is

$$\begin{aligned} T(\boldsymbol{x}^*) &= \left\{\boldsymbol{y} \in \mathbb{R}^3 \mid \nabla \boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{y} = 0\right\} = \left\{\boldsymbol{y} \in \mathbb{R}^3 \mid [1, 1, 1]\boldsymbol{y} = 0\right\} \\ &= \left\{\boldsymbol{y} \in \mathbb{R}^3 \mid \boldsymbol{y} = k_1[-1, 1, 0]^\top + k_2[-1, 0, 1]^\top,\ \forall k_1, k_2 \in \mathbb{R}\right\} \\ &= \left\{\boldsymbol{y} \in \mathbb{R}^3 \mid \boldsymbol{y} = \begin{bmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix},\ \forall k_1, k_2 \in \mathbb{R}\right\} \\ &= \left\{\boldsymbol{y} \in \mathbb{R}^3 \mid \boldsymbol{y} = \boldsymbol{P}\boldsymbol{k},\ \forall \boldsymbol{k} \in \mathbb{R}^2\right\} \end{aligned}$$

$\therefore\ \boldsymbol{y}^\top \boldsymbol{L}(\boldsymbol{x}^*, \lambda^*)\boldsymbol{y} = \boldsymbol{k}^\top \boldsymbol{P}^\top \boldsymbol{L}(\boldsymbol{x}^*, \lambda^*)\boldsymbol{P}\boldsymbol{k} = \boldsymbol{k}^\top \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \boldsymbol{k} > 0,\ \forall 0 \neq \boldsymbol{y} \in T(\boldsymbol{x}^*).$

$\therefore\ \boldsymbol{L}(\boldsymbol{x}^*, \lambda^*) \succ 0$ on $T(\boldsymbol{x}^*)$

$\therefore\ \boldsymbol{x}^* = [1, 1, 1]^\top$ is a strict local minimizer.

Example

$\max \frac{\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x}}$, where $\boldsymbol{Q} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$, $\boldsymbol{P} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$. (hint: $\iff \max\limits_{\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x}=1} \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}$)

Ans: Lagrangian function: $l(\boldsymbol{x}, \lambda) = \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \lambda(1 - \boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x})$.

Lagrangian condition (eigenvalue problem): $(\lambda \boldsymbol{I} - \boldsymbol{P}^{-1}\boldsymbol{Q})\boldsymbol{x} = \boldsymbol{0}$.

Eigenvalues of $\boldsymbol{P}^{-1}\boldsymbol{Q}$ are: $\mu_1 = 2$ and $\mu_2 = 1$.

As prior discussed, $\boldsymbol{x}^* = \pm \left[\frac{1}{\sqrt{2}}, 0\right]^\top$ and $\lambda^* = 2$ satisfies the Lagrange conditions.

We now check that whether $\pm\boldsymbol{x}^*$ are strict local optima.

$\because$ Hessian of $l(\boldsymbol{x}, \lambda)$: $\boldsymbol{L}(\boldsymbol{x}^*, \lambda^*) = 2\boldsymbol{Q} - 2\lambda \boldsymbol{P} = \begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix}$.

Second-Order Conditions

The tangent space $T(\boldsymbol{x}^*)$ of constraint $\{\boldsymbol{x} \in \mathbb{R}^2 \mid 1 - \boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} = 0\}$ is

$$T(\boldsymbol{x}^*) = \left\{\boldsymbol{y} \in \mathbb{R}^2 \mid \boldsymbol{x}^{*\top} \boldsymbol{P}\boldsymbol{y} = 0\right\} = \left\{\boldsymbol{y} \in \mathbb{R}^2 \mid [\sqrt{2}, 0]\boldsymbol{y} = 0\right\}$$
$$= \left\{\boldsymbol{y} \in \mathbb{R}^2 \mid \boldsymbol{y} = [0, a]^\top, a \in \mathbb{R}\right\}.$$

$\because$ $\forall \boldsymbol{y} \in T(\boldsymbol{x}^*)$, $\boldsymbol{y} \neq 0$, we have

$$\boldsymbol{y}^\top \boldsymbol{L}(\boldsymbol{x}^*, \lambda^*)\boldsymbol{y} = [0, a] \begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 0 \\ a \end{bmatrix} = -2a^2 < 0.$$

$\therefore$ $\boldsymbol{L}(\boldsymbol{x}^*, \lambda^*) \prec 0$ on $T(\boldsymbol{x}^*)$

$\therefore$ $\boldsymbol{x}^* = [\frac{1}{\sqrt{2}}, 0]^\top$ is a strict local maximizer.

The same operation for the point $-\boldsymbol{x}^*$.

Second-Order Conditions

Lemma (a useful lemma)

Let $\boldsymbol{P}$ and $\boldsymbol{Q}$ be symmetric matrices. Assume that $\boldsymbol{Q} \succeq 0$, and $\boldsymbol{P} \succ 0$ on $\mathcal{N}(\boldsymbol{Q})$ (i.e., $\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq 0$ with $\boldsymbol{Q}\boldsymbol{x} = 0$). Then, there exists $c \in \mathbb{R}$ such that

$$\boldsymbol{P} + c\boldsymbol{Q} \succ 0, \ \forall c > \bar{c}.$$

proof. If contrary, $\exists \boldsymbol{x}^k \in \mathbb{R}^n$ with $\|\boldsymbol{x}^k\| = 1$ such that

$$(\boldsymbol{x}^k)^\top \boldsymbol{P}\boldsymbol{x}^k + k(\boldsymbol{x}^k)^\top \boldsymbol{Q}\boldsymbol{x}^k < 0, \quad \forall k.$$

Let $\{\boldsymbol{x}^{k_j}\}$ be a subsequence converging to some $\boldsymbol{x}$ with $\|\boldsymbol{x}\| = 1$.

By taking the limit superimum over j,
$\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} + \limsup_{j\to\infty}(j(\boldsymbol{x}^{k_j})^\top \boldsymbol{Q}\boldsymbol{x}^{k_j}) \le 0 \implies (\boldsymbol{x}^{k_j})^\top \boldsymbol{Q}\boldsymbol{x}^{k_j} \le 0.$

$\because \ \boldsymbol{Q} \succeq 0 \implies \lim_{j\to\infty}(\boldsymbol{x}^{k_j})^\top \boldsymbol{Q}\boldsymbol{x}^{k_j} = 0 \implies \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} = 0.$

By the hypothesis, $\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} > 0$, a contradiction.

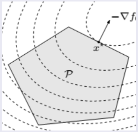
Quadratic Programming

Definition (quadratic programming, QP)

$\min\ \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}$, s.t. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$, $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $m < n$, $\operatorname{rank} \boldsymbol{A} = m$.

properties of QP:

- QP $\Longleftrightarrow$ minimize/maximize quadratic function on polyhedron.
- if $\boldsymbol{Q} \succeq 0$, then QP is convex optimization.



Lagrange condition of QP with equality constraint (i.e., $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$)

Lagrangian function: $l(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{\lambda}^\top(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})$.

Lagrange condition:

$$\begin{cases} \nabla_{\boldsymbol{x}} l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{Q}\boldsymbol{x}^* - \boldsymbol{A}^\top \boldsymbol{\lambda}^* = \boldsymbol{0} \\ \nabla_{\boldsymbol{\lambda}} l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0} \end{cases} \xrightarrow{\text{if } \boldsymbol{Q} \succ 0} \begin{cases} \boldsymbol{x}^* = \boldsymbol{Q}^{-1}\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^\top)^{-1}\boldsymbol{b} \\ \boldsymbol{\lambda}^* = (\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^\top)^{-1}\boldsymbol{b} \end{cases}$$

★ To check whether $\boldsymbol{x}^*$ is a minimizer, we can use the SOSC. Indeed, as the Hessian matrix $\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \equiv \boldsymbol{Q} \succ 0$, $\boldsymbol{x}^*$ is a global minimizer.

one linear equality constraint optimization

$\min\ f(\boldsymbol{x}),\ $ s.t. $\boldsymbol{a}^\top \boldsymbol{x} = b$, where $\boldsymbol{a} \in \mathbb{R}^n$, $b \in \mathbb{R}$.

Question: if b is perturbed as $b + \delta_b$, the local minimizer $\boldsymbol{x}^*$ will change to $\boldsymbol{x}^* + \delta_x$. So, what δ_x should be?

Analysis:

$\because\ b + \delta_b = \boldsymbol{a}^\top(\boldsymbol{x}^* + \delta_x) = \boldsymbol{a}^\top \boldsymbol{x}^* + \boldsymbol{a}^\top \delta_x = b + \boldsymbol{a}^\top \delta_x.$

$\therefore\ \boldsymbol{a}^\top \delta_x = \delta_b.$

By the optimality condition $\nabla f(\boldsymbol{x}^*) + \lambda^* \boldsymbol{a} = 0$,

$$\therefore \delta_f = f(\boldsymbol{x}^* + \delta_x) - f(\boldsymbol{x}^*) = \nabla f(\boldsymbol{x}^*)^\top \delta_x + o(\|\delta_x\|)$$
$$= -\lambda^* \boldsymbol{a}^\top \delta_x + o(\|\delta_x\|) = -\lambda^* \delta_b + o(\|\delta_x\|).$$

$\therefore$ up to the 1st order term: $\lambda^* = -\frac{\delta_f}{\delta_b}$.

★ For multiple constraints $\boldsymbol{a}_i^\top \boldsymbol{x} = b_i$, $(i = 1, \ldots, m)$, we have

$$\delta_f = -\sum_{i=1}^{m} \lambda_i^* \delta_{b_i} + o(\|\delta_x\|).$$

Sensitivity Theorem

linear equality constraint optimization

$$\min \ f(\boldsymbol{x}), \quad \text{s.t.} \ \ \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{u}, \quad \text{where} \ \ \boldsymbol{u} \in \mathbb{R}^m \ \ \text{is perturbation.} \tag{3}$$

When $\boldsymbol{u} = 0$, let $\boldsymbol{x}^*$ be the local minimizer and $\boldsymbol{\lambda}^*$ be the associated Lagrange multiplier.

Theorem

Assume that $\boldsymbol{x}^*$ is regular and Lagrange multiplier $\boldsymbol{\lambda}^*$ is unique. Then, there exists a $\gamma > 0$ such that:

- (3) *has KKT solution/pair* $(\boldsymbol{x}(\boldsymbol{u}), \boldsymbol{\lambda}(\boldsymbol{u}))$, $\forall \boldsymbol{u} \in \mathcal{B}(0, \gamma)$.
- $\boldsymbol{x}(\cdot)$ *and* $\boldsymbol{\lambda}(\cdot)$ *are continuously differentiable in* $\mathcal{B}(0, \gamma)$.
- $\boldsymbol{x}(0) = \boldsymbol{x}^*$, $\boldsymbol{\lambda}(0) = \boldsymbol{\lambda}^*$, *and* $\nabla p(\boldsymbol{u}) = -\boldsymbol{\lambda}(\boldsymbol{u})$ *for all* $\boldsymbol{u} \in \mathcal{B}(0, \gamma)$, *where* $p(\boldsymbol{u})$ *is the primal function* $p(\boldsymbol{u}) := f(\boldsymbol{x}(\boldsymbol{u})) = \min\{f(\boldsymbol{x}) \mid \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{u}\}$.

proof. By applying implicit function theorem to the system

$$\nabla f(\boldsymbol{x}) + \nabla \boldsymbol{h}(\boldsymbol{x})^\top \boldsymbol{\lambda} = 0, \quad \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{u}. \tag{4}$$

- When $\boldsymbol{u} = 0$, system (4) has solution $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$, and its Jacobian is
$$\boldsymbol{J}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \begin{bmatrix} \nabla^2 f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h_i(\boldsymbol{x}) & \nabla \boldsymbol{h}(\boldsymbol{x}^*) \\ \nabla \boldsymbol{h}(\boldsymbol{x}^*)^\top & 0 \end{bmatrix} \in \mathbb{R}^{(n+m)\times(n+m)}.$$
By using the sufficiency conditions, it can be shown $\boldsymbol{J}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ is nonsingular. $\therefore\ \forall \boldsymbol{u} \in \mathcal{B}(0, \gamma)$, $\exists(\boldsymbol{x}(\boldsymbol{u}), \boldsymbol{\lambda}(\boldsymbol{u}))$ such that $(\boldsymbol{x}(0), \boldsymbol{\lambda}(0)) = (\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$. The mapping $\boldsymbol{x}(\cdot)$ and $\boldsymbol{\lambda}(\cdot)$ are continuously differentiable, and
$$\nabla f(\boldsymbol{x}(\boldsymbol{u})) + \nabla \boldsymbol{h}(\boldsymbol{x}(\boldsymbol{u}))\boldsymbol{\lambda}(\boldsymbol{u}) = 0, \quad \boldsymbol{h}(\boldsymbol{x}(\boldsymbol{u})) = \boldsymbol{u}.$$
- When $\boldsymbol{u} \to 0$, by sufficient conditions, $(\boldsymbol{x}(\boldsymbol{u}), \boldsymbol{\lambda}(\boldsymbol{u}))$ is a KKT solution of (3).
- To drive $\nabla p(\boldsymbol{u})$, by differentiating $\boldsymbol{h}(\boldsymbol{x}(\boldsymbol{u})) = \boldsymbol{u}$ in $\boldsymbol{u}$, we obtain
$$\nabla \boldsymbol{x}(\boldsymbol{u}) \nabla \boldsymbol{h}(\boldsymbol{x}(\boldsymbol{u})) = \boldsymbol{I}.$$
By combining the relations $\nabla \boldsymbol{x}(\boldsymbol{u}) \nabla f(\boldsymbol{x}(\boldsymbol{u})) + \nabla \boldsymbol{x}(\boldsymbol{u}) \nabla \boldsymbol{h}(\boldsymbol{x}(\boldsymbol{u}))\lambda(\boldsymbol{u}) = 0$.
$\therefore\ \nabla p(\boldsymbol{u}) = \nabla_{\boldsymbol{u}} \{f(\boldsymbol{x}(\boldsymbol{u}))\} = \nabla \boldsymbol{x}(\boldsymbol{u}) \nabla f(\boldsymbol{x}(\boldsymbol{u}))$.

Sensitivity Theorem

$\min\ f(\boldsymbol{x}) = \frac{1}{2}(x_1^2 - x_2^2) - x_2, \quad$ s.t. $h(\boldsymbol{x}) = x_2 = 0.$

Primal function is give by

$$p(u) = \min_{h(\boldsymbol{x})=u} f(\boldsymbol{x}) = -\frac{1}{2}u^2 - u \ \text{ and } \ \lambda^* = -\nabla p(0) = 1.$$

The result is consistent with the sensitivity theorem

★ Need for regularity of $\boldsymbol{x}^*$.

if change constraint to $h(\boldsymbol{x}) = x_2^2 = 0$. Then $p(u) = \begin{cases} -\frac{u}{2} - \sqrt{u}, & \text{if } u \geq 0 \\ +\infty, & \text{if } u < 0 \end{cases}$

Homework

Exercise in textbook: 20.3, 20.8, 20.10, 20.15

Why is convexity so special?

- convex function has no local minima that are not global
- convex set has a nonempty relative interior
- convex set is connected and has feasible directions at any point
- nonconvex function can be "convexified" while maintaining the optimality of its global minima
- The existence of a global minimum of a convex function over a convex set is conveniently characterized in terms of directions of recession
- polyhedral convex set is characterized in terms of a finite set of extreme points and extreme directions
- convex function is continuous and has nice differentiability properties
- closed convex cones are self-dual with respect to polarity Convex, lower semicontinuous functions are self-dual with respect to conjugacy

Theorem

Let $f : \Omega \to \mathbb{R}$*. If* Ω *is convex set and* f *is convex function, then:*

- *A local minimum of* f *over* Ω *is also a global minimum of* f *over* Ω*.*
- *If* f *is strictly convex, then there exists at most one global minimum of f over* Ω*.*

proof. handwriting