Prerequisites

Calculus (multivariate functions, gradients)

Linear algebra (norms)

Sparse regression via the lasso

Convexity
Epigraph

The epigraph of $f : \mathbb{R}^n \to \mathbb{R}$ is a set in $\mathbb{R}^{n+1}$

$$\text{epi}(f) := \left\{ x \mid f \left( \begin{bmatrix} x[1] \\ \vdots \\ x[n] \end{bmatrix} \right) \leq x[n + 1] \right\}$$
A hyperplane $\mathcal{H}$ is a supporting hyperplane of a set $S$ at $x$ if

- $\mathcal{H}$ and $S$ intersect at $x$
- $S$ is contained in one of the half-spaces bounded by $\mathcal{H}$
Supporting hyperplane
Convexity

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if its epigraph has a supporting hyperplane at every point.

It is strictly convex if and only for all $x \in \mathbb{R}^n$ it only intersects with the supporting hyperplane at one point.
Subgradients

The subgradient of $f : \mathbb{R}^n \to \mathbb{R}$ at $x \in \mathbb{R}^n$ is a vector $g \in \mathbb{R}^n$ such that

$$f(y) \geq f(x) + g^T(y - x), \quad \text{for all } y \in \mathbb{R}^n$$

The hyperplane

$$\mathcal{H}_g := \left\{ y \mid y[n+1] = f(x) + g^T \begin{pmatrix} y[1] \\ \vdots \\ y[n] \end{pmatrix} - x \right\}$$

is a supporting hyperplane of the epigraph of $f$ at $\begin{bmatrix} x \\ f(x) \end{bmatrix}$.
If a function is differentiable, the only subgradient at each point is the gradient.
Proof

Assume \( g \) is a subgradient at \( x \), for any \( \alpha \geq 0 \)

\[
f(x + \alpha e_i) \geq f(x) + g^T \alpha e_i \\
= f(x) + g[i] \alpha 
\]

\[
f(x) \leq f(x - \alpha e_i) + g^T \alpha e_i \\
= f(x - \alpha e_i) + g[i] \alpha 
\]

Combining both inequalities

\[
\frac{f(x) - f(x - \alpha e_i)}{\alpha} \leq g[i] \leq \frac{f(x + \alpha e_i) - f(x)}{\alpha} 
\]

Letting \( \alpha \to 0 \), implies \( g[i] = \frac{\partial f(x)}{\partial x[i]} \)
Optimality condition for nondifferentiable functions

$x$ is a minimum of $f$ if and only if the zero vector is a subgradient of $f$ at $x$

\[ f(y) \geq f(x) + 0^T(y - x) \]

\[ = f(x) \]

for all $y \in \mathbb{R}^n$

Under strict convexity the minimum is unique
Let $g_1$ and $g_2$ be subgradients at $x \in \mathbb{R}^n$ of $f_1 : \mathbb{R}^n \to \mathbb{R}$ and $f_2 : \mathbb{R}^n \to \mathbb{R}$.

$g := g_1 + g_2$ is a subgradient of $f := f_1 + f_2$ at $x$.

**Proof:** For any $y \in \mathbb{R}^n$

\[
\begin{align*}
f (y) &= f_1 (y) + f_2 (y) \\
&\geq f_1 (x) + g_1^T (y - x) + f_2 (y) + g_2^T (y - x) \\
&\geq f (x) + g^T (y - x)
\end{align*}
\]
Let $g_1$ be a subgradient at $x \in \mathbb{R}^n$ of $f_1 : \mathbb{R}^n \rightarrow \mathbb{R}$.

For any $\alpha \geq 0$ $g_2 := \alpha g_1$ is a subgradient of $f_2 := \alpha f_1$ at $x$.

**Proof:** For any $y \in \mathbb{R}^n$

\[
f_2(y) = \alpha f_1(y) \\
\geq \alpha \left( f_1(x) + g_1^T (y - x) \right) \\
\geq f_2(x) + g_2^T (y - x)
\]
Subdifferential of absolute value

At $x \neq 0$, $f(x) = |x|$ is differentiable, so $g = \text{sign}(x)$

At $x = 0$, we need

$$f(0 + y) \geq f(0) + g(y - 0)$$

$$|y| \geq gy$$

Holds if and only if $|g| \leq 1$
Subdifferential of absolute value

\[ f(x) = |x| \]
Subdifferential of $\ell_1$ norm

$g$ is a subgradient of the $\ell_1$ norm at $x \in \mathbb{R}^n$ if and only if

$$g[i] = \text{sign}(x[i]) \quad \text{if} \quad x[i] \neq 0$$

$$|g[i]| \leq 1 \quad \text{if} \quad x[i] = 0$$
Proof (one direction)

Assume $g[i]$ is a subgradient of $|\cdot|$ at $|x[i]|$ for $1 \leq i \leq n$

For any $y \in \mathbb{R}^n$

$$||y||_1 = \sum_{i=1}^{n} |y[i]|$$

$$\geq \sum_{i=1}^{n} |x[i]| + g[i] (y[i] - x[i])$$

$$= ||x||_1 + g^T (y - x)$$
Subdifferential of $\ell_1$ norm
Subdifferential of $\ell_1$ norm
Subdifferential of $\ell_1$ norm
What have we learned?

Definition of subgradients

Optimality condition for nondifferentiable convex functions

Subgradients of $\ell_1$ norm