CHICAGO

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

$$\Omega=\left\{x \mid c_i(x)=0, i\in\mathcal{E};\quad c_i(x)\geq 0, i\in\mathcal{I}\right\}$$

- Compact formulation

$$\min_{x\in\Omega} f(x)$$

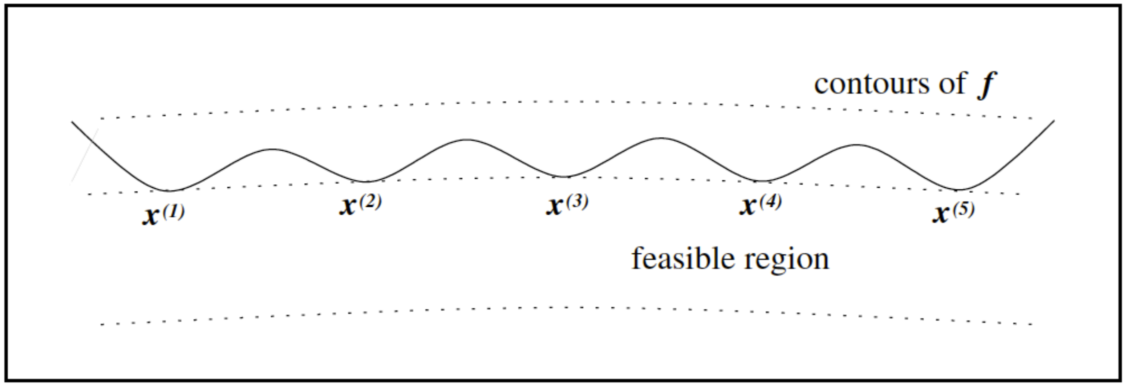
CHICAGO

Local and Global Solutions

- Constraints make make the problem simpler since the search space is smaller.
- But it can also make things more complicated.

$$\min\left(x_2+100\right)^2+0.01x_1^2 \text{ subject to } x_2-\cos x_1\geq 0$$

- 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 $x \neq x^*$.

Smoothness

- 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.

$$\max\left\{f_1(x), f_2(x)\right\} \le a \Leftrightarrow \begin{cases} f_1(x) \le a \\ f_2(x) \le a \end{cases}$$

Examples of max nonsmoothness removal

- In Constraints:

$$\|x\|_1 = |x_1| + |x_2| \le 1 \Leftrightarrow \max\{-x_1, x_1\} + \max\{-x_2, x_2\} \le 1 \Leftrightarrow$$
$$-x_1 - x_2 \le 1, \quad x_1 - x_2 \le 1, \quad -x_1 + x_2 \le 1, \quad x_1 + x_2 \le 1$$

- In Optimization:

$$\min f(x); \quad f(x) = \max\{x^2, x\}; \quad \Leftrightarrow \begin{cases} \min & t \\ \text{subject to} & \max\{x^2, x\} \le t \end{cases}$$

$$\Leftrightarrow \begin{cases} \min & t \\ \text{subject to} & x^2 \le t, x \le t \end{cases}$$

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)

$$\min x_1 + x_2 \quad \text{subject to} \quad -\left(x_1^2 + x_2^2 - 2\right) \geq 0$$

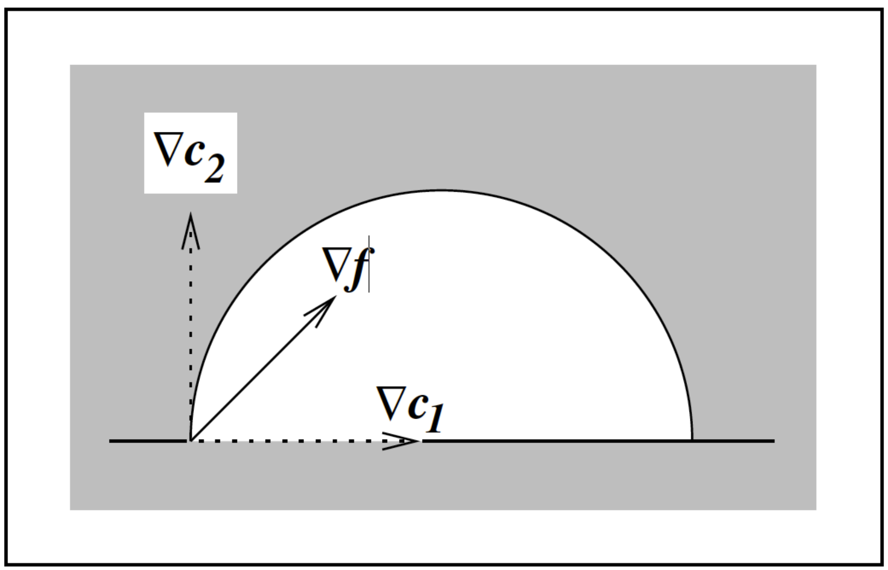
- Two inequality constraints (KKT, complementarity relationship, sign of the multiplier)

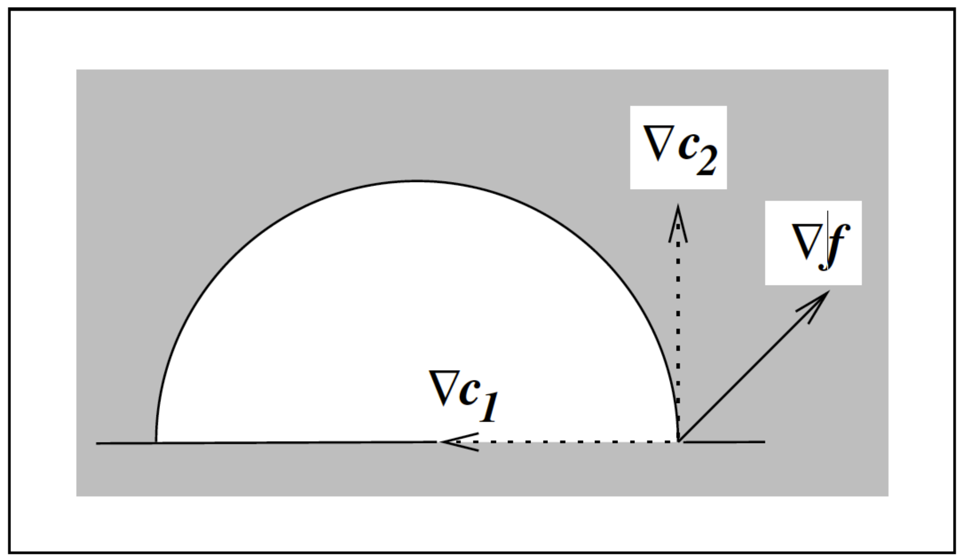
$$\min x_1 + x_2 \quad \text{subject to} \quad -\left(x_1^2 + x_2^2 - 2\right) \geq 0, x_1 \geq 0$$

CHICAGO

Multiplier Sign Example

- There are two solutions for the Lagrangian equation, but only one is the right.





8.3 IMPLICIT FUNCTION THEOREM REVIEW

Refresher (Marsden and Tromba)

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 $(\mathbf{x}_0, z) \in \mathbb{R}^{n+1}$, $F(\mathbf{x}_0, z) = 0$ and $\frac{\partial F}{\partial z}(\mathbf{x}_0, z_0) \neq 0$, then locally near $(\mathbf{x}_0, z_0)$ there is a unique solution $z = g(\mathbf{x})$ of the equation $F(\mathbf{x}, z) = 0$. We say that $F(\mathbf{x}, z) = 0$ ***implicitly defines*** z as a function of $\mathbf{x} = (x_1, \ldots, x_n)$.

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 $\nabla g(\mathbf{x}_0) \neq \mathbf{0}$, then the level set $g = c$ is a smooth surface near $\mathbf{x}_0$, a fact needed in the proof of the Lagrange multiplier theorem.

5. The general implicit function theorem deals with solving m equations

$$\begin{aligned} F_1(x_1,\ldots,x_n,z_1,\ldots,z_m) &= 0 \\ \vdots \quad & \quad \vdots \\ F_m(x_1,\ldots,x_n,z_1,\ldots,z_m) &= 0 \end{aligned}$$

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

$$\begin{vmatrix} \frac{\partial F_1}{\partial z_1} & \cdots & \frac{\partial F_1}{\partial z_m} \\ \vdots & & \vdots \\ \frac{\partial F_m}{\partial z_1} & \cdots & \frac{\partial F_m}{\partial z_m} \end{vmatrix} \neq 0$$

at $(\mathbf{x}_0, \mathbf{z}_0)$, then these equations define $(z_1,\ldots,z_m)$ as functions of $(x_1,\ldots,x_n)$. The partial derivatives $\partial z_i/\partial x_j$ 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.

$$x \in \Omega(x) \quad \mathcal{A}(x) = \mathcal{E} \cup \{ i \in \mathcal{I}; c_i(x) = 0 \}$$

“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)

$$T_\Omega(x) = \left\{ d \,\middle|\, \exists \{z_k\} \in \Omega, z_k \to x, \exists \{t_k\} \in \mathbb{R}_+, t_k \to 0, \lim_{k\to\infty} \frac{z_k - x}{t_k} = d \right\}$$

- Linearized feasible direction set (EXPAND)

$$\mathcal{F}(x) = \left\{ d \,\middle|\, d^T \nabla c_i(x) = 0, i \in \mathcal{E}; d^T \nabla c_i(x) \ge 0, i \in \mathcal{A}(x) \cap \mathcal{I} \right\} \Rightarrow T_\Omega(x) \subset \mathcal{F}(x)$$

- Essence of constraint qualification at a point x ("geometry=algebra"):

$$T_\Omega(x) = \mathcal{F}(x)$$

CHICAGO

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

$$d \in \mathcal{F}(x) \Rightarrow c_{\mathcal{A}(x)}\left(\tilde{x}(t)\right) = t\nabla c_{\mathcal{A}(x)} d \Rightarrow \exists \tau > 0, \forall 0 < t < \tau;$$

$$c_{\bar{\mathcal{A}}(x)}\left(\tilde{x}(t)\right) > 0; c_{\mathcal{A}(x)\cap\mathcal{I}}\left(\tilde{x}(t)\right) \geq 0; c_{\mathcal{E}}\left(\tilde{x}(t)\right) = 0 \Rightarrow \tilde{x}(t) \in \Omega \Rightarrow d \in T_{\Omega}(x)$$

CHICAGO

8.4.1 OPTIMALITY CONDITIONS FOR EQUALITY CONSTRAINTS

CHICAGO

IFT for optimality conditions in the equality-only case

- Problem: $\boxed{(NLP)\ \min f(x) \text{ subject to } c(x)=0;\ c:\mathbb{R}^n \to \mathbb{R}^m}$
- Assumptions:
  1. x^* is a solution
  2. LICQ: $\nabla c(x)$ has full row rank.
- From LICQ: $\exists\, x^* = \left(\overbrace{x_{\mathcal{D}}^*}^{n-m}, \overbrace{x_{\mathcal{H}}^*}^{m} \right); \nabla c_{\mathcal{H}}(x^*) \in \mathbb{R}^{m\times m}; \nabla c_{\mathcal{H}}(x^*)$ invertible.
- From IFT:

$$\exists \mathcal{N}(x^*), \Psi(x_{\mathcal{D}}), \mathcal{N}(x_{\mathcal{D}}^*) \text{ such that } x \in \mathcal{N}(x^*) \cap \Omega \Leftrightarrow x_{\mathcal{H}} = \Psi(x_{\mathcal{D}})$$

- As a result x^* is a solution of NLP iff $x_{\mathcal{D}}^*$ solves unconstrained problem:
  – $$\boxed{\min_{x_{\mathcal{D}}} f\left(x_{\mathcal{D}}, \Psi(x_{\mathcal{D}})\right)}$$

Properties of Mapping

- From IFT:

$$c\left(x_{\mathcal{D}},\Psi\left(x_{\mathcal{D}}\right)\right)=0\Rightarrow\nabla_{x_{\mathcal{D}}}c\left(x_{\mathcal{D}},\Psi\left(x_{\mathcal{D}}\right)\right)+\nabla_{x_{\mathcal{H}}}c\left(x_{\mathcal{D}},\Psi\left(x_{\mathcal{D}}\right)\right)\nabla_{x_{\mathcal{D}}}\Psi\left(x_{\mathcal{D}}\right)=0$$

- Two important consequences

$$(1)\ \nabla_{x_{\mathcal{D}}}\Psi\left(x_{\mathcal{D}}\right)=-\left[\nabla_{x_{\mathcal{H}}}c\left(x_{\mathcal{D}},\Psi\left(x_{\mathcal{D}}\right)\right)\right]^{-1}\nabla_{x_{\mathcal{D}}}c\left(x_{\mathcal{D}},\Psi\left(x_{\mathcal{D}}\right)\right)$$

$$(2) Z=\begin{bmatrix} I_{n-m} \\ \nabla_{x_{\mathcal{D}}}\Psi\left(x_{\mathcal{D}}\right)\end{bmatrix}\Rightarrow\nabla c(x)Z=0\Rightarrow\operatorname{Im}[Z]=\ker\left[\nabla c(x)\right]$$

CHICAGO

First-order optimality conditions

- Optimality of unconstrained optimization problem

$$\nabla_{x_{\mathcal{D}}} f\left(x_{\mathcal{D}}^*,\Psi\left(x_{\mathcal{D}}^*\right)\right)=0 \Rightarrow \nabla_{x_{\mathcal{D}}} f\left(x_{\mathcal{D}}^*,\Psi\left(x_{\mathcal{D}}^*\right)\right)+\nabla_{x_{\mathcal{H}}} f\left(x_{\mathcal{D}}^*,\Psi\left(x_{\mathcal{D}}^*\right)\right)\nabla_{x_{\mathcal{D}}}\Psi\left(x_{\mathcal{D}}^*\right)=0 \Rightarrow$$

$$\nabla_{x_{\mathcal{D}}} f\left(x_{\mathcal{D}}^*,\Psi\left(x_{\mathcal{D}}^*\right)\right)-\underbrace{\nabla_{x_{\mathcal{H}}} f\left(x_{\mathcal{D}}^*,\Psi\left(x_{\mathcal{D}}^*\right)\right)\left[\nabla_{x_{\mathcal{H}}} c\left(x_{\mathcal{D}},\Psi\left(x_{\mathcal{D}}\right)\right)\right]^{-1}}_{\lambda^T}\nabla_{x_{\mathcal{D}}} c\left(x_{\mathcal{D}},\Psi\left(x_{\mathcal{D}}\right)\right)=0$$

- The definition of the Lagrange Multiplier Result in the first-order (Lagrange, KKT) conditions:

$$\begin{bmatrix} \nabla_{x_{\mathcal{D}}} f\left(x_{\mathcal{D}}^*,\Psi\left(x_{\mathcal{D}}^*\right)\right) & \nabla_{x_{\mathcal{H}}} f\left(x_{\mathcal{D}}^*,\Psi\left(x_{\mathcal{D}}^*\right)\right) \end{bmatrix}-\lambda^T\begin{bmatrix} \nabla_{x_{\mathcal{D}}} c\left(x_{\mathcal{D}},\Psi\left(x_{\mathcal{D}}\right)\right) & \nabla_{x_{\mathcal{H}}} c\left(x_{\mathcal{D}}^*,\Psi\left(x_{\mathcal{D}}^*\right)\right) \end{bmatrix}=0$$

$$\nabla f\left(x^*\right)-\lambda^T\nabla c\left(x^*\right)=0$$

A more abstract and general proof

- Optimality of unconstrained optimization problem

$$D_{x_{\mathcal{D}}} f\left(x_{\mathcal{D}}^*, \Psi\left(x_{\mathcal{D}}^*\right)\right) = 0 \Rightarrow \nabla_{x_{\mathcal{D}}} f\left(x_{\mathcal{D}}^*, \Psi\left(x_{\mathcal{D}}^*\right)\right) + \nabla_{x_{\mathcal{H}}} f\left(x_{\mathcal{D}}^*, \Psi\left(x_{\mathcal{D}}^*\right)\right) \nabla_{x_{\mathcal{D}}} \Psi\left(x_{\mathcal{D}}^*\right) = 0 \Rightarrow \nabla_x f\left(x^*\right) Z = 0$$

- Using $\ker M \perp \operatorname{Im} M^T; \dim\left(\ker M\right) + \dim\left(\operatorname{Im} M^T\right) = \text{nr cols M}$
- We obtain: $\nabla_x f\left(x^*\right) Z = 0 \Rightarrow \nabla_x f\left(x^*\right)^T \in \ker\left(Z^T\right) = \operatorname{Im}\left[\nabla c\left(x^*\right)^T\right]$
- We thus obtain the optimality conditions:

$$\exists \lambda \in \mathbb{R}^m \text{ s.t. } \nabla_x f\left(x^*\right)^T = \nabla_x c\left(x^*\right)^T \lambda \Rightarrow \nabla_x f\left(x^*\right) - \lambda^T \nabla_x c\left(x^*\right) = 0$$

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: $\boxed{\nabla\mathcal{L}(x,\lambda)=0}$

Second-order conditions

- First, note that: $Z^T \nabla^2_{xx} L\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right) Z = D^2_{x_{\mathcal{D}} x_{\mathcal{D}}} f\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right) \succ= 0$
- Sketch of proof: total derivatives in : $x_{\mathcal{D}}$

$$D_{x_{\mathcal{D}}} f\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right) = \nabla_{x_{\mathcal{D}}} f\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right) - \lambda\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right)^T \nabla_{x_{\mathcal{D}}} c\left(x^*_{\mathcal{D}}, \Psi\left(x^*_{\mathcal{D}}\right)\right) =$$

$$\nabla_{x_{\mathcal{D}}} \mathcal{L}\left(\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right), \lambda\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right)\right);$$

$$\nabla_{x_{\mathcal{H}}} f\left(x^*_{\mathcal{D}}, \Psi\left(x^*_{\mathcal{D}}\right)\right) = \lambda\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right)^T \nabla_{x_{\mathcal{H}}} c\left(x^*_{\mathcal{D}}, \Psi\left(x^*_{\mathcal{D}}\right)\right)$$

- Second derivatives:

$$D_{x_{\mathcal{D}} x_{\mathcal{D}}} f\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right) = \nabla_{x_{\mathcal{D}}} f\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right) - \lambda\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right)^T \nabla_{x_{\mathcal{D}}} c\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right) =$$

$$\nabla_{x_{\mathcal{D}} x_{\mathcal{D}}} \mathcal{L}\left(\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right), \lambda\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right)\right) + \nabla_{x_{\mathcal{D}}} \Psi\left(x_{\mathcal{D}}\right)^T \nabla_{x_{\mathcal{H}} x_{\mathcal{D}}} \mathcal{L}\left(\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right), \lambda\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right)\right)$$

$$-D_{\mathcal{D}}\left(\lambda\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right)^T\right) \nabla_{x_{\mathcal{D}}} c\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right)$$

Computing Second-Order Derivatives

- Expressing the second derivatives of Lagrangian

$$\nabla_{x_{\mathcal{H}}} f\left(x_{\mathcal{D}}^{*}, \Psi\left(x_{\mathcal{D}}^{*}\right)\right) = \lambda\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right)^{T} \nabla_{x_{\mathcal{H}}} c\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right) \Rightarrow$$

$$D_{x_{\mathcal{D}}}\left[\lambda\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right)^{T}\right] \nabla_{x_{\mathcal{H}}} c\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right) = D_{x_{\mathcal{D}}}\left[\nabla_{x_{\mathcal{H}}} f\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right) - \underbrace{\lambda\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right)^{T}}_{inactive} \nabla_{x_{\mathcal{H}}} c\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right)\right] =$$

$$D_{x_{\mathcal{D}}} \nabla_{x_{\mathcal{H}}} \mathcal{L}\left(\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right), \underbrace{\lambda\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right)^{T}}_{inactive}\right) = \nabla_{x_{\mathcal{D}}} \nabla_{x_{\mathcal{H}}} \mathcal{L}\left(\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right), \lambda\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right)^{T}\right) +$$

$$\nabla_{x_{\mathcal{D}}} \Psi\left(x_{\mathcal{D}}\right)^{T} \nabla_{x_{\mathcal{H}}} \nabla_{x_{\mathcal{H}}} \mathcal{L}\left(\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right), \lambda\left(x_{\mathcal{D}}, \Psi\left(x_{\mathcal{D}}\right)\right)^{T}\right)$$

- Solve for total derivative of multiplier and replace conclusion follows.

Summary: Necessary Optimality Conditions

- Summary: $\boxed{\nabla \mathcal{L}\left(x^*,\lambda^*\right)=0;\ Z^T \nabla^2_{xx} L\left(x^*_{\mathcal{D}},\Psi\left(x^*_{\mathcal{D}}\right)\right) Z \succ= 0}$

- Rephrase first order:

$$\nabla_x \mathcal{L}\left(x^*,\lambda^*\right)=0$$

- Rephrase second order necessary conditions.

$$\nabla_x c\left(x^*\right) w = 0 \Rightarrow w^T \nabla^2_{xx} \mathcal{L}\left(x^*,\lambda^*\right) w \geq 0$$

CHICAGO

Sufficient Optimality Conditions

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

$$(1)\nabla_x \mathcal{L}\left(x^*,\lambda^*\right)=0;\ (2)\nabla_x c\left(x^*\right)w=0 \Rightarrow \exists\sigma>0\ w^T\nabla^2_{xx}\mathcal{L}\left(x^*,\lambda^*\right)w \geq \sigma\|w\|^2$$

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

$$u\in\mathcal{N}(0)\subset\mathbb{R}^{n-n_c}u\leftrightarrow x(u);\ \tilde{x}\in\mathcal{N}\left(x^*\right),c(\tilde{x})=0\Leftrightarrow\exists\tilde{u}\in\mathcal{N}(0);\ \tilde{x}=x(\tilde{u})$$

$$\nabla_x c\left(x^*\right)\nabla_u x(\tilde{u})\Big|_{\tilde{u}=0}=0;\quad Z=\nabla_u x(\tilde{u})$$

- The original problem can be phrased as

$$\min_u f\left(x(u)\right)$$

Sufficient Optimality Conditions

- We can now piggy back on theory of unconstrained optimization, noting that.

$$\nabla_u f\left(x(u)\right)\Big|_{u=0} = \nabla_x \mathcal{L}\left(x^*,\lambda^*\right) = 0;$$

$$\nabla^2_{uu} f\left(x(u)\right)\Big|_{u=0} = Z^T \nabla^2_{xx} \mathcal{L}\left(x^*,\lambda^*\right) Z \succ 0;\ Z = \nabla_u x(u)$$

- 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).

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

- The system of nonlinear equations has an invertible Jacobian,

$$\begin{bmatrix} \nabla_x\mathcal{L}\left(x^*,\lambda^*\right) \\ c\left(x^*\right) \end{bmatrix} = 0$$

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

$$\mathcal{L}(x) = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{A}} \lambda_i c_i(x)$$

CHICAGO

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 λ^*, with components λ_i^*, $i \in \mathcal{E} \cup \mathcal{I}$, such that the following conditions are satisfied at (x^*, λ^*)

$$
\begin{aligned}
\nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, & & \qquad (12.34a)\\
c_i(x^*) &= 0, & \text{for all } i \in \mathcal{E}, & \qquad (12.34b)\\
c_i(x^*) &\geq 0, & \text{for all } i \in \mathcal{I}, & \qquad (12.34c)\\
\lambda_i^* &\geq 0, & \text{for all } i \in \mathcal{I}, & \qquad (12.34d)\\
\lambda_i^* c_i(x^*) &= 0, & \text{for all } i \in \mathcal{E} \cup \mathcal{I}. & \qquad (12.34e)
\end{aligned}
$$

Equivalent Form:

$$\nabla f\left(x^*\right) - \lambda^T_{\mathcal{A}(x^*)} \nabla c_{\mathcal{A}(x^*)}\left(x^*\right) = 0 \Rightarrow \text{Multipliers are unique !!}$$

Sketch of the Proof

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

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

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

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

- 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

$$j \in \mathcal{A}\left(x^*\right) \bigcap \mathcal{I}, \quad \lambda_j < 0$$

- Since LICQ holds, we can construct a feasible path that "takes off" from that constraint (inactive constraints do not matter locally)
- $$c_{\mathcal{A}\left(x^*\right)}\left(\tilde{x}(t)\right) = te_j \Rightarrow \tilde{x}(t) \in \Omega \quad \text{Define } b = \frac{d}{dt}\tilde{x}(t)_{t=0} \Rightarrow \nabla c_{\mathcal{A}(x)} b = e_j$$

$$\frac{d}{dt} f\left(\tilde{x}(t)\right)_{t=0} = \nabla f\left(x^*\right)^T b = \lambda^T_{c_{\mathcal{A}(x)}} \nabla c_{\mathcal{A}(x)} b = \lambda_j < 0 \quad \Rightarrow$$

$$\exists t_1 > 0, \quad f\left(\tilde{x}(t_1)\right) < f\left(\tilde{x}(0)\right) = f\left(x^*\right), \quad \text{CONTRADICTION!!}$$

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 λ^* satisfying (12.34), we say that the strict complementarity condition *holds if* exactly one *of λ_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 $i \in \mathcal{I} \cap \mathcal{A}(x^*)$.*

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.

$$\mathcal{C}\left(x^*, \lambda^*\right) = \left\{ w \in T_\Omega\left(x^*\right) \middle| \nabla f\left(x^*\right)^T w = 0 \right\}$$

Rephrasing of the Critical Cone

- By investigating the definition

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

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

$$w \in \mathcal{C}\left(x^*,\lambda^*\right) \Leftrightarrow \nabla c_i\left(x^*\right) w = 0 \ \forall\, i \in \mathcal{A}\left(x^*\right)$$

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 λ^* be the Lagrange multiplier vector for which the KKT conditions (12.34) are satisfied. Then

$$w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w \geq 0, \quad \text{for all } w \in \mathcal{C}(x^*, \lambda^*). \tag{12.57}$$

- 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 λ^* such that the KKT conditions (12.34) are satisfied. Suppose also that

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

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.

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

How to derive those conditions in the other case?

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

$$
\begin{array}{llll}
\min_{x\in\mathbb{R}^n, z\in\mathbb{R}^{n_I},} & f(x) & & \\
s.t. & c_E(x) & = 0 & \\
 & \left[c_I(x)\right]_j - z_j^2 & = 0 & j = 1,2,\dots n_I
\end{array}
$$

- 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.

$$\begin{bmatrix} \nabla_x \mathcal{L}\left(x^*, \lambda^*\right) \\ c_{\mathcal{A}}\left(x^*\right) \end{bmatrix} = 0; \ \det \begin{bmatrix} \nabla_{xx}^2 \mathcal{L}\left(x^*, \lambda^*\right) & \nabla_x c_{\mathcal{A}}\left(x^*\right) \\ \nabla_x^T c_{\mathcal{A}}\left(x^*\right) & 0 \end{bmatrix} \neq 0$$

- I can use concepts from nonlinear equations such as Newton's for the algorithmics.