Min \( f(x) \)

\[ \begin{align*}
\text{s.t. } & h(x) = 0 \quad \text{m equality constraints} \\
& g(x) \leq 0 \quad \text{k inequality constraints}
\end{align*} \]

Theorem: Consider a regular point such that \( h(x^*) = 0 \) and \( g(x^*) \leq 0 \).

- First order necessary condition (KKT):
  If \( x^* \) is a local solution, then there exist constants \( \lambda_1, ..., \lambda_m \) and \( \mu_1, ..., \mu_k \) such that:
  \[
  \begin{align*}
  \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{i=1}^{k} \mu_i \nabla g_i(x^*) &= 0 \\
  \mu_i g_i(x^*) &= 0 \quad i = 1, ..., k \\
  \mu_i &\geq 0 \quad i = 1, ..., k
  \end{align*}
  \]

- Second order necessary condition:
  If \( x^* \) is a local solution, then
  \[
  \Delta x^T \left( \nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla^2 h_i(x^*) + \sum_{i=1}^{k} \mu_i \nabla^2 g_i(x^*) \right) \Delta x \geq 0
  \]
  for every \( \Delta x \) that is orthogonal to the gradients of all active constraints at \( x^* \).

- Second order sufficiency condition:
  \( x^* \) is a local solution if there exists \( \lambda_1, ..., \lambda_m \) and \( \mu_1, ..., \mu_k \) such that:
1. is satisfied

2. is satisfied as strict inequality as long as $\Delta x \neq 0$ and $\Delta x$ is orthogonal to the gradients of all active constraints at $x^*$.

Example: \[ \min_{x \in \mathbb{R}^2} \quad 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2 \]

s.t. \[ x_1^2 + x_2^2 \leq 5 \]
\[ 3x_1 + x_2 \leq 6 \]

- It was shown last time that $x_1^* = 1$, $x_2^* = 2$, $\mu_1 = 1$, $\mu_2 = 0$ is a KKT point.
- Let's show that $x^* = [1, 2]'$ is a local min.

\[ \Delta x^T \left( \nabla^2 f(x^*) + \mu_1 \nabla^2 g_1(x^*) + \mu_2 \nabla^2 g_2(x^*) \right) \Delta x \]

\[ = \Delta x^T \left( \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} + (1) \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \right) \Delta x \]

\[ = \Delta x^T \begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix} \Delta x \]
\( g_1(x^*) = 0 \) and \( g_2(x^*) < 0 \)

- Due to second order sufficiency condition, \( x^* \) is a local min if

\[
\Delta x^T \begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix} \Delta x > 0 \quad \text{for all } \Delta x \text{ such that } \Delta x \neq 0 \text{ and } \nabla g_1(x^*) \Delta x = 0
\]

- Note that \( \begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix} > 0 \)

\[
\Rightarrow \Delta x^T \begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix} \Delta x > 0 \quad \text{for all nonzero } \Delta x.
\]

\[
\Rightarrow x^* = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \text{ is a local minimum.}
\]

---

Sensitivity analysis:

- Consider

\[
\begin{align*}
\min_{x} & \quad f(x) \\
\text{subject to} & \quad h(x) = 0 \\
& \quad g(x) \leq 0
\end{align*}
\]

Assume \( x^* \) is a local minimum, is a regular point, and satisfies second order sufficiency condition.
- Denote the optimal objective value as \( f^* = f(x^*). \)

- Perturb the problem as

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & \quad f(x) \\
\text{s.t.} & \quad h_i(x) = \varepsilon_i \\
& \quad h_m(x) = \varepsilon_m \\
& \quad g_i(x) \leq \bar{\varepsilon}_i \\
& \quad g_k(x) \leq \bar{\varepsilon}_k
\end{align*}
\]

\[
\text{optimal objective value} = f^* - \sum_{i=1}^{m} \lambda_i \varepsilon_i - \sum_{i=1}^{k} \mu_i \bar{\varepsilon}_i
\]

- If \( \mu_i \) or \( \lambda_i \) is zero, a small perturbation of the corresponding constraint doesn't affect the solution around \( x^* \). \( \Rightarrow \) Those constraints are unimportant and may be removed without changing the local solution.
Exact penalty method:

- Consider \( \min_{x \in \mathbb{R}^n} f(x) \) s.t. \( h(x) = 0 \) \( \Rightarrow \) Assume that \( g(x) \leq 0 \)

\( x_* \) is a local solution, is a regular point and satisfies second order sufficiency condition.

- Does this constrained optimization have an unconstrained optimization counterpart?

- Consider the unconstrained problem:

\[
\min_{x \in \mathbb{R}^n} f(x) + c \left[ \sum_{i=1}^{m} |h_i(x)| + \sum_{i=1}^{k} \max(0, g_i(x)) \right]
\]

Note that \( c |h_i(x)| \) and \( c \times \max(0, g_i(x)) \) are very large if \( h_i(x) \neq 0 \) or \( g_i(x) > 0 \), as long as \( \xi \) is a large penalty term.

**Theorem:** If \( c > 1 |h_1|, \ldots, 1 |h_m|, \mu_1, \ldots, \mu_k \),

then \( x_* \) is a local solution of the unconstrained optimization problem too.
- How can we guarantee that every local solution is a global one?
- Note that first and second order conditions are just about local solutions.
- Convexity helps.

- Recall that $f(x)$ is convex if
  1. For every two points $y$ and $z$, the segment connecting $(y, f(y))$ to $(z, f(z))$ is above the function.

$2.01$, if

$$\alpha f(y) + (1-\alpha) f(z) \geq f(\alpha y + (1- \alpha) z)$$

for all $y, z \in \mathbb{R}^n$ and $\alpha \in [0,1]$.  

[Diagram showing a convex function with points A and B on the curve and line segments AB and AC, where A and C are points on the curve, and AC is above the line segment AB.]  

$$A = \alpha f(y) + (1-\alpha) f(z)$$
$$B = f(\alpha y + (1- \alpha) z)$$
Consider the problem:

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & \quad f(x) \\
\text{s.t.} & \quad h_i(x) = 0 \\
& \quad h_m(x) = 0 \\
& \quad g_j(x) \leq 0 \\
& \quad g_{i_c}(x) \leq 0
\end{align*}
\]

**Theorem:** If \( f(x) \) is convex, \( h_i(x), \ldots, h_m(x) \) are linear, and \( g_j(x), \ldots, g_{i_c}(x) \) are convex, then every local solution is a global solution.

**Note:** Such an optimization (convex functions for objective and constraints) is called a **convex optimization**.

**Proof:** By contradiction, assume that \( x^* \) is a local solution that is not global.

\[ f(x^*) > f(\bar{x}) \]
\[ h(x^*) = h(x) = 0 \quad \text{but} \quad f(x^*) > f(x) \]

- We perturb \( x^* \) a bit as \( x = (1-\varepsilon) x^* + \varepsilon \bar{x} \)

Illustration for \( n = 2 \):

- Note that

\[ h_i(x) = h_i((1-\varepsilon) x^* + \varepsilon \bar{x}) = (1-\varepsilon) h_i(x^*) + \varepsilon h_i(\bar{x}) \]

\[ = 0 \quad \text{Linearity of } h_i(.) \]

\[ g_i(x) = g_i((1-\varepsilon) x^* + \varepsilon \bar{x}) \leq (1-\varepsilon) g_i(x^*) + \varepsilon g_i(\bar{x}) \]

\[ \leq 0 \quad \text{Convexity of } g_i(.) \]

\[ \Rightarrow x \text{ is feasible.} \]

- Also, \( f(x) = f((1-\varepsilon) x^* + \varepsilon \bar{x}) \leq (1-\varepsilon) f(x^*) + \varepsilon f(\bar{x}) \)

\[ \leq (1-\varepsilon + \varepsilon) f(x^*) = f(x^*) \]
So, for a small $\epsilon$, $x$ is a point close to $x^*$ but has a better objective value.

$\Rightarrow$ $x^*$ can't be a local solution $\Rightarrow$ Contradiction.

Example:

$$\begin{align*}
\min_{x \in \mathbb{R}^2} & \quad e^{x_1 + x_2} + (x_1 - x_2)^2 + x_4^4 \\
\text{s.t.} & \quad x_1 - x_2 = 2 \\
& \quad e^{x_1} + e^{x_2} + (x_1 - 4x_2)^4 \leq 5 \\
\end{align*}$$

$\Rightarrow$ This is a convex optimization.

$\Rightarrow$ Every local solution is a global one.