Convexity, Local and Global Optimality, etc.

More of Basic Subgradient Calculus

- Scaling: $\partial(af) = a \cdot \partial f$ provided $a > 0$. The condition $a > 0$ makes function f remain convex.
- Addition: $\partial(f_1 + f_2) = \partial(f_1) + \partial(f_2)$
- Affine composition: if $g(\mathbf{x}) = f(A\mathbf{x} + \mathbf{b})$, then $\partial g(\mathbf{x}) = A^T \partial f(A\mathbf{x} + b)$
- Norms: important special case, $f(\mathbf{x}) = ||\mathbf{x}||_p$ =max over inner product <z,x> such that ||z||_q <= 1

More of Basic Subgradient Calculus

- Scaling: $\partial(af) = a \cdot \partial f$ provided $a > 0$. The condition $a > 0$ makes function f remain convex.
- Addition: $\partial(f_1 + f_2) = \partial(f_1) + \partial(f_2)$
- Affine composition: if $g(\mathbf{x}) = f(A\mathbf{x} + \mathbf{b})$, then $\partial g(\mathbf{x}) = A^T \partial f(A\mathbf{x} + b)$
- Norms: important special case, $f(\mathbf{x}) = ||\mathbf{x}||_p = \max_{||\mathbf{z}||_q \leq 1} \mathbf{z}^T\mathbf{x}$ where q is such that $1/p + 1/q = 1$. Then

Subdifferential of f(x) = z with ||z||_q <= 1 and for which the max is attained...

More of Basic Subgradient Calculus

- Scaling: $\partial(af) = a \cdot \partial f$ provided $a > 0$. The condition $a > 0$ makes function f remain convex.
- Addition: $\partial(f_1 + f_2) = \partial(f_1) + \partial(f_2)$
- Affine composition: if $g(\mathbf{x}) = f(A\mathbf{x} + \mathbf{b})$, then $\partial g(\mathbf{x}) = A^T \partial f(A\mathbf{x} + b)$
- Norms: important special case, $f(\mathbf{x}) = ||\mathbf{x}||_p = \max_{||\mathbf{z}||_q \leq 1} \mathbf{z}^T\mathbf{x}$ where q is such that $1/p + 1/q = 1$. Then
$$\partial f(\mathbf{x}) = \left\{ \mathbf{y} : ||\mathbf{y}||_q \leq 1 \text{ and } \mathbf{y}^T x = \max_{||\mathbf{z}||_q \leq 1} \mathbf{z}^T\mathbf{x} \right\}$$

Homework (Optional)

Subgradients for the 'Lasso' Problem in Machine Learning

We use Lasso ($\min_{\mathbf{x}} f(\mathbf{x})$) as an example to illustrate subgradients of affine composition:

$$f(\mathbf{x}) = \frac{1}{2}||\mathbf{y} - \mathbf{x}||^2 + \lambda||\mathbf{x}||_1$$

The subgradients of $f(\mathbf{x})$ are

Generalizing from midsem where y=0

x-y + \lambda s(x)

s_i(x) = sign(x_i) if x_i not equal to 0
s_i(x) anyvalue in [-1,+1]

Subgradients for the ‘Lasso’ Problem in Machine Learning

We use Lasso ($\min_{\mathbf{x}} f(\mathbf{x})$) as an example to illustrate subgradients of affine composition:

$$f(\mathbf{x}) = \frac{1}{2}||\mathbf{y} - \mathbf{x}||^2 + \lambda||\mathbf{x}||_1$$

The subgradients of $f(\mathbf{x})$ are

$$\mathbf{h} = \mathbf{x} - \mathbf{y} + \lambda\mathbf{s},$$

where $s_i = sign(x_i)$ if $x_i \neq 0$ and $s_i \in [-1, 1]$ if $x_i = 0$.

More Subgradient Calculus: Composition

Following functions, though convex, may not be differentiable everywhere. How does one compute their subgradients? (what holds for subgradient also holds for gradient)

- **Composition with functions:** Let $p : \Re^k \to \Re$ with $q(x) = \infty, \forall\ \mathbf{x} \notin \mathbf{dom}\ h$ and $q : \Re^n \to \Re^k$. Define $f(\mathbf{x}) = p(q(\mathbf{x}))$. f is convex if
  - q_i is convex, p is convex and nondecreasing in each argument
  - or q_i is concave, p is convex and nonincreasing in each argument

More Subgradient Calculus: Composition

Following functions, though convex, may not be differentiable everywhere. How does one compute their subgradients? (what holds for subgradient also holds for gradient)

- **Composition with functions:** Let $p : \Re^k \rightarrow \Re$ with $q(x) = \infty, \forall\ \mathbf{x} \notin \mathbf{dom}\ h$ and $q : \Re^n \rightarrow \Re^k$. Define $f(\mathbf{x}) = p(q(\mathbf{x}))$. f is convex if (concave, p is concave)
  - q_i is convex, p is convex and nondecreasing in each argument (q_i convex, p nonincreasing)
  - or q_i is concave, p is convex and nonincreasing in each argument (q_i concave, p nondecreasing)

  Some examples illustrating this property are:
  - $\exp q(\mathbf{x})$ is convex if q is convex
  - $\sum_{i=1}^{m} \log q_i(\mathbf{x})$ is concave if q_i are concave and positive p = summation of logs is concave and nondecreasing
  - $\log \sum_{i=1}^{m} \exp q_i(\mathbf{x})$ is convex if q_i are convex p = -log of summation of exps... (convex and nonincrea
  - $1/q(\mathbf{x})$ is convex if q is concave and positive p = 1/...is convex and non-increasing

More Subgradient Calculus: Composition (contd)

- **Composition with functions:** Let $p : \Re^k \rightarrow \Re$ with $q(x) = \infty, \forall\ \mathbf{x} \notin \mathbf{d}om\ h$ and $q : \Re^n \rightarrow \Re^k$. Define $f(\mathbf{x}) = p(q(\mathbf{x}))$. f is convex if
  - q_i is convex, p is convex and nondecreasing in each argument
  - or q_i is concave, p is convex and nonincreasing in each argument
- Subgradients for the first case (second one is homework):

f(y) = p(q(y)) >= p(q_i(y) + h_{q_i}^T (y-x))

More Subgradient Calculus: Composition (contd)

- **Composition with functions:** Let $p : \Re^k \rightarrow \Re$ with $q(x) = \infty, \forall\ \mathbf{x} \notin \mathbf{d}om\ h$ and $q : \Re^n \rightarrow \Re^k$. Define $f(\mathbf{x}) = p(q(\mathbf{x}))$. f is convex if
  - q_i is convex, p is convex and nondecreasing in each argument
  - or q_i is concave, p is convex and nonincreasing in each argument
- Subgradients for the first case (second one is homework): q_k(x) +
  - $f(\mathbf{y}) = p\left(q_1(\mathbf{y}), \ldots, q_k(\mathbf{y})\right) \geq p\left(q_1(\mathbf{x}) + \mathbf{h}_{q_1}^T(\mathbf{y}-\mathbf{x}), \ldots, q_k(\mathbf{x})\mathbf{h}_{q_k}^T(\mathbf{y}-\mathbf{x})\right)$

    Where $\mathbf{h}_{q_i} \in \partial q_i(\mathbf{x})$ for $i = 1..k$ and since $p(.)$ is non-decreasing in each argument.

    Lower bound the RHS above with the linear underestimator of p obtained using p's subgradient at x.

More Subgradient Calculus: Composition (contd)

- **Composition with functions:** Let $p: \Re^k \rightarrow \Re$ with $q(x) = \infty, \forall\ \mathbf{x} \notin \mathbf{dom}\ h$ and $q: \Re^n \rightarrow \Re^k$. Define $f(\mathbf{x}) = p(q(\mathbf{x}))$. f is convex if
  - q_i is convex, p is convex and nondecreasing in each argument
  - or q_i is concave, p is convex and nonincreasing in each argument
- Subgradients for the first case (second one is homework):
  - $f(\mathbf{y}) = p\left(q_1(\mathbf{y}), \ldots, q_k(\mathbf{y})\right) \geq p\left(q_1(\mathbf{x}) + \mathbf{h}_{q_1}^T(\mathbf{y}-\mathbf{x}), \ldots, q_k(\mathbf{x})\mathbf{h}_{q_k}^T(\mathbf{y}-\mathbf{x})\right)$
    Where $\mathbf{h}_{q_i} \in \partial q_i(\mathbf{x})$ for $i = 1..k$ and since $p(.)$ is non-decreasing in each argument.
  - $p\left(q_1(\mathbf{x}) + \mathbf{h}_{q_1}^T(\mathbf{y}-\mathbf{x}), \ldots, q_k(\mathbf{x}) + \mathbf{h}_{q_k}^T(\mathbf{y}-\mathbf{x})\right) \geq$
    $p\left(q_1(\mathbf{x}), \ldots, q_k(\mathbf{x})\right) + \mathbf{h}_p^T\left(\mathbf{h}_{q_1}^T(\mathbf{y}-\mathbf{x}), \ldots, \mathbf{h}_{q_k}^T(\mathbf{y}-\mathbf{x})\right)$
    Where $\mathbf{h}_p \in \partial p\left(q_1(\mathbf{x}), \ldots, q_k(\mathbf{x})\right)$

Wrt what we have on board:

q' = [q_1(x)....q_k(x)]

h_q_1(y-x)..... h_q_k(y-x) = q - q'

More Subgradient Calculus: Composition (contd)

- **Composition with functions:** Let $p : \Re^k \rightarrow \Re$ with $q(x) = \infty, \forall\ \mathbf{x} \notin \mathbf{dom}\ h$ and $q : \Re^n \rightarrow \Re^k$. Define $f(\mathbf{x}) = p(q(\mathbf{x}))$. f is convex if
  - q_i is convex, p is convex and nondecreasing in each argument
  - or q_i is concave, p is convex and nonincreasing in each argument
- Subgradients for the first case (second one is homework):
  - $f(\mathbf{y}) = p\left(q_1(\mathbf{y}), \ldots, q_k(\mathbf{y})\right) \geq p\left(q_1(\mathbf{x}) + \mathbf{h}_{q_1}^T(\mathbf{y}-\mathbf{x}), \ldots, q_k(\mathbf{x})\mathbf{h}_{q_k}^T(\mathbf{y}-\mathbf{x})\right)$
    Where $\mathbf{h}_{q_i} \in \partial q_i(\mathbf{x})$ for $i = 1..k$ and since $p(.)$ is non-decreasing in each argument.
  - $p\left(q_1(\mathbf{x}) + \mathbf{h}_{q_1}^T(\mathbf{y}-\mathbf{x}), \ldots, q_k(\mathbf{x}) + \mathbf{h}_{q_k}^T(\mathbf{y}-\mathbf{x})\right) \geq$
    $p\left(q_1(\mathbf{x}), \ldots, q_k(\mathbf{x})\right) + \mathbf{h}_p^T\left(\mathbf{h}_{q_1}^T(\mathbf{y}-\mathbf{x}), \ldots, \mathbf{h}_{q_k}^T(\mathbf{y}-\mathbf{x})\right)$
    Where $\mathbf{h}_p \in \partial p\left(q_1(\mathbf{x}), \ldots, q_k(\mathbf{x})\right)$
  - $p\left(q_1(\mathbf{x}), \ldots, q_k(\mathbf{x})\right) + h_p^T\left(h_{q_1}^T(\mathbf{y}-\mathbf{x}), \ldots, h_{q_k}^T(\mathbf{y}-\mathbf{x})\right) = f(\mathbf{x}) + \sum_{i=1}^{k}(h_p)_i q_i(\mathbf{x})$

That is, $\sum_{i=1}^{k}(h_p)_i q_i(\mathbf{x})$ is a subgradient of the composite function at $\mathbf{x}$.

More Subgradient Calculus: Proximal Operator

Following functions are again convex, but again, may not be differentiable everywhere. How does one compute their subgradients at points of non-differentiability?

- **Infimum:** If $c(x, y)$ is convex in (x, y) and $\mathcal{C}$ is a convex set, then $d(x) = \inf_{y \in \mathcal{C}} c(x, y)$ is convex. For example:

H/w (from first principles)

  - Let $d(\mathbf{x}, \mathcal{C})$ that returns the distance of a point $\mathbf{x}$ to a convex set $\mathcal{C}$. That is $d(\mathbf{x}, \mathcal{C}) = \inf_{y \in \mathcal{C}} ||\mathbf{x} - \mathbf{y}||^2$. Then $d(\mathbf{x}, \mathcal{C})$ is a convex function.
  - $\underset{y \in \mathcal{C}}{\operatorname{argmin}}\, d(\mathbf{x}, \mathcal{C})$ is a special case of the proximity operator: $prox_f(\mathbf{x}) = \underset{y}{\operatorname{argmin}}\, PROX_f(\mathbf{x})$ of a convex function $f(\mathbf{x})$. Here, $PROX_f(\mathbf{x}) = f(\mathbf{y}) + \frac{1}{2}||\mathbf{x} - \mathbf{y}||^2$ The special case is when

f(y) is the indicator function on the set S

More Subgradient Calculus: Proximal Operator

Following functions are again convex, but again, may not be differentiable everywhere. How does one compute their subgradients at points of non-differentiability?

- **Infimum:** If $c(x, y)$ is convex in (x, y) and $\mathcal{C}$ is a convex set, then $d(x) = \inf_{y \in \mathcal{C}} c(x, y)$ is convex. For example:
  - Let $d(\mathbf{x}, \mathcal{C})$ that returns the distance of a point $\mathbf{x}$ to a convex set $\mathcal{C}$. That is $d(\mathbf{x}, \mathcal{C}) = \inf_{y \in \mathcal{C}} ||\mathbf{x} - \mathbf{y}||^2$. Then $d(\mathbf{x}, \mathcal{C})$ is a convex function.
  - $\underset{y \in \mathcal{C}}{\text{argmin}}\, d(\mathbf{x}, \mathcal{C})$ is a special case of the proximity operator: $prox_f(\mathbf{x}) = \underset{y}{\text{argmin}}\, PROX_f(\mathbf{x})$ of a convex function $f(\mathbf{x})$. Here, $PROX_f(\mathbf{x}) = f(\mathbf{y}) + \frac{1}{2}||\mathbf{x} - \mathbf{y}||^2$ The special case is when $f(\mathbf{y})$ is the indicator function $I_C(\mathbf{y})$ introduced earlier to remove the contraints of an optimization problem.
    - ★ Recall that $\partial I_C(\mathbf{y}) = N_C(\mathbf{y}) = \{\mathbf{h} \in \Re^n : \mathbf{h}^T\mathbf{y} \geq \mathbf{h}^T\mathbf{z} \text{ for any } \mathbf{z} \in C\}$
    - ★ The subdifferential $partialPROX_f(\mathbf{x}) = \partial f(\mathbf{y}) + \mathbf{x} - \mathbf{y}$ which can now be obtained for the special case $f(\mathbf{y}) = I_C(\mathbf{y})$.
    - ★ We will invoke this when we discuss the **proximal gradient descent** algorithm

More Subgradient Calculus: Perspective (Advanced)

Following functions are again convex, but again, may not be differentiable everywhere. How does one compute their subgradients at points of non-differentiability?

- **Perspective Function:** The perspective of a function $f: \Re^n \to \Re$ is the function $g: R^n \times \Re \to \Re$, $g(x,t) = tf(x/t)$. Function g is convex if f is convex on $\mathbf{dom} g = \left\{(x,t) | x/t \in \mathbf{dom} f, t > 0\right\}$. For example,
  - The perspective of $f(x) = x^T x$ is (quadratic-over-linear) function $g(x,t) = \frac{x^T x}{t}$ and is convex.
  - The perspective of negative logarithm $f(x) = -\log x$ is the relative entropy function $g(x,t) = t \log t - t \log x$ and is convex.

More on SubGradient kind of functions: Monotonicity

A differentiable function $f : \Re \to \Re$ is (strictly) convex, iff and only if $f'(x)$ is (strictly) increasing. Is there a closer analog for $f : \Re^n \to \Re$?

What is the notion of (monotonically) increasing vector (sub)gradient
Subgradient h of f will lie in R^n

More on SubGradient kind of functions: Monotonicity

A differentiable function $f : \Re \to \Re$ is (strictly) convex, iff and only if $f'(x)$ is (strictly) increasing. Is there a closer analog for $f : \Re^n \to \Re$? View subgradient as an instance of a general function $\mathbf{h} : \mathcal{D} \to \Re^n$ and $\mathcal{D} \subseteq \Re^n$. Then

If h is monotone then function is convex and vice versa

More on SubGradient kind of functions: Monotonicity

A differentiable function $f : \Re \to \Re$ is (strictly) convex, iff and only if $f'(x)$ is (strictly) increasing. Is there a closer analog for $f : \Re^n \to \Re$? View subgradient as an instance of a general function $\mathbf{h} : \mathcal{D} \to \Re^n$ and $\mathcal{D} \subseteq \Re^n$. Then

Definition

1. $\mathbf{h}$ is *monotone* on $\mathcal{D}$ if for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$,

$$\left(\mathbf{h}(\mathbf{x}_1) - \mathbf{h}(\mathbf{x}_2)\right)^T (\mathbf{x}_1 - \mathbf{x}_2) \geq 0 \tag{14}$$

More on SubGradient kind of functions: Monotonicity (contd)

Definition

2. $\mathbf{h}$ is *strictly monotone* on $\mathcal{D}$ if for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$ with $\mathbf{x}_1 \neq \mathbf{x}_2$,

$$\left(\mathbf{h}(\mathbf{x}_1) - \mathbf{h}(\mathbf{x}_2)\right)^T (\mathbf{x}_1 - \mathbf{x}_2) > 0 \tag{15}$$

3. $\mathbf{h}$ is *uniformly* or *strongly monotone* on $\mathcal{D}$ if for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$, there is a constant $c > 0$ such that

$$\left(\mathbf{h}(\mathbf{x}_1) - \mathbf{h}(\mathbf{x}_2)\right)^T (\mathbf{x}_1 - \mathbf{x}_2) \geq c||\mathbf{x}_1 - \mathbf{x}_2||^2 \tag{16}$$

(Sub)Gradients and Convexity

Based on the definition of monotonic functions, we next show the relationship between convexity of a function and monotonicity of its (sub)gradient:

Theorem

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

1. *f is convex on $\mathcal{D}$* **iff** *its gradient ∇f is monotone. That is, for all $\mathbf{x}, \mathbf{y} \in \Re$:* $\left(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\right)^T (\mathbf{x} - \mathbf{y}) \geq 0$
2. *f is strictly convex on $\mathcal{D}$* **iff** *its gradient ∇f is strictly monotone. That is, for all $\mathbf{x}, \mathbf{y} \in \Re$ with $\mathbf{x} \neq \mathbf{y}$:* $\left(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\right)^T (\mathbf{x} - \mathbf{y}) > 0$
3. *f is uniformly or strongly convex on $\mathcal{D}$* **iff** *its gradient ∇f is uniformly monotone. That is, for all $\mathbf{x}, \mathbf{y} \in \Re$,* $\left(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\right)^T (\mathbf{x} - \mathbf{y}) \geq c||\mathbf{x} - \mathbf{y}||^2$ *for some constant $c > 0$.*

While these results also hold for subgradients $\mathbf{h}$, we will show them only for gradients ∇f

(Sub)Gradients and Convexity (contd)

Proof:
Necessity: Suppose f is strongly convex on $\mathcal{D}$. Then we know from an earlier result that for any $\mathbf{x}, \mathbf{y} \in \mathcal{D}$,

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

Adding the two inequalities,

(Sub)Gradients and Convexity (contd)

Proof:
Necessity: Suppose f is strongly convex on $\mathcal{D}$. Then we know from an earlier result that for any $\mathbf{x}, \mathbf{y} \in \mathcal{D}$,

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

Adding the two inequalities, we get uniform/strong monotonicity in definition (3). If f is convex, the inequalities hold with $c = 0$, yielding monotonicity in definition (1). If f is strictly convex, the inequalities will be strict, yielding strict monotonicity in definition (2).

(Sub)Gradients and Convexity (contd)

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

(Sub)Gradients and Convexity (contd)

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

$$\phi(1) - \phi(0) = \phi'(t) \tag{17}$$

Letting $\mathbf{z} = \mathbf{x} + t(\mathbf{y} - \mathbf{x})$, (17) translates to

$$f(\mathbf{y}) - f(\mathbf{x}) = \nabla^T f(\mathbf{z})(\mathbf{y} - \mathbf{x}) \tag{18}$$

Also, by definition of monotonicity of ∇f,

$$\big(\nabla f(\mathbf{z}) - \nabla f(\mathbf{x})\big)^T (\mathbf{y} - \mathbf{x}) = \frac{1}{t}\big(\nabla f(\mathbf{z}) - \nabla f(\mathbf{x})\big)^T (\mathbf{z} - \mathbf{x}) \geq 0 \tag{19}$$

(Sub)Gradients and Convexity (contd)

Combining (18) with (19), we get,

$$
\begin{aligned}
f(\mathbf{y}) - f(\mathbf{x}) &= \left(\nabla f(\mathbf{z}) - f(\mathbf{x})\right)^T (\mathbf{y} - \mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \\
&\geq \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x})
\end{aligned}
\tag{20}
$$

By a previous foundational result, this inequality proves that f is convex. Strict convexity can be similarly proved by using the strict inequality in (19) inherited from strict monotonicity, and letting the strict inequality follow through to (20).

(Sub)Gradients and Convexity (contd)

For the case of strong convexity, we have

$$\phi'(t) - \phi'(0) = \left(\nabla f(\mathbf{z}) - f(\mathbf{x})\right)^T (\mathbf{y} - \mathbf{x})$$
$$= \frac{1}{t}\left(\nabla f(\mathbf{z}) - f(\mathbf{x})\right)^T (\mathbf{z} - \mathbf{x}) \geq \frac{1}{t} c||\mathbf{z} - \mathbf{x}||^2 = ct||\mathbf{y} - \mathbf{x}||^2 \quad (21)$$

Therefore,

....Applying fundamental theorem of calculus

(Sub)Gradients and Convexity (contd)

For the case of strong convexity, we have

$$\phi'(t) - \phi'(0) = \left(\nabla f(\mathbf{z}) - f(\mathbf{x})\right)^T (\mathbf{y} - \mathbf{x})$$

$$= \frac{1}{t}\left(\nabla f(\mathbf{z}) - f(\mathbf{x})\right)^T (\mathbf{z} - \mathbf{x}) \geq \frac{1}{t} c||\mathbf{z} - \mathbf{x}||^2 = ct||\mathbf{y} - \mathbf{x}||^2 \tag{21}$$

Therefore,

....Applying fundamental theorem of calculus

$$\phi(1) - \phi(0) - \phi'(0) = \int_0^1 [\phi'(t) - \phi'(0)]dt \geq \frac{1}{2} c||\mathbf{y} - \mathbf{x}||^2 \tag{22}$$

which translates to

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

Thus, f must be strongly convex.

Local and Global Minima, Gradients and Convexity

Recall that for functions of single variable, at local extreme points, the tangent to the curve is a line with a constant component in the direction of the function and is therefore parallel to the x-axis. If the function is differentiable at the extreme point, then the derivative must vanish. This idea can be extended to functions of multiple variables. The requirement in this case turns out to be that the tangent plane to the function at any extreme point must be parallel to the plane $z = 0$. This can happen if and only if the gradient ∇F is parallel to the $z-$axis at the extreme point, or equivalently, the gradient to the function f must be the zero vector at every extreme point.