2.2.3 Optimisation theory and convex analysis

Convexity

A set $A \subseteq \mathbb{R}^d$ is convex if

$$x, y \in A \Rightarrow (1-t)x + ty \in A \quad \forall t \in (0,1)$$

In certain settings it will be convenient to consider functions that take, in addition to real values, the value ∞. Denote $\bar{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$. A function $f: \mathbb{R}^d \to \bar{\mathbb{R}}$ is convex if

$$f((1-t)x + ty) \leq (1-t)f(x) + t f(y) \quad \forall x, y \in \mathbb{R}^d, \ t \in (0,1)$$

and $f(x) < \infty$ for at least one x [proper convex function]. It is strictly convex if the equality is strict for all $x, y \in \mathbb{R}^d$, $x \neq y$ and $t \in (0,1)$. Define the domain of f, to be $\operatorname{dom} f = \{x : f(x) < \infty\}$. Note that when f is convex, $\operatorname{dom} f$ must be a convex set.

Prop 16

i) Let $f_1, \ldots, f_m : \mathbb{R}^d \to \bar{\mathbb{R}}$ be convex functions with $\operatorname{dom} f_1 \cap \ldots \cap \operatorname{dom} f_m \neq \phi$. Then if $c_1, \ldots, c_m \geq 0$, $c_1 f_1 + \ldots + c_m f_m$ is a convex function.

ii) If $f: \mathbb{R}^d \to \mathbb{R}$ is twice its differentiable then

a) f is convex iff its Hessian $H(x)$ is pos semi-definite $\forall x$

b) f is strictly convex if $H(x)$ is pos definite $\forall x$.

The Lagrangian method

Consider an optimisation problem of the form

$$\text{minimise } f(x), \text{ subject to } g(x) = 0 \quad \text{where } g: \mathbb{R}^d \to \mathbb{R}^b. \quad (*)$$

Suppose the optimal value is $c^* \in \mathbb{R}$. The Lagrangian for this problem is def as

$$L(x, \theta) = f(x) + \theta^T g(x) \quad \text{where } \theta \in \mathbb{R}^b.$$

Note that

$$\inf_{x \in \mathbb{R}^d} L(x, \theta) \leq \inf_{x \in \mathbb{R}^d : g(x) = 0} L(x, \theta) = c^* \quad \forall \theta$$

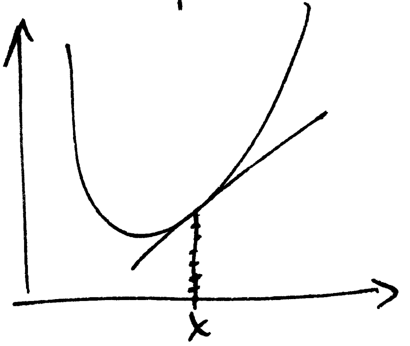
The Lagrangian method involves finding a $\theta^* \in \mathbb{R}^b$ s.t. the minimising x^* on the LHS ($L(x, \theta^*)$) satisfies $g(x^*) = 0$. Then x^* must minimise $(*)$.

Subgradients

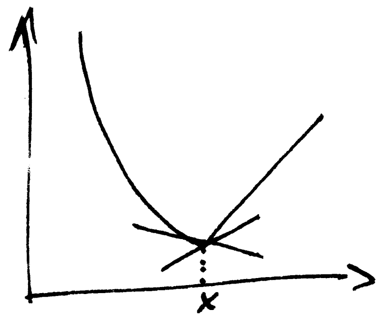
Def^n 5 A vector $v \in \mathbb{R}^d$ is a subgradient of a convex f^n $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ at x if,

$$f(y) \geq f(x) + v^T (y - x) \quad \forall y \in \mathbb{R}^d$$

The set of subgradients of f at x is called the subdifferential, and is denoted $\partial f(x)$.



$\partial f(x) = \{\nabla f(x)\}$



$\partial f(x)$ has many elements

Prop 17 Let $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ be convex and differentiable at $x \in \operatorname{int}(\operatorname{dom} f)$.
Then $\partial f(x) = \{\nabla f(x)\}$

Prop 18 Let $f, g : \mathbb{R}^d \to \overline{\mathbb{R}}$ be convex fn with $\operatorname{int}(\operatorname{dom} f) \cap \operatorname{int}(\operatorname{dom} g) \neq \phi$,
and let $\alpha > 0$. Then

$$\partial(\alpha f)(x) = \alpha\, \partial f(x) = \{\alpha v : v \in \partial f(x)\}$$

$$\partial(f+g)(g) = \partial f(x) + \partial g(x) = \{v + w : v \in \partial f(x),\ w \in \partial g(x)\}$$

Prop 19 $x^* = \arg\min f(x)$ iff $0 \in \partial f(x^*)$

Proof: $f(y) \geq f(x^*)\ \forall y \iff f(y) \geq f(x^*) + 0^T (y - x)$

$$\iff 0 \in \partial f(x^*) \quad \square.$$

Notation

For $x \in \mathbb{R}^d$, x_A where $A \subseteq \{1, \ldots, d\}$, $A \neq \phi$ is the subvector of x formed from components of x indexed by A. In this context $A^c = \{1, \ldots, d\} \setminus A$. Also, $-j$, $-jk$ in subscripts is shorthand for $\{j\}^c$, $\{j, k\}^c$. All subsetting ops occur first, so $x_A^T = (x_A)^T$.

Define,

$$\operatorname{sgn}(x_i) = \begin{cases} 1 & : x_i > 0 \\ 0 & : x_i = 0 \\ -1 & : x_i < 0 \end{cases} \quad \text{and} \quad \operatorname{sgn}(x) = (\operatorname{sgn}(x_1), \dots, \operatorname{sgn}(x_n))^T$$

We now compute the subdifferential of the l_1-norm. Note, $\|\cdot\| : \mathbb{R}^d \to \mathbb{R}$ since it is a norm.

$$\|tx + (1-t)y\| < \|tx\|_1 + \|(1-t)y\|_1 = t\|x\|_1 + (1-t)\|y\|_1$$

<u>Prop 20</u> For $x \in \mathbb{R}^d$, let $A = \{k : x_k \neq 0\}$. Then

$$\partial \|x\|_1 = \{v \in \mathbb{R}^d : \|v\|_\infty < 1 \text{ and } v_A = \operatorname{sgn}(x_A)\}$$

<u>Proof</u>: For $j = 1, \dots, d$ let $g_j : \mathbb{R}^d \to \mathbb{R}$, $x \mapsto |x_j|$ so $\|\cdot\|_1 = \sum_j g_j(\cdot)$

and by prop 18, $\partial\|x\|_1 = \sum_j \partial g_j(x)$. When $x_j \neq 0$ then g_j is differentiable at x, so by prop 17, $\partial g_j(x) = \{\nabla g_j(x)\} = \{\operatorname{sgn}(x_j) e_j\}$ where e_j is the j^{th} unit vector.

When $x_j = 0$ if $v \in \partial g_j(x)$ then

$$g_j(y) \geq g_j(x) + v^T(y - x) \quad \forall y$$

So $|y_j| \geq v^T(y-x) \quad \forall y$.

We claim that this holds for $v \iff v_{-j} = 0$ and $v_j \in [-1, 1]$

"$\Leftarrow$" $v^T(y-x) = v_j y_j \leq |y_j|$ ✓

"$\Rightarrow$" Set $y_{-j} = x_{-j} + v_{-j}$ and $y_j = 0$. Have,

$$0 \geq v_{-j}^T v_{-j} = \|v_{-j}\|_2^2 \Rightarrow v_{-j} = 0$$

Take y with $y_{-j} = x_{-j}$. Then $|y_j| \geq v_j y_j \Rightarrow |v_j| \leq 1$ ✓ □

2.2.4 Properties of the Lasso

Recall $Q_\lambda(\beta) = \frac{1}{2n}\|Y - X\beta\|_2^2 + \lambda\|\beta\|_1$

$\hat{\beta}_\lambda^L$ is a Lasso soln iff $0 \in \partial Q_\lambda(\hat{\beta}_\lambda^L)$. This is equiv to

$$-\frac{1}{n}X^T(Y - X\hat{\beta}_\lambda^L) = \lambda\hat{\nu} \qquad \text{— KKT conditions for the Lasso}$$

for $\|\hat{\nu}\|_\infty \leq 1$ and $\hat{S}_\lambda = \{k : \hat{\beta}_{\lambda,k}^L \neq 0\}$ then $\hat{\nu}_{\hat{S}_\lambda} = \operatorname{sgn}\{\hat{\beta}_{\lambda,\hat{S}_\lambda}^L\}$