Mod 2: Convex Optimization and Coordinate Descent

Ruben Zamar
Department of Statistics
UBC

September 30, 2018

Convex Sets in R^p

$$\mathbf{a}, \mathbf{b} \quad \in \quad A \Rightarrow \alpha \mathbf{a} + (1-\alpha)\, \mathbf{b} \in A, \quad \text{for all} \quad 0 \leq \alpha \leq 1$$

Convex Functions

Let $A \subset R^p$ be a convex set and $f : A \to R$

f is **convex** if for all $0 \leq \alpha \leq 1$ and all $\mathbf{a}, \mathbf{b} \in A$ we have

$$f(\alpha \mathbf{a} + (1-\alpha)\mathbf{b}) \leq \alpha f(\mathbf{a}) + (1-\alpha) f(\mathbf{b})$$

f is **strictly convex** if for all $0 < \alpha < 1$ and all $\mathbf{a} \neq \mathbf{b} \in A$ we have

$$f(\alpha \mathbf{a} + (1-\alpha)\mathbf{b}) < \alpha f(\mathbf{a}) + (1-\alpha) f(\mathbf{b})$$

- **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) $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

Differentiable Function on R^p

The function $f(\mathbf{x})$ **is differentiable at x** if

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

The function $f(\mathbf{x})$ **is differentiable** if the gradient $\nabla f(\mathbf{x})$ exists at each interior point of its domain.

First Order Condition for Convexity

Suppose the function $f(\mathbf{x})$ is differentiable on an open domain A. Then $f(\mathbf{x})$ is convex if and only if the first order condition below holds

$$f(\mathbf{y}) \quad \geq \quad f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \quad \text{for all} \quad \mathbf{x}, \mathbf{y} \in A.$$

Example

Consider the convex function

$$f(x) = x^2$$

In this case

$$\nabla f(x) = 2x$$

The first order condition for convexity is:

$$\begin{aligned} f(y) &\geq f(x) + \nabla f(x)(y - x) \\ y^2 &\geq x^2 + 2x(y - x) \\ (y - x)^2 &\geq 0 \end{aligned}$$

Global minimization of a differentiable convex function

Suppose

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

   $\mathbf{x}_0$ is a global minimizer of $f(\mathbf{x})$ if and only if $\nabla f(\mathbf{x}_0) = 0$.

Proof: Sufficiency follows directly from (**??**) 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

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 $\mathbf{y}$.

Back-Fitting:

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:

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

By construction

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

and

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

Moreover:

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

If $f(\mathbf{x},\mathbf{y})$ is convex and differentiable, $f\left(\mathbf{x}^{k},\mathbf{y}^{k}\right)$ converges to a global minimum, $f(\mathbf{x}^{*},\mathbf{y}^{*})$ (next theorem)

Later on, we will show that the differentiability condition can be relaxed to *sub-differentiability.*

Theorem 1. Suppose that $f(\mathbf{x}, \mathbf{y})$ is

(i) convex,

(ii) differentiable,

(iii) $\nabla f(\mathbf{x}^*, \mathbf{y}^*) = \mathbf{0}$ for some $(\mathbf{x}^*, \mathbf{y}^*) \in R^{p+q}$.

(iv) If the domain of f is unbounded then

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

Then,

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

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

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