Ch.20 Problems with Equality Constraints

20.1 Introduction

- Solving a class of nonlinear constrained optimization problems

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h(x) = 0 \\ & g(x) \le 0 \end{aligned}$$

- **Def. 20.1** *Any point satisfying the constraints is called a feasible point. The set of all feasible points*

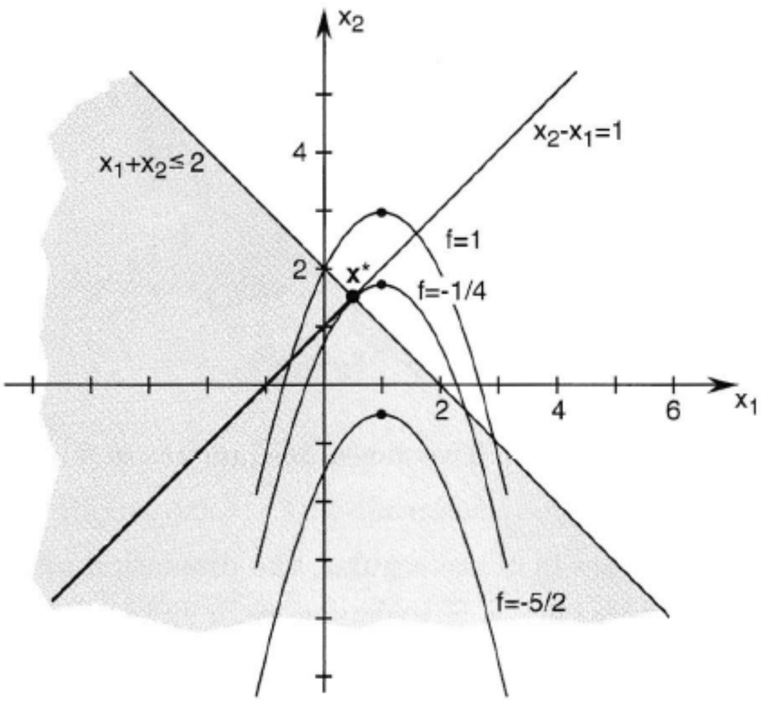
$$\{x \in \Re^n : h(x) = 0,\ g(x) \le 0\}$$

is called the feasible set.

- **Ex. 20.1**

$$\begin{aligned} \text{minimize} \quad & (x_1 - 1)^2 + x_2 - 2 \\ \text{subject to} \quad & x_2 - x_1 = 1 \\ & x_1 + x_2 \le 2 \end{aligned}$$

♣ Fig. 20.1 on page 367.



- We discuss constrained optimization problem with only equality constraints.

20.2 Problem Formulation

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

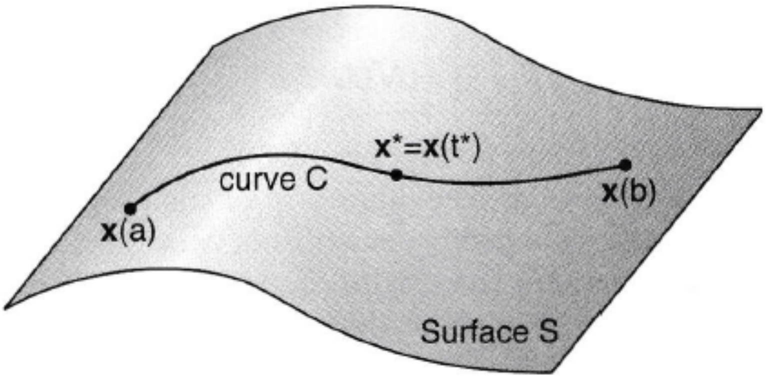
where $x \in \Re^n$, $f : \Re^n \to \Re$, $h : \Re^n \to \Re^m$,
$h = [h_1, \cdots, h_m]^T$, and $m \leq n$.

- **Def. 20.2** *A point x^* satisfying the constraints $h_1(x^*) = 0, \cdots, h_m(x^*) = 0$ is said to be a regular point of the constraints if the gradient vectors $\nabla h_1(x^*), \cdots, \nabla h_m(x^*)$ are linearly independent.*

20.3 Tangent and Normal Spaces

- **Def. 20.3** *A curve C on a surface S is a set of points $\{x(t) \in S : t \in (a,b)\}$, continuously parameterized by $t \in (a,b)$; that is, $x : (a,b) \to S$ is a continuous function.*

  ♣ Fig. 20.4 on page 369.



- **Def. 20.4** *The curve $C = \{x(t) : t \in (a,b)\}$ is differentiable if*

$$\dot{x}(t) = \begin{bmatrix} \dot{x}_1(t) \\ \vdots \\ \dot{x}_n(t) \end{bmatrix}$$

  exists for all $t \in (a,b)$.

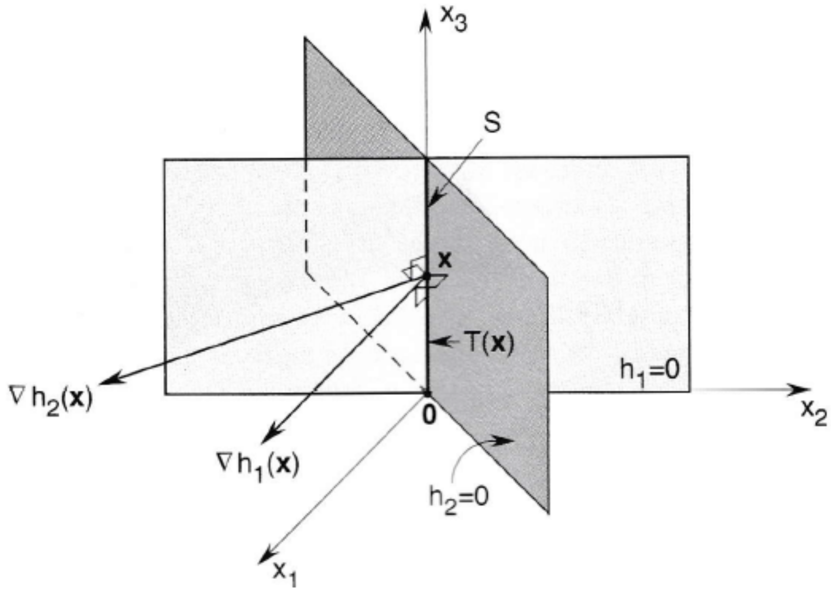
- **Def. 20.5** *The tangent space at a point x^* on the surface $S = \{x \in \Re^n : h(x) = 0\}$ is the set*

$$T(x^*) = \{y : Dh(x^*)y = 0\}$$

- **Ex. 20.4** Let

$S = \{x \in \Re^3 : h_1(x) = x_1 = 0,\ h_2(x) = x_1 - x_2 = 0\}$

♣ Fig. 20.8 on page 373.



$$Dh(x) = \begin{bmatrix} \nabla h_1(x)^T \\ \nabla h_2(x)^T \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix}$$

The tangent space at an arbitrary point of S is

$$\begin{aligned} T(x) &= \{y : \nabla h_1(x)^T y = 0,\ \nabla h_2(x)^T y = 0\} \\ &= \left\{ y : \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = 0 \right\} \\ &= \{[0, 0, \alpha]^T : \alpha \in \Re\} \\ &= \text{the } x_3\text{-axis in } \Re^3 \end{aligned}$$

- **Def. 20.6** *The normal space* $N(x^*)$ *at a point* x^* *on the surface* $S = \{x \in \Re^n : h(x) = 0\}$ *is the set*

$$N(x^*) = \{x \in \Re^n : x = Dh(x^*)^T z,\ z \in \Re^m\}$$

- **Lemma 20.1** *we have* $T(x^*) = N(x^*)^{\perp}$ *and* $T(x^*)^{\perp} = N(x^*)$.

20.4 Lagrange Condition

- First-order necessary condition for extremum problems with constraints. $\Rightarrow$ Lagrange's theorem
- **Theorem 20.2** *Lagrarange's Thoerem for* $n = 2,\ m = 1$. *Let the point* x^* *be a minimizer of* $f : \Re^2 \to \Re$ *subject to the constraint* $h(x) = 0,\ h : \Re^2 \to \Re$. *Then,* $\nabla f(x^*)$ *and* $\nabla h(x^*)$ *are parallel. That is, if* $\nabla h(x^*) \neq 0$, *then there exists a scalar* λ^* *such that*

$$\nabla f(x^*) + \lambda^* \nabla h(x^*) = 0$$

We refer to λ^* as the Lagrange multiplier.

- Lagrange condition

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

Note that the Lagrange condition is only necessary but not sufficient.

- **Theorem 20.3 Lagrange's Theorem.** *Let x^* be local minimizer (or maximizer) of $f : \Re^n \to \Re$, subject to $h(x) = 0$, $h : \Re^n \to \Re^m$, $m \le n$. Assume that x^* is a regular point. Then, there exists $\lambda^* \in \Re^m$ such that*

$$Df(x^*) + \lambda^{*T} Dh(x^*) = 0^T$$

- *We define the Lagrangian function $l : \Re^n \times \Re^m \to \Re$, given by*

$$l(x, \lambda) \equiv f(x) + \lambda^T h(x)$$

- The Lagrange condition for a local minimizer x^* is

$$Dl(x^*, \lambda^*) = 0^T$$

for some λ^*, where the derivative operation D is with repect to the entire argument $[x^T, \lambda^T]^T$.

- Denote the derivative of l with respect to x as $D_x l$, and the derivative of l with respect to λ as $D_\lambda l$.

$$Dl(x, \lambda) = [D_x l(x, \lambda), D_\lambda l(x, \lambda)]$$

- The Lagrange's theorem for a local minimizer x^* is

$$\begin{aligned} D_x l(x^*, \lambda^*) &= 0^T \\ D_\lambda l(x^*, \lambda^*) &= 0^T \end{aligned}$$

- Again, the Lagrange condition is only necessary, not sufficient.

- **Ex. 20.6** We wish to construct a closed card-board box with maximum volume.

$$\begin{array}{rl} \text{maximize} & x_1x_2x_3 \\ \text{subject to} & x_1x_2 + x_2x_3 + x_3x_1 = \dfrac{A}{2} \end{array}$$

From the Lagarange theorem,

$$\begin{aligned} l(x,\lambda) &\equiv -x_1x_2x_3 + \lambda(x_1x_2 + x_2x_3 + x_3x_1 - \frac{A}{2}) \\ D_{x_1}l &= x_2x_3 - \lambda(x_2 + x_3) = 0 \\ D_{x_2}l &= x_1x_3 - \lambda(x_1 + x_3) = 0 \\ D_{x_3}l &= x_1x_2 - \lambda(x_1 + x_2) = 0 \\ D_{\lambda}l &= x_1x_2 + x_2x_3 + x_3x_1 - \frac{A}{2} = 0 \end{aligned}$$

The solution that satisfies the above conditions is $x_1 = x_2 = x_3 = \sqrt{A/6},\ \lambda = \sqrt{A/24}$.

- **Ex. 20.7**

$$\begin{aligned} f(x) &= x_1^2 + x_2^2 \\ \{[x_1, x_2]^T : h(x) &= x_1^2 + 2x_2^2 - 1 = 0\} \end{aligned}$$

Following the Lagrange theorem,

$$\begin{aligned} l(x_1, x_2, \lambda) &= x_1^2 + x_2^2 + \lambda[x_1^2 + 2x_2^2 - 1] \\ D_{x_1}l &= 2x_1 + 2\lambda x_1 = 0 \\ D_{x_2}l &= 2x_2 + 4\lambda x_2 = 0 \\ D_{\lambda}l(x_1, x_2, \lambda) &= x_1^2 + 2x_2^2 - 1 = 0 \end{aligned}$$

The points that satisfy the Lagrange condition for extrema are

$$x^{(1)} = \begin{bmatrix} 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad x^{(2)} = \begin{bmatrix} 0 \\ -1/\sqrt{2} \end{bmatrix},$$

$$x^{(3)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad x^{(4)} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

$x^{(1)}$ and $x^{(2)}$ are minimizers, and $x^{(3)}$ and $x^{(4)}$ are maximizers.

20.5 Second-Order Conditions

- We assume that $f : \Re^n \to \Re$ and $h : \Re^n \to \Re^m$ are twice continuously differentiable, that is, $f, h \in C^2$.

  $l(x, \lambda) = f(x) + \lambda^T h(x) = f(x) + \lambda_1 h_1(x) + \cdots + \lambda_m h_m(x)$

- Let $L(x, \lambda)$ be the Hessian matrix of $l(x, \lambda)$ with respect to x,

$$L(x, \lambda) = F(x) + \lambda_1 H_1(x) + \cdots + \lambda_m H_m(x)$$

where $F(x)$ is the Hessian matrix of f at x, and $H_k(x)$

is the Hessian matrix of h_k at x, $k = 1, \cdots, m$,

$$H_k(x) = \begin{bmatrix} \frac{\partial^2 h_k}{\partial x_1^2}(x) & \cdots & \frac{\partial^2 h_k}{\partial x_n \partial x_1}(x) \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 h_k}{\partial x_1 \partial x_n}(x) & \cdots & \frac{\partial^2 h_k}{\partial^2 x_n}(x) \end{bmatrix}$$

or

$$L(x, \lambda) = F(x) + [\lambda H(x)]$$

$$\lambda H(x) = \lambda_1 H_1(x) + \cdots + \lambda_m H_m(x)$$

- **Theorem 20.4 Second-Order Necessary Conditions.** *Let x^* be a local minimizer of $f : \Re^n \to \Re$ subject to $h(x) = 0$, $h : \Re^n \to \Re^m$, $m \leq n$, and $f, h \in C^2$. Suppose x^* is regular. Then, there exists $\lambda^* \in \Re^m$ such that*

  1. $Df(x^*) + \lambda^{*T} Dh(x^*) = 0^T$

  2. for all $y \in T(x^*)$, we have $y^T L(x^*, \lambda^*) y \geq 0$.

- **Theorem 20.5 Second-Order Sufficient Condition.** *Suppose $f, h \in C^2$ and there exist a point $x^* \in \Re^n$ and $\lambda^* \in \Re^n$ such that*

  1. $Df(x^*) + \lambda^* Dh(x^*) = 0^T$

  2. for all $y \in T(x^*)$, $y \neq 0$, we have $y^T L(x^*, \lambda^*) y \geq 0$.

  Then, x^* is a strict local minimizer of f subject to $h(x) = 0$.

20.6 Minimizing Quadratic Subject to Linear Constraints

$$\begin{array}{rl} \text{minimize} & \frac{1}{2}x^T Q x \\ \text{subject to} & Ax = b \end{array}$$

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

- Lagrangian function

$$\begin{aligned} l(x,\lambda) &= \frac{1}{2}x^T Q x + \lambda^T(b - Ax) \\ D_x l(x^*,\lambda^*) &= x^{*T}Q - \lambda^{*T}A = 0^T \\ x^* &= Q^{-1}A^T\lambda^* \\ Ax^* &= AQ^{-1}A^T\lambda^* \\ \lambda^* &= (AQ^{-1}A^T)^{-1}b \\ x^* &= Q^{-1}A^T(AQ^{-1}A^T)^{-1}b \end{aligned}$$

- **Ex. 20.10** linear quadratic regulator (LQR) problem

$$\begin{array}{rl} \text{minimize} & \frac{1}{2}\sum_{i=1}^{N}(Qx_i^2 + ru_i^2) \\ \text{subject to} & x_k = ax_{k-1} + bu_k, \;\; k = 1,\cdots,N, \;\; x_0 \text{ given} \end{array}$$

$$Q = \begin{bmatrix} qI_N & O \\ O & rI_N \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & \cdots & 0 & -b & \cdots & 0 \\ -a & 1 & \cdots & 0 & -b & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & -a & 1 & 0 & \cdots & -b \end{bmatrix}$$

$$b = \begin{bmatrix} ax_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad z = [x_1, \cdots, x_N, u_1, \cdots, u_N]^T$$

- The solution can be

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} z^T Q z \\ \text{subject to} \quad & Az = b \end{aligned}$$

where Q is $2N \times 2N$, A is $N \times 2N$, and $b \in \Re^N$.

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