Nonlinear Optimization

Derek Li

Contents

1 Review **2**
1.1 One-Variable Calculus 2
1.2 Multi-variable Calculus 3
1.2.1 Mean Value Theorems and Taylor Approximations 3
1.2.2 Implicit Function Theorem 4
1.2.3 Convexity 4
1.2.4 Extreme Value Theorem 8
1.3 Matrix Calculus 8
1.3.1 Matrix Multiplication 8
1.3.2 Partitioned Matrices 9
1.3.3 Matrix Differentiation 9

2 Finite Dimensional Optimization **11**
2.1 Unconstrained Optimization 11
2.2 Equality Constraints 15
2.2.1 Tangent Space 15
2.2.2 Lagrange Multiplier 15
2.3 Inequality Constraints 15

1 Review

1.1 One-Variable Calculus

Theorem 1.1 (Mean Value Theorem)**.** Let $g \in C^1$ on $\mathbb{R}$. We have

$$\frac{g(x+h) - g(x)}{h} = g'(x + \theta h)$$

for some $\theta \in (0,1)$ and $\frac{g(x+h)-g(x)}{h}$ is the slope of secant line between $(x, g(x))$ and $(x+h, g(x+h))$. Or we can write $g(x+h) = g(x) + hg'(x+\theta h)$.

Theorem 1.2 (First Order Taylor Approximation)**.** Let $g \in C^1$ on $\mathbb{R}$. We have

$$g(x+h) = g(x) + hg'(x) + o(h)$$

where $o(h)$ is the error and we say a function $f(h) = o(h)$ to mean

$$\lim_{h \to 0} \frac{f(h)}{h} = 0$$

Proof. Want to show $g(x+h) - g(x) - hg'(x) = o(h)$.

We have

$$\begin{aligned} \lim_{h \to 0} \frac{g(x+h) - g(x) - hg'(x)}{h} &= \lim_{h \to 0} \frac{hg'(x+\theta h) - hg'(x)}{h} \\ &= \lim_{h \to 0} g'(x+\theta h) - g'(x) = 0 \end{aligned}$$

□

Theorem 1.3 (Second Order Mean Value Theorem)**.** Let $g \in C^2$ on $\mathbb{R}$. We have

$$g(x+h) = g(x) + hg'(x) + \frac{h^2}{2} g''(x+\theta h)$$

for some $\theta \in (0,1)$.

Theorem 1.4 (Second Order Taylor Approximation)**.** Let $g \in C^2$ on $\mathbb{R}$. We have

$$g(x+h) = g(x) + hg'(x) + \frac{h^2}{2} g''(x) + o(h^2)$$

Proof. W.T.S. $g(x+h) - g(x) - hg'(x) - \frac{h^2}{2} g''(x) = o(h^2)$.

We have

$$\begin{aligned} \lim_{h \to 0} \frac{g(x+h) - g(x) - hg'(x) - \frac{h^2}{2} g''(x)}{h^2} &= \lim_{h \to 0} \frac{\frac{h^2}{2} g''(x+\theta h) - \frac{h^2}{2} g''(x)}{h^2} \\ &= \lim_{h \to 0} \frac{1}{2}[g''(x+\theta h) - g''(x)] = 0 \end{aligned}$$

□

1.2 Multi-variable Calculus

1.2.1 Mean Value Theorems and Taylor Approximations

Definition 1.1 (Gradient)**.** Gradient of $f : \mathbb{R}^n \to \mathbb{R}$ at $\mathbf{x} \in \mathbb{R}^n, \nabla f(\mathbf{x})$, if exists is a vector characterized by the property

$$\lim_{\mathbf{v}\to\mathbf{0}} \frac{f(\mathbf{x}+\mathbf{v}) - f(\mathbf{x}) - \nabla f(\mathbf{x}) \cdot \mathbf{v}}{\|\mathbf{v}\|} = 0$$

and $\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \cdots, \frac{\partial f}{\partial x_n}(\mathbf{x})\right)$.

The instantaneous rate of change of f at $\mathbf{x}$ in direction $\mathbf{v}$ (suppose w.l.o.g. $\|\mathbf{v}\| = 1$) is

$$\begin{aligned}
\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f(\mathbf{x}+t\mathbf{v}) &= \nabla f(\mathbf{x}+t\mathbf{v}) \cdot \mathbf{v}|_{t=0} \\
&= \nabla f(\mathbf{x}) \cdot \mathbf{v} \\
&= |\nabla f(\mathbf{x})||\mathbf{v}| \cos\theta \\
&= |\nabla f(\mathbf{x})| \cos\theta
\end{aligned}$$

where θ is the angle between $\nabla f(\mathbf{x})$ and $\mathbf{v}$. Obviously, the instantaneous rate maximizes when $\theta = 0$. Therefore, when it is not equal to zero, $\nabla f(\mathbf{x})$ points in the direction of steepest ascent.

Theorem 1.5 (Mean Value Theorem in $\mathbb{R}^n$)**.** Let $f \in C^1$ on $\mathbb{R}^n$, then for any $\mathbf{x}, \mathbf{v} \in \mathbb{R}^n$, we have

$$f(\mathbf{x}+\mathbf{v}) = f(\mathbf{x}) + \nabla f(\mathbf{x}+\theta\mathbf{v}) \cdot \mathbf{v}$$

for some $\theta \in (0, 1)$.

Proof. Consider $g(t) = f(\mathbf{x}+t\mathbf{v})$, where $t \in \mathbb{R}$ and $g \in C^1$ on $\mathbb{R}$. We have

$$\begin{aligned}
g'(t) &= \frac{\mathrm{d}}{\mathrm{d}t} f(\mathbf{x}+t\mathbf{v}) \\
&= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}+t\mathbf{v}) \cdot \frac{\mathrm{d}(\mathbf{x}+t\mathbf{v})_i}{\mathrm{d}t} \\
&= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}+t\mathbf{v}) \cdot \frac{\mathrm{d}(x_i+tv_i)}{\mathrm{d}t} \\
&= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}+t\mathbf{v}) \cdot v_i \\
&= \nabla f(\mathbf{x}+t\mathbf{v}) \cdot \mathbf{v}
\end{aligned}$$

By Mean Value Theorem in $\mathbb{R}$, we have

$$\begin{aligned}
g(0+1) &= g(0) + 1 \cdot g'(0+\theta \cdot 1) \\
&= g(0) + g'(\theta) \\
&= f(\mathbf{x}) + \nabla f(\mathbf{x}+\theta\mathbf{v}) \cdot \mathbf{v} \\
&= g(1) = f(\mathbf{x}+\mathbf{v})
\end{aligned}$$

for some $\theta \in (0, 1)$. □

Theorem 1.6 (First Order Taylor Approximation in $\mathbb{R}^n$)**.** Let $f \in C^1$ on $\mathbb{R}^n$. We have

$$f(\mathbf{x}+\mathbf{v}) = f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot \mathbf{v} + o(\|\mathbf{v}\|)$$

Proof. We have

$$\begin{aligned}\lim_{\|\mathbf{v}\|\to 0} \frac{f(\mathbf{x}+\mathbf{v}) - f(\mathbf{x}) - \nabla f(\mathbf{x})\cdot\mathbf{v}}{\|\mathbf{v}\|} &= \lim_{\|\mathbf{v}\|\to 0} \frac{\nabla f(\mathbf{x}+\theta\mathbf{v})\cdot\mathbf{v} - \nabla f(\mathbf{x})\cdot\mathbf{v}}{\|\mathbf{v}\|} \\ &= \lim_{\|\mathbf{v}\|\to 0} [\nabla f(\mathbf{x}+\theta\mathbf{v}) - \nabla f(\mathbf{v})]\cdot\frac{\mathbf{v}}{\|\mathbf{v}\|} = 0\end{aligned}$$

□

Theorem 1.7 (Second Order Mean Value Theorem in $\mathbb{R}^n$)**.** Let $f \in C^2$ on $\mathbb{R}^n$. We have

$$f(\mathbf{x}+\mathbf{v}) = f(\mathbf{x}) + \nabla f(\mathbf{x})\cdot\mathbf{v} + \frac{1}{2}\mathbf{v}^T\nabla^2 f(\mathbf{x}+\theta\mathbf{v})\cdot\mathbf{v}$$

for some $\theta \in (0,1)$.

Note 1. Hessian matrix

$$\nabla^2 f(\mathbf{x}) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x})\right)_{1\leqslant i,j\leqslant n}$$

is a symmetric matrix because of Clairaut's Theorem.

Note 2.

$$\mathbf{v}^T\nabla^2 f(\mathbf{x})\cdot\mathbf{v} = \sum_{1\leqslant i,j\leqslant n} \frac{\partial^2 f}{\partial x_i\partial x_j}(\mathbf{x})v_i v_j.$$

Theorem 1.8 (Second Order Taylor Approximation in $\mathbb{R}^n$)**.** Let $f \in C^2$ on $\mathbb{R}^n$. We have

$$f(\mathbf{x}+\mathbf{v}) = f(\mathbf{x}) + \nabla f(\mathbf{x})\cdot\mathbf{v} + \frac{1}{2}\mathbf{v}^T\nabla^2 f(\mathbf{x})\cdot\mathbf{v} + o(\|\mathbf{v}\|^2).$$

Proof. We have

$$\begin{aligned}&\lim_{\|\mathbf{v}\|\to 0} \frac{f(\mathbf{x}+\mathbf{v}) - f(\mathbf{x}) - \nabla f(\mathbf{x})\cdot\mathbf{v} - \frac{1}{2}\mathbf{v}^T\nabla^2 f(\mathbf{x})\cdot\mathbf{v}}{\|\mathbf{v}\|^2} \\ &= \lim_{\|\mathbf{v}\|\to 0} \frac{\frac{1}{2}\mathbf{v}^T\nabla^2 f(\mathbf{x}+\theta\mathbf{v})\cdot\mathbf{v} - \frac{1}{2}\mathbf{v}^T\nabla^2 f(\mathbf{x})\cdot\mathbf{v}}{\|\mathbf{v}\|^2} \\ &= \lim_{\|\mathbf{v}\|\to 0} \frac{1}{2}\left(\frac{\mathbf{v}}{\|\mathbf{v}\|}\right)^T\cdot[\nabla^2 f(\mathbf{x}+\theta\mathbf{v}) - \nabla^2 f(\mathbf{x})]\cdot\frac{\mathbf{v}}{\|\mathbf{v}\|} \\ &= 0\end{aligned}$$

□

1.2.2 Implicit Function Theorem

Theorem 1.9 (Implicit Function Theorem)**.** Let $f : \mathbb{R}^{n+1} \to \mathbb{R}$ be a C^1 function. Fix $(\mathbf{a}, b) \in \mathbb{R}^n \times \mathbb{R}$ s.t. $f(\mathbf{a}, b) = 0$ and $\nabla f(\mathbf{a}, b) \neq 0$. We have $\{(\mathbf{x}, y) \in \mathbb{R}^n \times \mathbb{R} | f(\mathbf{x}, y) = 0\}$ is locally the graph of a function.

1.2.3 Convexity

Definition 1.2 (Level Set)**.** $\{\mathbf{x} \in \mathbb{R}^n | f(\mathbf{x}) = c\}$ is called c-level set of f.

Theorem 1.10. Gradient $\nabla f(\mathbf{x}_0) \perp$ level curve through $\mathbf{x}_0$.

Definition 1.3 (Convex Set)**.** $\Omega \subseteq \mathbb{R}^n$ is a convex set if for all $\mathbf{x}_1, \mathbf{x}_2 \in \Omega$, we have line segment between $\mathbf{x}_1, \mathbf{x}_2 : s\mathbf{x}_1 + (1-s)\mathbf{x}_2 \in \Omega, s \in [0,1]$.

Definition 1.4 (Convex Function)**.** A function $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ is convex if

$$f(s\mathbf{x}_1 + (1-s)\mathbf{x}_2) \leqslant sf(\mathbf{x}_1) + (1-s)f(\mathbf{x}_2)$$

for all $\mathbf{x}_1, \mathbf{x}_2 \in \Omega$ and all $s \in [0,1]$, where Ω is a convex set.

Example 1.1. $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(\mathbf{x}) = \|\mathbf{x}\|$ is convex.

Proof. Let $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$. We have

$$\begin{aligned} f(s\mathbf{x}_1 + (1-s)\mathbf{x}_2) &= \|s\mathbf{x}_1 + (1-s)\mathbf{x}_2\| \leqslant \|s\mathbf{x}_1\| + \|(1-s)\mathbf{x}_2\| \\ &= s\|\mathbf{x}_1\| + (1-s)\|\mathbf{x}_2\| = s_1 f(\mathbf{x}_1) + (1-s)f(\mathbf{x}_2) \end{aligned}$$

□

Definition 1.5 (Concave Function)**.** A function f concave if $-f$ is convex.

Note. The linear function is both convex and concave.

Theorem 1.11 (Basic Properties of Convex Function)**.** Let $\Omega \subseteq \mathbb{R}^n$ be a convex set.

(1) f_1, f_2 are convex functions on $\Omega \Rightarrow f_1 + f_2$ is convex function on Ω.

(2) f is convex functions and $a \geqslant 0 \Rightarrow af$ is a convex function.

(3) f is a convex function on $\Omega \Rightarrow \mathrm{SL}_c := \{\mathbf{x} \in \mathbb{R}^n | f(x) \leqslant c\}$, the sub-level sets are convex.

Proof of (3). W.T.S. for $\mathbf{x}_1, \mathbf{x}_2 \in \mathrm{SL}_c$, $s\mathbf{x}_1 + (1-s)\mathbf{x}_2 \in \mathrm{SL}_c$ for any $s \in [0,1]$.

Since $\mathbf{x}_1, \mathbf{x}_2 \in \mathrm{SL}_c$, we have $f(\mathbf{x}_1) \leqslant c, f(\mathbf{x}_2) \leqslant c$. Because f is a convex function, we have

$$f(s\mathbf{x}_1 + (1-s)\mathbf{x}_2) \leqslant sf(\mathbf{x}_1) + (1-s)f(\mathbf{x}_2) \leqslant sc + (1-s)c = c$$

Thus, $s\mathbf{x}_1 + (1-s)\mathbf{x}_2 \in \mathrm{SL}_c$. □

Theorem 1.12 (C^1 Criterion for Convexity)**.** Let $f : \Omega \to \mathbb{R}$ be a C^1 function and Ω is a convex subset of $\mathbb{R}^n$. Then f is convex on Ω iff

$$f(\mathbf{y}) \geqslant f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x})$$

for all $\mathbf{x}, \mathbf{y} \in \Omega$.

Proof. ($\Rightarrow$) Suppose f is convex. By definition,

$$\begin{aligned} &f(s\mathbf{x}_1 + (1-s)\mathbf{x}_2) \leqslant sf(\mathbf{x}_1) + (1-s)f(\mathbf{x}_2), 0 \leqslant s \leqslant 1 \\ &\Rightarrow \frac{f(s\mathbf{x}_1 + (1-s)\mathbf{x}_2) - f(\mathbf{x}_2)}{s} \leqslant f(\mathbf{x}_1) - f(\mathbf{x}_2), 0 < s \leqslant 1 \\ &\Rightarrow \lim_{s \to 0} \frac{f(\mathbf{x}_2 + s(\mathbf{x}_1 - \mathbf{x}_2)) - f(\mathbf{x}_2)}{s} \leqslant f(\mathbf{x}_1) - f(\mathbf{x}_2) \end{aligned}$$

Recall that $\partial_{\mathbf{x}_1 - \mathbf{x}_2} f(\mathbf{x}_2) := \lim\limits_{s \to 0} \frac{f(\mathbf{x}_2 + s(\mathbf{x}_1 - \mathbf{x}_2)) - f(\mathbf{x}_2)}{s}$, i.e., the directional derivative of f at $\mathbf{x}_2$ in the direction $\mathbf{x}_1 - \mathbf{x}_2$.

Since f is C^1, we have

$$\nabla f(\mathbf{x}_2) \cdot (\mathbf{x}_1 - \mathbf{x}_2) \leqslant f(\mathbf{x}_1) - f(\mathbf{x}_2) \Rightarrow f(\mathbf{x}_1) \geqslant f(\mathbf{x}_2) + \nabla f(\mathbf{x}_2) \cdot (\mathbf{x}_1 - \mathbf{x}_2)$$

i.e.,

$$f(\mathbf{y}) \geqslant f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x})$$

($\Leftarrow$) Fix $\mathbf{x}_0, \mathbf{x}_1 \in \Omega, s \in (0,1)$. Let $\mathbf{x} = s\mathbf{x}_0 + (1-s)\mathbf{x}_1$. We have

$$\begin{aligned} f(\mathbf{x}_0) &\geqslant f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{x}_0 - \mathbf{x}) = f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (1-s)(\mathbf{x}_0 - \mathbf{x}_1) \\ f(\mathbf{x}_1) &\geqslant f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{x}_1 - \mathbf{x}) = f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot s(\mathbf{x}_1 - \mathbf{x}_0) \end{aligned}$$

Therefore,

$$\begin{aligned} sf(\mathbf{x}_0) &\geqslant sf(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot s(1-s)(\mathbf{x}_0 - \mathbf{x}_1) \\ (1-s)f(\mathbf{x}_1) &\geqslant (1-s)f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot s(1-s)(\mathbf{x}_1 - \mathbf{x}_0) \end{aligned}$$

Thus,

$$sf(\mathbf{x}_0) + (1-s)f(\mathbf{x}_1) \geqslant sf(\mathbf{x}) + (1-s)f(\mathbf{x}) = f(\mathbf{x}) = f(s\mathbf{x}_0 + (1-s)\mathbf{x}_1)$$

for all $\mathbf{x}_0, \mathbf{x}_1 \in \Omega$ and $s \in (0,1)$.

When $s = 0$ or $s = 1, sf(\mathbf{x}_0) + (1-s)f(\mathbf{x}_1) = f(s\mathbf{x}_0 + (1-s)\mathbf{x}_1)$.

In conclusion, f is convex on Ω. □

Theorem 1.13 (C^2 Criterion for Convexity)**.** Let $f \in C^2$ on $\Omega \subseteq \mathbb{R}^n, \Omega$ is a convex set containing an interior point. Then f is convex on Ω iff $\nabla^2 f(\mathbf{x}) \geqslant 0, \forall \mathbf{x} \in \Omega$.

Proof. ($\Leftarrow$) Recall that $\mathbf{A} \geqslant 0$ means $\mathbf{A}$ is an $n \times n$ matrix which is positive semi-definite, $\mathbf{v}^T\mathbf{A}\mathbf{v} \geqslant 0, \forall \mathbf{v} \in \mathbb{R}^n$.

By second order MVT,

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}) + \frac{1}{2}(\mathbf{y} - \mathbf{x})^T \nabla^2 f(\mathbf{x} + s(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x})$$

for some $s \in [0,1]$. Therefore,

$$f(\mathbf{y}) \geqslant f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}), \forall \mathbf{x}, \mathbf{y} \in \Omega$$

i.e., f is convex.

($\Rightarrow$) Suppose $\nabla^2 f(\mathbf{x})$ is not positive semi-definite at some $\mathbf{x} \in \Omega$, then

$$\exists \mathbf{v} \neq \mathbf{0} \text{ s.t. } \mathbf{v}^T \nabla^2 f(\mathbf{x}) \mathbf{v} < 0$$

Note we can assume $\mathbf{x}$ is an interior point of Ω, because if $\mathbf{x}$ is on the boundary, by continuity, we can find $\mathbf{x}' \in B(\mathbf{x}, \varepsilon)$ s.t. $\mathbf{v}\nabla^2 f(\mathbf{x}')\mathbf{v} < 0$.

By continuity, $\exists \mathbf{v}$ s.t.

$$\mathbf{v}^T \nabla^2 f(\mathbf{x} + s\mathbf{v}) \mathbf{v} < 0, \forall s \in [0,1]$$

Let $\mathbf{y} = \mathbf{x} + \mathbf{v}$, then

$$(\mathbf{y} - \mathbf{x})^T \nabla^2 f(\mathbf{x} + s(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x}) < 0, \forall s \in [0,1]$$

By MVT,

$$f(\mathbf{y}) < f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x})$$

for some $\mathbf{x}, \mathbf{y} \in \Omega$, which contradicts C^1 criterion so that $\nabla^2 f(\mathbf{x})$ is positive semi-definite for all $\mathbf{x} \in \Omega$. $\square$

Note that in $\mathbb{R}, f$ is convex $\Leftrightarrow f''(x) \geqslant 0, \forall x \in \Omega$.

Theorem 1.14. Let $f : \Omega \to \mathbb{R}$ be a convex function, where $\Omega \subseteq \mathbb{R}^n$ is convex. Suppose $\Gamma := \left\{\mathbf{x} \in \Omega | f(\mathbf{x} = \min_{\Omega} f(\mathbf{x}))\right\} \neq \varnothing$, i.e., minimizer exists, then Γ is convex set and any local minimum of f is a global minimum.

Proof. Let $m = \min_{\Omega} f(\mathbf{x})$. $\Gamma = \{x \in \Omega | f(\mathbf{x}) = m\} = \{x \in \Omega | f(\mathbf{x}) \leqslant m\}$ is a sub-level set that is convex.

Let $\mathbf{x}$ be a local minimizer. Suppose $\exists \mathbf{y}$ s.t. $f(\mathbf{y}) < f(\mathbf{x})$. We have

$$f(\mathbf{x} + s(\mathbf{y} - \mathbf{x})) \leqslant sf(\mathbf{y}) + (1-s)f(\mathbf{x}) < f(\mathbf{x}), \forall s \in [0,1]$$

When $s \to 0, f(\mathbf{x} + s(\mathbf{y} - \mathbf{x})) \to \mathbf{x}$ with $f(\mathbf{x} + s(\mathbf{y} - \mathbf{x})) < f(\mathbf{x})$ and thus $\mathbf{x}$ cannot be a local minimizer. By contradiction, m is a global minimum. $\square$

Theorem 1.15. Let $f : \Omega \to \mathbb{R}$ be a convex function, where $\Omega \subseteq \mathbb{R}^n$ is convex and compact. Then

$$\max_{\mathbf{x} \in \Omega} f(\mathbf{x}) = \max_{\mathbf{x} \in \partial\Omega} f(\mathbf{x})$$

Proof. Since Ω is closed, $\partial\Omega \subseteq \Omega$. Hence

$$\max_{\mathbf{x} \in \Omega} f(\mathbf{x}) \geqslant \max_{\mathbf{x} \in \partial\Omega} f(\mathbf{x})$$

Suppose $f(\mathbf{x}_0) = \max_{\mathbf{x} \in \Omega} f(\mathbf{x})$, for some $\mathbf{x}_0 \notin \partial\Omega$. Let l be an arbitrary line through $\mathbf{x}_0$.

By convexity and compactness of Ω, l meets $\partial\Omega$ at two points $\mathbf{x}_1, \mathbf{x}_2$, and thus

$$\mathbf{x}_0 = s\mathbf{x}_1 + (1-s)\mathbf{x}_2$$

for some $s \in [0,1]$.

Hence,

$$\begin{aligned} f(\mathbf{x}_0) &= f(s\mathbf{x}_1 + (1-s)\mathbf{x}_2) \leqslant sf(\mathbf{x}_1) + (1-s)f(\mathbf{x}_2) \\ &\leqslant \max_{\mathbf{x} \in \partial\Omega} f(\mathbf{x}) + (1-s) \max_{\mathbf{x} \in \partial\Omega} f(\mathbf{x}) = \max_{\mathbf{x} \in \partial\Omega} f(\mathbf{x}) \end{aligned}$$

Therefore, $\max_{\mathbf{x} \in \Omega} f(\mathbf{x}) = \max_{\mathbf{x} \in \partial\Omega} f(\mathbf{x})$. $\square$

Example 1.2. $|ab| \leqslant \frac{1}{p}|a|^p + \frac{1}{g}|b|^g$, where $p, g > 1$ s.t. $\frac{1}{p} + \frac{1}{g} = 1$.

Proof. We know $-\ln$ is convex and thus

$$\begin{aligned} -\ln|ab| &= -\ln|a| - \ln|b| = -\frac{1}{p}\ln|a|^p - \frac{1}{g}\ln|b|^g \\ &\geqslant -\ln\left(\frac{1}{p}|a|^p + \frac{1}{g}|b|^g\right) \end{aligned}$$

Therefore,

$$\ln|ab| \leqslant \ln\left(\frac{1}{p}|a|^p + \frac{1}{g}|b|^g\right) \Rightarrow |ab| \leqslant \frac{1}{p}|a|^p + \frac{1}{g}|b|^g$$

$\square$

Note that if $p = g = 2$, then $|ab| \leqslant \frac{|a|^2+|b|^2}{2}$.

Remark that f is convex does not imply f is continuous.

1.2.4 Extreme Value Theorem

Recall that if $h_1, \cdots, h_k$ and $g_1, \cdots, g_m$ are continuous functions on $\mathbb{R}^n$, then the set of all points $\mathbf{x} \in \mathbb{R}^n$ that satisfy

$$\begin{cases} h_i(x) = 0, & \forall i \\ g_j(x) \leqslant 0, & \forall j \end{cases}$$

is a closed set.

Theorem 1.16 (Extreme Value Theorem)**.** Let $f : \mathbb{R}^m \to \mathbb{R}$ be continuous and $K \subseteq \mathbb{R}^n$ be a compact set, then the problem $\min\limits_{\mathbf{x} \to K} f(\mathbf{x})$ has a solution.

Proof. Let $m = \inf\limits_{\mathbf{x} \in K} f(\mathbf{x})$, (always exists but may be $-\infty$), then $\exists \{\mathbf{x}_i\} \subset K$ s.t. $f(\mathbf{x}_i) \to m$.

Since K is compact, then $\exists \{\mathbf{x}_{i_j}\}$ a subsequence s.t. $\mathbf{x}_{i_j} \to \mathbf{x}_\infty \in K$.

Since f is continuous, then

$$\begin{cases} f(\mathbf{x}_{i_j}) \to f(\mathbf{x}_\infty) \\ f(\mathbf{x}_{i_j}) \to m \end{cases} \Rightarrow f(\mathbf{x}_\infty) = m = \inf_{\mathbf{x} \in K} f(\mathbf{x}) = \min_{\mathbf{x} \in K} f(\mathbf{x})$$

□

Remark. Computer algorithms for solving minimum problems try to construct a sequence x_i s.t. $f(x_i)$ decreases to minimum value quickly.

Example 1.3. $\min f(x, y, z)$ subject to $\begin{cases} x + y + z = 5 \\ x, y, z \geqslant 0 \end{cases}$, where f is continuous.

There is a solution because $K := \{x + y + z - 5 = 0, -x \leqslant 0, -y \leqslant 0, -z \leqslant 0\}$ is closed and bounded, and hence K is compact.

1.3 Matrix Calculus

1.3.1 Matrix Multiplication

Definition 1.6. Let A be $m \times n, B$ be $n \times p$, and the product be $C = AB$, then C is a $m \times p$ matrix, with element (i, j) given by

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}, \forall i = 1, \cdots, m, j = 1, \cdots, p$$

Property 1.1. Let A be $m \times n, \mathbf{x}$ be $n \times 1$, then the element of the product $\mathbf{z} = A\mathbf{x}$ is given by

$$z_i = \sum_{k=1}^{n} a_{ik} x_k, \forall i = 1, \cdots, m$$

Let $\mathbf{y}$ be $m \times 1$, then the element of the product $\mathbf{z}^T = \mathbf{y}^T A$ is given by

$$z_i^T = \sum_{k=1}^{n} y_k a_{ki}, \forall i = 1, \cdots, n$$

The scalar resulting from the product $\alpha = \mathbf{y}^T A\mathbf{x}$ is given by

$$\alpha = \sum_{j=1}^{m} \sum_{k=1}^{n} y_j a_{jk} x_k$$

1.3.2 Partitioned Matrices

Property 1.2. Let A be a square, non-singular matrix of order m. Partition A as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

so that A_{11} and A_{22} are invertible, then

$$A^{-1} = \begin{pmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & -A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \\ -A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \end{pmatrix}$$

1.3.3 Matrix Differentiation

Property 1.3. Let A be a matrix,

$$\frac{\partial A}{\partial x} = \frac{\partial A^T}{\partial x}$$

Property 1.4. Let $\mathbf{y} = A\mathbf{x}$ where $\mathbf{y}$ is $m \times 1, \mathbf{x}$ is $n \times 1, A$ is $m \times n$, and A does not depend on $\mathbf{x}$. Suppose that $\mathbf{x}$ is a function of $\mathbf{z}$, while A is independent of $\mathbf{z}$, then

$$\frac{\partial \mathbf{y}}{\partial \mathbf{z}} = A\frac{\partial \mathbf{x}}{\partial \mathbf{z}}$$

Property 1.5. Let the scalar α be $\alpha = \mathbf{y}^T A\mathbf{x}$, where $\mathbf{y}$ is $m \times 1, \mathbf{x}$ is $n \times 1, A$ is $m \times n$, and A is independent of $\mathbf{x}$ and $\mathbf{y}$, then

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{y}^T A$$

and

$$\frac{\partial \alpha}{\partial \mathbf{y}} = \mathbf{x}^T A^T$$

Property 1.6. Let the scalar α is given by the quadratic form $\alpha = \mathbf{x}^T A\mathbf{x}$, where $\mathbf{x}$ is $n \times 1, A$ is $n \times n$, and A does not depend on $\mathbf{x}$, then

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{x}^T(A + A^T)$$

Property 1.7. Let A be a symmetric matrix and $\alpha = \mathbf{x}^T A\mathbf{x}$, where $\mathbf{x}$ is $n \times 1, A$ is $n \times n$, and A does not depend on $\mathbf{x}$, then

$$\frac{\partial \alpha}{\partial \mathbf{x}} = 2\mathbf{x}^T A$$

Property 1.8. Let the scalar α be $\alpha = \mathbf{y}^T\mathbf{x}$, where $\mathbf{y}$ is $n \times 1, \mathbf{x}$ is $n \times 1$, and both $\mathbf{y}$ and $\mathbf{x}$ are functions for $\mathbf{z}$, then

$$\frac{\partial \alpha}{\partial \mathbf{z}} = \mathbf{x}^T\frac{\partial \mathbf{y}}{\partial \mathbf{z}} + \mathbf{y}^T\frac{\partial \mathbf{x}}{\partial \mathbf{z}}$$

Property 1.9. Let the scalar α be $\alpha = \mathbf{x}^T\mathbf{x}$, where $\mathbf{x}$ is $n \times 1$, and $\mathbf{x}$ is a function of $\mathbf{z}$, then

$$\frac{\partial \alpha}{\partial \mathbf{z}} = 2\mathbf{x}^T \frac{\partial \mathbf{y}}{\partial \mathbf{z}}$$

Property 1.10. Let the scalar α be $\alpha = \mathbf{y}^T A\mathbf{x}$, where $\mathbf{y}$ is $m \times 1, A$ is $m \times n, \mathbf{x}$ is $n \times 1$, both $\mathbf{y}$ and $\mathbf{x}$ are functions of $\mathbf{z}$, and A does not depend on $\mathbf{z}$, then

$$\frac{\partial \alpha}{\partial \mathbf{z}} = \mathbf{x}^T A^T \frac{\partial \mathbf{y}}{\partial \mathbf{z}} + \mathbf{y}^T A \frac{\partial \mathbf{x}}{\partial \mathbf{z}}$$

Property 1.11. Let A be an invertible, $m \times m$ matrix whose elements are functions of the scalar parameter α, then

$$\frac{\partial A^{-1}}{\partial \alpha} = -A^{-1} \frac{\partial A}{\partial \alpha} A^{-1}$$

Theorem 1.17 (Young's Theorem)**.** We have the symmetry of second derivatives

$$[\nabla_{\mathbf{xy}} f(\mathbf{x}, \mathbf{y})]^T = \nabla_{\mathbf{yx}} f(\mathbf{x}, \mathbf{y})$$

2 Finite Dimensional Optimization

2.1 Unconstrained Optimization

The optimization problems can be written as

$$\min_{x\in\Omega\subseteq\mathbb{R}^n} f(\mathbf{x}).$$

Typically, $\Omega \subseteq \mathbb{R}^n, \Omega = \mathbb{R}^n, \Omega$ is an open set, or Ω is the closure of an open set.

Property 2.1. $\max f(\mathbf{x}) = -\min -f(\mathbf{x})$ and $\min f(\mathbf{x}) = -\max -f(\mathbf{x})$.

Definition 2.1 ((Strictly) Local Minimum)**.** f has a local minimum at a point $\mathbf{x}_0 \in \Omega$ if

$$\exists\varepsilon > 0 \text{ s.t. } f(\mathbf{x}_0) \leqslant f(\mathbf{x}), \forall \mathbf{x} \in B_\Omega(\mathbf{x}_0, \varepsilon).$$

f has a strictly local minimum at $\mathbf{x}_0$ if

$$\exists\varepsilon > 0 \text{ s.t. } f(\mathbf{x}_0) < f(\mathbf{x}), \forall \mathbf{x} \in B_\Omega(\mathbf{x}_0, \varepsilon)\backslash\{\mathbf{x}_0\}.$$

Definition 2.2 ((Strictly) Global Minimum)**.** f has a global minimum at $\mathbf{x}_0 \in \Omega$ if

$$f(\mathbf{x}_0) \leqslant f(\mathbf{x}), \forall \mathbf{x} \in \Omega.$$

f has a strictly global minimum at $\mathbf{x}_0$ if

$$f(\mathbf{x}_0) < f(\mathbf{x}), \forall \mathbf{x} \in \Omega\backslash\{\mathbf{x}_0\}.$$

Note that strictly global minimum is always unique.

Definition 2.3 (Feasible Direction)**.** $\mathbf{v} \in \mathbb{R}^n$ is a feasible direction at $\mathbf{x}_0$ if

$$\mathbf{x}_0 + s\mathbf{v} \in \Omega, \forall 0 \leqslant s \leqslant \overline{s},$$

where $\overline{s} \in \mathbb{R}$.

Theorem 2.1 (First Order Necessary Condition for Local Minimum)**.** Let f be C^1 function on $\Omega \subseteq \mathbb{R}^n$. If f has a local minimum at $\mathbf{x}_0 \in \Omega$, then

$$\nabla f(\mathbf{x}_0) \cdot \mathbf{v} \geqslant 0,$$

for all feasible directions $\mathbf{v}$ at $\mathbf{x}_0$.

Proof. Let $\mathbf{v}$ be a feasible direction at $\mathbf{x}_0 \in \Omega$. Let $g(s) = f(\mathbf{x}_0 + s\mathbf{v}), s \geqslant 0$. We have $g(s) \geqslant g(0)$, for small $s \geqslant 0$. Thus,

$$g'(0) = \lim_{s\to 0} \frac{g(s) - g(0)}{s - 0} \geqslant 0.$$

Besides,

$$g'(0) = \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} f(\mathbf{x_0} + s\mathbf{v}) = \nabla f(\mathbf{x}_0) \cdot \mathbf{v}.$$

Hence, $\nabla f(\mathbf{x_0}) \cdot \mathbf{v} \geqslant 0$. □

Corollary 2.1. When Ω is an open set, if f has a local minimum at $\mathbf{x}_0$ on Ω, then $\nabla f(\mathbf{x}_0) = \mathbf{0}$.

Proof. Since Ω is an open set, then all directions are feasible.

By theorem, we have $\nabla f(\mathbf{x}_0) \cdot \mathbf{v} \geqslant 0$, and $\nabla f(\mathbf{x}_0) \cdot (-\mathbf{v}) \geqslant 0 \Leftrightarrow \nabla f(\mathbf{x}_0) \cdot \mathbf{v} \leqslant 0$, $\forall \mathbf{v} \in \mathbb{R}^n$, i.e.,

$$\nabla f(\mathbf{x}_0) \cdot \mathbf{v} = 0, \forall \mathbf{v} \in \mathbb{R}^n.$$

Therefore, $\nabla f(\mathbf{x}_0) = \mathbf{0}$. □

Example 2.1. $\min f(x, y) = x^2 - xy + y^2 - 3y$ over $\mathbb{R}^2$.

Solution. Let $\nabla f(x_0, y_0) = \begin{pmatrix} 2x_0 - y_0 \\ -x_0 + 2y_0 - 3 \end{pmatrix} = \mathbf{0}$. We have $x_0 = 1, y_0 = 2$, i.e., $(1, 2)$ is the only candidate for a local minimizer. In fact $f(1, 2)$ is the global minimum because $f(x, y) = \left(x - \frac{y}{2}\right)^2 + \frac{3}{4}(y-2)^2 - 3$.

Example 2.2. $\min f(x, y) = x^2 - x + y + xy$ over $\Omega = \{(x, y) | x, y \geqslant 0\}$.

Solution. We have $\nabla f(x, y) = \begin{pmatrix} 2x - 1 + y \\ x + 1 \end{pmatrix}$.

Consider $(x_0, y_0) \in \operatorname{int}(\Omega), (x_0, y_0) \in \partial\Omega$.

(i) Let $\nabla f(x_0, y_0) = \mathbf{0}$. We have $x_0 = -1$ that is outside Ω. So there is no interior point can be local minimizer.

(ii) $(x, 0)$: The feasible direction is (v, w) s.t. $w \geqslant 0$.

If $(x_0, 0)$ is a local minimizer, we have $\nabla f(\mathbf{x}_0) \cdot \begin{pmatrix} v \\ w \end{pmatrix} \geqslant 0, \forall v$ and $\forall w \geqslant 0$. Therefore,

$$(2x_0 - 1)v + (x_0 + 1)w \geqslant 0, \forall v \text{ and } \forall w \geqslant 0 \Rightarrow (2x_0 - 1)v \geqslant 0, \forall v \text{ and } w = 0$$

Therefore,

$$2x_0 - 1 = 0 \Rightarrow x_0 = \frac{1}{2}.$$

(iii) $(y, 0)$: The feasible direction is (v, w) s.t. $v \geqslant 0$.

Similarly, we have

$$(-1 + y_0)v + w \geqslant 0, \forall v \geqslant 0 \text{ and } \forall w \Rightarrow w \geqslant 0, v = 0 \text{ and } \forall w,$$

which is a contradiction.

(iv) $(0, 0)$: The feasible direction is (v, w) s.t. $v, w \geqslant 0$.

Similarly, we have

$$-v + w \geqslant 0, \forall v, w \geqslant 0,$$

but nothing can satisfy this inequation.

Thus, $\left(\frac{1}{2}, 0\right)$ is the only candidate for local minimizer.

Theorem 2.2 (Second Order Necessary Condition for Local Minimum)**.** Let f be C^2 function on $\Omega \subseteq \mathbb{R}^n$. If f has a local minimum at $\mathbf{x}_0 \in \Omega$, then

$$\text{if } \nabla f(\mathbf{x}_0) \cdot \mathbf{v} = 0, \text{ then } \mathbf{v}^T \nabla^2 f(\mathbf{x}_0) \mathbf{v} \geqslant 0.$$

Proof. Fix feasible direction $\mathbf{v}$ at $\mathbf{x}_0$. We have $f(\mathbf{x}_0) \leqslant f(\mathbf{x}_0 + s\mathbf{v})$, for small $s \geqslant 0$.

By second order Taylor theorem, we have

$$0 \leqslant f(\mathbf{x}_0 + s\mathbf{v}) - f(\mathbf{x}) = s\nabla f(\mathbf{x}_0) \cdot \mathbf{v} + \frac{1}{2}s^2\mathbf{v}^T\nabla^2 f(\mathbf{x}_0)\mathbf{v} + o(s^2).$$

Since $\nabla f(\mathbf{x}_0) \cdot \mathbf{v} = 0$, then

$$\frac{1}{2}s^2\mathbf{v}^T\nabla^2 f(\mathbf{x}_0)\mathbf{v} + o(s^2) = \frac{1}{2}\mathbf{v}^T\nabla^2 f(\mathbf{x}_0)\mathbf{v} + \frac{o(s^2)}{s^2} \geqslant 0, \forall s > 0.$$

Therefore,

$$\lim_{s \to 0} \left(\frac{1}{2}\mathbf{v}^T\nabla^2 f(\mathbf{x}_0)\mathbf{v} + \frac{o(s^2)}{s^2}\right) = \frac{1}{2}\mathbf{v}^T\nabla^2 f(\mathbf{x}_0)\mathbf{v} \geqslant 0,$$

i.e., $\mathbf{v}^T\nabla^2 f(\mathbf{x}_0)\mathbf{v} \geqslant 0$. □

Note that in first and second order necessary condition for local minimum, if $\mathbf{x}_0 \in \text{int}(\Omega), \nabla f(\mathbf{x}_0) = \mathbf{0}, \nabla^2 f(\mathbf{x}_0) \geqslant 0$.

Definition 2.4 (Positive Definite)**.** We say an $n \times n$ matrix A is positive definite iff $\mathbf{v}^T A\mathbf{v} > 0, \forall \mathbf{v} \neq \mathbf{0} \Leftrightarrow$ all eigenvalues is greater than 0.

Definition 2.5 (Positive Semi-Definite)**.** We say an $n \times n$ matrix A is positive definite iff $\mathbf{v}^T A\mathbf{v} \geqslant 0, \forall \mathbf{v} \Leftrightarrow$ all eigenvalues is greater than or equals 0.

Theorem 2.3 (Sylvester's Criterion)**.** (1) A is positive definite iff determinant of all leading minors is greater than 0.

(2) A is positive semi-definite iff determinant of all principle minors is greater than 0.

Example 2.3. Suppose A is a 3×3 matrix, the principle minors are all matrix $Q_{I,I}$, where $I \subseteq \{1, 2, 3\}$ s.t. $|I| = k$.

Let

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}.$$

When $k = 1, I = \{1\}, \{2\}, \{3\}$ and thus the determinant of principle minors are a, c, i.

When $k = 2, I = \{1, 2\}, \{1, 3\}, \{2, 3\}$ and thus the determinant of principle minors are $ae - bd, ai - cg, ei - fh$.

When $k = 3, I = \{1, 2, 3\}$ and thus the determinant of principle minors are $\det(A)$.

Example 2.4. $f(x, y) = x^2 - xy + y^2 - 3y, \Omega = \mathbb{R}^2$.

Solution. We have $\nabla f(x, y) = \begin{pmatrix} 2x - y \\ -x + 2y - 3 \end{pmatrix}, \nabla f(1, 2) = \mathbf{0}$. Besides,

$$\nabla^2 f(x, y) = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}.$$

We want to check $\mathbf{v}^T\nabla^2 f(x, y)\mathbf{v} \geqslant 0, \forall \mathbf{v}$, which is equivalent to check if $\nabla^2 f(1, 2) \geqslant 0$. Since the determinant of all principle minors is greater than 0, then $\nabla^2 f(1, 2) \geqslant 0$.

Example 2.5. $f(x,y) = x^2 - x + y + xy, \Omega = \{(x,y)|x,y \geqslant 0\}$.

Solution. We have $\nabla f(x,y) = \begin{pmatrix} 2x-1+y \\ 1+x \end{pmatrix}$, but $(\frac{1}{2},0)$ is the only candidate for local minimum and

$$\nabla f(\frac{1}{2},0) = \begin{pmatrix} 0 \\ \frac{2}{3} \end{pmatrix} \Rightarrow \begin{pmatrix} 0 \\ \frac{2}{3} \end{pmatrix} \begin{pmatrix} v \\ w \end{pmatrix} = \frac{3}{2}w = 0 \Rightarrow \mathbf{v} = (v,0).$$

Also

$$\nabla^2 f(\frac{1}{2},0) = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix} \geqslant 0.$$

Lemma 1. $\nabla^2 f(\mathbf{x}_0)$ is positive definite, then $\exists a > 0$ s.t.

$$\mathbf{v}^T \nabla^2 f(\mathbf{x}_0)\mathbf{v} \geqslant a\|\mathbf{v}\|^2, \forall \mathbf{v}.$$

Proof. Suppose Q is orthogonal, i.e.,

$$Q^T Q = I \Leftrightarrow Q^{-1} = Q^T \Leftrightarrow \|Q\mathbf{v}\| = \|\mathbf{v}\|, \forall \mathbf{v}.$$

Hence,

$$Q^T \nabla^2 f(\mathbf{x}_0) Q = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \lambda_n \end{pmatrix},$$

and

$$\begin{aligned} \mathbf{v}^T \nabla^2 f(\mathbf{x}_0)\mathbf{v} &= (Q\mathbf{w})^T \nabla^2 f(\mathbf{x}_0)(Q\mathbf{w}) = \mathbf{w}^T Q^T \nabla^2 f(\mathbf{x}_0) Q\mathbf{w} \\ &= \sum_{i=1}^n \lambda_i w_i^2 \geqslant a \sum_{i=1}^n w_i^2 = a\|\mathbf{w}\|^2 = a\|Q^T\mathbf{v}\|^2 = a\|\mathbf{v}\|^2. \end{aligned}$$

where $\mathbf{w} = Q^{-1}\mathbf{v}, 0 < a = \min\{\lambda_1, \cdots, \lambda_n\}$. □

Theorem 2.4 (Second Order Sufficient Conditions for Interior Points)**.** Let $f \in C^2$ on Ω. If $\nabla f(x_0) = 0, \nabla^2 f(x_0) > 0$, then x_0 is a strict local minimum.

Proof. From the second order Taylor approximation, we have

$$\begin{aligned} f(\mathbf{x}_0 + \mathbf{v}) - f(\mathbf{x}_0) &= \nabla f(\mathbf{x}_0)^T \mathbf{v} + \frac{1}{2}\mathbf{v}^T \nabla^2 f(\mathbf{x}_0)\mathbf{v} + o(\|\mathbf{v}\|^2) \\ &\geqslant \frac{1}{2}a\|\mathbf{v}\|^2 + o(a\|\mathbf{v}\|^2) = \|\mathbf{v}\|^2 \left(\frac{a}{2} + \frac{o(\|\mathbf{v}\|^2)}{\|\mathbf{v}\|^2}\right) > 0, \end{aligned}$$

for small $\|\mathbf{v}\|$. Therefore, $f(\mathbf{x}_0 + \mathbf{v}) > f(\mathbf{x}_0)$ for all small $\mathbf{v} \in B(\mathbf{x}_0, \varepsilon)$. □

Example 2.6. $\min f(x,y) = xy$ on $\Omega = \mathbb{R}^2$.

Solution. Suppose (x_0,y_0) is the local minimum. Let $\nabla f(x_0,y_0) = 0$, then $(x_0,y_0) = (0,0)$. However $\nabla^2 f(0,0) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is not positive positive definite.

2.2 Equality Constraints

2.2.1 Tangent Space

We define a surface $M := \{\mathbf{x} \in \mathbb{R}^n | h_1(\mathbf{x}) = 0, \cdots, h_k(\mathbf{x}) = 0\}, h_i \in C^1$.

Definition 2.6 (Differentiable Curve)**.** A differentiable curve on surface $M \subseteq \mathbb{R}^n$ is a C^1 function $x : (-\varepsilon, \varepsilon) \to M$ given by $s \mapsto \mathbf{x}(s)$.

Let $\mathbf{x}(s)$ be a differentiable curve on M that passes through $\mathbf{x}_0 \in M$, say $\mathbf{x}(0) = \mathbf{x}_0$. The vector $\mathbf{v} = \frac{\mathrm{d}}{\mathrm{d}s}\big|_{s=0}\mathbf{x}(s)$ touches M tangentially, we say $\mathbf{v}$ is generated by $\mathbf{x}(s)$.

Definition 2.7 (Tangent Vector)**.** Any vector $\mathbf{v}$ which is generated by some differentiable curve on M through $\mathbf{x}_0$ is called a tangent vector.

Tangent space to M at $\mathbf{x}_0$ is

$$T_{\mathbf{x}_0}M = \left\{ \mathbf{v} \in \mathbb{R}^n | \mathbf{v} = \frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=0} \mathbf{x}(s), \text{ some differentiable curve } \mathbf{x}(s) \text{ on } M \text{ s.t. } \mathbf{x}(0) = \mathbf{x}_0 \right\}.$$

2.2.2 Lagrange Multiplier

2.3 Inequality Constraints