Note on Optimization Methods

Chunpai Wang

October 2015

1 Subgradient Method

1.1 Recall

Epigraph: the epigraph, denoted $epi(f)$, describes the set of input-output pairs that f can achieve, as well as "anything above"

$$epi(f) := \{(x, t) \mid x \in dom(f), f(x) \leq t\} \tag{1}$$

Level sets: level sets are sets of points that achieve exactly a certain value for f. Precisely, the t-level set of f is defined by

$$L_t(f) := \{x \in dom(f) \mid f(x) = t\} \tag{2}$$

Sub-level Sets: t-sub-level set of f is defined by

$$S_t(f) := \{x \in dom(f) \mid f(x) \leq t\} \tag{3}$$

Notice the difference between definitions of epigraph and sub-level sets.
First Order Condition: for convex differentiable f

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \forall y \in dom(f) \tag{4}$$

- the first order approximation of f at x is a global lower bound.
- $\nabla f(x)$ defines non-vertical supporting hyperplane to $epi(f)$ at $(x, f(x))$

$$f(x) + \nabla f(x)(y - x) \leq t \quad \forall (y, t) \in epi(f) \tag{5}$$

$$\Leftrightarrow \begin{bmatrix} \nabla f(x) \\ -1 \end{bmatrix} \left(\begin{bmatrix} y \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leq 0 \quad \forall (y, t) \in epi(f) \tag{6}$$

- note that, $y \in dom(f)$ as well, not y-axis. $\forall (y, t) \in epi(f)$

1.2 Why Subgradients?

Subgradients are important for two reasons:

- Convex analysis: optimality characterization via subgradients, monotonicity, relationship to duality
- Convex optimization: if you can compute subgradients, then you can minimize (almost) any convex function

1.3 Subgradients

Subgradient: A vector $g \in \mathbb{R}^n$ is a subgradient of function $f : \mathbb{R}^n \to \mathbb{R}$ at $x \in dom(f)$ if

$$f(y) \geq f(x) + g^T(y - x) \quad \forall y \in dom(f) \tag{7}$$

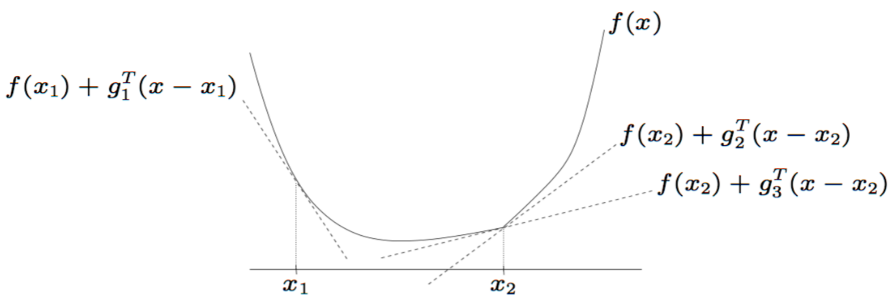

Figure 1: At x_1, the convex function f is differentiable, and g_1 (which is the derivative of f at x_1) is the unique subgradient at x_1. At the point x_2, f is not differentiable. At this point, f has many subgradients: two subgradients, g_2 and g_3 are shown.

Properties:

- The affine function (of y) $f(x) + g^T(y - x)$ is a global lower bound on $f(y)$.
- Geometrically, g define non-vertical supporting hyperplane to $epi(f)$ at $(x, f(x))$ or $(g, -1)$ supports $epi(f)$ at $(x, f(x))$.

$$f(x) + g(y - x) \leq t \quad \forall (y, t) \in epi(f) \tag{8}$$

$$\Leftrightarrow \begin{bmatrix} g \\ -1 \end{bmatrix} \left(\begin{bmatrix} y \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leq 0 \quad \forall (y, t) \in epi(f) \tag{9}$$

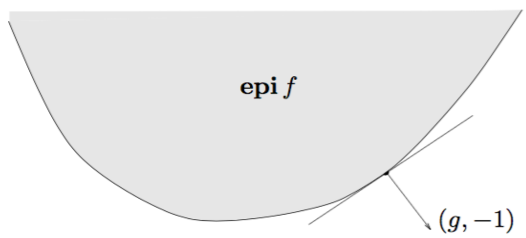

Figure 2: A vector $g \in \mathbb{R}^n$ is a subgradient of f at x if and only if $(g, -1)$ defines a supporting hyperplane to $epi(f)$ at $(x, f(x))$

.

- If f is convex and differentiable, then $\nabla f(x)$ is a subgradient of f at x
- If f is not differentiable at x, there can be more than one subgradient of a function f at a point x.

1.4 Subdifferential

Subdifferential: the subdifferential $\partial f(x)$ of f at x is the set of all subgradients:

$$\partial f(x) = \{g | f(y) \geq f(x) + g^T(y - x) \quad \forall y \in dom(f)\} \tag{10}$$

- A function f is called subdifferentiable at x if there exists at least one subgradient at x.
- A function f is called subdifferentiable if it is subdifferentiable at all $x \in dom(f)$.

Properties:

- The subdifferentiable $\partial f(x)$ is always a closed convex set, even if f is not convex. This follows from the fact that $\partial f(x)$ is the intersection of an infinite set of halfspaces:

$$\partial f(x) = \bigcap_{y \in dom(f)} \{g | f(y) \geq f(x) + g^T(y-x)\} \tag{11}$$

- If f is continuous at x, then the subdifferential $\partial f(x)$ is bounded. Choose some $\epsilon > 0$ such that $-\infty < \underline{f} \leq f(y) \leq \bar{f} < \infty$ for all $y \in \mathbb{R}^n$ such that $||y-x||_2 \leq \epsilon$. If $\partial f(x)$ is unbounded, then there is a sequence $g_n \in \partial f(x)$ such that $||g_n||_2 \to \infty$. Taking the sequence $y_n = x + \epsilon \frac{g_n}{||g_n||_2}$, we find that $f(y_n) \geq f(x) + g_n^T(y_n - x) = f(x) + \epsilon ||g_n||_2 \to \infty$, which is a contradiction to $f(y_n)$ being bounded.

- If f is convex and differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$, i.e., its gradient is its only subgradient. Conversely, if f is convex and $\partial f(x) = \{g\}$, then f is differentiable at x and $g = \nabla f(x)$.

- Existence of Subgradients: If f is convex and $x \in$ int $dom(f)$, then $\partial f(x)$ is nonempty and bounded.

  Proof. □

- Monotonicity: subdifferential of a convex function is a monotone operator:

$$(u-v)^T(y-x) \geq 0 \quad \forall x, y, \quad u \in \partial f(x),\ v \in \partial f(y), \tag{12}$$

  which means, if y is greater than x, then subdifferential $\partial f(y) \geq \partial f(x)$.

  Proof. Because $u \in \partial f(x)$, we can get by definition

$$f(y) \geq f(x) + u^T(y-x)$$

  and because $v \in \partial f(y)$, we can get by definition

$$f(x) \geq f(y) + v^T(x-y)$$

  Combining these two inequalities shows monotonicity

$$f(y) + f(x) \geq f(x) + f(y) + u^T(y-x) + v^T(x-y)$$
$$(u-v)^T(x-y) \geq 0$$

  □

1.5 Examples

(Example 1) Absolute value $f(x) = |x|$

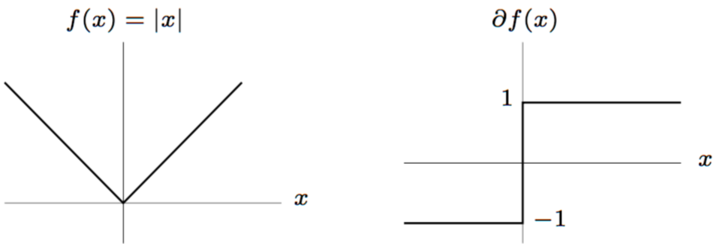

Figure 3: Consider $f(x) = |x|$. For $x < 0$ the subgradient is unique: $\partial f(x) = \{-1\}$. Similarly, for $x > 0$ we have $\partial f(x) = 1$. At $x = 0$ the subdifferential is defined by the inequality $|y| \geq g(y - 0)$ for all y, which is satisfied if and only if $g \in [-1, 1]$.Therefore, we have $\partial f(0) = [-1, 1]$.

(Example 2) $f(x) = \max\{f_1(x), f_2(x)\}$ $\quad f_1, f_2$ convex and differentiable

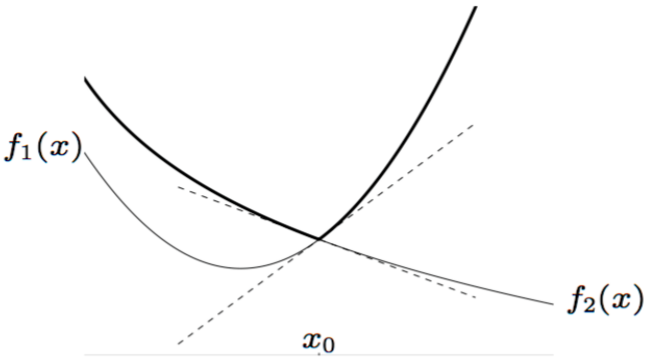

Figure 4: For $f_1(x) > f_2(x)$, unique subgradient $g = \nabla f_1(x)$; for $f_2(x) > f_1(x)$, unique subgradient $g = \nabla f_2(x)$; for $f_1(x) = f_2(x)$, subgradient g is any point on the line segment between $\nabla f_1(x)$ and $\nabla f_2(x)$. Note, $g \in \mathbb{R}^n$, when $n = 1$, g is scalar and line segment is just an interval; when n $>$1, for example, $g \in \mathbb{R}^2$, $f_1(x) = (1, 0)$ and $f_2(x) = (2, 3)$, g is any point between line segment from point $(1, 0)$ to $(2, 3)$

(Example 3) Euclidean norm $f(x) = ||x||_2$

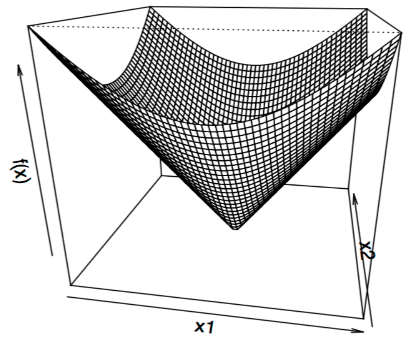

Figure 5: For $x \neq 0$, $\partial f(x) = \frac{x}{||x||_2}$; for $x = 0$, $\partial f(x) = \{g|||g||_2 \leq 1\}$

(Example 4) l_1 norm $f(x) = ||x||_1$

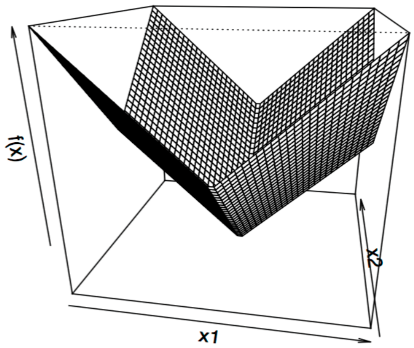

Figure 6: For $x_i \neq 0$, unique i^{th} component $g_i = sign(x_i)$; for $x_i = 0$, i^{th} component g_i is any element of [-1,1]

Questions:
If a function has subgradient at every point, can we prove the function is convex ?
Think about the supporting hyperplane theory and subgradient.

1.6 Connection to Convex Geometry

Now we try to derive subgradient from indication function of convex set.
Convex set $C \subset \mathbb{R}^n$, consider indicator function $I_C : \mathbb{R}^n \to \mathbb{R}$,

$$I_C(x) = I(x \in C) = \begin{cases} 0 & \text{if } \ x \in C \\ \infty & \text{if } \ x \notin C \end{cases} \tag{13}$$

For $x \in C, \partial I_C(x) = N_C(x)$, the normal cone of C at x

$$N_C(x) = \{g \in R^n : g^T x \geq g^T y\} \ \text{ for any } \ y \in C \tag{14}$$

Proof. By definition of subgradient g.

$$I_C(y) \geq I_C(x) + g^T(y - x) \quad \text{for all } \ y \tag{15}$$

- for $y \notin C$, $I_C(y) = \infty$
- for $y \in C$, this means $0 \geq g^T(y - x)$

□

For $x \notin C$, $I_C(x) = \infty$. You cannot find any points $y \in C$ to make inequality (15) satisfied. This is also a proof of existence of subgradient that $x \in$ int $dom(f)$.

1.7 Optimality Condition

Subgradient Optimality Condition: A point x^* is a minimizer of a function f (convex or not) if and only if f is subdifferentiable at x^* and $0 \in \partial f(x^*)$, i.e., $g = 0$ is a subgradient of f at x^*.

$$f(x^*) = \min_x f(x) \quad \Leftrightarrow \quad 0 \in \partial f(x^*) \tag{16}$$

Proof. From the fact that $f(x) \geq f(x^*)$ for all $x \in dom(f)$. Clearly, if f is subdifferentiable at x^* with $0 \in \partial f(x^*)$, then $f(x) \geq f(x^*) + 0^T(x - x^*) = f(x^*)$ for all x. Let $g = 0$ being a subgradient means that for all y □

Remark: while this simple characterization of optimality via the subdifferential holds for nonconvex functions, it is not particularly useful in that case, since we generally cannot find the subdifferential of a nonconvex function.

Theorem 1.1. *For f convex and differentiable, the problem*

$$\min_x f(x) \quad \textit{subject to} \quad x \in C \tag{17}$$

is solved at x if and only if

$$\nabla f(x)^T(y - x) \geq 0 \quad \textit{for all} \quad y \in C \tag{18}$$

Proof. First recast problem as

$$\min_x f(x) + I_C(x) \tag{19}$$

Now we apply subgradient optimality: $0 \in \partial(f(x) + I_C(x))$. We get,

$$\begin{aligned}
0 \in \partial(f(x) + I_C(x)) &\Leftrightarrow 0 \in \{\partial f(x)\} + N_C(x) \\
&\Leftrightarrow -\nabla f(x) \in N_C(x) \qquad \text{(because } f \text{ is convex and differentiable)} \\
&\Leftrightarrow -\nabla f(x)^T x \geq -\nabla f(x)^T y \quad \text{for all} \quad y \in C \\
&\Leftrightarrow \nabla f(x)^T(y - x) \geq 0 \quad \text{for all} \quad y \in C
\end{aligned}$$

□

Example: *Lasso Optimality Conditions.* Given $A \in \mathbb{R}^{n \times p}, b \in \mathbb{R}^n$, lasso problem can be parametrized as:

$$\min_{x \in \mathbb{R}^p} \frac{1}{2}||Ax - b||_2^2 + \lambda||x||_1 \tag{20}$$

where $\lambda \geq 0$. And we can get from subgradient optimality that:

$$0 \in \partial\Big(\frac{1}{2}||Ax - b||_2^2 + \lambda||x||_1\Big) \tag{21}$$

$$\Leftrightarrow 0 \in A^T(Ax - b) + \lambda\partial||x||_1 \tag{22}$$

$$\Leftrightarrow A^T(Ax - b) = -\lambda v \tag{23}$$

for some $v \in \partial||x||_1$, i.e., (check the subgradient of l-1 norm on page 48)

$$v_i \in \begin{cases} \{1\} & if \quad x_i > 0 \\ \{-1\} & if \quad x_i < 0, i = 1, ..., p \\ [-1, 1] & if \quad x_i = 0 \end{cases} \tag{24}$$

Write $A_1, A_2, ..., A_p$ for columns of A. Then subgradient optimality of lasso becomes:

$$\begin{cases} A_i^T(Ax - b) = -\lambda \ sign(x_i) & if \quad x_i \neq 0 \\ |A_i^T(Ax - b)| \leq \lambda & if \quad x_i = 0 \end{cases} \tag{25}$$

Note: the subgradient optimality conditions do not directly lead to an expression for a lasso solution. However, they do provide a way to check lasso optimality, to check if it has converged or not?? They are also helpful in understanding the lasso estimator; e.g., if $|A_i^T(Ax - b)| < \lambda$, then $x_i = 0$.(What this useful ???????????????????)

Example: *Soft-Thresholding.* Simplified Lasso problem with $A = I$:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}||x - b||_2^2 + \lambda||x||_1 \tag{26}$$

This we can solve directly using subgradient optimality. The closed form solution is $x = S_\lambda(b)$, where S_λ is the soft-thresholding operator:

$$\begin{cases} b_i - \lambda & if \quad b_i > \lambda \\ 0 & if - \lambda \leq b_i \leq \lambda, i = 1, ..., n \\ b_i + \lambda & if \quad b_i < -\lambda \end{cases} \tag{27}$$

Check: for Lasso problem, subgradient optimality conditions are

$$\begin{cases} A_i^T(Ax - b) = -\lambda \; sign(x_i) & if \quad x_i \neq 0 \\ |A_i^T(Ax - b)| \leq \lambda & if \quad x_i = 0 \end{cases}$$

Now plug in $x = S_\lambda(b)$ and check these are satisfied:

- when $b_i > \lambda$, $x_i = b_i - \lambda > 0$, so $x_i - b_i = -\lambda = -\lambda \cdot 1$
- when $b_i < -\lambda$,$x_i = b_i + \lambda < 0$
- when $|b_i| \leq \lambda$, $x_i = 0$, and $|x_i - b_i| = |b_i| \leq \lambda$

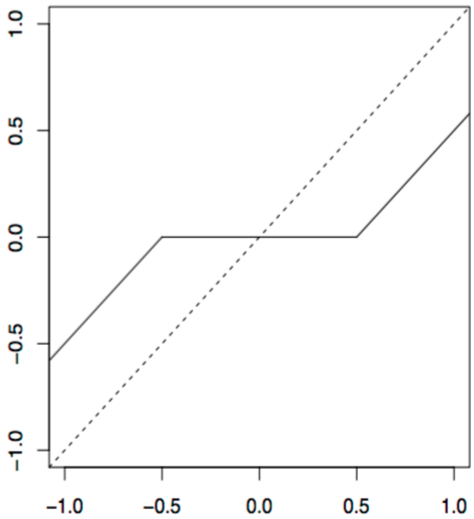

Figure 7: Soft-thresholding in one variable

1.8 Subgradient Calculus

weak subgradient calculus: rules for finding one subgradient

- sufficient for most non-differentiable convex optimization algorithms
- if you can evaluate $f(x)$, you can usually compute a subgradient

strong subgradient calculus: rules for finding $\partial f(x)$ (all subgradients)

- some algorithms, optimality conditions, etc., need entire subdifferential
- can be quite complicated

1.8.1 Basic rules for convex functions

- Scaling: $\partial(af) = a \cdot \partial f$ provided $a > 0$ to assure convexity.
- Addition: $\partial(f_1 + f_2) = \partial f_1 + \partial f_2$
- Affine Composition: if $g(x) = f(Ax + b)$, then

$$\partial g(x) = A^T \partial f(Ax + b)$$

1.8.2 Finite pointwise maximum

If $f(x) = \max_{i=1,\dots,m} f_i(x)$, then

$$\partial f(x) = conv\Big(\bigcup \partial f_i(x)\Big),$$

the convex hull of union of subdifferentials of all active functions at x, since subdifferentials are always convex.
Convex Hull: the convex hull of a set C, is the set of all convex combinations of points in C:

$$\text{conv}(C) = \{\theta_1 x_1 + ... + \theta_k x_k | x_i \in C, \theta_i \geq 0, i = 1, ..., k, \theta_1 + ... + \theta_k = 1\} \tag{28}$$

The convex hull is always convex. It is the smallest convex set that contains C.

Example: $l_1 - norm$. The l_1-norm

$$f(x) = ||x||_1 = |x_1| + ... + |x_n| \tag{29}$$

is a nondifferentiable convex function of x. To find its subgradients, we note that f can expressed as the maximum of 2^n linear functions:

$$||x||_1 = \max\{s^T x | s_i \in \{-1, 1\}\}, \tag{30}$$

so we can apply the rules for the subgradient of the maximum. The first step is to identify an active function $s^T x$, i.e., find an $s \in \{-1, +1\}^n$ such that $s^T x = ||x||_1$. Since the function $s^T x$ is differentiable and has a unique subgradient s. We can therefore take

$$s_i = g_i = \begin{cases} +1 & x_i > 0 \\ -1 & x_i < 0 \\ -1 \text{ or } +1 & x_i = 0 \end{cases} \tag{31}$$

The subdifferential is the convex hull of all subgradients that can be generated this way:

$$\partial f(x) = \{g | ||g||_\infty \leq 1, g^T x = ||x||_1\} WHY??? \tag{32}$$

1.8.3 Pointwise Supremum

We consider the extension to the supremum over an infinite number of functions,

$$f(x) = \sup_{\alpha \in A} f_\alpha(x), \tag{33}$$

where the functions f_α are subdifferentiable.
Weak Result: assume maximum is attained, i.e., $\sup_{\alpha \in A} f_\alpha(x) = \max_{\alpha \in A} f_\alpha(x)$, we can find a subgradient at x.

- find any β for which $f_\beta(x) = f(x)$
- choose any $g \in \partial f_\beta(x)$

(Partial) Strong Result: define $I(x) = \{\alpha \in A | f_\alpha = f(x)\}$

$$\partial f(x) \supseteq conv\Big(\bigcup_{\alpha \in I(x)} \partial f_\alpha(x) \Big) \tag{34}$$

If A is compact and f_α continuous in α, then

$$\partial f(x) = conv\Big(\bigcup_{\alpha \in I(x)} \partial f_\alpha(x) \Big) \tag{35}$$

Example1: *maximum eigenvalue of a symmetric matrix.* Recall, a real scalar λ is said to be an eigenvalue of symmetric matrix S if there exist a non-zero vector $u \in \mathbb{R}^n$ such that

$$Su = \lambda u,$$

where vector u is referred to as an eigenvector associated with the eigenvalue λ. The eigenvector u is said to be normalized if $||u||_2 = 1$. In this case, we have

$$u^T Su = u^T \lambda u = \lambda u^T u = \lambda ||u||_2^2 = \lambda$$

The interpretation of u is that it defines a direction along S behaves just like scalar multiplication. And we can find the smallest and largest eigenvalues of S, denoted λ_{min} and λ_{max} respectively,

$$\lambda_{min} = \min_x \{u^T Su | u^T u = 1\} \tag{36}$$

$$\lambda_{max} = \max_x \{u^T Su | u^T u = 1\} \tag{37}$$

Now let $f(x) = \lambda_{max}(S(x))$, where $S(x) = S_0 + x_1 S_1 + ... + x_n S_n$ with symmetric coefficients S_i. We can express f as the pointwise supremum of convex functions, (why convex ??)

$$f(x) = \lambda_{max}(S(x)) = \sup_{||u||_2 = 1} u^T S(x) u \tag{38}$$

Since sup means we may not find the maximum of this function by satisfying $||u||_2 = 1$, hence the index set A is

$$A = \{u \in \mathbb{R}^n | ||u||_2 \leq 1\} \tag{39}$$

here has infinite number of u, therefore we are solving the supremum over an infinite number of functions. Each of the functions $f_u(x) = u^T S(x) u$ is affine in x for fixed u, as can be easily seen from

$$u^T S(x) u = u^T S_0 u + x_1 u^T S_1 u + ... + x_n u^T S_n u \tag{40}$$

so it is differentiable with gradient

$$\nabla f_u(x) = (u^T S_1 u + ... + u^T S_n u) \tag{41}$$

Hence to find a subgradient, we compute an eigenvector u with eigenvalue λ_{max}, normalized to have unit norm, and take

$$g = (u^T S_1 u + ... + u^T S_n u) \tag{42}$$

The index set in this example is $A = \{u | ||u||_2 = 1\}$ is a compact set (closed and bounded). Therefore,

$$\partial f(x) = conv\{\nabla f_u(x) | u^T S(x) u = \lambda_{max}(S(x)), ||u||_2 = 1\} \tag{43}$$

Example2: *maximum eigenvalue of a symmetric matrix,revisited.* Let $f(S) = \lambda_{max}(S)$,where S is a $n \times n$ symmetric matrix. Then as above, $f(S) = \lambda_{max}(S) = \sup_{||u||_2=1} u^T Su$, but we note that $u^T Su = Trace(Suu^T)$, so that each of the functions $f_u(A) = u^T Su$ is linear in S with gradient $\nabla f_u(A) = uu^T$. Then using an identical argument to that above, we find that

$$\partial f(S) = conv\{uu^T | ||u||_2 = 1, u^T Su = \lambda_{max}(S)\} = conv\{uu^T = 1, Su = \lambda_{max}(S) u\} \tag{44}$$

1.8.4 Minimization Over Some Variables

Now, we consider the function with the form

$$f(x) = \inf_y H(x, y) \tag{45}$$

where $H(x, y)$ is subdifferentiable and jointly convex in $x \in R^n$ and $y \in R^n$. Suppose that the infimum over y in the definition of $f(x)$ is attained on the set $Y_x \subset \mathbb{R}^m$ (where $Y_x \neq 0$), so that $H(x, y) = f(x)$ for $y \in Y_x$. By definition, a vector $g \in \mathbb{R}^n$ is a subgradient of f is and if

$$f(x') \geq f(x) + g^T(x' - x) = H(x, y) + g^T(x' - x) \tag{46}$$

for all $x' \in \mathbb{R}^n$ and any $y \in Y_x$. This is equivalent to

$$H(x', y') \geq H(x, y) + g^T(x' - x) = H(x, y) + \begin{bmatrix} g \\ 0 \end{bmatrix} \left(\begin{bmatrix} x' \\ y' \end{bmatrix} - \begin{bmatrix} x \\ y \end{bmatrix} \right) \tag{47}$$

for all $(x', y') \in R^n \times R^m$ and $x, y \in Y_x$.

Weak Result: to find a subgradient at x,

- find y that minimize $H(x, y)$
- find subgradient $(g, 0) \in \partial H(x, y)$

In particular we have the result that

$$\partial f(x) = \{g \in \mathbb{R}^n | (g, 0) \in \partial H(x, y) \text{ for some } y \in Y_x\} \tag{48}$$

Example:*Euclidean Distance to Convex Set.* Now we are trying to find the a subgradient of

$$f(x) = \inf_{y \in C} ||x - y||_2 \tag{49}$$

where C is a closed convex set. To find a subgradient at x, we can conclude the solution as following,

- if $f(x) = 0$, that is $x \in C$ and $f(x)$ is the minimum of $||x - y||_2$, thus $g = 0$
- if $f(x) > 0$, find projection $y = P(x)$ on C

$$g = \frac{1}{||x - y||_2}(x - y) = \frac{1}{||x - P(x)||_2}(x - P(x)) \quad (WHY???????) \tag{50}$$

The gradient points in the direction of the greatest rate of increase of the function and its magnitude is the slope of the graph in that direction .

1.8.5 Optimal Value Function of a Convex Optimization Problem

Suppose $f : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ is defined as the optimal value of a convex optimization problem in standard form, with $z \in \mathbb{R}^n$ as optimization variable,

$$\begin{aligned} &\text{minimize} \quad f_0(z) \\ &\text{subject to} \quad f_i(z) \leq x_i, \quad i = 1, ..., m \text{ and } \quad Az = y \end{aligned} \tag{51}$$

In other words, $f(x, y) = \inf_z H(x, y, z)$ where

$$H(x, y, z) = \begin{cases} f_0(z) & f_i(z) \leq x_i, i = 1, ..., m, Az = y \\ +\infty & otherwise \end{cases} \tag{52}$$

which is jointly convex in x, y, z. Subgradients of f can be related to the dual problem of (43) as follows. Suppose we are interested in subdifferentiating f at (x, y). We can express the dual problem of (43) as

$$\begin{aligned} &\text{maximize} \quad g(\lambda) - x^T\lambda - y^T v \\ &\text{subject to} \quad \lambda \succeq 0 \end{aligned} \tag{53}$$

where

$$g(\lambda) = \inf_z \left(f_0(z) + \sum_{i=1}^m \lambda_i f_i(z) + v^T Az \right) \tag{54}$$

1.9 Directional Derivatives and Subgradients

Directional derivative of f at x in the direction v is

$$f'(x;v) = \lim_{\alpha \to 0} \frac{f(x+tv) - f(x)}{t} \tag{55}$$

$$\tag{56}$$

This quantity always exists for <u>convex</u> f, though it may be $+\infty$ or $-\infty$. To see the existence of the limit, we use that the ratio

$$\frac{f(x+tv) - f(x)}{t} \tag{57}$$

is non-decreasing in t. For $0 < t_1 \le t_2$, we have $0 \le t_1/t_2 \le 1$, and

$$\frac{f(x+t_1 v) - f(x)}{t_1} = \frac{f(\frac{t_1}{t_2}(x+t_2 v) + (1-\frac{t_1}{t_2})x) - f(x)}{t_1} \tag{58}$$

$$\text{(convex definition: } f(\alpha x_1 + (1-\alpha)x_2) \le \alpha f(x_1) + (1-\alpha) f(x_2)) \tag{59}$$

$$\le \frac{\frac{t_1}{t_2} f(x+t_2 v)}{t_1} + \frac{(1-\frac{t_1}{t_2}) f(x) - f(x)}{t_1} \tag{60}$$

$$= \frac{f(x+t_2 v) - f(x)}{t_2} \tag{61}$$

so the limit in the definition of $f'(x;v)$ exists.
<u>Properties:</u> Several properties of directional derivative $f'(x;v)$

- it is convex in v, and if f is finite in a neighborhood of x, then $f'(x;v)$ exists.