General constrained problems (§19.1)

- General problem with functional constraints:

$$\begin{array}{lll} \text{minimize} & f(\boldsymbol{x}) & \\ \text{subject to} & h_i(\boldsymbol{x}) = 0, & i = 1, \ldots, m \\ & g_j(\boldsymbol{x}) \leq 0, & j = 1, \ldots, p, \end{array}$$

  where $f : \mathbb{R}^n \to \mathbb{R}$, $h_i : \mathbb{R}^n \to \mathbb{R}$, $g_j : \mathbb{R}^n \to \mathbb{R}$, and $m \leq n$.

- LP problem is an example of such a problem.
- We will develop techniques for solving the above problems (similar to FONC, SONC, SOSC).

Example (19.1):

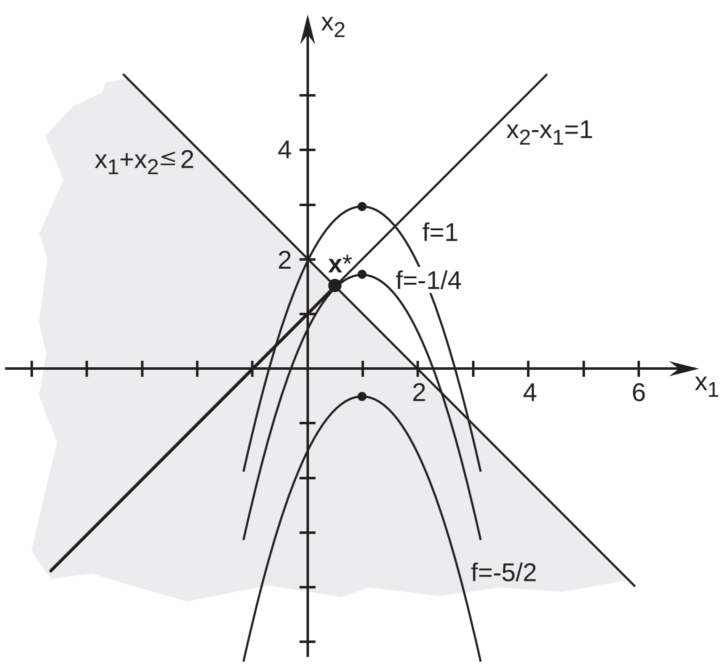
- Consider the problem

$$\begin{array}{ll} \text{minimize} & (x_1 - 1)^2 + x_2 - 2 \\ \text{subject to} & x_2 - x_1 = 1, \\ & x_1 + x_2 \leq 2. \end{array}$$

- The constraint (feasible) set is

$$S = \{\boldsymbol{x} \in \mathbb{R}^2 : x_2 - x_1 = 1,\ x_1 + x_2 \leq 2\}.$$

- We can solve this problem graphically.
- In general, the graphical approach will not suffice. We need more powerful tools.



Problems with equality constraints (§19.2)

- We now focus on problems with only equality constraints:

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & h_i(\boldsymbol{x}) = 0, \qquad i = 1, \ldots, m. \end{array}$$

- Writing $\boldsymbol{h} = [h_1, \ldots, h_m]^T$, we can use vector notation:

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \end{array}$$

  where $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, $m \leq n$.

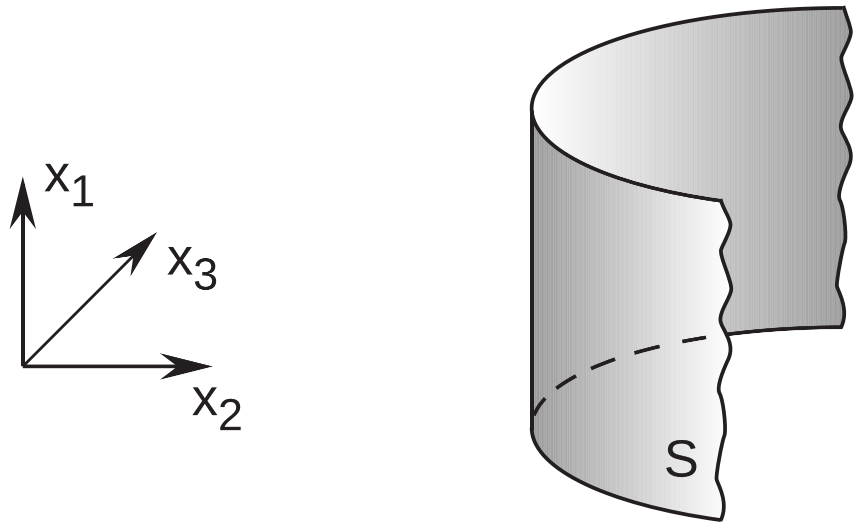
- We always assume that $f, \boldsymbol{h} \in \mathcal{C}^1$.
- For simplicity, we first consider the case where $m = 1$. The constraint is $h(\boldsymbol{x}) = 0$ (scalar).
- Definition ($m = 1$ case): A feasible point $\boldsymbol{x}^*$ is said to be *regular* if $\nabla h(\boldsymbol{x}^*) \neq \boldsymbol{0}$.
- Geometrically, if all points in the constraint set S are regular, then the dimension of the surface S is $n - 1$.

Example (19.2):

- Consider the constraint set

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

- Here, $n = 3$ and $m = 1$.
- We have $\nabla h_1(\boldsymbol{x}) = [0, 1, -2x_3]^T$, which is nonzero everywhere. Hence, any point in S is regular.
- The dimension of S is $3 - 1 = 2$.



$$S=\{[x_1,x_2,x_3]^T : x_2-x_3^2=0\}$$

Lagrange conditions (§19.4)

- We now give a FONC type necessary condition for problems with equality constraints.
- First consider the simple case where $m = 1$:

$$\begin{array}{rl} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & h(\boldsymbol{x}) = 0, \end{array}$$

  where $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}$.

- Lagrange's Theorem ($m = 1$ case): Suppose $\boldsymbol{x}^*$ is a local minimizer and is regular. Then, there exists a scalar λ^* such that

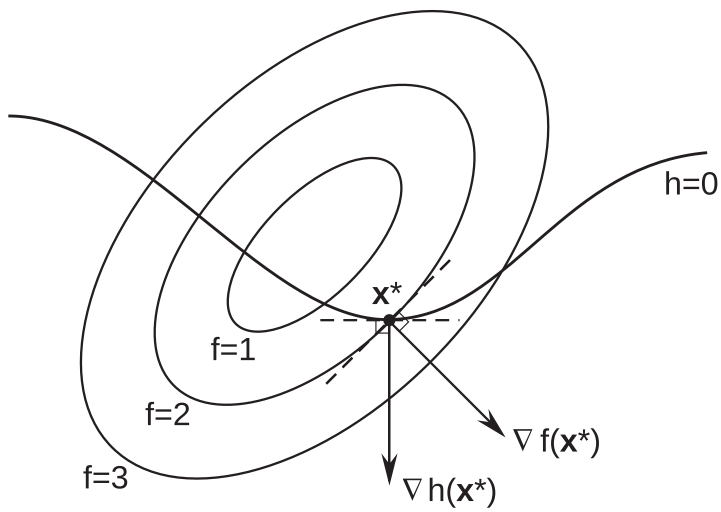
$$\nabla f(\boldsymbol{x}^*) + \lambda^* \nabla h(\boldsymbol{x}^*) = \mathbf{0}.$$

- In other words, $\nabla f(\boldsymbol{x}^*)$ and $\nabla h(\boldsymbol{x}^*)$ are parallel; i.e., $\nabla f(\boldsymbol{x}^*)$ is normal to S at $\boldsymbol{x}^*$.
- λ^* is called the *Lagrange multiplier*.

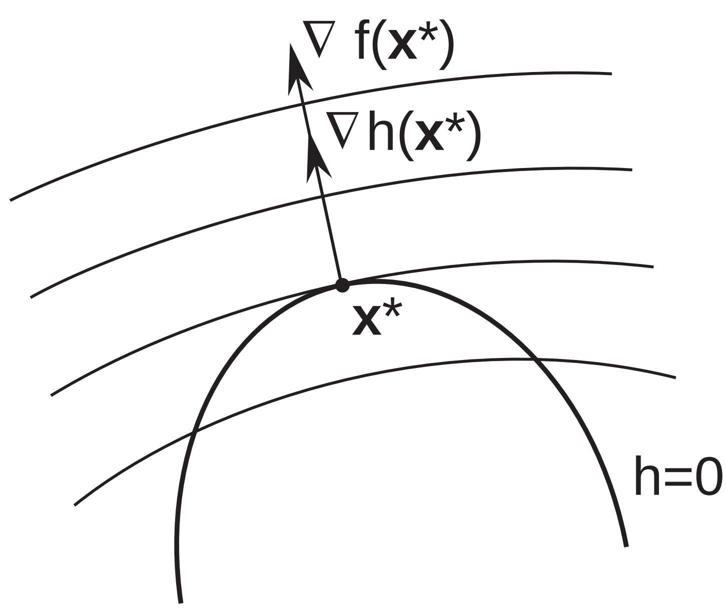
Idea of proof of theorem:

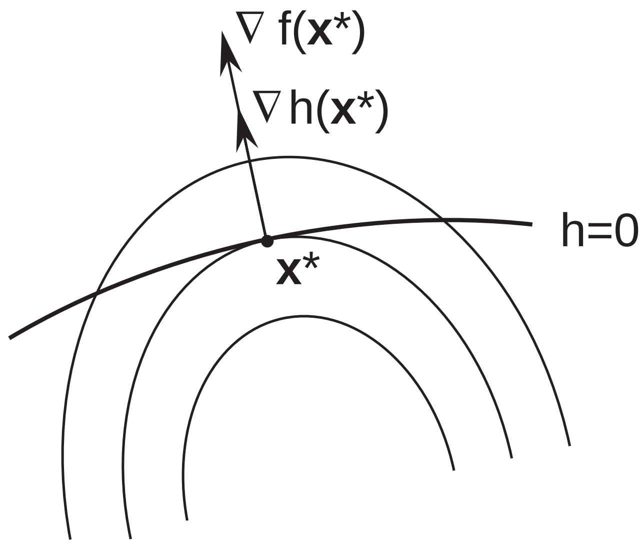
- Note that $\nabla f(\boldsymbol{x}^*)$ is orthogonal to the level set of f.
- Also, $\nabla h(\boldsymbol{x}^*)$ is orthogonal to the constraint set S (it is normal to S).

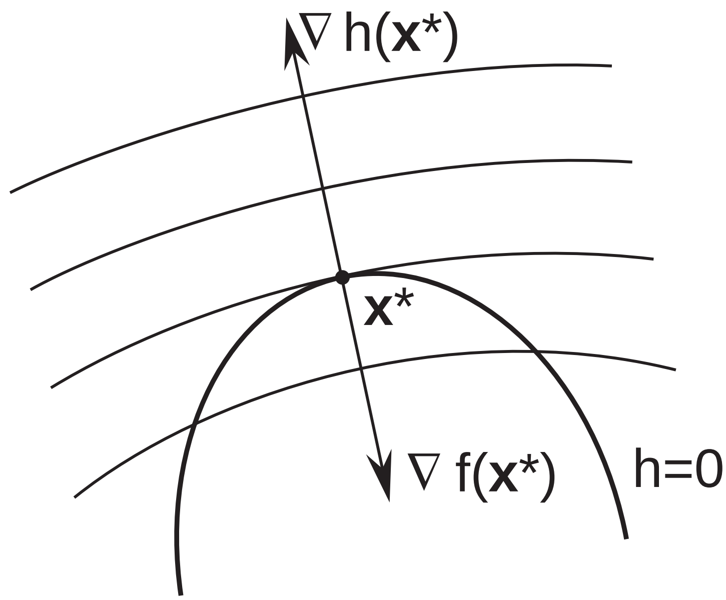
- If $\nabla f(\boldsymbol{x}^*)$ and $\nabla h(\boldsymbol{x}^*)$ were not parallel, then we can move in a direction along S in the opposite direction to $\nabla f(\boldsymbol{x}^*)$, and the objective function decreases.

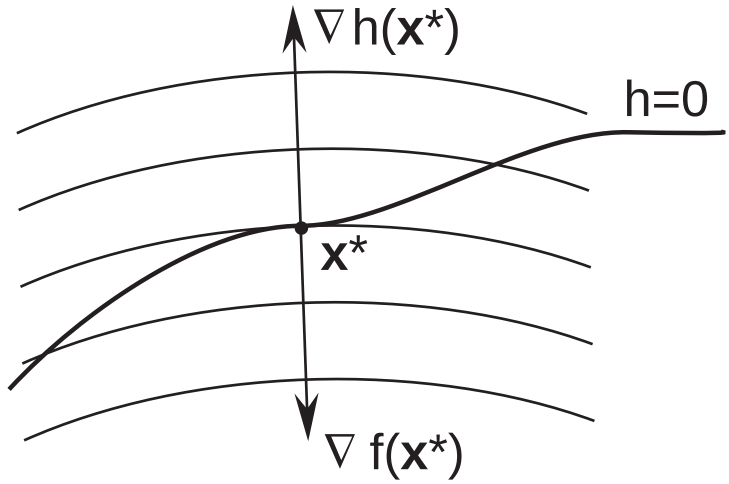


- Note that the Lagrange condition is only a necessary condition, not sufficient in general.
- Since it is only a first order condition, both minimizers and maximizers satisfy it.
- There may also be points that are neither minimizers nor maximizers that satisfy it.




∇ f(x*)
∇ h(x*)
h=0
x*


∇ h(x*)
x*
∇ f(x*)
h=0



- To apply the Lagrange theorem, it is convenient to define the Lagrangian function

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

  Note that l is a function from $\mathbb{R}^{n+1}$ to $\mathbb{R}$.

- Then, the condition in Lagrange's Theorem is equivalent to the FONC for l; i.e., we just take $\nabla l(\boldsymbol{x}^*, \lambda^*) = \mathbf{0}$.

- To see this, note that

$$\nabla l(\boldsymbol{x}, \lambda) = \begin{bmatrix} \nabla_x l(\boldsymbol{x}, \lambda) \\ \nabla_\lambda l(\boldsymbol{x}, \lambda) \end{bmatrix},$$

  where $\nabla_\lambda \equiv \frac{\partial}{\partial \lambda}$.

- We have

$$\begin{aligned} \nabla_x l(\boldsymbol{x}, \lambda) &= \nabla f(\boldsymbol{x}) + \lambda \nabla h(\boldsymbol{x}) \\ \nabla_\lambda l(\boldsymbol{x}, \lambda) &= h(\boldsymbol{x}). \end{aligned}$$

- Therefore, the condition $\nabla l(\boldsymbol{x}^*, \lambda^*) = \mathbf{0}$ is equivalent to the two conditions

$$\begin{aligned} \nabla f(\boldsymbol{x}^*) + \lambda^* \nabla h(\boldsymbol{x}^*) &= \mathbf{0} \\ h(\boldsymbol{x}^*) &= 0. \end{aligned}$$

  We call the above the *Lagrange conditions*.

Example (19.5):

- Consider the optimization problem where

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

and

$$h(\boldsymbol{x}) = x_1^2 + 2x_2^2 - 1.$$

- The level sets are circles, and the constraint set is an ellipse.
- We have $\nabla f(\boldsymbol{x}) = [2x_1, 2x_2]^T$, $\nabla h(\boldsymbol{x}) = [2x_1, 4x_2]^T$.
- Note that all feasible points are regular.

- To solve the problem, we first write down the Lagrange conditions:

$$\begin{aligned} 2x_1 + 2\lambda x_1 &= 0 \\ 2x_2 + 4\lambda x_2 &= 0 \\ x_1^2 + 2x_2^2 &= 1. \end{aligned}$$

- We find that there are four points that satisfy the Lagrange conditions:

$$\begin{aligned} \lambda_1^* = -\frac{1}{2} : \quad & \boldsymbol{x}^{*(1)} = \begin{bmatrix} 0 \\ 1/\sqrt{2} \end{bmatrix}, \qquad \boldsymbol{x}^{*(2)} = \begin{bmatrix} 0 \\ -1/\sqrt{2} \end{bmatrix}, \\ \lambda_2^* = -1 : \quad & \boldsymbol{x}^{*(3)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{x}^{*(4)} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}. \end{aligned}$$

- All four points are regular.

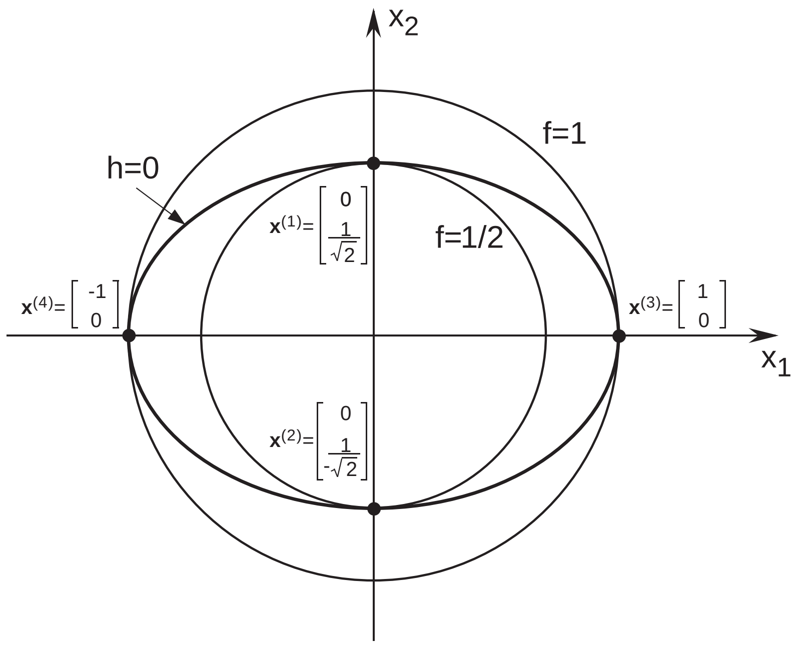
- Note that

$$f(\boldsymbol{x}^{*(1)}) = f(\boldsymbol{x}^{*(2)}) = \frac{1}{2}$$

and

$$f(\boldsymbol{x}^{*(3)}) = f(\boldsymbol{x}^{*(4)}) = 1$$

- Thus, if there are minimizers, then they are located at $\boldsymbol{x}^{*(1)}$ and $\boldsymbol{x}^{*(2)}$, and if there are maximizers, then they are located at $\boldsymbol{x}^{*(3)}$ and $\boldsymbol{x}^{*(4)}$.



Example

- Given a fixed area of cardboard (2 sq.ft.), find the dimensions of the (closed) box that has maximum volume.
- Denote the dimensions by x_1, x_2, and x_3.
- The problem is then

$$\begin{array}{ll} \text{maximize} & x_1x_2x_3 \\ \text{subject to} & x_1x_2 + x_2x_3 + x_3x_1 = 1. \end{array}$$

- Denote

$$\begin{aligned} f(\boldsymbol{x}) &= -x_1x_2x_3, \\ h(\boldsymbol{x}) &= x_1x_2 + x_2x_3 + x_3x_1 - 1. \end{aligned}$$

- We have

$$\nabla f(\boldsymbol{x}) = -\begin{bmatrix} x_2x_3 \\ x_1x_3 \\ x_1x_2 \end{bmatrix}, \qquad \nabla h(\boldsymbol{x}) = \begin{bmatrix} x_2 + x_3 \\ x_1 + x_3 \\ x_1 + x_2 \end{bmatrix}.$$

- By the Lagrange conditions, the dimensions of the box with maximum volume satisfies

$$\begin{aligned} x_2x_3 - \lambda(x_2 + x_3) &= 0 \\ x_1x_3 - \lambda(x_1 + x_3) &= 0 \\ x_1x_2 - \lambda(x_1 + x_2) &= 0 \\ x_1x_2 + x_2x_3 + x_3x_1 - 1 &= 0, \end{aligned}$$

  where $\lambda \in \mathbb{R}$.

- Claim: the quantities above are all nonzero. (Why?)

- To solve the Lagrange conditions, first, multiply the first equation by x_1 and the second by x_2, and subtract one from the other.
- We get $x_3\lambda(x_1 - x_2) = 0$, which implies that $x_1 = x_2$ (because $x_3, \lambda \neq 0$).
- We similarly deduce that $x_2 = x_3$.
- Hence, from the constraint equation, we deduce that

$$x_1^* = x_2^* = x_3^* = \frac{1}{\sqrt{3}}.$$

General Lagrange theorem

- We now consider the general m case. The constraint is $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}$ ($\mathbb{R}^m$ vector).
- Definition: A feasible point $\boldsymbol{x}^*$ is said to be *regular* if $D\boldsymbol{h}(\boldsymbol{x}^*)$ is of full rank (rank m).
- The definition for $m = 1$ is a special case of the above.
- Regular means that the $\nabla h_i(\boldsymbol{x}^*)$, $i = 1, \ldots, m$, are linearly independent.
- Geometrically, if all points in the constraint set S are regular, then the dimension of S is $n - m$.

Example (19.3):

- Consider the constraint $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}$, where

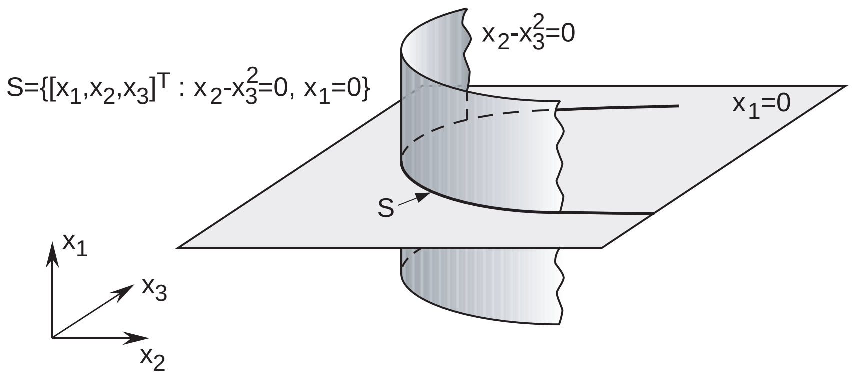
$$\boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} x_2 - x_3^2 \\ x_1 \end{bmatrix}.$$

- Here, $n = 3$ and $m = 2$.
- We have

$$D\boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} 0 & 1 & -2x_3 \\ 1 & 0 & 0 \end{bmatrix},$$

  which has rank 2 everywhere. Hence, any point in S is regular.

- The dimension of S is $3 - 2 = 1$.



Tangent space (§19.3)

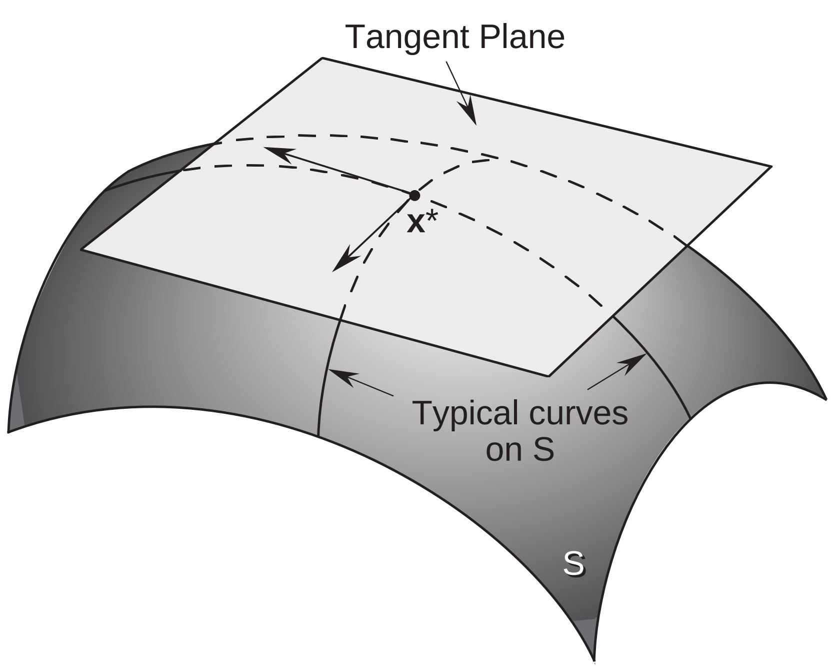
- The *tangent space* at a point $\boldsymbol{x}^*$ on the surface $S = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$ is the set

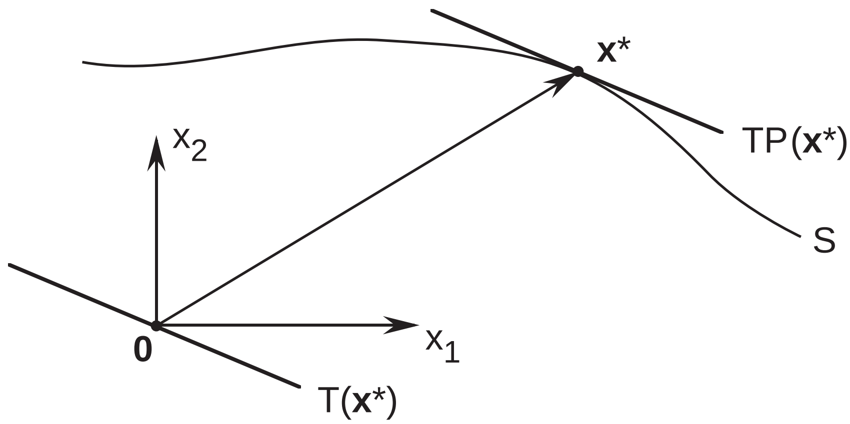
$$T(\boldsymbol{x}^*) = \{\boldsymbol{y} : D\boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{y} = \boldsymbol{0}\}.$$

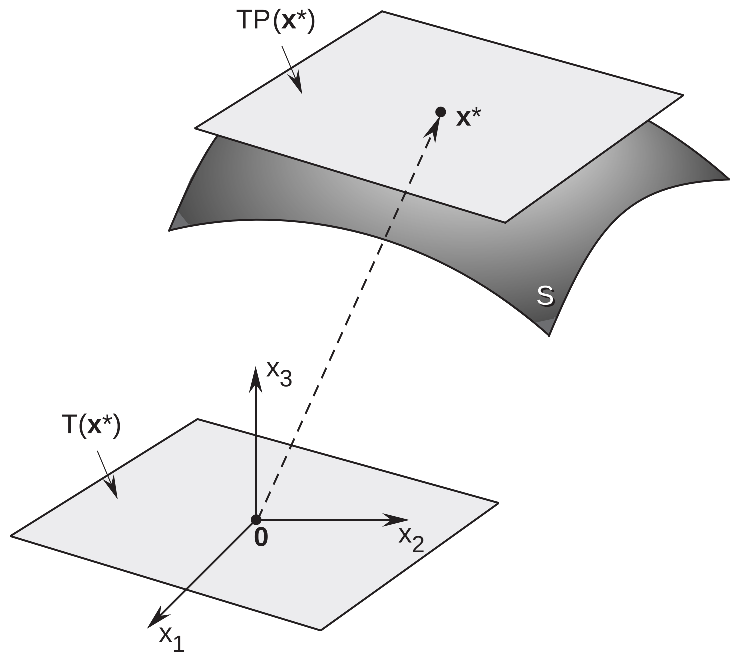
- Note that the tangent space $T(\boldsymbol{x}^*)$ is the nullspace of the matrix $D\boldsymbol{h}(\boldsymbol{x}^*)$; i.e.,

$$T(\boldsymbol{x}^*) = \mathcal{N}(D\boldsymbol{h}(\boldsymbol{x}^*)).$$

- The tangent space is a subspace of $\mathbb{R}^n$ (plane passing through the origin).
- The dimension of $T(\boldsymbol{x}^*)$ is $n - m$ (assuming regularity of $\boldsymbol{x}^*$).
- Geometric view: if we assume that $\boldsymbol{x}^*$ is regular, and we shift $T(\boldsymbol{x}^*)$ so that it touches $\boldsymbol{x}^*$, then the resulting plane is *tangent* to S at $\boldsymbol{x}^*$. We call this plane the *tangent plane*.







Normal space (§19.3)

- The *normal space* $N(\boldsymbol{x}^*)$ at a point $\boldsymbol{x}^*$ on the surface $S = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$ is the set

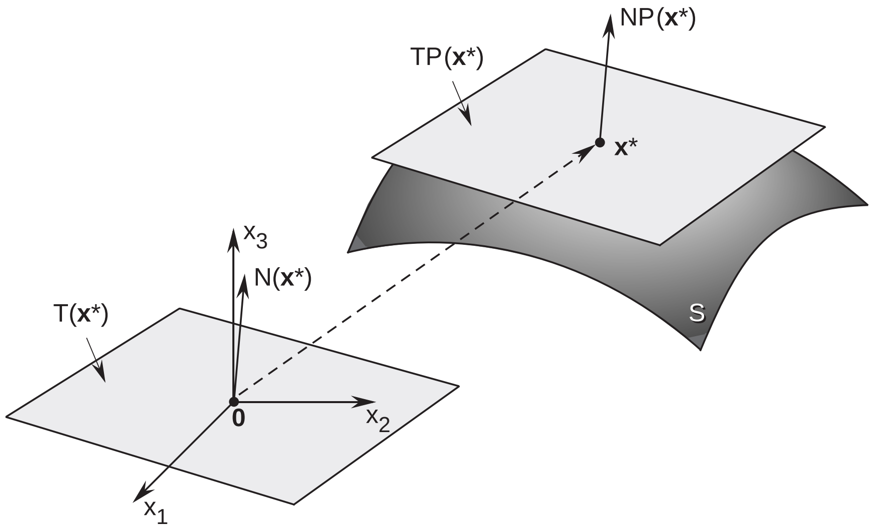
$$N(\boldsymbol{x}^*) = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} = D\boldsymbol{h}(\boldsymbol{x}^*)^T \boldsymbol{z},\ \boldsymbol{z} \in \mathbb{R}^m\}.$$

- Note that the normal space is the range of the matrix $D\boldsymbol{h}(\boldsymbol{x}^*)^T$; i.e.,

$$N(\boldsymbol{x}^*) = \mathcal{R}\left(D\boldsymbol{h}(\boldsymbol{x}^*)^T\right).$$

- The normal space is a subspace. Its dimension is m (assuming regularity of $\boldsymbol{x}^*$).

- Geometric view: if we assume that $\boldsymbol{x}^*$ is regular, and we shift $N(\boldsymbol{x}^*)$ so that it touches $\boldsymbol{x}^*$, then the resulting space is *normal* to S at $\boldsymbol{x}^*$.



- Lagrange Multiplier Theorem (19.3): Suppose $\boldsymbol{x}^*$ is a local minimizer and is regular. Then, there exists $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that

$$Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*T} D\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}^T.$$

- In other words, $\nabla f(\boldsymbol{x}^*) \in \mathcal{R}(D\boldsymbol{h}(\boldsymbol{x}^*)^T)$ (the normal space to S at $\boldsymbol{x}^*$).

- $\boldsymbol{\lambda}^*$ is called the *Lagrange multiplier vector*.

- As before, it is convenient to define the Lagrangian function

$$l(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^T \boldsymbol{h}(\boldsymbol{x}).$$

  Note that l is a function from $\mathbb{R}^{n+m}$ to $\mathbb{R}$.

- Then, the Lagrange condition is equivalent to the FONC for l; i.e., we just take $\nabla l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{0}$.

- The condition $\nabla l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{0}$ is equivalent to

$$\begin{aligned} Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*T} D\boldsymbol{h}(\boldsymbol{x}^*) &= \boldsymbol{0}^T \\ \boldsymbol{h}(\boldsymbol{x}^*) &= \boldsymbol{0}. \end{aligned}$$

Example:

- Consider the problem

$$\begin{array}{rl} \text{minimize} & (1-x_1)^3 - (x_2+1)^2 + 3x_3^2 \\ \text{subject to} & x_1 = 0 \\ & x_2 = x_3^2. \end{array}$$

- Write

$$\begin{aligned} f(\boldsymbol{x}) &= (1-x_1)^3-(x_2+1)^2+3x_3^2, \\ \boldsymbol{h}(\boldsymbol{x}) &= \begin{bmatrix} x_1 \\ x_2-x_3^2 \end{bmatrix}. \end{aligned}$$

- We have

$$\begin{aligned} Df(\boldsymbol{x}) &= [-3(1-x_1)^2, -2(x_2+1), 6x_3], \\ D\boldsymbol{h}(\boldsymbol{x}) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2x_3 \end{bmatrix}. \end{aligned}$$

  Note that all feasible points are regular.

- Introduce the vector $[\lambda_1, \lambda_2]^T$. The Lagrange conditions are:

$$\begin{aligned} -3(1-x_1)^2+\lambda_1 &= 0 \\ -2(x_2+1)+\lambda_2 &= 0 \\ 6x_3-2\lambda_2 x_3 &= 0 \\ x_1 &= 0 \\ x_2-x_3^2 &= 0. \end{aligned}$$

- We immediately deduce that $x_1=0$, $\lambda_1=3$.
- The third equation implies that either $x_3=0$ or $\lambda_2=3$.
- Hence, there are three solutions:

$$\boldsymbol{x}^{*(1)}=\mathbf{0}, \text{ with } \boldsymbol{\lambda}^{*(1)}=\begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$$\boldsymbol{x}^{*(2)}=\begin{bmatrix} 0 \\ 1/2 \\ 1/\sqrt{2} \end{bmatrix}, \boldsymbol{x}^{*(3)}=\begin{bmatrix} 0 \\ 1/2 \\ -1/\sqrt{2} \end{bmatrix}, \text{ with } \boldsymbol{\lambda}^{*(2)}=\begin{bmatrix} 3 \\ 3 \end{bmatrix}.$$

Another example: Quadratic programming problem

- Consider the special optimization problem:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}, \end{aligned}$$

  where $\boldsymbol{Q}>0$, $\boldsymbol{A}\in\mathbb{R}^{m\times n}$, $m<n$, rank $\boldsymbol{A}=m$.

- Name: *Quadratic programming* (QP) problem.
  (Actually, ours is a special case; the general case includes linear *inequality* constraints).

- We have $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}$, $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}$.

- We have
$$Df(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{Q}, \qquad D\boldsymbol{h}(\boldsymbol{x}) = -\boldsymbol{A}.$$

- The Lagrange conditions are
$$\begin{aligned} \boldsymbol{x}^{*T}\boldsymbol{Q} - \boldsymbol{\lambda}^{*T}\boldsymbol{A} &= \boldsymbol{0}^T \\ \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^* &= \boldsymbol{0}. \end{aligned}$$

- From the first equation we get
$$\boldsymbol{x}^* = \boldsymbol{Q}^{-1}\boldsymbol{A}^T\boldsymbol{\lambda}^*.$$

- Multiplying both sides by $\boldsymbol{A}$ and using the second equation (constraint), we get
$$\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T\boldsymbol{\lambda}^* = \boldsymbol{b}.$$

- Since $\boldsymbol{Q} > 0$ and $\boldsymbol{A}$ is of full rank, we can write
$$\boldsymbol{\lambda}^* = (\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T)^{-1}\boldsymbol{b}.$$

- Hence,
$$\boldsymbol{x}^* = \boldsymbol{Q}^{-1}\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T)^{-1}\boldsymbol{b}.$$

Example: Simple optimal control

- Consider the discrete-time system model
$$x_k = ax_{k-1} + bu_k, \qquad x_0 \text{ given}$$

- Think of $\{x_k\}$ as a discrete-time signal that is controlled by an external input signal $\{u_k\}$.
- Given x_0, we wish to drive x_k to as small a value as possible without too much control "effort", over a time interval $[1, N]$.

- We can formulate the problem as:
$$\begin{array}{ll} \text{minimize} & \displaystyle\frac{1}{2}\sum_{i=1}^{N}\left(qx_i^2 + ru_i^2\right) \\ \text{subject to} & x_k = ax_{k-1} + bu_k,\ k = 1, \ldots, N. \end{array}$$

- The above problem is an instance of the *linear quadratic regulator* (LQR) problem.

- The parameters q and r reflect the relative importance of driving the signal x_k to zero versus minimizing the control effort in u_k.
- To solve the problem, we can rewrite it as a QP problem.
- Define

$$
\begin{aligned}
\boldsymbol{Q} &= \begin{bmatrix} q\boldsymbol{I}_N & \boldsymbol{O} \\ \boldsymbol{O} & r\boldsymbol{I}_N \end{bmatrix} \\
\boldsymbol{A} &= \begin{bmatrix} 1 & & \cdots & 0 & -b & & \cdots & 0 \\ -a & 1 & & \vdots & & -b & & \vdots \\ & \ddots & \ddots & & \vdots & & \ddots & \\ 0 & & -a & 1 & 0 & \cdots & & -b \end{bmatrix} \\
\boldsymbol{b} &= \begin{bmatrix} ax_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad \boldsymbol{z} = [x_1, \ldots, x_N, u_1, \ldots, u_N]^T.
\end{aligned}
$$

- With the previous definitions, the problem reduces to the previous QP problem

$$
\begin{array}{ll} \text{minimize} & \frac{1}{2}\boldsymbol{z}^T\boldsymbol{Q}\boldsymbol{z} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{z} = \boldsymbol{b}, \end{array}
$$

  where $\boldsymbol{Q}$ is $2N \times 2N$, $\boldsymbol{A}$ is $N \times 2N$, and $\boldsymbol{b} \in \mathbb{R}^N$.
- The solution is

$$
\boldsymbol{z}^* = \boldsymbol{Q}^{-1}\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T)^{-1}\boldsymbol{b}.
$$

- In practice, the computation of the above inverse may be too costly. There are other ways to tackle the problem by exploiting the special structure. This is the study of *optimal control*.

Other variations:

- Higher order systems: $x_k = a_1 x_{k-1} + a_2 x_{k-2} + bu_k$.
- Vary the q and r parameters with k:

$$
\frac{1}{2}\sum_{i=1}^{N}\left(q_i x_i^2 + r_i u_i^2\right)
$$

- Vector signals: $\boldsymbol{x}(k) = \boldsymbol{A}\boldsymbol{x}(k-1) + \boldsymbol{B}\boldsymbol{u}(k)$, with

$$
\frac{1}{2}\sum_{i=1}^{N}\left(\boldsymbol{x}(i)^T\boldsymbol{Q}_i\boldsymbol{x}(i) + \boldsymbol{u}(i)^T\boldsymbol{R}_i\boldsymbol{u}(i)\right).
$$

- Infinite horizon: $N \to \infty$.

Second order conditions (§19.5)

- We now develop a SONC and SOSC for problems with equality constraints.
- We assume that $f, \boldsymbol{h} \in \mathcal{C}^2$.
- Recall: Lagrangian function

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

- Given $\boldsymbol{\lambda}$, the Hessian of $l(\boldsymbol{x}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{x}$ is denoted

$$\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{F}(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i \boldsymbol{H}_i(\boldsymbol{x}),$$

  where $\boldsymbol{F}$ is the Hessian of f, and $\boldsymbol{H}_i$ is the Hessian of h_i, $i = 1, \ldots, m$.

- Theorem (19.4): (SONC) Suppose $\boldsymbol{x}^*$ is a local minimizer and is regular. Then, there exists $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that the Lagrange conditions hold, and

$$\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \boldsymbol{y} \geq 0 \text{ for all } \boldsymbol{y} \in T(\boldsymbol{x}^*).$$

- Vectors $\boldsymbol{y} \in T(\boldsymbol{x}^*)$ play the role of "feasible directions".
- Often, we use the phrase "$\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \geq 0$ on $T(\boldsymbol{x}^*)$" to refer to the above condition.

Side note: Definiteness on a subspace

- $\boldsymbol{M} \geq 0$ (> 0) on $\mathcal{V}$ means $\boldsymbol{y}^T \boldsymbol{M} \boldsymbol{y} \geq 0$ (> 0) for all nonzero $\boldsymbol{y} \in \mathcal{V}$.
- Even though a matrix $\boldsymbol{M}$ may not be positive semidefinite, it is possible that $\boldsymbol{M} \geq 0$ on some subspace $\mathcal{V}$ (or even $\boldsymbol{M} > 0$ on $\mathcal{V}$).
- Examples:

$$\begin{aligned} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} &> 0 \qquad \text{on } \{\boldsymbol{y} : y_2 = 0\} \\ \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} &\geq 0 \qquad \text{on } \{\boldsymbol{y} : y_1 = y_2\} \\ \boldsymbol{M} &> 0 \qquad \text{on } \{\boldsymbol{0}\} \text{ for any } \boldsymbol{M}. \end{aligned}$$

- If $\boldsymbol{M} = \boldsymbol{O}$, then it is trivially true that $\boldsymbol{M} \geq 0$ on $\mathcal{V}$, regardless of $\mathcal{V}$.
- If $\mathcal{V} = \{\boldsymbol{0}\}$, then it is trivially true that $\boldsymbol{M} \geq 0$ on $\mathcal{V}$, regardless of $\boldsymbol{M}$.

- In fact, if $\mathcal{V} = \{\mathbf{0}\}$, then $\boldsymbol{M} > 0$ on $\mathcal{V}$, regardless of $\boldsymbol{M}$.
- If $\boldsymbol{M} = \boldsymbol{O}$ but $\mathcal{V} \neq \{\mathbf{0}\}$, then $\boldsymbol{M} \not> 0$ on $\mathcal{V}$.

Example of SONC:

- Recall the example where

$$\begin{aligned} f(\boldsymbol{x}) &= (1-x_1)^3 - (x_2+1)^2 + 3x_3^2, \\ \boldsymbol{h}(\boldsymbol{x}) &= \begin{bmatrix} x_1 \\ x_2 - x_3^2 \end{bmatrix}. \end{aligned}$$

- We have

$$l(\boldsymbol{x}, \boldsymbol{\lambda}) = (1-x_1)^3 - (x_2+1)^2 + 3x_3^2 + \lambda_1 x_1 + \lambda_2(x_2 - x_3^2).$$

- The Hessian (w.r.t. $\boldsymbol{x}$) of the Lagrangian is

$$\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \begin{bmatrix} 6(1-x_1) & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 2(3-\lambda_2) \end{bmatrix}.$$

- Recall that there are three solutions (all regular):

$$\boldsymbol{x}^{*(1)} = \mathbf{0}, \text{ with } \boldsymbol{\lambda}^{*(1)} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$$\boldsymbol{x}^{*(2)} = \begin{bmatrix} 0 \\ 1/2 \\ 1/\sqrt{2} \end{bmatrix}, \boldsymbol{x}^{*(3)} = \begin{bmatrix} 0 \\ 1/2 \\ -1/\sqrt{2} \end{bmatrix}, \text{ with } \boldsymbol{\lambda}^{*(2)} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}.$$

- We will check if each of these three solutions satisfies the SONC.
- Recall that

$$D\boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2x_3 \end{bmatrix}.$$

- Consider the first solution $\boldsymbol{x}^{*(1)} = \mathbf{0}$ with $\boldsymbol{\lambda}_2^{*(1)} = 2$. In this case,

$$\boldsymbol{L}(\boldsymbol{x}^{*(1)}, \boldsymbol{\lambda}^{*(1)}) = \begin{bmatrix} 6 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

- The tangent space is $T(\boldsymbol{x}^{*(1)}) = \mathcal{N}(D\boldsymbol{h}(\boldsymbol{x}^{*(1)}))$, where

$$D\boldsymbol{h}(\boldsymbol{x}^{*(1)}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

- Hence, $T(\boldsymbol{x}^{*(1)}) = \{\boldsymbol{y} \in \mathbb{R}^3 : y_1 = y_2 = 0\}$, i.e., the x_3-axis (makes sense, geometrically).
- Hence, the SONC holds at $\boldsymbol{x}^{*(1)} = \boldsymbol{0}$.

- Consider now the second solution $\boldsymbol{x}^{*(2)} = [0, 1/2, 1/\sqrt{2}]^T$ with $\boldsymbol{\lambda}_2^{*(2)} = [3, 3]^T$. In this case,

$$\boldsymbol{L}(\boldsymbol{x}^{*(2)}, \boldsymbol{\lambda}^{*(2)}) = \begin{bmatrix} 6 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

- The tangent space is $T(\boldsymbol{x}^{*(2)}) = \mathcal{N}(D\boldsymbol{h}(\boldsymbol{x}^{*(2)}))$, where

$$D\boldsymbol{h}(\boldsymbol{x}^{*(2)}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -\sqrt{2} \end{bmatrix}.$$

- Hence, $T(\boldsymbol{x}^{*(2)}) = \{\boldsymbol{y} : y_1 = 0, y_2 = y_3\sqrt{2}\}$, (does this makes sense, geometrically?).

- In this case, we see that the SONC does not hold. Indeed, consider $\boldsymbol{y} = [0, \sqrt{2}, 1]^T \in T(\boldsymbol{x}^{*(2)})$. We have

$$\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}^{*(2)}, \boldsymbol{\lambda}^{*(2)})\boldsymbol{y} = -4 \not\geq 0.$$

- The same argument can be made about the third point, $\boldsymbol{x}^{*(3)} = [0, 1/2, -1/\sqrt{2}]^T$ with $\boldsymbol{\lambda}_2^{*(2)} = [3, 3]^T$.
- Hence, $\boldsymbol{x}^{*(2)}$ and $\boldsymbol{x}^{*(3)}$ are not local minimizers.

Example: QP problem

- Consider the QP problem. We have

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

- We have

$$\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{Q} > 0.$$

- Hence, the SONC holds at the solution point we found previously.

- What about sufficient conditions?

- Theorem (19.5): (SOSC) Suppose $\boldsymbol{x}^*$ (feasible) and $\boldsymbol{\lambda}^*$ satisfy
  1. $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*T} D\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}^T$; and
  2. $\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)\boldsymbol{y} > 0$ for all nonzero $\boldsymbol{y} \in T(\boldsymbol{x}^*)$.

Then, $\boldsymbol{x}^*$ is a strict local minimizer.

- Often, we say "$\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) > 0$ on $T(\boldsymbol{x}^*)$" to refer to condition 2 above.

Example

- Recall the example where

$$\begin{aligned} f(\boldsymbol{x}) &= (1-x_1)^3 - (x_2+1)^2 + 3x_3^2, \\ \boldsymbol{h}(\boldsymbol{x}) &= \begin{bmatrix} x_1 \\ x_2 - x_3^2 \end{bmatrix}. \end{aligned}$$

- The only point satisfying the SONC is $\boldsymbol{x}^* = \boldsymbol{0}$, $\boldsymbol{\lambda}^* = [3,2]^T$.

- The Hessian of the Lagrangian at $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ is

$$\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \begin{bmatrix} 6 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

- Consider $\boldsymbol{y} \in T(\boldsymbol{x}^*) = \{\boldsymbol{y} : y_1 = y_2 = 0\}$, $\boldsymbol{y} \neq \boldsymbol{0}$.

- Then, $y_3 \neq 0$. Hence,

$$\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)\boldsymbol{y} = 2y_3^2 > 0.$$

- Hence, by the SOSC, $\boldsymbol{x}^*$ is a strict local minimizer.

Example: QP problem

- Recall that in the QP problem,

$$\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{Q} > 0.$$

- Hence, by the SOSC, the solution $\boldsymbol{x}^* = \boldsymbol{Q}^{-1}\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T)^{-1}\boldsymbol{b}$ is a strict local minimizer.

- In fact, we shall see later that it is a global minimizer.

