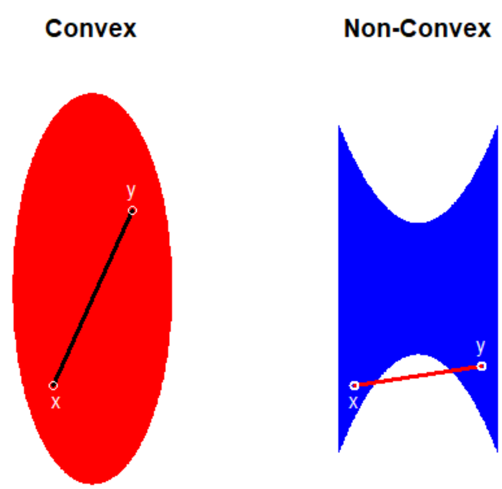
Convex sets, functions, subgradient and subdifferential

October 8, 2018

Convex Sets and Functions

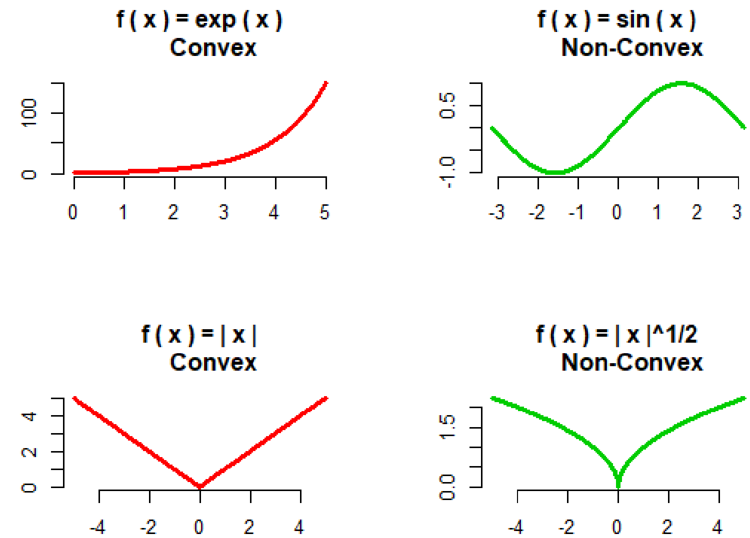
Convex Set: A subset $C \subset R^p$ is convex if for all $x, y \in C$ and $0 \leq \alpha \leq 1$ we have $\alpha x + (1-\alpha) y \in C$. That is, the line segment from x to y is fully contained in C.



Convex Function: a function $f : A \subset R^n \to R$ (where A is a convex set) is convex if for all $\mathbf{x}_1, \mathbf{x}_2 \in A$ and all $0 \leq \alpha \leq 1$, we have

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

and it is strictly convex if the inequality is strict for all $0 < \alpha < 1$ and $\mathbf{x}_1 \neq \mathbf{x}_2$.



Examples of convex functions in R**:**

- $f(x) = e^{ax}$, for all a
- $f(x) = -\log(x)\,,\ x > 0$
- $f(x) = |x|^{\alpha}\,,$ for $\alpha \geq 1$
- $f(x) = x\log(x)\,,\ x > 0$

Examples of convex functions in R^p

- All affine functions: $f(\mathbf{x}) = \mathbf{a}'\mathbf{x}+b$, (but not strictly convex)
- Some quadratic functions: $f(\mathbf{x}) = \mathbf{x}Q\mathbf{x} + \mathbf{a}'\mathbf{x}+b$, provided Q is non-negative definite, $Q \succeq 0$. Strictly convex if Q is positive definite $Q \succ 0$
- All norms $f(\mathbf{x}) = \|\mathbf{x}\|$. Recall that a norm is a function that satisfies a) $\|\mathbf{x}\| \geq 0$, b) $\|\mathbf{x}\| = 0$ iff $\mathbf{x} = \mathbf{0}$, c) $\|\alpha\mathbf{x}\| = |\alpha|\,\|\mathbf{x}\|$, and d) $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

First Order Condition

Definition. A function $f(\mathbf{x})$ is differentiable at $\mathbf{x}$ if the gradient

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \partial f(\mathbf{x}) / \partial x_1 \\ \partial f(\mathbf{x}) / \partial x_2 \\ \vdots \\ \partial f(\mathbf{x}) / \partial x_p \end{pmatrix}$$

exists. A function $f(\mathbf{x})$ is differentiable if $\nabla f(\mathbf{x})$ exists at every interior point of its domain.

First Order Condition. A differentiable function $f(\mathbf{x})$ with convex domain is convex iff

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x}). \tag{2}$$

The resader may wish to prove the equivalence of (1) and (2) for differentiable convex functions.

Example: let $f(x)$ be the convex function $f(x) = x^2$, with $\nabla f(x) = 2x$. In this case (2) becomes

$$y^2 \geq x^2 + 2x(y - x)$$

which is equivalent to $(y - x)^2 \geq 0$.

Global minimization of a differentiable convex function

A simple but important result. Suppose

1. $f(\mathbf{x})$ is convex and differentiable
2. $\mathbf{x}_0$ belongs to the interior of the domain of f.

   A sufficient and necessary condition for $\mathbf{x}_0$ to be a global minimizer of $f(\mathbf{x})$ is that $\nabla f(\mathbf{x}_0) = 0$.

   Proof: Sufficiency follows directly from (2) and the fact that $\nabla f(\mathbf{x}_0) = 0$. The necessity follows because $f(\mathbf{x})$ is differentiable and $\mathbf{x}_0$ belongs to the interior of the domain of f.

Remark: if $f(\mathbf{x})$ is **strictly convex** then the global minimizer $\mathbf{x}_0$ is **unique**. To see that suppose that there is another global minimizer $\mathbf{x}_1$. Then for all $0 < \alpha < 1$, $f(\alpha \mathbf{x}_0 + (1 - \alpha)\mathbf{x}_1) < \alpha f(\mathbf{x}_0) + (1 - \alpha) f(\mathbf{x}_1) = f(\mathbf{x}_0)$, contradicting the fact that $\mathbf{x}_0$ is a global minimizer.

Coordinate-descent algorithm

In this section we will introduce the *back-fitting algorithm*, which in the context of *regularization* is known as the *coordinate-descent algorithm.* Let $f(\mathbf{x},\mathbf{y})$ be a real valued function with $\mathbf{x} \in R^p$ and $\mathbf{y} \in R^q$. Suppose that we have a way for minimizing $f(\mathbf{x},\mathbf{y})$ in $\mathbf{y}$ for each fixed $\mathbf{x}$, and also for minimizing $f(\mathbf{x},\mathbf{y})$ in $\mathbf{x}$ for each fixed $\mathbf{y}$. Starting from some initial value $\mathbf{x}^0$ (e.g. $\mathbf{x}^0 = 0$) we form a decreasing sequence $\left\{f\left(\mathbf{x}^k,\mathbf{y}^k\right)\right\}$ as follows:

$$
\begin{aligned}
&f\left(\mathbf{x}^0,\mathbf{y}\right) \geq f\left(\mathbf{x}^0,\mathbf{y}^0\right) \rightarrow f\left(\mathbf{x}^0,\mathbf{y}^0\right), \\
&f\left(\mathbf{x},\mathbf{y}^0\right) \geq f\left(\mathbf{x}^1,\mathbf{y}^0\right) \rightarrow f\left(\mathbf{x}^1,\mathbf{y}\right) \geq f\left(\mathbf{x}^1,\mathbf{y}^1\right) \rightarrow f\left(\mathbf{x}^1,\mathbf{y}^1\right), \\
&f\left(\mathbf{x},\mathbf{y}^1\right) \geq f\left(\mathbf{x}^2,\mathbf{y}^1\right) \rightarrow f\left(\mathbf{x}^2,\mathbf{y}\right) \geq f\left(\mathbf{x}^2,\mathbf{y}^2\right) \rightarrow f\left(\mathbf{x}^2,\mathbf{y}^2\right), \\
&f\left(\mathbf{x},\mathbf{y}^2\right) \geq f\left(\mathbf{x}^3,\mathbf{y}^2\right) \rightarrow f\left(\mathbf{x}^3,\mathbf{y}\right) \geq f\left(\mathbf{x}^3,\mathbf{y}^3\right) \rightarrow f\left(\mathbf{x}^3,\mathbf{y}^3\right), \quad \text{etc.}
\end{aligned}
$$

Hence, by construction

$$f\left(\mathbf{x},\mathbf{y}^k\right) \geq f\left(\mathbf{x}^{k+1},\mathbf{y}^{k+1}\right), \quad \text{for all } \mathbf{x} \tag{3}$$

and

$$f\left(\mathbf{x}^k,\mathbf{y}\right) \geq f\left(\mathbf{x}^{k+1},\mathbf{y}^{k+1}\right), \quad \text{for all } \mathbf{y} \tag{4}$$

In particular

$$f\left(\mathbf{x}^k,\mathbf{y}^k\right) \geq f\left(\mathbf{x}^{k+m},\mathbf{y}^{k+m}\right), \quad m = 1, 2, ... \tag{5}$$

The following theorem shows that if $f(\mathbf{x},\mathbf{y})$ is convex and differentiable, $f\left(\mathbf{x}^k,\mathbf{y}^k\right)$ converges to a global minimum, $f(\mathbf{x}^*,\mathbf{y}^*)$. Later on, we will show that the differentiability condition can be relaxed to *sub-differentianility.*

Theorem 1. Suppose that $f : R^{p+q} \rightarrow R$ is

(i) convex and differentiable

(ii) There exists $(\mathbf{x}^*,\mathbf{y}^*) \in R^{p+q}$ such that $\nabla f(\mathbf{x}^*,\mathbf{y}^*) = \mathbf{0}$

(iii) $\lim_{\|(\mathbf{x},\mathbf{y})\|\rightarrow\infty} f(\mathbf{x},\mathbf{y}) = \infty$

(iv) $\lim_{\|\mathbf{y}\|\rightarrow\infty} f(\mathbf{x},\mathbf{y}) = \infty$ for all $\mathbf{x}$ and $\lim_{\|\mathbf{x}\|\rightarrow\infty} f(\mathbf{x},\mathbf{y}) = \infty$ for all $\mathbf{y}$

Then,

(a) $\lim_{k\rightarrow\infty} f\left(\mathbf{x}^k,\mathbf{y}^k\right) = f(\mathbf{x}^*,\mathbf{y}^*)$.

If $f(\mathbf{x},\mathbf{y})$ is strictly convex, then

(b) $\lim_{k\rightarrow\infty} \left(\mathbf{x}^k,\mathbf{y}^k\right) = (\mathbf{x}^*,\mathbf{y}^*)$.

Proof. By (iii) the sequence $\left(\mathbf{x}^k, \mathbf{y}^k\right)$ is bounded. Therefore, every subsequence $\left(\mathbf{x}^{k_m}, \mathbf{y}^{k_m}\right)$ has a sub-subsequence $\left(\mathbf{x}^{k_{m_j}}, \mathbf{y}^{k_{m_j}}\right) \to (\widetilde{\mathbf{x}},\widetilde{\mathbf{y}})$ as $j \to \infty$. Now, by (3) and (5),

$$f\left(\mathbf{x}, \mathbf{y}^{k_{m_j}}\right) \geq f\left(\mathbf{x}^{k_{m_j}+1}, \mathbf{y}^{k_{m_j}+1}\right) \geq f\left(\mathbf{x}^{k_{m_{j+1}}}, \mathbf{y}^{k_{m_{j+1}}}\right) \quad \text{for all } \mathbf{x}. \tag{6}$$

Taking limit for $j \to \infty$ in (6) we obtain

$$f(\mathbf{x}, \widetilde{\mathbf{y}}) \geq f(\widetilde{\mathbf{x}},\widetilde{\mathbf{y}}) \quad \text{for all } \mathbf{x}.$$

By (i) and (iv) the partial gradient, $\nabla_{\mathbf{x}} f(\mathbf{x}, \widetilde{\mathbf{y}})$, satisfies

$$\nabla_{\mathbf{x}} f(\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}}) = \mathbf{0}. \tag{7}$$

Similarly,

$$\nabla_{\mathbf{y}} f(\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}}) = \mathbf{0}. \tag{8}$$

Therefore,

$$\nabla f(\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}}) = \begin{pmatrix} \nabla_{\mathbf{x}} f(\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}}) \\ \nabla_{\mathbf{y}} f(\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}}) \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} = \mathbf{0} \tag{9}$$

By (i) $(\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}})$ is a global minimizer of $f(\mathbf{x}, \mathbf{y})$ and so

$$f(\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}}) = f(\mathbf{x}^*, \mathbf{y}^*).$$

Therefore, Part (a) follows [all subsequence of $f\left(\mathbf{x}^k, \mathbf{y}^k\right)$ has a sub-subsequence that converges to $f(\mathbf{x}^*, \mathbf{y}^*)$]. For Part (b) just notice that the global minimizer $(\mathbf{x}^*, \mathbf{y}^*)$ is unique and so $(\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}}) = (\mathbf{x}^*, \mathbf{y}^*)$ [now we have that all subsequence of $\left(\mathbf{x}^k, \mathbf{y}^k\right)$ has a sub-subsequence that converges to $(\mathbf{x}^*, \mathbf{y}^*)$].

Subgradient and Subdifferential

Suppose that $f(\mathbf{x})$ is real valued and defined on R^p. A *subgradient* of $f(\mathbf{x})$ is any vector $\mathbf{g} \in R^p$ with the property

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \mathbf{g}'(\mathbf{y} - \mathbf{x}).$$

The function $f(\mathbf{x})$ is *subdifferentiable at* $\mathbf{x}$ if there exists at least one subgradient of $f(\mathbf{x})$ at $\mathbf{x}$. The set of subgradients of $f(\mathbf{x})$ at $\mathbf{x}$ is called *subdifferential of* $f(\mathbf{x})$ at $\mathbf{x}$ and denoted $\partial f(\mathbf{x})$. The function f is called *subdifferentiable* if it is subdifferentiable a all $\mathbf{x}$.

Some Notes:

1. If $f(\mathbf{x})$ is convex and continuous then it is subdifferentiable ($\partial f(\mathbf{x}) \neq \phi$ for all $\mathbf{x}$).

2. If $\mathbf{g}$ is a subgradient of $f(\mathbf{x})$ then the affine function (of $\mathbf{y}$) $f(\mathbf{x}) + \mathbf{g}'(\mathbf{y}-\mathbf{x})$ is a global lower bound for $f(\mathbf{y})$. Geometrically, $(\mathbf{g},-1)$ supports the epigraph at $(\mathbf{x}, f(\mathbf{x}))$:

$$(\mathbf{y}-\mathbf{x},\ f(\mathbf{y})-f(\mathbf{x}))\begin{pmatrix} \mathbf{g} \\ -1 \end{pmatrix} \geq 0 \quad \text{for all } \mathbf{y} \in R^p$$

3. The subdifferential $\partial f(\mathbf{x})$ is convex and closed.

   Proof. Suppose $\mathbf{g}_1, \mathbf{g}_2 \in \partial f(\mathbf{x})$ and let $0 \leq \alpha \leq 1$. Then

$$\alpha f(\mathbf{y}) \geq \alpha f(\mathbf{x}) + \alpha \mathbf{g}_1'(\mathbf{y}-\mathbf{x})$$
$$(1-\alpha) f(\mathbf{y}) \geq (1-\alpha) f(\mathbf{x}) + (1-\alpha)\mathbf{g}_2'(\mathbf{y}-\mathbf{x})$$

   $\Rightarrow$

$$f(\mathbf{y}) \geq f(\mathbf{x}) + [\alpha \mathbf{g}_1 + (1-\alpha)\mathbf{g}_2]'(\mathbf{y}-\mathbf{x})$$

   $\Rightarrow$

$$\alpha \mathbf{g}_1 + (1-\alpha)\mathbf{g}_2 \in \partial f(\mathbf{x}).$$

   This proves the convexity of $\partial f(\mathbf{x})$. Closeness is shown in a similar fashion.

4. Suppose that $f(\mathbf{x})$ is convex and differentiable at $\mathbf{x}$.Then $\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$.

   Proof. We consider the univariate case (w.l.g.)

   Suppose $g \in \partial f(x)$. Then,

$$g \leq \lim_{y \downarrow x} \frac{f(y)-f(x)}{y-x} = \Delta f_+(x) = \nabla f(x) \Rightarrow g \leq \Delta f(x).$$

   On the other hand, for

$$g \geq \lim_{y \uparrow x} \frac{f(y)-f(x)}{y-x} = \Delta f_-(x) = \Delta f(x) \Rightarrow g \geq \Delta f(x).$$

   Therefore,

$$g = f'(x).$$

5. Suppose that $\partial f(\mathbf{x}) = \{\mathbf{g}\}$. Then f is differentiable at $\mathbf{x}$ and $\mathbf{g} = \nabla f(\mathbf{x})$.

   Proof. We consider the univariate case (w.l.g.). By assumption $g \in \partial f(x)$ and so

$$g \leq \lim_{y \downarrow x} \frac{f(y)-f(x)}{y-x} = \Delta f_+(x) \quad \text{and } g \geq \lim_{y \uparrow x} \frac{f(y)-f(x)}{y-x} = \nabla f_-(x)$$

Therefore, for convex functions **we always have** $\nabla f_- (x) \leq \nabla f_+ (x)$. Moreover, it can be shown that in the case

of convex functions the left and right derivatives are left and right subgradients (left as an exercise). That is

$$f (y) \geq f (x) + (y - x) \nabla f_- (x) , \quad \text{for all } y \leq x \tag{10}$$
$$f (y) \geq f (x) + (y - x) \nabla f_+ (x) , \quad \text{for all } y \geq x \tag{11}$$

If $\nabla f_- (x) = \nabla f_+ (x) = \nabla f (x)$ then f is differentiable at x and by Point 3 above $\partial f (x) = \{\Delta f (x)\}$. If

$$\nabla f_+ (x) > \nabla f_- (x) , \tag{12}$$

then

$$f (y) \geq f (x) + (y - x) \Delta f_- (x) \geq f (x) + (y - x) \nabla f_+ (x) \quad \text{for all } y \leq x \tag{13}$$

and by (10) and (13)

$$\nabla f_+ (x) \in \partial f (x) . \tag{14}$$

Moreover,

$$f (y) \geq f (x) + (y - x) \nabla f_+ (x) \geq f (x) + (y - x) \nabla f_- (x) \quad \text{for all } y \geq x \tag{15}$$

and by (10), and (15) we have

$$\nabla f_+ (x) \in \partial f (x) . \tag{16}$$

Finally (12), (14) and (16) contradict the uniqueness of g.

6. Suppose that $f_1 (\mathbf{x})$ and $f_2 (\mathbf{x})$ are subdifferentiable at $\mathbf{x}$. Then

$$\partial (f_1 (\mathbf{x}) + f_2 (\mathbf{x})) = \partial f_1 (\mathbf{x}) + \partial f_2 (\mathbf{x}) .$$

The meaning of the latest is as follows. If A and B are subsets of R^p, $A + B = \{c : c = a + b, a \in A, b \in B\}$.

7. A point $\mathbf{x}^*$ is a (global) minimizer of a convex function f iff $\mathbf{0} \in \partial f (\mathbf{x}^*)$.

Proof.

Sufficiency: $\mathbf{0} \in \partial f (\mathbf{x}^*) \Rightarrow f (\mathbf{x}) \geq f (\mathbf{x}^*) + \mathbf{0}' (\mathbf{x} - \mathbf{x}^*) = f (\mathbf{x}^*)$ for all $\mathbf{x}$.

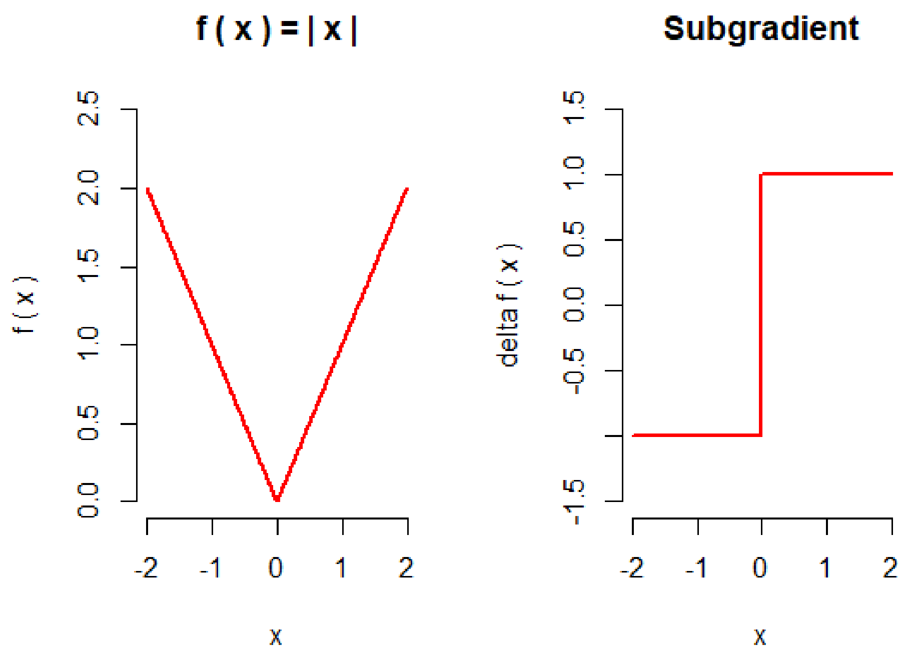
Necessity: $f (\mathbf{x}) \geq f (\mathbf{x}^*)$ for all $\mathbf{x} \Rightarrow f (\mathbf{x}) \geq f (\mathbf{x}^*) + \mathbf{0}' (\mathbf{x} - \mathbf{x}^*)$ for all $\mathbf{x} \Rightarrow \mathbf{0} \in \partial f (\mathbf{x}^*)$.

Example 1

Let $f(x) = |x|$. In this case

$$\partial f(x) = \begin{cases} \{-1\} & x < 0 \\ [-1, 1] & x = 0 \\ \{1\} & x > 0 \end{cases}$$

Since $0 \in \partial f(0)$, we have that $x = 0$ is the unique global minimizer of $f(x)$

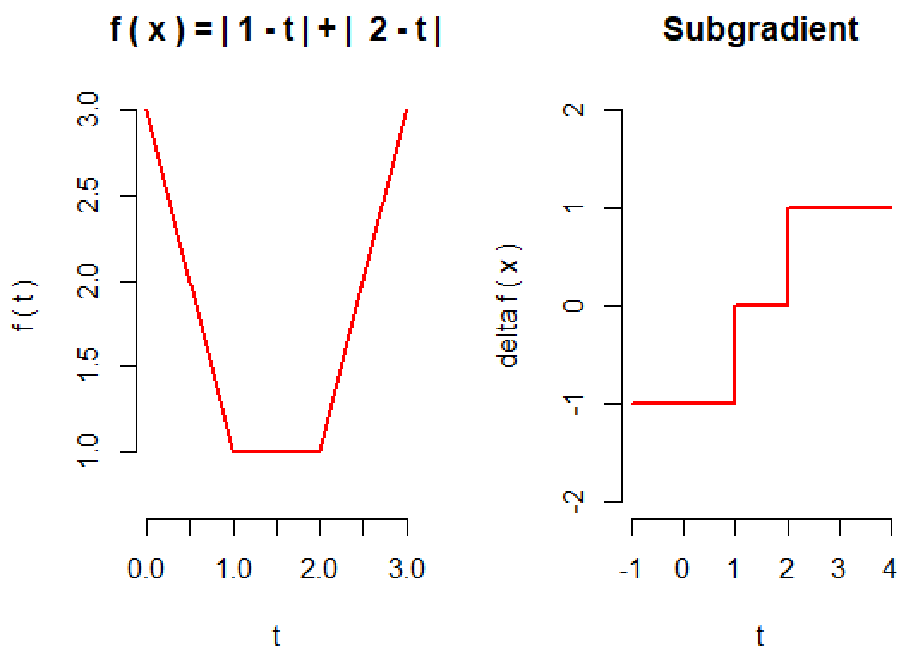


Example 2:

Let $f(t) = |1 - t| + |2 - t|$. In this case

$$\partial f(t) = \begin{cases} \{-1\} + \{-1\} = \{-2\} & t < 1 \\ [-1, 0] + \{-1\} = \{\delta - 1 : -1 \le \delta \le 0\} & t = 1 \\ \{1\} + -\{1\} = \{0\} & 1 < t < 2 \\ \{1\} + [0, 1] = \{1 + \delta : 0 \le \delta \le 1\} & t = 2 \\ \{1\} + \{1\} = \{-2\} & t > 2 \end{cases}$$

and $0 \in \partial f(t)$ for $1 \le t \le 2$, the entire interval $[1, 2]$ is a global minimizer of $f(t) = |1 - t| + |2 - t|$.



The reader can verify that in the case of three terms, $f(t) = |1-t| + |2-t| + |3-t|$ say, we have

$$\partial f(t) = \begin{cases} \{-3\} & t < 1 \\ [-3,-1] & t = 1 \\ \{-1\} & 1 < t < 2 \\ [-1,1] & t = 2 \\ \{1\} & 2 < t < 3 \\ [1,3] & t = 3 \\ \{3\} & t > 3 \end{cases}$$

and $0 \in \partial f(t)$ if and only if $t = 2$. More generally, the reader can verify that if

$$f(t) = \sum_{i=1}^{n} |x_i - t|.$$

We can assume w.l.g. that $x_1 < x_2 < \cdots < x_n$ and notice that we have two cases:

Case 1: n **is odd**

In this case

$$0 \in \partial f(t) \Longleftrightarrow t = x_{(n+1)/2}.$$

Case 1: n **is even**

In this case

$$\partial f(t) = 0 \Longleftrightarrow t \in \left[x_{n/2}, x_{1+n/2}\right].$$

In summary:

$$\boxed{\begin{array}{c}\text{The median minimizes} \\ \sum_{i=1}^{n} |x_i - t|.\end{array}} \qquad (17)$$

Example 3:

Suppose that $f_1(\mathbf{x}), ..., f_m(\mathbf{x})$ are convex subdifferentiable functions at $\mathbf{x}$. Define

$$f(\mathbf{x}) = \max_{1 \le j \le m} \{f_j(\mathbf{x})\}.$$

Then $f(\mathbf{x})$ is also convex and subdifferentiable at $\mathbf{x}$, and

$$\partial f(\mathbf{x}) = CO\{\cup_{j \in A} \partial f_j(\mathbf{x})\},$$

where $A = \{j : f_j(\mathbf{x})$ is active at $\mathbf{x}\}$ and $CO(A)$ stand for "convex hull of A", which is defined as the smallest convex set that contains A.

For this example we take

$$f_1(x) = e^{-x} \quad , \quad f_2(x) = e^{x}$$

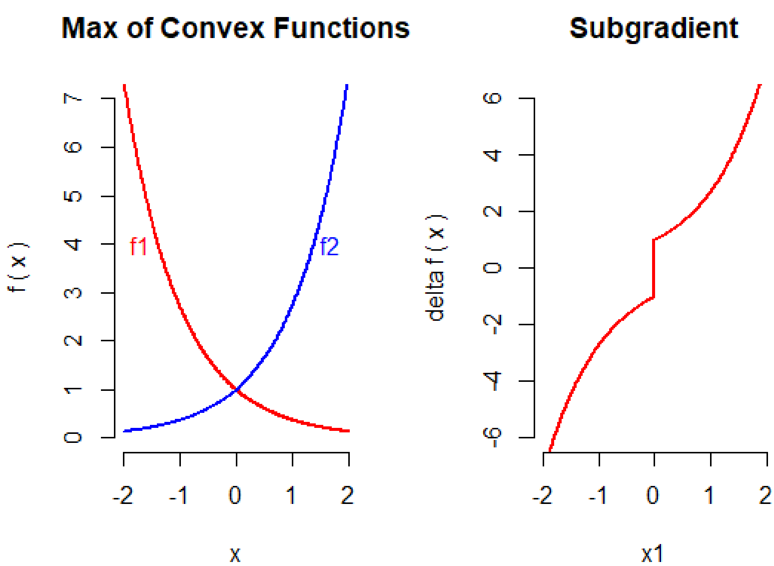
Then

$$f(x) = \max\left\{e^{-x}, e^{x}\right\} = \begin{cases} e^{-x} & x \le 0 \\ e^{x} & x \ge 0 \end{cases}$$

and

$$\partial f(x) = \begin{cases} -e^{-x} & x < 0 \\ [-1, 1] & x = 0 \\ e^{x} & x > 0 \end{cases}$$

Note that the minimizer, 0, belongs to $\partial f(0)$.


Max of Convex Functions
f (x)
x
f1
f2
-2
-1
0
1
2
0
1
2
3
4
5
6
7
Subgradient
delta f (x)
x1
-6
-4
-2
0
2
4
6