8.1 INTRODUCTION IN CONSTRAINED OPTIMIZATION
Notations

• Problem Formulation

\[
\min_{x \in \mathbb{R}^n} f(x) \text{ subject to } \begin{cases} 
  c_i(x) = 0 & i \in \mathcal{E} \\
  c_i(x) \geq 0 & i \in \mathcal{I} 
\end{cases}
\]

• Feasible set

\[ \Omega = \{ x | c_i(x) = 0, i \in \mathcal{E}; \ c_i(x) \geq 0, i \in \mathcal{I} \} \]

• Compact formulation

\[
\min_{x \in \Omega} f(x) 
\]
Local and Global Solutions

• Constraints make the problem simpler since the search space is smaller.
• But it can also make things more complicated.
  \[
  \min (x_2 + 100)^2 + 0.01x_1^2 \quad \text{subject to } x_2 - \cos x_1 \geq 0
  \]
• Unconstrained problem has one minimum, constrained problem has MANY minima.
Types of Solutions

- Similar as the unconstrained case, except that we now restrict it to a neighborhood of the solution.
- Recall, we aim only for local solutions.

A vector $x^*$ is a *local solution* of the problem (12.3) if $x^* \in \Omega$ and there is a neighborhood $\mathcal{N}$ of $x^*$ such that $f(x) \geq f(x^*)$ for $x \in \mathcal{N} \cap \Omega$.

A point $x^*$ is an *isolated local solution* if $x^* \in \Omega$ and there is a neighborhood $\mathcal{N}$ of $x^*$ such that $x^*$ is the only local solution in $\mathcal{N} \cap \Omega$.

A vector $x^*$ is a *strict local solution* (also called a *strong local solution*) if $x^* \in \Omega$ and there is a neighborhood $\mathcal{N}$ of $x^*$ such that $f(x) > f(x^*)$ for all $x \in \mathcal{N} \cap \Omega$ with $x \neq x^*$.
• It is ESSENTIAL that the problem be formulated with smooth constraints and objective function (since we will take derivatives).

• Sometimes, the problem is just badly phrased. For example, when it is done in terms of max function. Sometimes the problem can be rephrased as a constrained problem with SMOOTH constrained functions.

\[
\max \{ f_1(x), f_2(x) \} \leq a \iff \begin{cases} 
    f_1(x) \leq a \\ 
    f_2(x) \leq a 
\end{cases}
\]
Examples of max nonsmoothness removal

- In Constraints:

\[ \|x\|_1 = |x_1| + |x_2| \leq 1 \iff \max \{-x_1, x_1\} + \max \{-x_2, x_2\} \leq 1 \iff \\
- x_1 - x_2 \leq 1, \quad x_1 - x_2 \leq 1, \quad -x_1 + x_2 \leq 1, \quad x_1 + x_2 \leq 1 \]

- In Optimization:

\[ \min f(x); \quad f(x) = \max \{x^2, x\}; \quad \iff \left\{ \begin{array}{l}
\min \\
\text{subject to} \quad \max \{x^2, x\} \leq t
\end{array} \right. \]

\[ \iff \left\{ \begin{array}{l}
\min \\
\text{subject to} \quad x^2 \leq t, \; x \leq t
\end{array} \right. \]
8.2 EXAMPLES
Examples

• Single equality constraint (put in KKT form)
  \[ \min x_1 + x_2 \quad \text{subject to} \quad x_1^2 + x_2^2 - 2 = 0 \]

• Single inequality constraint (put in KKT form, point out complementarity relationship)
  \[ \min x_1 + x_2 \quad \text{subject to} \quad -(x_1^2 + x_2^2 - 2) \geq 0 \]

• Two inequality constraints (KKT, complementarity relationship, sign of the multiplier)
  \[ \min x_1 + x_2 \quad \text{subject to} \quad -(x_1^2 + x_2^2 - 2) \geq 0, \quad x_1 \geq 0 \]
There are two solutions for the Lagrangian equation, but only one is the right.
8.3 IMPLICIT FUNCTION
THEOREM REVIEW
3.5 The Implicit Function Theorem

Key Points in this Section.

1. **One-Variable Version.** If \( f : (a, b) \to \mathbb{R} \) is \( C^1 \) and if \( f'(x_0) \neq 0 \), then locally near \( x_0 \), \( f \) has a \( C^1 \) inverse function \( x = f^{-1}(y) \). If \( f'(x) > 0 \) on all of \((a, b)\) and is continuous on \([a, b]\), then \( f \) has an inverse defined on \([f(a), f(b)]\). This result is used in one-variable calculus to define, for example, the log function as the inverse of \( f(x) = e^x \) and \( \sin^{-1} \) as the inverse of \( f(x) = \sin x \).

2. **Special \( n \)-variable Version.** If \( F : \mathbb{R}^{n+1} \to \mathbb{R} \) is \( C^1 \) and at a point \((x_0, z) \in \mathbb{R}^{n+1}\), \( F(x_0, z) = 0 \) and \( \frac{\partial F}{\partial z}(x_0, z_0) \neq 0 \), then locally near \((x_0, z_0)\) there is a unique solution \( z = g(x) \) of the equation \( F(x, z) = 0 \). We say that \( F(x, z) = 0 \) *implicitly defines* \( z \) as a function of \( x = (x_1, \ldots, x_n) \).
3. The partial derivatives are computed by *implicit differentiation*:

\[
\frac{\partial F}{\partial x_i} + \frac{\partial F}{\partial z} \frac{\partial z}{\partial x_i} = 0,
\]

so

\[
\frac{\partial z}{\partial x_i} = -\frac{\partial F/\partial x_i}{\partial F/\partial z}
\]

4. The special implicit function theorem guarantees that if \( \nabla g(x_0) \neq 0 \), then the level set \( g = c \) is a smooth surface near \( x_0 \), a fact needed in the proof of the Lagrange multiplier theorem.
5. The general implicit function theorem deals with solving \( m \) equations

\[
\begin{align*}
F_1(x_1, \ldots, x_n, z_1, \ldots, z_m) &= 0 \\
\vdots & \quad \vdots \\
F_m(x_1, \ldots, x_n, z_1, \ldots, z_m) &= 0
\end{align*}
\]

for \( m \) unknowns \( \mathbf{z} = (z_1, \ldots, z_m) \). If

\[
\begin{vmatrix}
\frac{\partial F_1}{\partial z_1} & \ldots & \frac{\partial F_1}{\partial z_m} \\
\frac{\partial F_2}{\partial z_1} & \ldots & \frac{\partial F_2}{\partial z_m} \\
\vdots & \quad \vdots & \quad \vdots \\
\frac{\partial F_m}{\partial z_1} & \ldots & \frac{\partial F_m}{\partial z_m}
\end{vmatrix} \neq 0
\]

at \( (x_0, \mathbf{z}_0) \), then these equations define \( (z_1, \ldots, z_m) \) as functions of \( (x_1, \ldots, x_n) \). The partial derivatives \( \frac{\partial z_i}{\partial x_j} \) may again be computed by using implicit differentiation.
8.4 FIRST-ORDER OPTIMALITY CONDITIONS FOR NONLINEAR PROGRAMMING
Inequality Constraints: Active Set

\[
\min_{x \in \mathbb{R}^n} f(x) \text{ subject to } \begin{cases} 
    c_i(x) = 0 & i \in \mathcal{E} \\
    c_i(x) \geq 0 & i \in \mathcal{I}
\end{cases}
\]

- One of the key differences with equality constraints.
- Definition at a feasible point x.

\[
x \in \Omega(x) \quad A(x) = \mathcal{E} \cup \{ i \in \mathcal{I}; c_i(x) = 0 \}
\]
“Constraint Qualifications” for inequality constraints

• We need the equivalent of the “Jacobian has full rank” condition for the case with equality-only.
• This is called “the constraint qualification”.
• Intuition: “geometry of feasible set” = “algebra of feasible set”
Tangent and linearized cone

- Tangent Cone at \( x \) (can prove it is a cone)
  \[
  T_{\Omega}(x) = \left\{ d \mid \exists \{z_k\} \in \Omega, z_k \to x, \exists \{t_k\} \in \mathbb{R}_+, t_k \to 0, \lim_{k \to \infty} \frac{z_k - x}{t_k} = d \right\}
  \]

- Linearized feasible direction set (EXPAND)
  \[
  \mathcal{F}(x) = \left\{ d \mid d^T \nabla c_i(x) = 0, i \in \mathcal{E}; d^T \nabla c_i(x) \geq 0, i \in \mathcal{A}(x) \cap \mathcal{I} \right\} \Rightarrow T_{\Omega}(x) \subseteq \mathcal{F}(x)
  \]

- Essence of constraint qualification at a point \( x \) 
  ("geometry=algebra"): 
  \[
  T_{\Omega}(x) = \mathcal{F}(x)
  \]
What are sufficient conditions for constraint qualification?

- The most common (and only one we will discuss in the class): the linear independence constraint qualification (LICQ).
- We say that LICQ holds at a point $x \in \Omega$ if $\nabla c_{A(x)}$ has full row rank.
- How do we prove equality of the cones? If LICQ holds, then, from IFT

\[
d \in F(x) \Rightarrow c_{A(x)}(\tilde{x}(t)) = t\nabla c_{A(x)}d \Rightarrow \exists \tau > 0, \forall 0 < t < \tau; \\
c_{A(x)}(\tilde{x}(t)) > 0; c_{A(x) \cap I}(\tilde{x}(t)) \geq 0; c_{E}(\tilde{x}(t)) = 0 \Rightarrow \tilde{x}(t) \in \Omega \Rightarrow d \in T_{\Omega}(x)
\]
8.4.1 OPTIMALITY CONDITIONS FOR EQUALITY CONSTRAINTS
IFT for optimality conditions in the equality-only case

- **Problem:** \((NLP) \min f(x) \text{ subject to } c(x) = 0; \ c : \mathbb{R}^n \to \mathbb{R}^m\)

- **Assumptions:**
  1. \(x^*\) is a solution
  2. LICQ: \(\nabla c(x)\) has full row rank.

- From LICQ:
  \[\exists x^* = \begin{pmatrix} x_D^* & x_H^* \end{pmatrix}; \ \nabla c_H(x^*) \in \mathbb{R}^{m \times m}; \ \nabla c_H(x^*) \text{ invertible.}\]

- From IFT:
  \[\exists \mathcal{N}(x^*), \Psi(x_D), \mathcal{N}(x_D^*)\text{ such that } x \in \mathcal{N}(x^*) \cap \Omega \iff x_H = \Psi(x_D)\]

- As a result \(x^*\) is a solution of NLP iff \(x_D^*\) solves unconstrained problem:
  \[\min_{x_D} f\left(x_D, \Psi(x_D)\right)\]
Properties of Mapping

• From IFT:

\[ c(x_D, \Psi(x_D)) = 0 \Rightarrow \nabla_{x_D} c(x_D, \Psi(x_D)) + \nabla_{x_H} c(x_D, \Psi(x_D)) \nabla_{x_D} \Psi(x_D) = 0 \]

• Two important consequences

\[
(1) \nabla_{x_D} \Psi(x_D) = -\left[ \nabla_{x_H} c(x_D, \Psi(x_D)) \right]^{-1} \nabla_{x_D} c(x_D, \Psi(x_D)) \\
(2) Z = \begin{bmatrix} I_{n-m} \\ \nabla_{x_D} \Psi(x_D) \end{bmatrix} \Rightarrow \nabla c(x) Z = 0 \Rightarrow \text{Im}[Z] = \ker[\nabla c(x)]
\]
First-order optimality conditions

- Optimality of unconstrained optimization problem
  \[ \nabla_{x_D} f(x^*_D, \Psi(x^*_D)) = 0 \Rightarrow \nabla_{x_D} f(x^*_D, \Psi(x^*_D)) + \nabla_{x_\mathcal{H}} f(x^*_D, \Psi(x^*_D)) \nabla_{x_D} \Psi(x^*_D) = 0 \Rightarrow \]
  \[ \nabla_{x_D} f(x^*_D, \Psi(x^*_D)) - \nabla_{x_\mathcal{H}} f(x^*_D, \Psi(x^*_D)) \left[ \nabla_{x_\mathcal{H}} c(x_D, \Psi(x_D)) \right]^{-1} \nabla_{x_D} c(x_D, \Psi(x_D)) = 0 \]

- The definition of the Lagrange Multiplier Result in the first-order (Lagrange, KKT) conditions:
  \[
  \begin{bmatrix}
  \nabla_{x_D} f(x^*_D, \Psi(x^*_D)) & \nabla_{x_\mathcal{H}} f(x^*_D, \Psi(x^*_D))
  \end{bmatrix}
  - \lambda^T
  \begin{bmatrix}
  \nabla_{x_D} c(x_D, \Psi(x_D)) & \nabla_{x_\mathcal{H}} c(x_D, \Psi(x_D))
  \end{bmatrix}
  = 0
  \]

  \[
  \nabla f(x^*) - \lambda^T \nabla c(x^*) = 0
  \]
A more abstract and general proof

• Optimality of unconstrained optimization problem

\[ D_{x_D} f\left(x_D^*, \Psi(x_D^*)\right) = 0 \Rightarrow \nabla_{x_D} f\left(x_D^*, \Psi(x_D^*)\right) + \nabla_{\mathcal{H}} f\left(x_D^*, \Psi(x_D^*)\right) \nabla_{x_D} \Psi(x_D^*) = 0 \Rightarrow \nabla_x f(x^*) Z = 0 \]

• Using \( \ker M \perp \text{Im } M^T \); \( \dim(\ker M) + \dim(\text{Im } M^T) = \text{nr cols } M \)

• We obtain: \( \nabla_x f(x^*) Z = 0 \Rightarrow \nabla_x f(x^*)^T \in \ker(Z^T) = \text{Im}\left[\nabla c(x^*)^T\right] \)

• We thus obtain the optimality conditions:

\[ \exists \lambda \in \mathbb{R}^m \text{ s.t. } \nabla_x f(x^*)^T = \nabla_x c(x^*)^T \lambda \Rightarrow \nabla_x f(x^*) - \lambda^T \nabla_x c(x^*) = 0 \]
The Lagrangian

- Definition: $\mathcal{L}(x, \lambda) = f(x) - \lambda^T c(x)$
- Its gradient:
  $\nabla \mathcal{L}(x, \lambda) = \left[ \nabla f(x) - \lambda^T \nabla c(x), c(x)^T \right]$
- Its Hessian:
  $\nabla^2 \mathcal{L}(x, \lambda) = \begin{bmatrix} \nabla^2_{xx} \mathcal{L}(x, \lambda) & \nabla c(x)^T \\ \nabla c(x) & 0 \end{bmatrix}$
- Where:
  $\nabla^2_{xx} \mathcal{L}(x, \lambda) = \nabla^2_{xx} f(x, \lambda) - \sum_{i=1}^{m} \lambda_i \nabla^2_{xx} c_i(x, \lambda)$
- Optimality conditions:
  $\nabla \mathcal{L}(x, \lambda) = 0$
Second-order conditions

- First, note that:
  \[ Z^T \nabla_{xx}^2 L(x_D, \Psi(x_D)) Z = D_{x_D x_D}^2 f(x_D, \Psi(x_D)) \geq 0 \]

- Sketch of proof: total derivatives in \( x_D \)
  \[ D_{x_D} f(x_D, \Psi(x_D)) = \nabla_{x_D} f(x_D, \Psi(x_D)) - \lambda(x_D, \Psi(x_D))^T \nabla_{x_D} c(x_D^*, \Psi(x_D)) = \]
  \[ \nabla_{x_D} L((x_D, \Psi(x_D)), \lambda(x_D, \Psi(x_D))); \]
  \[ \nabla_{x_D^H} f(x_D^*, \Psi(x_D^*)) = \lambda(x_D, \Psi(x_D))^T \nabla_{x_D^H} c(x_D^*, \Psi(x_D^*)) \]

- Second derivatives:
  \[ D_{x_D x_D} f(x_D, \Psi(x_D)) = \nabla_{x_D} f(x_D, \Psi(x_D)) - \lambda(x_D, \Psi(x_D))^T \nabla_{x_D} c(x_D, \Psi(x_D)) = \]
  \[ \nabla_{x_D x_D} L((x_D, \Psi(x_D)), \lambda(x_D, \Psi(x_D))) + \nabla_{x_D} \Psi(x_D)^T \nabla_{x_D^H x_D} L((x_D, \Psi(x_D)), \lambda(x_D, \Psi(x_D))) \]
  \[ -D_D \left( \lambda(x_D, \Psi(x_D))^T \right) \nabla_{x_D} c(x_D, \Psi(x_D)) \]
Computing Second-Order Derivatives

• Expressing the second derivatives of Lagrangian

$\nabla_{x_H} f(x^*_D, \Psi(x^*_D)) = \lambda(x_D, \Psi(x_D))^T \nabla_{x_H} c(x_D, \Psi(x_D)) \Rightarrow$

$$D_{x_p} \left[ \lambda(x_D, \Psi(x_D))^T \right] \nabla_{x_H} c(x_D, \Psi(x_D)) = D_{x_p} \left[ \nabla_{x_H} f(x_D, \Psi(x_D)) - \lambda(x_D, \Psi(x_D))^T \nabla_{x_H} c(x_D, \Psi(x_D)) \right] =$$

$$D_{x_p} \nabla_{x_H} \mathcal{L} \left( (x_D, \Psi(x_D), \lambda(x_D, \Psi(x_D))^T \right) = \nabla_{x_D} \nabla_{x_H} \mathcal{L} \left( (x_D, \Psi(x_D), \lambda(x_D, \Psi(x_D))^T \right) +$$

$$\nabla_{x_D} \Psi(x_D)^T \nabla_{x_H} \nabla_{x_H} \mathcal{L} \left( (x_D, \Psi(x_D), \lambda(x_D, \Psi(x_D))^T \right)$$

• Solve for total derivative of multiplier and replace conclusion follows.
Summary: Necessary Optimality Conditions

- Summary: \( \nabla \mathcal{L}(x^*, \lambda^*) = 0; \ Z^T \nabla^2_{xx} \mathcal{L}(x_D^*, \Psi(x_D^*)) Z \succeq 0 \)

- Rephrase first order:
  \( \nabla_x \mathcal{L}(x^*, \lambda^*) = 0 \)

- Rephrase second order necessary conditions.
  \( \nabla_x c(x^*) w = 0 \Rightarrow w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w \geq 0 \)
Sufficient Optimality Conditions

- The point is a local minimum if LICQ and the following holds:

\[ \nabla_x \mathcal{L}(x^*, \lambda^*) = 0; \quad (2) \nabla_x c(x^*) w = 0 \implies \exists \sigma > 0 \ n^T \nabla^2 x \mathcal{L}(x^*, \lambda^*) w \geq \sigma \|w\|^2 \]

- Proof: By IFT, there is a change of variables such that

\[
\begin{align*}
    u \in \mathcal{N}(0) \subset \mathbb{R}^{n-n_c} &\leftrightarrow x(u); \\
    \tilde{x} \in \mathcal{N}(x^*), c(\tilde{x}) = 0 &\iff \exists \tilde{u} \in \mathcal{N}(0); \tilde{x} = x(\tilde{u}) \\
    \nabla_x c(x^*) \nabla_u x(\tilde{u}) \bigg|_{\tilde{u}=0} &= 0; \quad Z = \nabla_u x(\tilde{u})
\end{align*}
\]

- The original problem can be phrased as

\[ \min_u f(x(u)) \]
Sufficient Optimality Conditions

- We can now piggy back on theory of unconstrained optimization, noting that.
  \[
  \nabla_u f(x(u))|_{u=0} = \nabla_x \mathcal{L}(x^*, \lambda^*) = 0; \\
  \nabla^2_{uu} f(x(u))|_{u=0} = Z^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z > 0; \ Z = \nabla_u x(u)
  \]

- Then from theory of unconstrained optimization we have a local isolated minimum at 0 and thus the original problem at $x^*$. (following the local isomorphism above)
Another Essential Consequence

• If LICQ+ second-order conditions hold at the solution $x^*$, then the following matrix must be nonsingular

• (EXPAND).

$$
\begin{bmatrix}
\nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) & \nabla_x c(x^*) \\
\n\nabla^T_x c(x^*) & 0
\end{bmatrix}
$$

• The system of nonlinear equations has an invertible Jacobian,

$$
\begin{bmatrix}
\nabla_x \mathcal{L}(x^*, \lambda^*) \\
c(x^*)
\end{bmatrix} = 0
$$
8.4.2 FIRST-ORDER OPTIMALITY CONDITIONS FOR MIXED EQ AND INEQ CONSTRAINTS
The Lagrangian

- Even in the general case, it has the same expression

\[ \mathcal{L}(x) = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{A}} \lambda_i c_i(x) \]
First-Order Optimality Condition Theorem

Suppose that \( x^* \) is a local solution of (12.1), that the functions \( f \) and \( c_i \) in (12.1) are continuously differentiable, and that the LICQ holds at \( x^* \). Then there is a Lagrange multiplier vector \( \lambda^* \), with components \( \lambda_i^* \), \( i \in \mathcal{E} \cup \mathcal{I} \), such that the following conditions are satisfied at \( (x^*, \lambda^*) \)

\[
\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \quad (12.34a)
\]
\[
c_i(x^*) = 0, \quad \text{for all } i \in \mathcal{E}, \quad (12.34b)
\]
\[
c_i(x^*) \geq 0, \quad \text{for all } i \in \mathcal{I}, \quad (12.34c)
\]
\[
\lambda_i^* \geq 0, \quad \text{for all } i \in \mathcal{I}, \quad (12.34d)
\]
\[
\lambda_i^* c_i(x^*) = 0, \quad \text{for all } i \in \mathcal{E} \cup \mathcal{I}. \quad (12.34e)
\]

Equivalent Form:

\[
\nabla f(x^*) - \lambda^T_{\mathcal{A}(x^*)} \nabla c_{\mathcal{A}(x^*)}(x^*) = 0 \Rightarrow \text{Multipliers are unique} \]
Sketch of the Proof

• If $x^*$ is a solution of the original problem, it is also a solution of the problem.

$$\min f(x) \text{ subject to } c_{A(x^*)}(x) = 0$$

• From the optimality conditions of the problem with equality constraints, we must have (since LICQ holds)

$$\exists \{\lambda_i\}_{i \in A(x^*)} \text{ such that } \nabla f(x^*) - \sum_{i \in A(x^*)} \lambda_i \nabla c_i(x^*) = 0$$

• But I cannot yet tell by this argument

$$\lambda_i \geq 0$$
Sketch of the Proof: The sign of the multiplier

- Assume now one multiplier has the “wrong” sign. That is:
  \[ j \in A(x^*) \cap \mathcal{I}, \quad \lambda_j < 0 \]

- Since LICQ holds, we can construct a feasible path that “takes off” from that constraint (inactive constraints do not matter locally)

- Define \( b = \frac{d}{dt} \tilde{x}(t) \rvert_{t=0} \Rightarrow \nabla c_{A(x)} b = e_j \)

\[
\begin{align*}
  c_{A(x^*)}(\tilde{x}(t)) &= te_j \Rightarrow \tilde{x}(t) \in \Omega \\
  \frac{d}{dt} f(\tilde{x}(t))_{t=0} &= \nabla f(\tilde{x})^T b = \lambda^T_{c_{A(x)}} \nabla c_{A(x)} b = \lambda_j < 0 \\
  \exists t_1 > 0, \quad f(\tilde{x}(t_1)) &= f(\tilde{x}(0)) = f(x^*), \quad \text{CONTRADICTION}!!
\end{align*}
\]
Strict Complementarity

• It is a notion that makes the problem look “almost” like an equality.

**Definition 12.5** (Strict Complementarity).

*Given a local solution* $x^*$ *of* (12.1) *and a vector* $\lambda^*$ *satisfying* (12.34), *we say that the strict complementarity condition holds if exactly one of* $\lambda_i^*$ *and* $c_i(x^*)$ *is zero for each index* $i \in \mathcal{I}$. *In other words, we have that* $\lambda_i^* > 0$ *for each* $i \in \mathcal{I} \cap A(x^*)$. 
8.5 SECOND-ORDER CONDITIONS
Critical Cone

- The subset of the tangent space, where the objective function does not vary to first-order.
- The book definition.

\[ \mathcal{C}(x^*, \lambda^*) = \{ w \in \mathcal{F}(x^*) \mid \nabla c_i(x^*)^T w = 0, \text{ all } i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ with } \lambda^*_i > 0 \}. \]

- An even simpler equivalent definition.

\[ \mathcal{C}(x^*, \lambda^*) = \left\{ w \in T_\Omega(x^*) \mid \nabla f(x^*)^T w = 0 \right\} \]
Rephrasing of the Critical Cone

• By investigating the definition

\[
\begin{align*}
    w \in C(x^*, \lambda^*) &\iff \\
    \begin{cases}
        \nabla c_i(x^*)^T w = 0 & i \in \mathcal{E} \\
        \nabla c_i(x^*)^T w = 0 & i \in A(x^*) \cap \mathcal{I} \quad \lambda_i^* > 0 \\
        \nabla c_i(x^*)^T w \geq 0 & i \in A(x^*) \cap \mathcal{I} \quad \lambda_i^* = 0
    \end{cases}
\end{align*}
\]

• In the case where strict complementarity holds, the cones has a MUCH simplex expression.

\[
w \in C(x^*, \lambda^*) \iff \nabla c_i(x^*) w = 0 \ \forall \ i \in A(x^*)
\]
Statement of the Second-Order Conditions

**Theorem 12.5** (Second-Order Necessary Conditions).

Suppose that $x^*$ is a local solution of (12.1) and that the LICQ condition is satisfied. Let $\lambda^*$ be the Lagrange multiplier vector for which the KKT conditions (12.34) are satisfied. Then

$$w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w \geq 0, \quad \text{for all } w \in \mathcal{C}(x^*, \lambda^*).$$

(12.57)

• How to prove this? In the case of Strict Complementarity the critical cone is the same as the problem constrained with equalities on active index.

• Result follows from equality-only case.
Statement of second-order sufficient conditions

**Theorem 12.6** (Second-Order Sufficient Conditions).

Suppose that for some feasible point $x^* \in \mathbb{R}^n$ there is a Lagrange multiplier vector $\lambda^*$ such that the KKT conditions (12.34) are satisfied. Suppose also that

$$ w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w > 0, \quad \text{for all } w \in C(x^*, \lambda^*), \ w \neq 0. \quad (12.65) $$

Then $x^*$ is a strict local solution for (12.1).

- How do we prove this? In the case of strict complementarity again from reduction to the equality case.

$$ x^* = \arg \min_x f(x) \text{ subject to } c_A(x) = 0 $$
How to derive those conditions in the other case?

- Use the slacks to reduce the problem to one with equality constraints.

\[
\begin{align*}
\min_{x \in \mathbb{R}^n, z \in \mathbb{R}^{n_I}} & \quad f(x) \\
\text{s.t.} & \quad c_E(x) = 0 \\
& \quad [c_I(x)]_j - z_j^2 = 0 \quad j = 1, 2, \ldots, n_I
\end{align*}
\]

- Then, apply the conditions for equality constraints.

- I will assign it as homework.
Summary: Why should I care about Lagrange Multipliers?

• Because it makes the optimization problem in principle equivalent to a nonlinear equation.

\[
\begin{bmatrix}
\nabla_x \mathcal{L}(x^*, \lambda^*) \\
c_A(x^*)
\end{bmatrix} = 0; \quad \det \begin{bmatrix}
\nabla^2_x \mathcal{L}(x^*, \lambda^*) & \nabla_x c_A(x^*) \\
\nabla^T_x c_A(x^*) & 0
\end{bmatrix} \neq 0
\]

• I can use concepts from nonlinear equations such as Newton’s for the algorithmics.