Numerical Optimization

Constrained Optimization

Shirish Shevade

Computer Science and Automation
Indian Institute of Science
Bangalore 560 012, India.

NPTEL Course on Numerical Optimization

Constrained Optimization

- Constrained Optimization Problem:

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \leq 0, \quad j = 1, \ldots, l \\ & e_i(\boldsymbol{x}) = 0, \quad i = 1, \ldots, m \\ & \boldsymbol{x} \in S \end{array}$$

- Inequality constraint functions: $h_j : \mathbb{R}^n \to \mathbb{R}$
- Equality constraint functions: $e_i : \mathbb{R}^n \to \mathbb{R}$
- Assume all functions (f, h_j's and e_i's) are sufficiently smooth
- Feasible set: $X = \{\boldsymbol{x} \in S : h_j(\boldsymbol{x}) \leq 0, e_i(\boldsymbol{x}) = 0, j = 1, \ldots, l, i = 1, \ldots, m\}$
- Given problem: *Minimize* $f(\boldsymbol{x})$ *subject to* $\boldsymbol{x} \in X$
- Assume X to be nonempty set in $\mathbb{R}^n$

Local and Global Minimum

Definition

A point $\boldsymbol{x}^* \in X$ is said to be a *global minimum* point of f over X if $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$ for all $\boldsymbol{x} \in X$. If $f(\boldsymbol{x}) > f(\boldsymbol{x}^*)$ for all $\boldsymbol{x} \in X, \boldsymbol{x} \neq \boldsymbol{x}^*$, then $\boldsymbol{x}^*$ is said to be a *strict global minimum point* of f over X.

Definition

A point $\boldsymbol{x}^* \in X$ is said to be a *local minimum* point of f over Xif there exists $\epsilon > 0$ such that $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$ for all $\boldsymbol{x} \in X \cap B(\boldsymbol{x}^*, \epsilon)$. $\boldsymbol{x}^* \in X$ is said to be a *strict local minimum* point of f over Xif there exists $\epsilon > 0$ such that $f(\boldsymbol{x}) > f(\boldsymbol{x}^*)$ for all $\boldsymbol{x} \in X \cap B(\boldsymbol{x}^*, \epsilon), \boldsymbol{x} \neq \boldsymbol{x}^*$.

Convex Programming Problem

$$
\begin{aligned}
\min \quad & f(\boldsymbol{x}) \\
\text{s.t.} \quad & h_j(\boldsymbol{x}) \leq 0, \;\; j = 1, \ldots, l \\
& e_i(\boldsymbol{x}) = 0, \;\; i = 1, \ldots, m \\
& \boldsymbol{x} \in S
\end{aligned}
$$

- $f(\boldsymbol{x})$ is a convex function
- $e_i(\boldsymbol{x})$ is affine ($e_i(\boldsymbol{x}) = \boldsymbol{a}_i^T \boldsymbol{x} + b_i, \;\; i = 1, \ldots, m$)
- $h_j(\boldsymbol{x})$ is a convex function for $j = 1, \ldots, l$
- S is a convex set
- Any local minimum is a global minimum
- The set of global minima form a convex set

Consider the problem:

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{x} \in X \end{array}$$

Different ways of solving this problem:

- Reformulation to an unconstrained problem needs to be done with care
- Solve the constrained problem directly

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{x} \in X \end{array}$$

- An iterative optimization algorithm generates a sequence $\{\boldsymbol{x}^k\}_{k\geq 0}$, which converges to a local minimum.

Constrained Minimization Algorithm

---

(1) Initialize $\boldsymbol{x}^0 \in X, k := 0$.

(2) **while** *stopping condition is not satisfied at* $\boldsymbol{x}^k$

   (a) Find $\boldsymbol{x}^{k+1} \in X$ such that $f(\boldsymbol{x}^{k+1}) < f(\boldsymbol{x}^k)$.

   (b) $k := k + 1$

endwhile

Output : $\boldsymbol{x}^* = \boldsymbol{x}^k$, a local minimum of $f(\boldsymbol{x})$ over the set X.

$$
\begin{aligned}
\min \quad & f(\boldsymbol{x}) \\
\text{s.t.} \quad & \boldsymbol{x} \in X
\end{aligned}
$$

Strict Local Minimum: There exists $\epsilon > 0$ such that

$$
f(\boldsymbol{x}^*) < f(\boldsymbol{x}) \; \forall \, \boldsymbol{x} \in X \cap B(\boldsymbol{x}^*, \epsilon), \; \boldsymbol{x} \neq \boldsymbol{x}^*
$$

At a local minimum of a constrained minimization problem:

the function does not decrease locally by moving along directions which contain feasible points

- How to convert this statement to an algebraic condition?

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{x} \in X \end{array}$$

Definition

A vector $\boldsymbol{d} \in \mathbb{R}^n, \boldsymbol{d} \neq \mathbf{0}$ is said to be a *feasible* direction at $\boldsymbol{x} \in X$ if there exists $\delta_1 > 0$ such that $\boldsymbol{x} + \alpha \boldsymbol{d} \in X$ for all $\alpha \in (0, \delta_1)$.

- Let $\mathcal{F}(\boldsymbol{x})$ = Set of *feasible* directions at $\boldsymbol{x} \in X$ (w.r.t. X)

Definition

A vector $\boldsymbol{d} \in \mathbb{R}^n, \boldsymbol{d} \neq \mathbf{0}$ is said to be a *descent* direction at $\boldsymbol{x} \in X$ if there exists $\delta_2 > 0$ such that $f(\boldsymbol{x} + \alpha \boldsymbol{d}) < f(\boldsymbol{x})$ for all $\alpha \in (0, \delta_2)$.

- Let $\mathcal{D}(\boldsymbol{x})$ = Set of *descent* directions at $\boldsymbol{x} \in X$ (w.r.t. f)

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & h_j(\boldsymbol{x}) \leq 0, \quad j = 1, \ldots, l \\ & e_i(\boldsymbol{x}) = 0, \quad i = 1, \ldots, m \\ & \boldsymbol{x} \in \mathbb{R}^n \end{aligned}$$

- $X = \{\boldsymbol{x} \in \mathbb{R}^n : h_j(\boldsymbol{x}) \leq 0, e_i(\boldsymbol{x}) = 0, j = 1, \ldots, l, i = 1, \ldots, m\}$
- At a local minimum $\boldsymbol{x}^* \in X$, *the function does not decrease by moving along feasible directions*

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{x} \in X \end{array}$$

Theorem

Let X be a nonempty set in $\mathbb{R}^n$ and $\boldsymbol{x}^* \in X$ be a local minimum of f over X. Then, $\mathcal{F}(\boldsymbol{x}^*) \cap \mathcal{D}(\boldsymbol{x}^*) = \phi$.

Proof.

Let $\boldsymbol{x}^* \in X$ be a local minimum.
By contradiction, assume that $\exists$ a nonzero $\boldsymbol{d} \in \mathcal{F}(\boldsymbol{x}^*) \cap \mathcal{D}(\boldsymbol{x}^*)$.
$\therefore \exists\ \delta_1 > 0 \ni \boldsymbol{x}^* + \alpha\boldsymbol{d} \in X\ \forall\ \alpha \in (0, \delta_1)$ and
$\exists\ \delta_2 > 0 \ni f(\boldsymbol{x}^* + \alpha\boldsymbol{d}) < f(\boldsymbol{x}^*)\ \forall\ \alpha \in (0, \delta_2)$.
Hence, $\exists\ \boldsymbol{x} \in B(\boldsymbol{x}^*, \alpha) \cap X \ni f(\boldsymbol{x}) < f(\boldsymbol{x}^*)$, for every $\alpha \in (0, \min(\delta_1, \delta_2))$.
This contradicts the assumption that $\boldsymbol{x}^*$ is a local minimum. $\square$

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{x} \in X \end{array}$$

- $\boldsymbol{x}^* \in X$ is a local minimum $\Rightarrow \mathcal{F}(\boldsymbol{x}^*) \cap \mathcal{D}(\boldsymbol{x}^*) = \phi$
- Consider any $\boldsymbol{x} \in X$ and assume $f \in \mathcal{C}^2$
- $\lim_{\alpha \to 0^+} \frac{f(\boldsymbol{x}+\alpha\boldsymbol{d})-f(\boldsymbol{x})}{\alpha} = \nabla f(\boldsymbol{x})^T\boldsymbol{d}$
- $\nabla f(\boldsymbol{x})^T\boldsymbol{d} < 0 \Rightarrow f(\boldsymbol{x}+\alpha\boldsymbol{d}) < f(\boldsymbol{x}) \Rightarrow \boldsymbol{d}$ is a descent direction $\Rightarrow \boldsymbol{d} \in \mathcal{D}(\boldsymbol{x})$
- Let $\tilde{\mathcal{D}}(\boldsymbol{x}) = \{\boldsymbol{d} : \nabla f(\boldsymbol{x})^T\boldsymbol{d} < 0\} \subseteq \mathcal{D}(\boldsymbol{x})$
- $\boldsymbol{x}^* \in X$ is a local minimum $\Rightarrow \mathcal{F}(\boldsymbol{x}^*) \cap \tilde{\mathcal{D}}(\boldsymbol{x}^*) = \phi$
- If $\mathcal{F}(\boldsymbol{x}^*) = \mathbb{R}^n$ (every direction in $\mathbb{R}^n$ is locally feasible),
  $\boldsymbol{x}^* \in X$ is a local minimum
  $\Rightarrow \{\boldsymbol{d} : \nabla f(\boldsymbol{x}^*)^T\boldsymbol{d} < 0\} = \phi \Rightarrow \nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$
- Can we characterize $\mathcal{F}(\boldsymbol{x}^*)$ algebraically for a constrained optimization problem?

Consider the problem:

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \leq 0, \quad j = 1, \ldots, l \\ & \boldsymbol{x} \in \mathbb{R}^n \end{array}$$

- Assume $f, \; h_j \in \mathcal{C}^2, \; j = 1, \ldots, l$
- $X = \{\boldsymbol{x} \in \mathbb{R}^n : h_j(\boldsymbol{x}) \leq 0, \;\; j = 1, \ldots, l\}$
- Active constraints:

$$\mathcal{A}(\boldsymbol{x}) = \{j : h_j(\boldsymbol{x}) = 0\}$$

Lemma

For any $\boldsymbol{x} \in X$,

$$\tilde{\mathcal{F}}(\boldsymbol{x}) \triangleq \{\boldsymbol{d} : \nabla h_j(\boldsymbol{x})^T \boldsymbol{d} < 0, \;\; j \in \mathcal{A}(\boldsymbol{x})\} \subseteq \mathcal{F}(\boldsymbol{x})$$

Lemma

For any $\boldsymbol{x} \in X$,

$$\tilde{\mathcal{F}}(\boldsymbol{x}) \triangleq \{\boldsymbol{d} : \nabla h_j(\boldsymbol{x})^T \boldsymbol{d} < 0, \;\; j \in \mathcal{A}(\boldsymbol{x})\} \subseteq \mathcal{F}(\boldsymbol{x})$$

Proof.

Suppose $\tilde{\mathcal{F}}(\boldsymbol{x})$ is nonempty and let $\boldsymbol{d} \in \tilde{\mathcal{F}}(\boldsymbol{x})$. Since $\nabla h_j(\boldsymbol{x})^T \boldsymbol{d} < 0 \; \forall \, j \in \mathcal{A}(\boldsymbol{x})$, $\boldsymbol{d}$ is a descent direction for $h_j, \; j \in \mathcal{A}(\boldsymbol{x})$ at $\boldsymbol{x}$. That is,

$$\exists \; \delta_1 > 0 \;\; \ni \;\; h_j(\boldsymbol{x} + \alpha \boldsymbol{d}) < h_j(\boldsymbol{x}) = 0 \; \forall \, j \in \mathcal{A}(\boldsymbol{x}).$$

Further, $h_j(\boldsymbol{x}) < 0 \; \forall \, j \notin \mathcal{A}(\boldsymbol{x})$. Therefore,

$$\exists \; \delta_3 > 0 \;\; \ni \;\; h_j(\boldsymbol{x} + \alpha \boldsymbol{d}) < 0 \; \forall \; \alpha \in (0, \delta_3), \;\; \forall j \notin \mathcal{A}(\boldsymbol{x})$$

Thus, $\boldsymbol{x} + \alpha \boldsymbol{d} \in X \; \forall \; \alpha \in (0, \min(\delta_1, \delta_3))$,
and $\therefore \boldsymbol{d} \in \mathcal{F}(\boldsymbol{x})$. $\square$

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \le 0, \ \ j = 1, \ldots, l \\ & \boldsymbol{x} \in \mathbb{R}^n \end{array}$$

Let $X = \{\boldsymbol{x} \in \mathbb{R}^n : h_j(\boldsymbol{x}) \le 0, \ \ j = 1, \ldots, l\}$.
For any $\boldsymbol{x} \in X$, $\tilde{\mathcal{F}}(\boldsymbol{x}) \triangleq \{\boldsymbol{d} : \nabla h_j(\boldsymbol{x})^T \boldsymbol{d} < 0, \ \ j \in \mathcal{A}(\boldsymbol{x})\} \subseteq \mathcal{F}(\boldsymbol{x})$
and $\tilde{\mathcal{D}}(\boldsymbol{x}) \triangleq \{\boldsymbol{d} : \nabla f(\boldsymbol{x})^T \boldsymbol{d} < 0\} \subseteq \mathcal{D}(\boldsymbol{x})$.

$$\begin{array}{lll} \boldsymbol{x}^* \in X \text{ is a local minimum} & \Rightarrow & \mathcal{F}(\boldsymbol{x}^*) \cap \mathcal{D}(\boldsymbol{x}^*) = \phi \\ & \Rightarrow & \tilde{\mathcal{F}}(\boldsymbol{x}^*) \cap \tilde{\mathcal{D}}(\boldsymbol{x}^*) = \phi \end{array}$$

$$\boldsymbol{x}^* \in X \text{ is a local minimum} \Rightarrow \tilde{\mathcal{F}}(\boldsymbol{x}^*) \cap \tilde{\mathcal{D}}(\boldsymbol{x}^*) = \phi$$

- This is only a necessary condition for a local minimum
- Utility of this condition depends on the constraint representation
- Cannot be directly used for equality constrained problems

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \le 0, \;\; j = 1, \ldots, l \\ & \boldsymbol{x} \in \mathbb{R}^n \end{array}$$

Let $X = \{\boldsymbol{x} \in \mathbb{R}^n : h_j(\boldsymbol{x}) \le 0, \;\; j = 1, \ldots, l\}$

$$\begin{array}{ll} & \boldsymbol{x}^* \in X \text{ is a local minimum} \\ \Rightarrow & \tilde{\mathcal{F}}(\boldsymbol{x}^*) \cap \tilde{\mathcal{D}}(\boldsymbol{x}^*) = \phi \\ \Rightarrow & \{\boldsymbol{d} : \nabla h_j(\boldsymbol{x}^*)^T \boldsymbol{d} < 0, \; j \in \mathcal{A}(\boldsymbol{x}^*)\} \cap \{\boldsymbol{d} : \nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} < 0\} = \phi \end{array}$$

Let $\boldsymbol{A} = \begin{pmatrix} \nabla f(\boldsymbol{x}^*)^T \\ \cdots \\ \nabla h_j(\boldsymbol{x}^*)^T, \; j \in \mathcal{A}(\boldsymbol{x}^*) \\ \cdots \end{pmatrix}_{(1+|\mathcal{A}(\boldsymbol{x}^*)|) \times n}$

$\therefore \boldsymbol{x}^* \in X$ is a local minimum $\Rightarrow \{\boldsymbol{d} : \boldsymbol{A}\boldsymbol{d} < 0\} = \phi$

Farkas' Lemma

Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{c} \in \mathbb{R}^n$. Then, exactly one of the following two systems has a solution:

(I) $\boldsymbol{Ax} \leq \boldsymbol{0}, \boldsymbol{c}^T\boldsymbol{x} > 0$ for some $\boldsymbol{x} \in \mathbb{R}^n$

(II) $\boldsymbol{A}^T\boldsymbol{y} = \boldsymbol{c}, \boldsymbol{y} \geq \boldsymbol{0}$ for some $\boldsymbol{y} \in \mathbb{R}^m$.

Corollary

Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. Then exactly one of the following systems has a solution:

(I) $\boldsymbol{Ax} < \boldsymbol{0}$ for some $\boldsymbol{x} \in \mathbb{R}^n$

(II) $\boldsymbol{A}^T\boldsymbol{y} = \boldsymbol{0}, \boldsymbol{y} \geq \boldsymbol{0}$ for some nonzero $\boldsymbol{y} \in \mathbb{R}^m$.

$\boldsymbol{x}^* \in X$ is a local minimum $\Rightarrow \{\boldsymbol{d} : \boldsymbol{Ad} < 0\} = \phi \Rightarrow$

$\exists\ \lambda_0 \geq 0$ and $\lambda_j \geq 0, j \in \mathcal{A}(\boldsymbol{x}^*)$ (not all λ's 0), such that

$$\lambda_0 \nabla f(\boldsymbol{x}^*) + \sum_{j \in \mathcal{A}(\boldsymbol{x}^*)} \lambda_j \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0}.$$

$\boldsymbol{x}^* \in X$ is a local minimum $\Rightarrow$ $\{\boldsymbol{d} : \boldsymbol{A}\boldsymbol{d} < 0\} = \phi$ $\Rightarrow$

$\exists\ \lambda_0 \geq 0$ and $\lambda_j \geq 0,\ j \in \mathcal{A}(\boldsymbol{x}^*)$ (not all λ's 0), such that

$$\lambda_0 \nabla f(\boldsymbol{x}^*) + \sum_{j \in \mathcal{A}(\boldsymbol{x}^*)} \lambda_j \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0}.$$

- Easy to satisfy these conditions if $\nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0}$ for some $j \in \mathcal{A}(\boldsymbol{x}^*)$ or $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$
- Regular point: A point $\boldsymbol{x}^* \in X$ is said to be a *regular point* if the gradient vectors, $\nabla h_j(\boldsymbol{x}^*),\ j \in \mathcal{A}(\boldsymbol{x}^*)$, are linearly independent.
- $\boldsymbol{x}^* \in X$ is a regular point $\Rightarrow$ $\lambda_0 \neq 0$

Letting $\lambda_j = 0 \; \forall \; j \notin \mathcal{A}(\boldsymbol{x}^*)$, we get the following conditions:

$$\begin{array}{rcl} \lambda_0 \nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{l} \lambda_j \nabla h_j(\boldsymbol{x}^*) & = & \mathbf{0} \\ \lambda_j h_j(\boldsymbol{x}^*) & = & 0 \; \forall \; j = 1, \ldots, l \\ \lambda_j & \geq & 0 \; \forall \; j = 0, \ldots, l \\ (\lambda_0, \boldsymbol{\lambda}) & \neq & (0, \mathbf{0}) \end{array}$$

where $\boldsymbol{\lambda}^T = (\lambda_1, \ldots, \lambda_l)$.

Consider the problem:

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \leq 0, \;\; j = 1, \ldots, l \\ & \boldsymbol{x} \in \mathbb{R}^n \end{array}$$

Assume $\boldsymbol{x}^* \in X$ to be a regular point.

$\boldsymbol{x}^*$ is a local minimum $\Rightarrow \; \exists \; \lambda_j^*, \; j = 1, \ldots, l$ such that

$$\begin{aligned} \nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{l} \lambda_j^* \nabla h_j(\boldsymbol{x}^*) &= \mathbf{0} \\ \lambda_j^* h_j(\boldsymbol{x}^*) &= 0 \; \forall \; j = 1, \ldots, l \\ \lambda_j^* &\geq 0 \; \forall \; j = 1, \ldots, l \end{aligned}$$

Karush-Kuhn-Tucker (KKT) Conditions

Consider the problem:

$$\begin{array}{rc} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \leq 0,\ \ j = 1, \ldots, l \\ & \boldsymbol{x} \in \mathbb{R}^n \end{array}$$

- $X = \{\boldsymbol{x} \in \mathbb{R}^n : h_j(\boldsymbol{x}) \leq 0,\ \ j = 1, \ldots, l\}$
- $\boldsymbol{x}^* \in X,\ \ \mathcal{A}(\boldsymbol{x}^*) = \{j : h_j(\boldsymbol{x}^*) = 0\}$

KKT necessary conditions (First Order) : If $\boldsymbol{x}^* \in X$ is a local minimum and a *regular* point, then there exists a unique vector $\boldsymbol{\lambda}^* (= (\lambda_1^*, \ldots, \lambda_l^*)^T)$ such that

$$\begin{aligned} \nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{l} \lambda_j^* \nabla h_j(\boldsymbol{x}^*) &= \mathbf{0} \\ \lambda_j^* h_j(\boldsymbol{x}^*) &= 0 \ \forall\, j = 1, \ldots, l \\ \lambda_j^* &\geq 0 \ \forall\, j = 1, \ldots, l \end{aligned}$$

KKT necessary conditions (First Order) : If $\boldsymbol{x}^* \in X$ is a local minimum and a *regular* point, then there exists a unique vector $\boldsymbol{\lambda}^*(= (\lambda_1^*, \ldots, \lambda_l^*)^T)$ such that

$$\begin{aligned} \nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{l} \lambda_j^* \nabla h_j(\boldsymbol{x}^*) &= \mathbf{0} \\ \lambda_j^* h_j(\boldsymbol{x}^*) &= 0 \;\forall\, j = 1, \ldots, l \\ \lambda_j^* &\geq 0 \;\forall\, j = 1, \ldots, l \end{aligned}$$

- *KKT point* : $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*),\ \boldsymbol{x}^* \in X,\ \boldsymbol{\lambda}^* \geq \mathbf{0}$
- Lagrangian function : $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{j=1}^{l} \lambda_j h_j(\boldsymbol{x})$
- $\nabla \mathcal{L}_{\boldsymbol{x}}(\boldsymbol{x}^*, \lambda^*) = 0$
- λ_j : Lagrange multipliers , $\lambda_j \geq 0$
- $\lambda_j^* h_j(\boldsymbol{x}^*) = 0$: *Complementary Slackness Condition*
- $\lambda_j^* = 0 \;\forall\, j \notin \mathcal{A}(\boldsymbol{x}^*)$

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \leq 0, \quad j = 1, \ldots, l \\ & \boldsymbol{x} \in \mathbb{R}^n \end{array}$$

- At a local minimum, *active set is unknown*
- Need to investigate all possible active sets for finding KKT points
  Example:
  $$\begin{array}{ll} \min & x_1^2 + x_2^2 \\ \text{s.t.} & x_2 \leq 1 \\ & x_1 + x_2 \geq 1 \end{array}$$
- A KKT point can be a local maximum
  Example:
  $$\begin{array}{ll} \min & -x^2 \\ \text{s.t.} & x \leq 0 \end{array}$$

Constraint Qualification

- Every local minimum need not be a KKT point
- Example [Kuhn and Tucker, 1951][1]

$$
\begin{array}{ll}
\min & -x_1 \\
\text{s.t.} & x_2 - (1 - x_1)^3 \leq 0 \\
 & x_2 \geq 0
\end{array}
$$

- *Linear Independence Constraint Qualification* (LICQ) : $\nabla h_j(\boldsymbol{x}^*),\ j \in \mathcal{A}(\boldsymbol{x}^*)$ are *linearly independent*
- *Mangasarian-Fromovitz Constraint Qualification* (MFCQ)

$$\{\boldsymbol{d} : \nabla h_j(\boldsymbol{x}^*)^T \boldsymbol{d} < 0,\ j \in \mathcal{A}(\boldsymbol{x}^*)\} \neq \phi$$

---

[1] H.W. Kuhn and A.W. Tucker, *Nonlinear Programming*, in Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, J. Neyman, ed., Berkeley, CA, 1951, University of California Press, pp. 481–492.

Consider the problem (**CP**):

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \leq 0, \quad j = 1, \ldots, l \\ & \boldsymbol{x} \in \mathbb{R}^n \end{array}$$

- Assumption: $f, h_j, j = 1, \ldots, l$ are differentiable convex functions
- **CP** is a *convex program*
- $X = \{\boldsymbol{x} \in \mathbb{R}^n : h_j(\boldsymbol{x}) \leq 0, \quad j = 1, \ldots, l\}$
- Every local minimum of a convex program is a global minimum
- The set of all optimal solutions to a convex program is convex

If $\boldsymbol{x}^* \in X$ is a *regular* point, then for $\boldsymbol{x}^*$ to be a global minimum of **CP**, first order KKT conditions are necessary and sufficient.

Proof.

Let $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ be a KKT point. We need to show that $\boldsymbol{x}^*$ is a global minimum of **CP**. We use the convexity of f and h_j to prove this. Consider any $\boldsymbol{x} \in X$. For a convex function f, $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*)$.

$$
\begin{aligned}
f(\boldsymbol{x}) &\geq f(\boldsymbol{x}) + \sum_j \lambda_j^* h_j(\boldsymbol{x}) \\
&\geq f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) \\
&\quad + \sum_j \lambda_j^* (h_j(\boldsymbol{x}^*) + \nabla h_j(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*)) \\
&= (f(\boldsymbol{x}^*) + \sum_j \lambda_j^* h_j(\boldsymbol{x}^*) \\
&\quad + (\nabla f(\boldsymbol{x}^*) + \sum_j \lambda_j^* \nabla h_j(\boldsymbol{x}^*))^T(\boldsymbol{x} - \boldsymbol{x}^*) \\
&= f(\boldsymbol{x}^*) \ \forall\, \boldsymbol{x} \in X \ \Rightarrow\ \boldsymbol{x}^* \text{ is a global minimum of } \textbf{CP}
\end{aligned}
$$

Consider the problem (**CP**):

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \leq 0, \;\; j = 1, \ldots, l \\ & \boldsymbol{x} \in \mathbb{R}^n \end{array}$$

- Assumption: $f, h_j, j = 1, \ldots, l$ are convex functions
- $X = \{\boldsymbol{x} \in \mathbb{R}^n : h_j(\boldsymbol{x}) \leq 0, \;\; j = 1, \ldots, l\}$
- Slater's Constraint Qualification: There exists $\boldsymbol{y} \in X$ such that

$$h_j(\boldsymbol{y}) < 0, \; j = 1, \ldots, l$$

- Useful when the constraint functions h_j are convex
- For example, the following program does not satisfy Slater's constraint qualification:

$$\begin{array}{ll} \min & x_1 + x_2 \\ \text{s.t.} & (x_1 + 1)^2 + x_2^2 \leq 1 \\ & (x_1 - 1)^2 + x_2^2 \leq 1 \end{array}$$

$(0, 0)^T$ is the global minimum; but it is *not a KKT point*.

Consider the problem:

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & e_i(\boldsymbol{x}) = 0, \quad i = 1, \ldots, m \\ & \boldsymbol{x} \in \mathbb{R}^n \end{array}$$

- Assumption: $f, e_i, i = 1, \ldots, m$ are smooth functions
- $X = \{\boldsymbol{x} \in \mathbb{R}^n : e_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, m\}$
- Let $\boldsymbol{x} \in X$, $\mathcal{A}(\boldsymbol{x}) = \{i : e_i(\boldsymbol{x}) = 0\} = \{1, \ldots, m\}$

Definition

A vector $\boldsymbol{d} \in \mathbb{R}^n$ is said to be a tangent of X at $\boldsymbol{x}$ if either $\boldsymbol{d} = \boldsymbol{0}$ or there exists a sequence $\{\boldsymbol{x}^k\} \subset X$, $\boldsymbol{x}^k \neq \boldsymbol{x} \ \forall\ k$ such that

$$\boldsymbol{x}^k \to \boldsymbol{x}, \quad \frac{\boldsymbol{x}^k - \boldsymbol{x}}{\|\boldsymbol{x}^k - \boldsymbol{x}\|} \to \frac{\boldsymbol{d}}{\|\boldsymbol{d}\|}.$$

The collection of all tangents of X at $\boldsymbol{x}$ is called the *tangent set* at $\boldsymbol{x}$ and is denoted by $\mathcal{T}(\boldsymbol{x})$.

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & e_i(\boldsymbol{x}) = 0, \quad i = 1, \ldots, m \\ & \boldsymbol{x} \in \mathbb{R}^n \end{array}$$

- $X = \{\boldsymbol{x} \in \mathbb{R}^n : e_i(\boldsymbol{x}) = 0, \quad i = 1, \ldots, m\}$
- Regular Point: A point $\bar{\boldsymbol{x}} \in X$ is said to be a regular point if $\nabla e_i(\bar{\boldsymbol{x}}),\ i = 1, \ldots, m$ are *linearly independent*.
- At a regular point $\bar{\boldsymbol{x}} \in X$,

$$\mathcal{T}(\bar{\boldsymbol{x}}) = \{\boldsymbol{d} : \nabla e_i(\bar{\boldsymbol{x}})^T \boldsymbol{d} = 0,\ i = 1, \ldots, m\}$$

- Let $\boldsymbol{x}^* \in X$ be a *regular point* and *local extremum* (minimum or maximum) of the problem
- Consider any $\boldsymbol{d} \in \mathcal{T}(\boldsymbol{x}^*)$.
- Let $\boldsymbol{x}(t)$ be any smooth curve such that
  - $\boldsymbol{x}(t) \in X$
  - $\boldsymbol{x}(0) = \boldsymbol{x}^*,\ \dot{\boldsymbol{x}}(0) = \boldsymbol{d}$
  - $\exists\, a > 0$ such that $e(\boldsymbol{x}(t)) = 0\ \forall\, t \in [-a, a]$
- $\boldsymbol{x}^*$ is a regular point
  $\Rightarrow\ \mathcal{T}(\boldsymbol{x}^*) = \{\boldsymbol{d} : \nabla e_i(\boldsymbol{x}^*)^T \boldsymbol{d} = 0,\ i = 1, \ldots, m\}$
- $\boldsymbol{x}^*$ is a constrained local extremum
  $\Rightarrow\ \frac{d}{dt} f(\boldsymbol{x}(t))|_{t=0} = 0 \Rightarrow\ \nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} = 0.$

If $\boldsymbol{x}^*$ is a regular point w.r.t. the constraints $e_i(\boldsymbol{x}) = 0$, $i = 1, \ldots, m$ and $\boldsymbol{x}^*$ is a local *extremum point* (a minimum or maximum) of f subject to these constraints, then $\nabla f(\boldsymbol{x}^*)$ is orthogonal to the tangent set, $\mathcal{T}(\boldsymbol{x}^*)$.

Theorem

Let $\boldsymbol{x}^* \in X$ be a regular point and be a local minimum. Then $\exists$ $\boldsymbol{\mu}^* \in \mathbb{R}^m$ such that

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i^* \nabla e_i(\boldsymbol{x}^*) = \boldsymbol{0}.$$

Proof.

Let $\boldsymbol{e}(\boldsymbol{x}) = (e_1(\boldsymbol{x}), \ldots, e_m(\boldsymbol{x}))$. $\boldsymbol{x}^* \in X$ is a local minimum.
$\therefore \{\boldsymbol{d} : \nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} < 0, \nabla e(\boldsymbol{x}^*)^T \boldsymbol{d} = 0\} = \phi$.
Let $C_1 = \{(y_1, \boldsymbol{y_2}) : y_1 = \nabla f(\boldsymbol{x}^*)^T \boldsymbol{d}, \boldsymbol{y_2} = \nabla e(\boldsymbol{x}^*)^T \boldsymbol{d}\}$ and
$C_2 = \{(y_1, \boldsymbol{y_2}) : y_1 < 0, \boldsymbol{y_2} = \boldsymbol{0}\}$
Note that C_1 and C_2 are convex and $C_1 \cap C_2 = \phi$.

If C_1 and C_2 are nonempty convex sets in $\mathbb{R}^n$ and $C_1 \cap C_2 = \phi$,
$\exists$ $\boldsymbol{\mu} \in \mathbb{R}^n (\boldsymbol{\mu} \neq \boldsymbol{0})$ such that $\boldsymbol{\mu}^T \boldsymbol{x}_1 \geq \boldsymbol{\mu}^T \boldsymbol{x}_2 \; \forall \; \boldsymbol{x}_1 \in C_1, \boldsymbol{x}_2 \in C_2$.

Proof. (continued)

Therefore, $\exists\ (\mu_0, \boldsymbol{\mu}) \in \mathbb{R}^{m+1}$ such that

$$\mu_0 \nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} + \boldsymbol{\mu}^T (\nabla e(\boldsymbol{x}^*)^T \boldsymbol{d}) \geq \mu_0 y_1 + \boldsymbol{\mu}^T \boldsymbol{y_2}\ \forall\ \boldsymbol{d} \in \mathbb{R}^n,\ (y_1, \boldsymbol{y_2}) \in C_2$$

Letting $\boldsymbol{y_2} = \mathbf{0}$, we get $\mu_0 \geq 0$.
Letting $(y_1, \boldsymbol{y_2}) = (0, \mathbf{0})$, we get

$$\mu_0 \nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} + \boldsymbol{\mu}^T (\nabla e(\boldsymbol{x}^*)^T \boldsymbol{d}) \geq 0\ \forall\ \boldsymbol{d} \in \mathbb{R}^n$$

If we take $\boldsymbol{d} = -(\mu_0 \nabla f(\boldsymbol{x}^*) + \boldsymbol{\mu}^T \nabla e(\boldsymbol{x}^*))$, we get
$-\|(\mu_0 \nabla f(\boldsymbol{x}^*) + \boldsymbol{\mu}^T \nabla e(\boldsymbol{x}^*))\|^2 \geq 0$.
Therefore,

$$\mu_0 \nabla f(\boldsymbol{x}^*) + \boldsymbol{\mu}^T \nabla e(\boldsymbol{x}^*) = \mathbf{0} \text{ where } (\mu_0, \boldsymbol{\mu}) \neq (0, \mathbf{0})$$

Note that, $\mu_0 > 0$ since $\boldsymbol{x}^*$ is a regular point.
Hence,

$$\nabla f(\boldsymbol{x}^*) + \boldsymbol{\mu}^{*T} \nabla e(\boldsymbol{x}^*) = \mathbf{0}$$

Examples:

1. $$\begin{array}{ll} \min & x_1 - 3x_2 \\ \text{s.t.} & (x_1-1)^2 + x_2^2 = 1 \\ & (x_1+1)^2 + x_2^2 = 1 \end{array}$$

   $(0,0)^T$ is the only feasible point; $(0,0)^T$ is not a regular point.

2. $$\begin{array}{ll} \min & x_1 + x_2 \\ \text{s.t.} & x_1^2 + x_2^2 = 1 \end{array}$$

   $\textit{local maximum} : (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^T$

   $\textit{local minimum} : (-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})^T$

General Nonlinear Programming Problems

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \leq 0,\ j = 1, \ldots, l \\ & e_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, m \end{array}$$

- $f, h_j (j = 1, \ldots, l), e_i (i = 1, \ldots, m)$ are sufficiently smooth
- $X = \{\boldsymbol{x} : h_j(\boldsymbol{x}) \leq 0, e_i(\boldsymbol{x}) = 0,\ \ j = 1, \ldots, l;\ i = 1, \ldots, m\}$
- $\boldsymbol{x}^* \in X$
- *Active set* of X at $\boldsymbol{x}^*$:
    - $\mathcal{I} = \{j : h_j(\boldsymbol{x}^*) = 0\}$
    - All the equality constraints, $\mathcal{E} = \{1, \ldots, m\}$

  $\mathcal{A}(\boldsymbol{x}^*) = \mathcal{I} \cup \mathcal{E}$
- Assumption: $\boldsymbol{x}^*$ is a *regular point*. That is, $\{\nabla h_j(\boldsymbol{x}^*) : j \in \mathcal{I}\} \cup \{\nabla e_i(\boldsymbol{x}^*) : i \in \mathcal{E}\}$ is a set of *linearly independent* vectors

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \le 0,\ j = 1, \ldots, l \\ & e_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, m \end{array}$$

- $X = \{\boldsymbol{x} : h_j(\boldsymbol{x}) \le 0, e_i(\boldsymbol{x}) = 0,\ \ j = 1, \ldots, l;\ i = 1, \ldots, m\}$

KKT necessary conditions (First Order) : If $\boldsymbol{x}^* \in X$ is a local minimum and a *regular* point, then there exist unique vectors $\boldsymbol{\lambda}^* \in \mathbb{R}^l_+$ and $\boldsymbol{\mu}^* \in \mathbb{R}^m$ such that

$$\begin{aligned} \nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{l} \lambda_j^* \nabla h_j(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i^* \nabla e_i(\boldsymbol{x}^*) &= \mathbf{0} \\ \lambda_j^* h_j(\boldsymbol{x}^*) &= 0\ \forall\, j = 1, \ldots, l \\ \lambda_j^* &\ge 0\ \forall\, j = 1, \ldots, l \end{aligned}$$

- KKT Point: $(\boldsymbol{x}^* \in X, \boldsymbol{\lambda}^* \in \mathbb{R}^l_+, \boldsymbol{\mu}^* \in \mathbb{R}^m)$ satisfying above conditions
- First order KKT conditions also satisfied at a local max

Consider the problem (**CP**):

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \leq 0,\ j = 1, \ldots, l \\ & e_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, m \end{array}$$

- Assumption: $f, h_j, j = 1, \ldots, l$ are smooth convex functions
- $e_i(\boldsymbol{x}) = \boldsymbol{a}^T \boldsymbol{x}_i - b_i,\ i = 1, \ldots, m$
- **CP** is a convex programming problem
- $X = \{\boldsymbol{x} : h_j(\boldsymbol{x}) \leq 0, e_i(\boldsymbol{x}) = 0,\ \ j = 1, \ldots, l;\ i = 1, \ldots, m\}$
- Assumption: Slater's Constraint Qualification holds for X.

There exists $\boldsymbol{y} \in X$ such that $h_j(\boldsymbol{y}) < 0,\ j = 1, \ldots, l$

- If X satisfies Slater's Constraint Qualification, then the first order KKT conditions are necessary and sufficient for a global minimum of a convex programming problem **CP**

Interpretation of Lagrange Multipliers

Consider the problem :

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \leq 0,\ j = 1, \ldots, l \end{array}$$

- $X = \{\boldsymbol{x} : h_j(\boldsymbol{x}) \leq 0,\ j = 1, \ldots, l; \}$
- Let $\boldsymbol{x}^* \in X$ be a regular point and a local minimum
- Let $\mathcal{A}(\boldsymbol{x}^*) = \{j : h_j(\boldsymbol{x}^*) = 0\}$
- $\nabla f(\boldsymbol{x}^*) + \sum_{j \in \mathcal{A}(\boldsymbol{x}^*)} \lambda_j^* \nabla h_j(\boldsymbol{x}^*) = 0$
- Suppose the constraint $h_{\tilde{j}}(\boldsymbol{x}),\ j \in \mathcal{A}(\boldsymbol{x}^*)$ is perturbed to
$$h_{\tilde{j}}(\boldsymbol{x}) \leq \epsilon \|\nabla h_{\tilde{j}}(\boldsymbol{x}^*)\| \qquad (\epsilon > 0)$$
- New problem:

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \leq 0,\ j = 1, \ldots, l,\ j \neq \tilde{j} \\ & h_{\tilde{j}}(\boldsymbol{x}) \leq \epsilon \|\nabla h_{\tilde{j}}(\boldsymbol{x}^*)\| \end{array}$$

For the new problem, let $\boldsymbol{x}_\epsilon^*$ be the solution.

- Assumption: $\mathcal{A}(\boldsymbol{x}^*) = \mathcal{A}(\boldsymbol{x}_\epsilon^*)$
- For the constraint $h_{\tilde{j}}(\boldsymbol{x})$,

$$\begin{aligned} h_{\tilde{j}}(\boldsymbol{x}_\epsilon^*) - h_{\tilde{j}}(\boldsymbol{x}^*) &= \epsilon \|\nabla h_{\tilde{j}}(\boldsymbol{x}^*)\| \\ \therefore (\boldsymbol{x}_\epsilon^* - \boldsymbol{x}^*)^T \nabla h_{\tilde{j}}(\boldsymbol{x}^*) &\approx \epsilon \|\nabla h_{\tilde{j}}(\boldsymbol{x}^*)\| \end{aligned}$$

- For other constraints, $h_j(\boldsymbol{x}), j \neq \tilde{j}$,

$$\begin{aligned} h_j(\boldsymbol{x}_\epsilon^*) - h_j(\boldsymbol{x}^*) &= 0 \\ \therefore (\boldsymbol{x}_\epsilon^* - \boldsymbol{x}^*)^T \nabla h_j(\boldsymbol{x}^*) &= 0 \end{aligned}$$

- Change in the objective function,

$$\begin{aligned} f(\boldsymbol{x}_\epsilon^*) - f(\boldsymbol{x}^*) &\approx (\boldsymbol{x}_\epsilon^* - \boldsymbol{x}^*)^T \nabla f(\boldsymbol{x}^*) \\ &= -\sum_{j \in \mathcal{A}(\boldsymbol{x}^*)} (\boldsymbol{x}_\epsilon^* - \boldsymbol{x}^*)^T (\lambda_j^* \nabla h_j(\boldsymbol{x}^*) \\ &= -\lambda_{\tilde{j}}^* \epsilon \|\nabla h_{\tilde{j}}(\boldsymbol{x}^*)\| \\ \therefore \left.\frac{df}{d\epsilon}\right|_{\boldsymbol{x}=\boldsymbol{x}^*} &\propto -\lambda_{\tilde{j}}^* \end{aligned}$$

Consider the problem (**NLP**):

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \leq 0,\ j = 1, \ldots, l \\ & e_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, m \end{array}$$

- Let $f, h_j, e_i \in \mathcal{C}^2$ for every j and i.
- $X = \{\boldsymbol{x} : h_j(\boldsymbol{x}) \leq 0, e_i(\boldsymbol{x}) = 0,\ \ j = 1, \ldots, l;\ i = 1, \ldots, m\}$
- $\boldsymbol{x}^* \in X$
- *Active set* of X at $\boldsymbol{x}^*$:
    - $\mathcal{I} = \{j : h_j(\boldsymbol{x}^*) = 0\}$
    - All the equality constraints, $\mathcal{E} = \{1, \ldots, m\}$

  $\mathcal{A}(\boldsymbol{x}^*) = \mathcal{I} \cup \mathcal{E}$
- Assumption: $\boldsymbol{x}^*$ is a *regular point.* That is, $\{\nabla h_j(\boldsymbol{x}^*) : j \in \mathcal{I}\} \cup \{\nabla e_i(\boldsymbol{x}^*) : i \in \mathcal{E}\}$ is a set of *linearly independent* vectors

Consider the problem (**NLP**):

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \leq 0,\ j = 1, \ldots, l \\ & e_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, m \end{array}$$

- Define the Lagrangian function, $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum_{j=1}^{l} \lambda_j h_j(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i e_i(\boldsymbol{x})$

KKT necessary conditions (Second Order) : If $\boldsymbol{x}^* \in X$ is a local minimum of **NLP** and a *regular* point, then there exist unique vectors $\boldsymbol{\lambda}^* \in \mathbb{R}^l_+$ and $\boldsymbol{\mu}^* \in \mathbb{R}^m$ such that

$$\begin{aligned} \nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) &= \mathbf{0} \\ \lambda_j^* h_j(\boldsymbol{x}^*) &= 0\ \forall\, j = 1, \ldots, l \\ \lambda_j^* &\geq 0\ \forall\, j = 1, \ldots, l \end{aligned}$$

and

$$\boldsymbol{d}^T \nabla^2_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \boldsymbol{d} \geq 0$$

for all $\boldsymbol{d} \ni \nabla h_j(\boldsymbol{x}^*)^T \boldsymbol{d} \leq 0, j \in \mathcal{I}$ and $\nabla e_i(\boldsymbol{x}^*)^T \boldsymbol{d} = 0,\ i \in \mathcal{E}$.

KKT sufficient conditions (Second Order) : If there exist $\boldsymbol{x}^* \in X$, $\boldsymbol{\lambda}^* \in \mathbb{R}^l_+$ and $\boldsymbol{\mu}^* \in \mathbb{R}^m$ such that such that

$$
\begin{aligned}
\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) &= \mathbf{0} \\
\lambda_j^* h_j(\boldsymbol{x}^*) &= 0 \;\forall\, j = 1, \ldots, l \\
\lambda_j^* &\geq 0 \;\forall\, j = 1, \ldots, l
\end{aligned}
$$

and

$$\boldsymbol{d}^T \nabla^2_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \boldsymbol{d} > 0$$

for all $\boldsymbol{d} \neq \mathbf{0}$ such that

$$
\begin{aligned}
\nabla h_j(\boldsymbol{x}^*)^T \boldsymbol{d} &= 0,\; j \in \mathcal{I} \text{ and } \lambda_j^* > 0 \\
\nabla h_j(\boldsymbol{x}^*)^T \boldsymbol{d} &\leq 0,\; j \in \mathcal{I} \text{ and } \lambda_j^* = 0 \\
\nabla e_i(\boldsymbol{x}^*)^T \boldsymbol{d} &= 0,\; i \in \mathcal{E},
\end{aligned}
$$

then $\boldsymbol{x}^*$ *is a strict local minimum of* **NLP**.

Existence and Uniqueness of Lagrange Multipliers

Example:

$$
\begin{array}{ll}
\min & -x_1 \\
\text{s.t.} & x_2 - (1 - x_1)^3 \leq 0 \\
 & x_1 \geq 0 \\
 & x_2 \geq 0
\end{array}
$$

- $\boldsymbol{x}^* = (1, 0)^T$ is the strict local minimum
- Cannot find a KKT point, $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$
- *Linear Independence Constraint Qualification does not hold at* $(1, 0)^T$
- Add an extra constraint

$$2x_1 + x_2 \leq 2$$

- Lagrange multipliers are *not unique*

Importance of Constraint Set Representation

$$\begin{aligned} \min \quad & (x_1 - \tfrac{9}{4})^2 + (x_2 - 2)^2 \\ \text{s.t.} \quad & x_1^2 - x_2 \leq 0 \\ & x_1 + x_2 \leq 6 \\ & x_1 \geq 0, x_2 \geq 0 \end{aligned}$$

- Convex Programming Problem
- Slater's Constraint Qualification holds
- First order KKT conditions are necessary and sufficient at a global minimum
- KKT point does not have $\boldsymbol{x}^* = (2, 4)^T$
- Solution : $\boldsymbol{x}^* = (\frac{3}{2}, \frac{9}{4})^T$
- Replace the first inequality in the constraints by

$$(x_1^2 - x_2)^3 \leq 0$$

- $(\frac{3}{2}, \frac{9}{4})^T$ is *not regular* for the new constraint representation!

Example: Find the point on the parabola $x_2 = \frac{1}{5}(x_1 - 1)^2$ that is closest to $(1, 2)^T$, in the Euclidean norm sense.

$$\begin{array}{ll} \min & (x_1 - 1)^2 + (x_2 - 2)^2 \\ \text{s.t.} & (x_1 - 1)^2 = 5x_2 \end{array}$$

- $\boldsymbol{x}^*, \mu^*$ is a KKT point : $\boldsymbol{x}^* = (1, 0)^T$ and $\mu^* = -\frac{4}{5}$
- Satisfies second order sufficiency conditions
- $\boldsymbol{x}^* = (1, 0)^T$ is a strict local minimum
- Reformulation to an unconstrained optimization problem

Unbounded problem

Example:

$$\begin{array}{ll} \min & -0.2(x_1-3)^2+x_2^2 \\ \text{s.t.} & x_1^2+x_2^2 \geq 1 \end{array}$$

- Unbounded objective function
- $(1,0)^T$ is a strict local minimum

Example:

$$\begin{array}{ll} \min & x_1^2 + x_2^2 + \frac{1}{4}x_3^2 \\ \text{s.t.} & -x_1 + x_3 = 1 \\ & x_1^2 + x_2^2 - 2x_1 = 1 \end{array}$$

- $(1-\sqrt{2}, 0, 2-\sqrt{2})^T$ is a strict local minimum.
- $(1+\sqrt{2}, 0, 2+\sqrt{2})^T$ is a strict local maximum.