2.098/15.093J: Recitation 10

Xuan Vinh Doan,

November 19, 2004

1 Convexity

Convex sets
A set $S \subset \mathbb{R}^n$ is convex if $\lambda x + (1-\lambda)y \in S$ for all $x, y \in S, \lambda \in [0,1]$.
Property
Intersection of two convex sets is a convex set.

Convex functions
A function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is convex if:

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \qquad \forall x, y \in \mathbb{R}^n, \lambda \in [0,1]$$

Properties
Sum of two convex functions is a convex function.
$f(x)$ is a convex function, then $g(y) := f(Ax+b)$ is a convex function.
The epigraph of a convex function f, epi$f = \{(x,\mu) \in \mathbf{R}^{n+1} | \mu \geq f(x)\}$ is a convex set.
How to recognize a convex function?
Consider a function f that is twice differentiable, that is the gradient vector $\nabla f(x)$ and Hessian matrix$\nabla^2 f(x)$ exists. We have: $[\nabla f(x)]_i = \frac{\partial f(x)}{\partial x_i}$ and $\left[\nabla^2 f(x)\right]_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$ and

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})'(x-\bar{x}) + \frac{1}{2}(x-\bar{x})'\nabla^2 f(\bar{x})(x-\bar{x}) + R(x)\|x-\bar{x}\|^2$$

where $R(x) \to 0$ when $x \to \bar{x}$.
The function $f(x)$ is convex if and only if the Hessian matrix $\nabla^2 f(x)$ is positive semi-definite (PSD) for all x.
For symmetric real matrix M, the followings are equivalent:

1. M is PSD
2. $x'Mx \geq 0$ for all $x \in \mathbf{R}^n$
3. All eigenvalues are nonnegative
4. All principal minors of M have nonnegative determinants

Characterization
The function $f(x)$ is convex if only if $\nabla f(\bar{x})'(x-\bar{x}) \leq f(x) - f(\bar{x})$.
Examples

1. Linear function $f(x) = a'x + b$
2. Quadratic function $f(x) = \frac{1}{2}x'Qx - c'x$, where Q is PSD
3. Log function $f(x) = -\sum_{i=1}^{m} \log(b_i - a_i'x)$, $Ax < b$

Convex optimization
Convex optimization problem is the problem of minimizing a convex function $f(x)$ over a convex set F.
Property
If x^* is a local minimum of a convex optimization problem, then x^* is the global minimum.

2 Unconstrained Optimization

Optimality conditions
Consider the unconstrained problem $\min_{x\in\mathbf{R}^n} f(x)$, where $f(x)$ is twice differentiable, the optimality conditions are:

1. Necessary conditions:
   If x^* is a local minimum then $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is PSD.
2. Sufficient conditions:
   If $\nabla f(\bar{x}) = 0$ and $\exists \epsilon > 0$: $\nabla^2 f(x)$ is PSD for all $x \in B(\bar{x}, \epsilon)$ then $\bar{x}$ is a local optimum.

For a continuously differentiable convex function f, the sufficient and necessary for x^* to be the minimum is $\nabla f(x^*) = 0$.
Example
Find a local minimum of the function $f(x_1, x_2) = -\log(1-x_1-x_2) - \log(x_1) - \log(x_2)$ with $x_1 + x_2 < 1$, $x_1 > 0$, and $x_2 > 0$.
We know that the function $f(x_1, x_2)$ is a convex function; therefore, the only condition we need to check is $\nabla f(x) = 0$.

$$\nabla f(x_1, x_2) = \left(\frac{1}{1-x_1-x_2} - \frac{1}{x_1}, \frac{1}{1-x_1-x_2} - \frac{1}{x_2}\right)'$$

Solving the system $\nabla f(x) = 0$, we will obtain the solution $(x_1, x_2) = \left(\frac{1}{3}, \frac{1}{3}\right)$ and this is the minimum of f.

Gradient methods
Given a differentiable function f in a high dimensional space, the derivative in a direction d at a point x is calculated as follows:

$$\nabla f(x)'d = \lim_{\lambda\to 0} \frac{f(x+\lambda d) - f(x)}{\lambda}$$

Descent direction: if we have $\nabla f(x)'d < 0$ then for sufficiently small $\lambda > 0$, $f(x + \lambda d) < f(x)$.
Generic algorithm elements:

1. Iterative update $x^{k+1} = x^k + \lambda^k d^k$
2. Descent direction $\nabla f(x^k)'d^k < 0$, principal example, $d^k = -D^k \nabla f(x^k)$, where D^k is SPD.
3. Best step length $\lambda^k = \text{argmin}_{\lambda>0} f(x^k + \lambda d^k)$.

Steepest descent method uses the descent direction with $D^k = I$ for all k, or $d^k = -\nabla f(x^k)$. The step length can be calculated using exact or inexact line search or can even follow other rules.