Convexity, Local and Global Optimality, etc.

Epigraph and Convexity

- Given a convex function $f(\mathbf{x})$ and a convex domain $\mathcal{D}$, the convex optimization problem

$$\min_{\mathbf{x}\in\mathcal{D}} \; f(\mathbf{x})$$

  can be equivalently expressed as

$$\min_{\mathbf{x}\in\mathcal{D}, t\in\Re, f(\mathbf{x})\le t} \; t = \min_{\mathbf{x}\in\mathcal{D}, (\mathbf{x},t)\in epi(f)} \; t$$

- Recall the first order condition for convexity of a differentiable function $f : \Re \to \Re$. Is there an equivalent for $f : \mathcal{D} \to \Re$? Let $f : \mathcal{D} \to \Re$ be a differentiable convex function on an open convex set $\mathcal{D}$. Then f is convex if and only if, for any $\mathbf{x}, \mathbf{y} \in \mathcal{D}$,

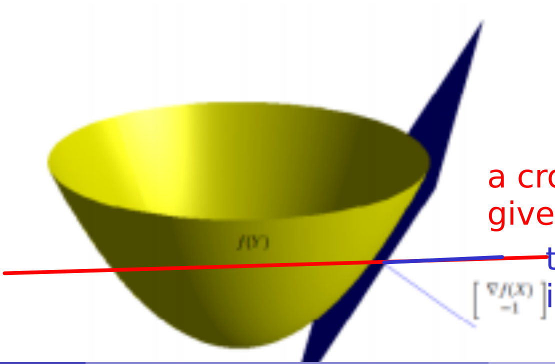
$$f(\mathbf{y}) \ge f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

Epigraph, Convexity and Gradients

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

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \quad (7)$$

If $\mathcal{D} \subseteq \Re^n$, this means that for each and every point $\mathbf{x} \in \mathcal{D}$ for a convex real function $f(\mathbf{x})$, there exists a hyperplane $H \in \Re^{n+1}$ having normal $[\nabla f(\mathbf{x}) \ \ -1]^T$ supporting the function epigraph at $[\mathbf{x} \ \ f(\mathbf{x})]^T$. See Figure below sourced from https://ccrma.stanford.edu/~dattorro/gcf.pdf



Epigraph, Convexity, Gradients and Level-sets

- **Revisiting level sets:** We can embed the graph of a function of n variables as the 0-level set of a function of $n+1$ variables

  F(x,z) = f(x) - z

  0-level set of F(x,z) is
  {(x,z) | F(x,z) = 0 } = {(x,z) | z = f(x)} = {(x,f(x))}

[1](that is, the tangent hyperplane to $f(\mathbf{x})$ at the point $\mathbf{x}$)

Epigraph, Convexity, Gradients and Level-sets

- **Revisiting level sets:** We can embed the graph of a function of n variables as the 0-level set of a function of $n+1$ variables
- More concretely, if $f : \mathcal{D} \rightarrow \Re,\ \mathcal{D} \subseteq \Re^n$ then we define $F : \mathcal{D}' \rightarrow \Re,\ \mathcal{D}' = \mathcal{D} \times \Re$ as $F(\mathbf{x}, z) = f(\mathbf{x}) - z$ with $\mathbf{x} \in \mathcal{D}'$.
- The gradient of F at any point $(\mathbf{x}, z)$ is simply, $\nabla F(\mathbf{x}, z) = \left[f_{x_1}, f_{x_2}, \ldots, f_{x_n}, -1\right]$ with the first n components of $\nabla F(\mathbf{x}, z)$ given by the n components of $\nabla f(\mathbf{x})$.
- The graph of f can be recovered as the $0-$level set of F given by $F(\mathbf{x}, z) = 0$.
- The equation of the tangent hyperplane $(\mathbf{y}, z)$ to the $0-$level set of F at the point $(\mathbf{x}, f(\mathbf{x}))$ is[1] $\nabla^T F(\mathbf{x}, f(\mathbf{x})).[\mathbf{y} - \mathbf{x}, z - f(\mathbf{x})]^T = [\nabla f(\mathbf{x}), -1]^T.[\mathbf{y} - \mathbf{x}, z - f(\mathbf{x})]^T = 0$.

[1](that is, the tangent hyperplane to $f(\mathbf{x})$ at the point $\mathbf{x}$)

Epigraph, Convexity, Gradients and Level-sets (contd.)

Substituting appropriate expression for $\nabla F(\mathbf{x})$, the equation of the tangent plane $(\mathbf{y}, z)$ can be written as

$$\left(\sum_{i=1}^{n} f_{x_i}(\mathbf{x})(y_i - x_i)\right) - \big(z - f(\mathbf{x})\big) = 0$$

or equivalently as,

$$\left(\nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x})\right) + f(\mathbf{x}) = z$$

f(y) >= z on the supporting hyperplane

Epigraph, Convexity, Gradients and Level-sets (contd.)

Substituting appropriate expression for $\nabla F(\mathbf{x})$, the equation of the tangent plane $(\mathbf{y}, z)$ can be written as

$$\left(\sum_{i=1}^{n} f_{x_i}(\mathbf{x})(y_i - x_i)\right) - \big(z - f(\mathbf{x})\big) = 0$$

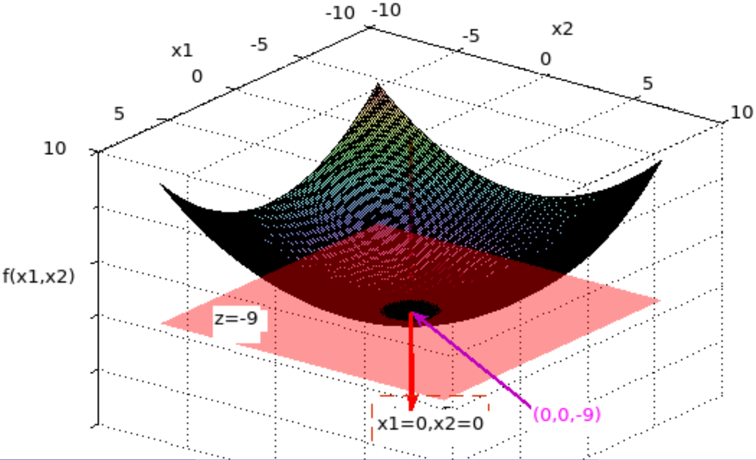
or equivalently as,

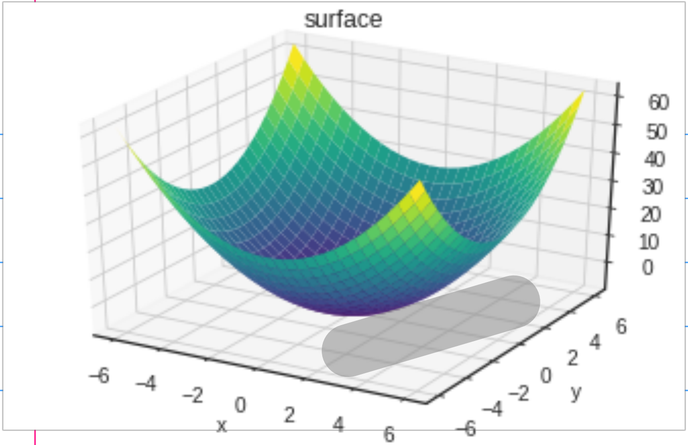
$$\left(\nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x})\right) + f(\mathbf{x}) = z$$

Revisiting the gradient-based condition for convexity in (7), we have that for a convex function, $f(\mathbf{y})$ is greater than each such z on the hyperplane: $f(\mathbf{y}) \geq z = f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x})$

Gradient and Epigraph (contd)

As an example, consider the paraboloid, $f(x_1, x_2) = x_1^2 + x_2^2 - 9$ that attains its minimum at $(0,0)$. We see below its epigraph.





Find equations of tangent hyperplane at any point y

$f(x_1,x_2) = x_1^2 + x_2^2 - 9$

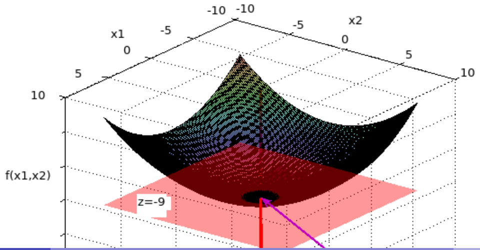
Illustrations to understand Gradient

- For the paraboloid, $f(x_1, x_2) = x_1^2 + x_2^2 - 9$, the corresponding $F(x_1, x_2, z) = x_1^2 + x_2^2 - 9 - z$ and the point $x^0 = (\mathbf{x}^0, z) = (1, 1, -7)$ which lies on the 0-level surface of F. The gradient $\nabla F(x_1, x_2, z)$ is $[2x_1,\ 2x_2,\ -1]$, which when evaluated at $x^0 = (1, 1, -7)$ is $[-2,\ -2,\ -1]$. The equation of the tangent plane to f at x^0 is therefore given by $2(x_1 - 1) + 2(x_2 - 1) - 7 = z$.
- The paraboloid attains its minimum at $(0, 0)$. Plot the tanget plane to the surface at $(0, 0, f(0, 0))$ as also the gradient vector ∇F at $(0, 0, f(0, 0))$. What do you expect?

We can expect the hyperplane to be parallel to the x1, x2 plane
(or rather, a constant value of z)
At point of min:
1) Expect Gradient of F to be vertical (down or up)
That is, gradient of f = 0
2) Tangent hyperplane is parallel to x1, x2 plane

Illustrations to understand Gradient

- For the paraboloid, $f(x_1, x_2) = x_1^2 + x_2^2 - 9$, the corresponding $F(x_1, x_2, z) = x_1^2 + x_2^2 - 9 - z$ and the point $x^0 = (\mathbf{x}^0, z) = (1, 1, -7)$ which lies on the 0-level surface of F. The gradient $\nabla F(x_1, x_2, z)$ is $[2x_1,\ 2x_2,\ -1]$, which when evaluated at $x^0 = (1, 1, -7)$ is $[-2,\ -2,\ -1]$. The equation of the tangent plane to f at x^0 is therefore given by $2(x_1 - 1) + 2(x_2 - 1) - 7 = z$.
- The paraboloid attains its minimum at $(0, 0)$. Plot the tanget plane to the surface at $(0, 0, f(0, 0))$ as also the gradient vector ∇F at $(0, 0, f(0, 0))$. What do you expect? Ans: A horizontal tanget plane and a vertical gradient! (of F)



Theorem

1. *For differentiable $f : \mathcal{D} \rightarrow \Re$ and open convex set $\mathcal{D}$, f is convex* **iff**, *for any* $\mathbf{x}, \mathbf{y} \in \mathcal{D}$,

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \tag{8}$$

2. *f is strictly convex* **iff**, *for any* $\mathbf{x}, \mathbf{y} \in \mathcal{D}$, *with* $\mathbf{x} \neq \mathbf{y}$,

The tangential hyperplane based lower bound is strict

$$f(\mathbf{y}) > f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \tag{9}$$

3. *f is strongly convex* **iff**, *for any* $\mathbf{x}, \mathbf{y} \in \mathcal{D}$, *and for some constant* $c > 0$,

The lower bound has a gap that increases atleast quadratically wrt L2 distance of x from y

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) + \frac{1}{2}c||\mathbf{y} - \mathbf{x}||^2 \tag{10}$$

First-Order Convexity Condition: Proof

Proof:

Sufficiency: The proof of sufficiency is very similar for all the three statements of the theorem. So we will prove only for statement (8). Suppose (8) 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,

$$\begin{aligned} f(\mathbf{x}_1) &\geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_1 - \mathbf{x}) \\ f(\mathbf{x}_2) &\geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_2 - \mathbf{x}) \end{aligned} \tag{11}$$

Multiply the first inequality by theta and second by (1-theta) and add the two inequalities

First-Order Convexity Condition: Proof

Proof:

Sufficiency: The proof of sufficiency is very similar for all the three statements of the theorem. So we will prove only for statement (8). Suppose (8) 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,

$$\begin{aligned} f(\mathbf{x}_1) &\geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_1 - \mathbf{x}) \\ f(\mathbf{x}_2) &\geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_2 - \mathbf{x}) \end{aligned} \tag{11}$$

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

First-Order Convexity Condition: Proof

Proof:

Sufficiency: The proof of sufficiency is very similar for all the three statements of the theorem. So we will prove only for statement (8). Suppose (8) 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,

$$\begin{aligned} f(\mathbf{x}_1) &\geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_1 - \mathbf{x}) \\ f(\mathbf{x}_2) &\geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_2 - \mathbf{x}) \end{aligned} \tag{11}$$

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, all inequalities will remain strict

First-Order Convexity Condition: Proof

Proof:

Sufficiency: The proof of sufficiency is very similar for all the three statements of the theorem. So we will prove only for statement (8). Suppose (8) 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,

$$\begin{aligned} f(\mathbf{x}_1) &\geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_1 - \mathbf{x}) \\ f(\mathbf{x}_2) &\geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_2 - \mathbf{x}) \end{aligned} \tag{11}$$

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 (11) and it follows through. In the case of strong convexity,

First-Order Convexity Condition: Proof

Proof:

Sufficiency: The proof of sufficiency is very similar for all the three statements of the theorem. So we will prove only for statement (8). Suppose (8) 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,

$$\begin{aligned} f(\mathbf{x}_1) &\geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_1 - \mathbf{x}) \\ f(\mathbf{x}_2) &\geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_2 - \mathbf{x}) \end{aligned} \tag{11}$$

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 (11) and it follows through. In the case of strong convexity, we need to additionally prove that

$$\theta \frac{1}{2} c||\mathbf{x} - \mathbf{x}_1||^2 + (1-\theta)\frac{1}{2} c||\mathbf{x} - \mathbf{x}_2||^2 =$$ theta (1-theta) c ||x1 - x2||^2

First-Order Convexity Condition: Proof

Proof:

Sufficiency: The proof of sufficiency is very similar for all the three statements of the theorem. So we will prove only for statement (8). Suppose (8) 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,

$$\begin{aligned} f(\mathbf{x}_1) &\geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_1 - \mathbf{x}) \\ f(\mathbf{x}_2) &\geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_2 - \mathbf{x}) \end{aligned} \tag{11}$$

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 (11) and it follows through. In the case of strong convexity, we need to additionally prove that

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

First-Order Convexity Conditions: Proofs

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,

$\nabla^T f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1) =$ Directional derivative at x1 in direction of x2 - x1
<= f(x2) - f(x1)

First-Order Convexity Conditions: Proofs

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,

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

First-Order Convexity Conditions: Proofs

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,

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

This proves necessity for (8). The necessity proofs for (9) and (10) 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

$$f(\mathbf{x}_2) = f(\mathbf{x}_1) + \nabla^T f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1) \qquad (12)$$

for some $\mathbf{x}_2 \neq \mathbf{x}_1$.

First-Order Convexity Conditions: Proofs

Necessity (contd for strict case):
Because f is stricly convex, for any $\theta \in (0,1)$ we can write

$$f\left(\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2\right) = f\left(\mathbf{x}_2 + \theta(\mathbf{x}_1 - \mathbf{x}_2)\right) < \theta f(\mathbf{x}_1) + (1-\theta) f(\mathbf{x}_2) \tag{13}$$

Since (8) is already proved for convex functions, we use it in conjunction with (12), and (13), to get

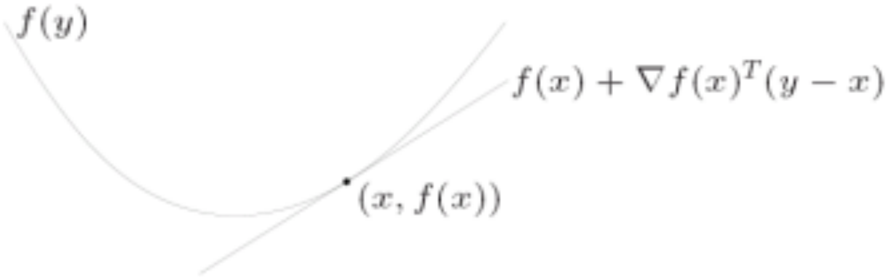
$$f(\mathbf{x}_2) + \theta \nabla^T f(\mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2) \leq f\left(\mathbf{x}_2 + \theta(\mathbf{x}_1 - \mathbf{x}_2)\right) < f(\mathbf{x}_2) + \theta \nabla^T f(\mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2)$$

which is a contradiction. Thus, equality can never hold in (8) for any $\mathbf{x}_1 \neq \mathbf{x}_2$. This proves the necessity of (9).

H/w: Understand the argument

First-Order Convexity Conditions: The complete statement

The geometrical interpretation of this theorem is that at any point, the linear approximation based on a local derivative gives a lower estimate of the function, *i.e.* the convex function always lies above the supporting hyperplane at that point. This is pictorially depicted below:



First-Order Convexity Conditions: Subgradients

The Theorem motivates the definition of the *subgradient* for non-differentiable convex functions, which has properties very similar to the gradient vector.

Definition

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

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

h(x) is a function of x

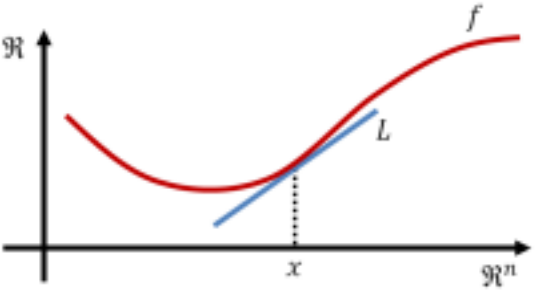
for all $\mathbf{y} \in \mathcal{D}$. The set of all such vectors is called the subdifferential of f at $\mathbf{x}$.

For a differentiable convex function, the gradient at point $\mathbf{x}$ is the only subgradient at that point. Most properties of differentiable convex functions that hold in terms of the gradient also hold in terms of the subgradient for non-differentiable convex functions.

Eg: Subgradient for $f(\mathbf{x}) = \|\mathbf{x}\|_1$ is ?

How do we compute such subgradients?

(Sub)Gradients and Convexity (contd)

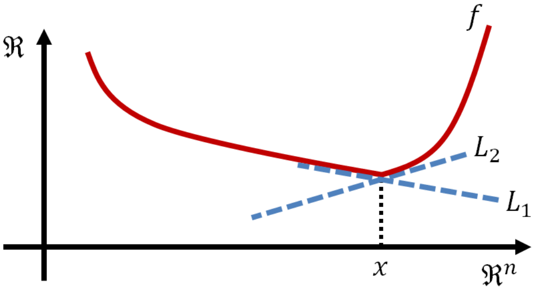


if the function is differentiable at x, the subgradient is unique and is called a gradient

To say that a function $f: \Re^n \mapsto \Re$ is differentiable at $\mathbf{x}$ is to say that there is a single unique linear tangent that under estimates the function:

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

(Sub)Gradients and Convexity (contd)



L1, and L2 are underestimators and so are all its convex (in fact conic) combinations (even if f is not convex)

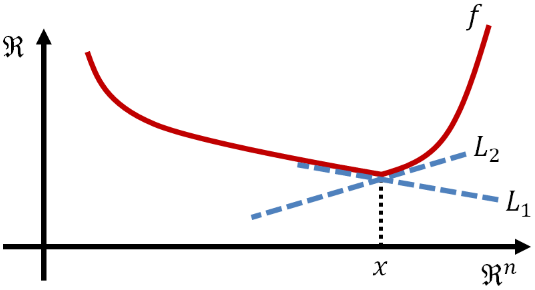
In this figure we see the function f at $\mathbf{x}$ has many possible linear tangents that may fit appropriately. Recall that a **subgradient** is any $\mathbf{h} \in \Re^n$ (same dimension as x) such that:

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

Thus, intuitively, if a function is differentiable at a point $\mathbf{x}$ then

the supporting hyperplane is uniquely specified by gradient

(Sub)Gradients and Convexity (contd)



Are we guaranteed to have a subgradient at each point for a convex function? Formal proof?

In this figure we see the function f at $\mathbf{x}$ has many possible linear tangents that may fit appropriately. Recall that a **subgradient** is any $\mathbf{h} \in \Re^n$ (same dimension as x) such that:

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

Thus, intuitively, if a function is differentiable at a point $\mathbf{x}$ then it has a unique subgradient at that point ($\nabla f(\mathbf{x})$). Formal Proof?

(Sub)Gradients and Convexity (contd)

- A **subdifferential** is the closed convex set of all subgradients of the convex function f:

$$\partial f(\mathbf{x}) = \{\mathbf{h} \in \Re^n : \mathbf{h} \text{ is a subgradient of } f \text{ at } \mathbf{x}\}$$

  Note that this set is guaranteed to be nonempty unless f is not convex.

- Often an indicator function, $I_C : \Re^n \mapsto \Re$, is employed to remove the contraints of an optimization problem (note that convex set $C \subseteq \Re^n$):

$$\min_{\mathbf{x}\in C} f(\mathbf{x}) \iff \min_{\mathbf{x}} f(\mathbf{x}) + I_C(\mathbf{x}), \quad \text{where} \quad I_C(\mathbf{x}) = I\{\mathbf{x} \in C\} = \begin{cases} 0 & \text{if } \mathbf{x} \in C \\ \infty & \text{if } \mathbf{x} \notin C \end{cases}$$

  The subdifferential of the indicator function at x is

Home work: Write expression for subdifferential for I_c