# Problem Set 2 for EE227C (Spring 2018): Convex Optimization and Approximation


Instructor: Moritz Hardt

Email: hardt+ee227c@berkeley.edu

Graduate Instructor: Max Simchowitz

Email: msimchow+ee227c@berkeley.edu


February 25, 2018

## Problem 1: Backtracking Line Search

Let $f : \mathbb{R}^n \to \mathbb{R}$ be an $m$-strongly convex, $M$-smooth (and thus differentiable) function with global minimum $x^*$. Consider the following algorithm:

Initialize with an arbitrary $x_0 \in \mathbb{R}^n$, and fix parameters $\alpha \in (0, 1/2), \beta \in (0, 1)$. Then at each step $t = 1, 2, \dots$, do the following:

(a) Let $g_t = \nabla f(x_t)$.

(b) For $k = \{0, 1, \dots\}$ in sequence, check if the following "sufficient decrease" condition holds:

$$f(x - tg_t) \leqslant f(x) - \alpha\beta^k \cdot \|g_t\|^2 \tag{1}$$

Assuming that this condition holds for some $k$ (you will show this), set $\eta_t = \beta^k$.

(c) Set $x_t \leftarrow x_{t-1} - \eta_t g_t$

**(A)** Show that condition 1 holds for all $t \in (0, 1/M]$.

**(B)** Show that $\eta_t \geqslant \min\{1, \beta/M\}$. Conclude that step (b) of the above algorithm aways terminates.

**(C)** Using part $b$, show that

$$f(x_t - \eta_t g_t) \leqslant f(x) - \alpha \min\{1, \frac{\beta}{M}\} \|\nabla f(x_t)^2\| \tag{2}$$

**(D)** Show that there is a constant $C = C(\alpha, \beta, M, m) < 1$ such

$$f(x_t - \eta_t g_t) - f(x) \leqslant C(\alpha, \beta, M, m) \cdot (f(x_t) - f(x_t)) \tag{3}$$

## Problem 2: Random Descent Directions

Let $f : \mathbb{R}^n \to \mathbb{R}$ be an $m$-strongly convex, $M$-smooth (and thus differentiable) function with global minimum $x^*$. Consider the following algorithm: Initialize with an arbitrary $x_0 \in \mathbb{R}^n$. Then at each step $t = 1, 2, \dots$, do the following:

(a) Choose $g_t \overset{\text{unif}}{\sim} \mathcal{S}^{n-1}$ (equivalently, $g_t$ has the distribution of $\frac{g}{\|g\|}$, where $g \sim \mathcal{N}(0, I_n)$).

(b) Compute a step size $\eta_t := \min_{\eta \geqslant 0} f(x_{t-1} - \eta g_t)$.

(c) set $x_t \leftarrow x_{t-1} - \eta g_t$

**(A)** Prove that the above algorithm is a (non-strict) descent method; that is $f(x_t)$ is non-increasing in $t$. Also prove that unless $x_t = x_*$, $f(x_{t+1}) < f(x_t)$ with probability 1/2.

**(B)** Prove that there exists a numerical constant $C$ such that, if

$$t \geqslant T(\epsilon) := Cn \cdot \frac{M}{m} \log\left(\frac{f(x_0) - f(x^*)}{\epsilon}\right), \tag{4}$$

then $\mathrm{Exp}[f(x_t) - f(x^*)] \leqslant \epsilon$.

**(C)** Ammend the stated algorithm to use line search instead of solving for the exactly-optimal step size. Are the rates qualitatively similar?

## Problem 3: Sh*t about Quadratics (in progress)

**(A)** Let $n = 1000$