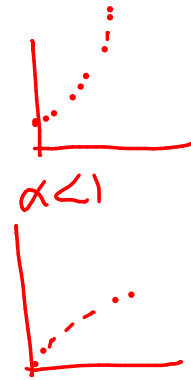
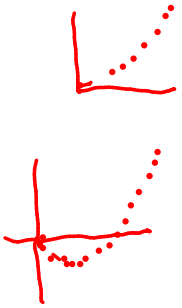


Examples on R

convex:

- affine: $ax + b$ on \mathbf{R} , for any $a, b \in \mathbf{R}$ ✓
- exponential: e^{ax} , for any $a \in \mathbf{R}$ $AM \geq GM$
- powers: x^α on \mathbf{R}_{++} , for $\alpha \geq 1$ or $\alpha \leq 0$ \rightarrow for $\alpha < 1$
- powers of absolute value: $|x|^p$ on \mathbf{R} , for $p \geq 1$
- negative entropy: $x \log x$ on \mathbf{R}_{++}



concave:

- affine: $ax + b$ on \mathbf{R} , for any $a, b \in \mathbf{R}$
- powers: x^α on \mathbf{R}_{++} , for $0 \leq \alpha \leq 1$
- logarithm: $\log x$ on \mathbf{R}_{++} $-\log x$ is convex

Most proofs by proving Jensen's inequality:
 $f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2)$

3-3

Examples on \mathbf{R}^n and $\mathbf{R}^{m \times n}$

affine functions are convex and concave; all norms are convex

examples on \mathbf{R}^n

- affine function $f(x) = a^T x + b$
- norms: $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ for $p \geq 1$; $\|x\|_\infty = \max_k |x_k|$

examples on $\mathbf{R}^{m \times n}$ ($m \times n$ matrices)

- affine function

$$f(X) = \underbrace{\text{tr}(A^T X)}_{\langle A, X \rangle_F} + b = \sum_{i=1}^m \sum_{j=1}^n A_{ij} X_{ij} + b$$

- spectral (maximum singular value) norm

$$f(X) = \|X\|_2 = \sigma_{\max}(X) = (\lambda_{\max}(X^T X))^{1/2}$$

$\left. \begin{array}{l} \text{max} \\ \sqrt{\frac{\|Xv\|_2}{\|v\|_2}} \end{array} \right\} = \max_{\|v\|_2=1} \|Xv\|_2$

First-order condition

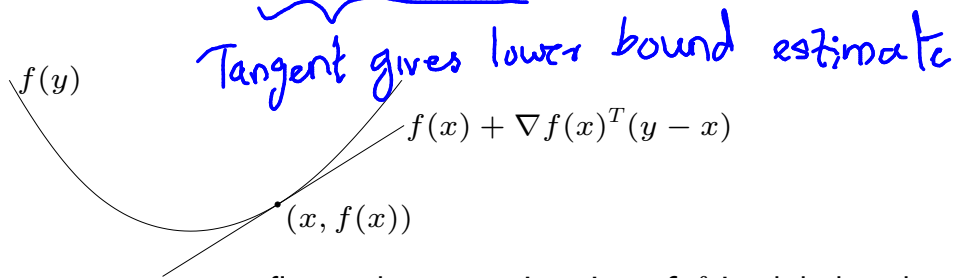
f is **differentiable** if $\text{dom } f$ is open and the gradient

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n} \right)$$

exists at each $x \in \text{dom } f$

1st-order condition: differentiable f with convex domain is convex iff

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \text{for all } x, y \in \text{dom } f$$



first-order approximation of f is global underestimator

Second-order conditions

f is **twice differentiable** if $\text{dom } f$ is open and the Hessian $\nabla^2 f(x) \in \mathbf{S}^n$,

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i, j = 1, \dots, n,$$

exists at each $x \in \text{dom } f$

2nd-order conditions: for twice differentiable f with convex domain

- f is convex if and only if

$$\nabla^2 f(x) \succeq 0 \quad \text{for all } x \in \text{dom } f$$

- if $\nabla^2 f(x) \succ 0$ for all $x \in \text{dom } f$, then f is strictly convex

Definition 35 [Convex Function]: A function $f : \mathcal{D} \rightarrow \mathbb{R}$ is **convex** if \mathcal{D} is a convex set and

$$f(\theta \mathbf{x} + (1 - \theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1 - \theta)f(\mathbf{y}) \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{D} \quad 0 \leq \theta \leq 1 \quad (4.31)$$

Figure 4.37 illustrates an example convex function. A function $f : \mathcal{D} \rightarrow \mathbb{R}$ is **strictly convex** if \mathcal{D} is convex and

$$f(\theta \mathbf{x} + (1 - \theta)\mathbf{y}) < \theta f(\mathbf{x}) + (1 - \theta)f(\mathbf{y}) \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{D} \quad 0 \leq \theta \leq 1 \quad (4.32)$$

A function $f : \mathcal{D} \rightarrow \mathbb{R}$ is called **uniformly or strongly convex** if \mathcal{D} is convex and there exists a constant $c > 0$ such that

$$f(\theta \mathbf{x} + (1 - \theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1 - \theta)f(\mathbf{y}) - \frac{1}{2}c\theta(1 - \theta)\|\mathbf{x} - \mathbf{y}\|^2 \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{D}$$

function at convex combination is less than convex combination of fn values & a factor dependent on distance between the pts

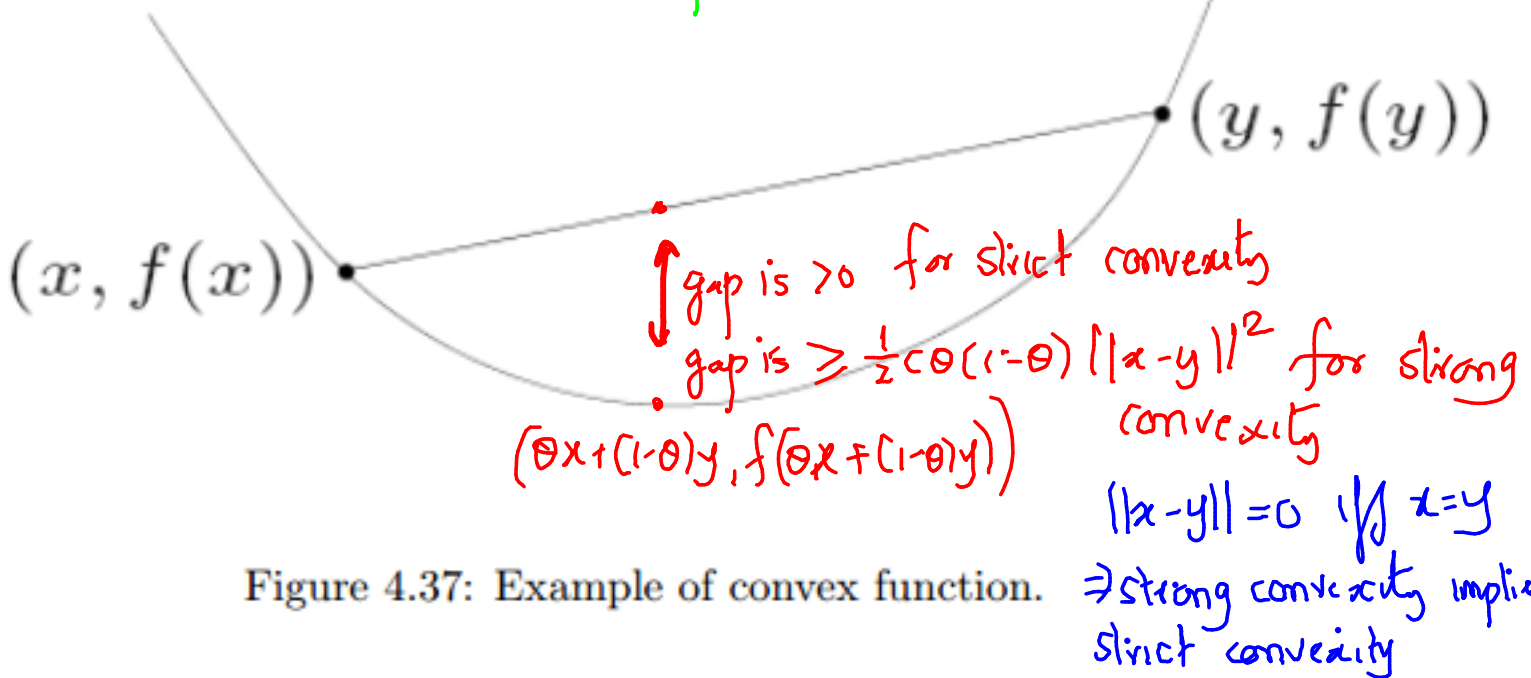


Figure 4.37: Example of convex function.

Theorem 75 Let $f : \mathcal{D} \rightarrow \mathbb{R}$ be a differentiable convex function on an open convex set \mathcal{D} . Then:

1. f is convex if and only if, for any $\mathbf{x}, \mathbf{y} \in \mathcal{D}$,

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \quad (4.44)$$

2. f is strictly convex on \mathcal{D} if and only if, for any $\mathbf{x}, \mathbf{y} \in \mathcal{D}$, with $\mathbf{x} \neq \mathbf{y}$,

$$f(\mathbf{y}) > f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \quad (4.45)$$

3. f is strongly convex on \mathcal{D} if and only if, for any $\mathbf{x}, \mathbf{y} \in \mathcal{D}$,

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) + \frac{1}{2}c\|\mathbf{y} - \mathbf{x}\|^2 \quad (4.46)$$

for some constant $c > 0$.

Proof:

Sufficiency: The proof of sufficiency is very similar for all the three statements of the theorem. So we will prove only for statement (4.44). Suppose (4.44) holds. Consider $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$ and any $\theta \in (0, 1)$. Let $\mathbf{x} = \theta\mathbf{x}_1 + (1 - \theta)\mathbf{x}_2$. Then,

$$\begin{aligned} f(\mathbf{x}_1) &\geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_1 - \mathbf{x}) \\ f(\mathbf{x}_2) &\geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_2 - \mathbf{x}) \end{aligned} \quad (4.47)$$

Adding $(1 - \theta)$ times the second inequality to θ times the first, we get,

$$\theta f(\mathbf{x}_1) + (1 - \theta)f(\mathbf{x}_2) \geq f(\mathbf{x})$$

which proves that $f(\mathbf{x})$ is a convex function. In the case of strict convexity, strict inequality holds in (4.47) and it follows through. In the case of strong convexity, we need to additionally prove that

$$\theta \frac{1}{2}c\|\mathbf{x} - \mathbf{x}_1\|^2 + (1 - \theta) \frac{1}{2}c\|\mathbf{x} - \mathbf{x}_2\|^2 = \frac{1}{2}c\theta(1 - \theta)\|\mathbf{x}_2 - \mathbf{x}_1\|^2$$

Necessity: Suppose f is convex. Then for all $\theta \in (0, 1)$ and $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$, we must have

$$f(\theta\mathbf{x}_2 + (1 - \theta)\mathbf{x}_1) \leq \theta f(\mathbf{x}_2) + (1 - \theta)f(\mathbf{x}_1)$$

Thus,

$$\nabla^T f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1) = \lim_{\theta \rightarrow 0} \frac{f(\mathbf{x}_1 + \theta(\mathbf{x}_2 - \mathbf{x}_1)) - f(\mathbf{x}_1)}{\theta} \leq f(\mathbf{x}_2) - f(\mathbf{x}_1)$$

This proves necessity for (4.44). The necessity proofs for (4.45) and (4.46) are very similar, except for a small difference for the case of strict convexity; the strict inequality is not preserved when we take limits. Suppose equality does hold in the case of strict convexity, that is for a strictly convex function f , let

$$f(\mathbf{x}_2) = f(\mathbf{x}_1) + \nabla^T f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1) \quad (4.48)$$

for some $\mathbf{x}_2 \neq \mathbf{x}_1$. Because f is strictly convex, for any $\theta \in (0, 1)$ we can write

$$f(\theta\mathbf{x}_1 + (1 - \theta)\mathbf{x}_2) = f(\mathbf{x}_2 + \theta(\mathbf{x}_1 - \mathbf{x}_2)) < \theta f(\mathbf{x}_1) + (1 - \theta)f(\mathbf{x}_2) \quad (4.49)$$

Since (4.44) is already proved for convex functions, we use it in conjunction with (4.48), and (4.49), to get

$$f(\mathbf{x}_2) + \theta \nabla^T f(\mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2) \leq f(\mathbf{x}_2 + \theta(\mathbf{x}_1 - \mathbf{x}_2)) < f(\mathbf{x}_2) + \theta \nabla^T f(\mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2)$$

which is a contradiction. Thus, equality can never hold in (4.44) for any $\mathbf{x}_1 \neq \mathbf{x}_2$.

This proves the necessity of (4.45). \square

Basic inequality

recall basic inequality for convex differentiable f :

$$\langle \underbrace{(\nabla f(x), -1)}_a, (y, f(y)) \rangle \geq \underbrace{\langle (\nabla f(x), -1), (x, f(x)) \rangle}_b = f(x) + \nabla f(x)^T (y - x)$$

- first-order approximation of f at x is global underestimator

- $(\nabla f(x), -1)$ supports $\text{epi } f$ at $(x, f(x))$

ie: equality holds for $y=x$ thus determining value of offset b

what if f is not differentiable?

Recall (we stated without proof)

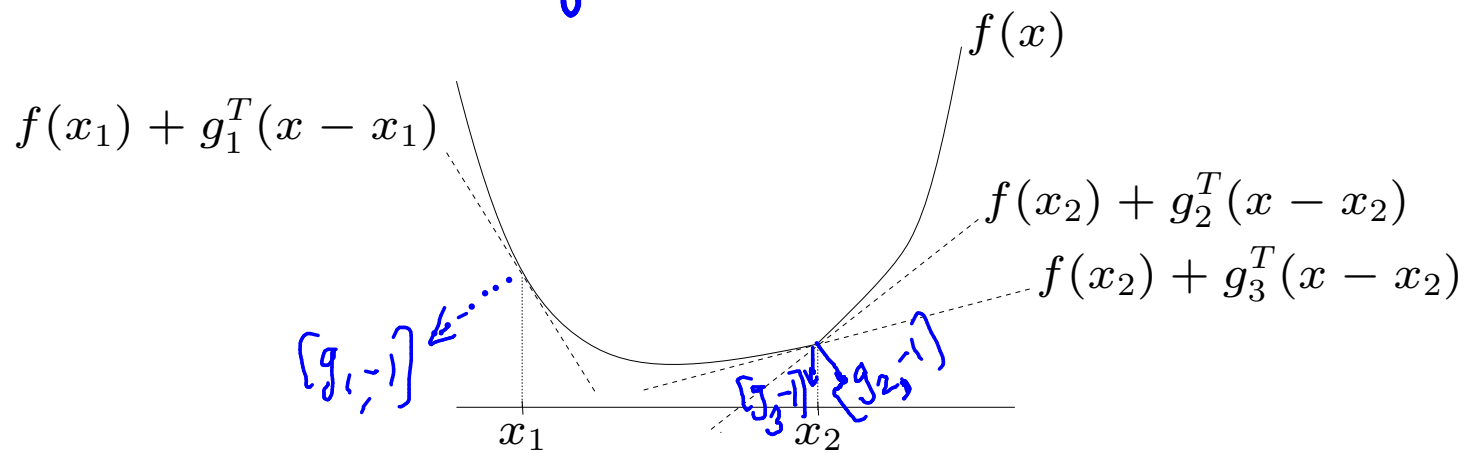
$$f^*(\nabla f(x)) = \langle x, \nabla f(x) \rangle - f(x) = \langle (\nabla f(x), -1), (x, f(x)) \rangle$$

Subgradient of a function

g is a **subgradient** of f (not necessarily convex) at x if

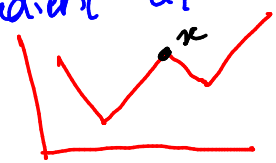
$$f(y) \geq f(x) + g^T(y - x) \quad \text{for all } y$$

global linear underestimator



g_2, g_3 are subgradients at x_2 ; g_1 is a subgradient at x_1

If f is NOT convex, there may be no global linear underestimator to f at a pt x ie there may exist no subgradient at x



Equivalent defn
motivated by $Df(x)$

- g is a subgradient of f at x iff $(g, -1)$ supports **epi** f at $(x, f(x))$
 - g is a subgradient iff $f(x) + g^T(y - x)$ is a global (affine) underestimator of f
 - if f is convex and differentiable, $\nabla f(x)$ is a subgradient of f at x
- If f is closed convex, subgradient is guaranteed to exist

subgradients come up in several contexts:

- algorithms for nondifferentiable convex optimization
- convex analysis, *e.g.*, optimality conditions, duality for nondifferentiable problems

(if $f(y) \leq f(x) + g^T(y - x)$ for all y , then g is a **supergradient**)

Definition 41 [Subgradient]: Let $f : \mathcal{D} \rightarrow \mathbb{R}$ be a convex function defined on a convex set \mathcal{D} . A vector $\mathbf{h} \in \mathbb{R}^n$ is said to be a subgradient of f at the point $\mathbf{x} \in \mathcal{D}$ if

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \mathbf{h}^T(\mathbf{y} - \mathbf{x})$$

for all $\mathbf{y} \in \mathcal{D}$. The set of all such vectors is called the subdifferential of f at \mathbf{x} .

Theorem 76 Let $f : \mathcal{D} \rightarrow \mathbb{R}$ be a convex function defined on a convex set \mathcal{D} . A point $\mathbf{x} \in \mathcal{D}$ corresponds to a minimum if and only if

$$\nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \geq 0$$

for all $\mathbf{y} \in \mathcal{D}$.

If $\nabla f(\mathbf{x})$ is nonzero, it defines a supporting hyperplane to \mathcal{D} at the point \mathbf{x} . Theorem 77 implies that for a differentiable convex function defined on an open set, every critical point must be a point of (global) minimum.

Theorem 77 Let $f : \mathcal{D} \rightarrow \mathbb{R}$ be differentiable and convex on an open convex domain $\mathcal{D} \subseteq \mathbb{R}^n$. Then \mathbf{x} is a critical point of f if and only if it is a (global) minimum.

Theorem 78 Let $f : \mathcal{D} \rightarrow \mathbb{R}$ with $\mathcal{D} \subseteq \mathbb{R}^n$ be differentiable on the convex set \mathcal{D} . Then,

1. f is convex on \mathcal{D} if and only if its gradient ∇f is monotone. That is, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}$

$$(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))^T (\mathbf{x} - \mathbf{y}) \geq 0 \quad (4.53)$$

2. f is strictly convex on \mathcal{D} if and only if its gradient ∇f is strictly monotone. That is, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}$ with $\mathbf{x} \neq \mathbf{y}$,

$$(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))^T (\mathbf{x} - \mathbf{y}) > 0 \quad (4.54)$$

3. f is uniformly or strongly convex on \mathcal{D} if and only if its gradient ∇f is uniformly monotone. That is, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}$,

$$(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))^T (\mathbf{x} - \mathbf{y}) \geq c\|\mathbf{x} - \mathbf{y}\|^2 \quad (4.55)$$

for some constant $c > 0$.

Necessity: Suppose f is uniformly convex on \mathcal{D} . Then from theorem 75, we know that for any $\mathbf{x}, \mathbf{y} \in \mathcal{D}$,

$$\begin{aligned} f(\mathbf{y}) &\geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) - \frac{1}{2}c\|\mathbf{y} - \mathbf{x}\|^2 \\ f(\mathbf{x}) &\geq f(\mathbf{y}) + \nabla^T f(\mathbf{y})(\mathbf{x} - \mathbf{y}) - \frac{1}{2}c\|\mathbf{x} - \mathbf{y}\|^2 \end{aligned}$$

Adding the two inequalities, we get (4.55). If f is convex, the inequalities hold with $c = 0$, yielding (4.54). If f is strictly convex, the inequalities will be strict, yielding (4.54).

Sufficiency: Suppose ∇f is monotone. For any fixed $\mathbf{x}, \mathbf{y} \in \mathcal{D}$, consider the function $\phi(t) = f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))$. By the mean value theorem applied to $\phi(t)$, we should have for some $t \in (0, 1)$,

$$\phi(1) - \phi(0) = \phi'(t) \quad (4.56)$$

Letting $\mathbf{z} = \mathbf{x} + t(\mathbf{y} - \mathbf{x})$, (4.56) translates to

$$f(\mathbf{y}) - f(\mathbf{x}) = \nabla^T f(\mathbf{z})(\mathbf{y} - \mathbf{x}) \quad (4.57)$$

Also, by definition of monotonicity of ∇f , (from (4.53)),

$$(\nabla f(\mathbf{z}) - \nabla f(\mathbf{x}))^T (\mathbf{y} - \mathbf{x}) = \frac{1}{t} (\nabla f(\mathbf{z}) - \nabla f(\mathbf{x}))^T (\mathbf{z} - \mathbf{x}) \geq 0 \quad (4.58)$$

Combining (4.57) with (4.58), we get,

$$\begin{aligned} f(\mathbf{y}) - f(\mathbf{x}) &= (\nabla f(\mathbf{z}) - \nabla f(\mathbf{x}))^T (\mathbf{y} - \mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \\ &\geq \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \end{aligned} \quad (4.59)$$

By theorem 75, this inequality proves that f is convex. Strict convexity can be similarly proved by using the strict inequality in (4.58) inherited from strict monotonicity, and letting the strict inequality follow through to (4.59). For the case of strong convexity, from (4.55), we have

$$\begin{aligned} \phi'(t) - \phi'(0) &= (\nabla f(\mathbf{z}) - \nabla f(\mathbf{x}))^T (\mathbf{y} - \mathbf{x}) \\ &= \frac{1}{t} (\nabla f(\mathbf{z}) - \nabla f(\mathbf{x}))^T (\mathbf{z} - \mathbf{x}) \geq \frac{1}{t}c\|\mathbf{z} - \mathbf{x}\|^2 = ct\|\mathbf{y} - \mathbf{x}\|^2 \end{aligned} \quad (4.60)$$

$$\phi(1) - \phi(0) - \phi'(0) = \int_0^1 [\phi'(t) - \phi'(0)]dt \geq \frac{1}{2}c\|\mathbf{y} - \mathbf{x}\|^2 \quad (4.61)$$

which translates to

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) + \frac{1}{2}c\|\mathbf{y} - \mathbf{x}\|^2$$