Solutions to Assignment 5 of CPSC 368/516 (Spring'23)

February 28, 2023

1 Problem 1

Problem 1.1. Gradient descent for strongly convex functions. *In this problem we analyze a gradient descent algorithm for minimizing a twice-differentiable convex function $f : \mathbb{R}^n \to \mathbb{R}$, which satisfies for every $x \in \mathbb{R}^n$, one has $mI \preceq \nabla^2 f(x) \preceq MI$ for some $0 < m \leq M$.*

The algorithm starts with some $x_0 \in \mathbb{R}^n$ and at every step $t = 0, 1, 2, \ldots$ it chooses the next point

$$x_{t+1} \coloneqq x_t - \alpha_t \nabla f(x_t),$$

where α_t is chosen to minimize the value $f(x_t - \alpha \nabla f(x_t))$ over all $\alpha \in \mathbb{R}$ while fixing x_t. Let $y^\star \coloneqq \min\{f(x) : x \in \mathbb{R}^n\}$.

1. *Prove that*

$$\forall x, y \in \mathbb{R}^n, \quad \frac{m}{2} \|y - x\|^2 \leq f(y) - f(x) + \langle \nabla f(x), x - y \rangle \leq \frac{M}{2} \|y - x\|^2 .$$

2. *Prove that*

$$\forall x \in \mathbb{R}^n, \quad f(x) - \frac{1}{2m} \|\nabla f(x)\|^2 \leq y^\star \leq f(x) - \frac{1}{2M} \|\nabla f(x)\|^2 .$$

3. *Prove that for every $t = 0, 1, 2 \ldots$*

$$f(x_{t+1}) \leq f(x_t) - \frac{1}{2M} \|\nabla f(x_t)\|^2 .$$

4. *Prove that for every $t = 0, 1, 2 \ldots$*

$$f(x_t) - y^\star \leq \left(1 - \frac{m}{M}\right)^t (f(x_0) - y^\star) .$$

   What is the number of iterations t required to reach $f(x_t) - y^\star \leq \varepsilon$?

5. *Consider a linear system $Ax = b$, where $b \in \mathbb{R}^n$ is a vector and $A \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix such that $\frac{L_n(A)}{L_1(A)} \leq \kappa$ (where $L_1(A)$ and $L_n(A)$ are the smallest and the largest eigenvalues of A respectively). Use the above framework to design an algorithm for approximately solving the system $Ax = b$ with logarithmic dependency on the error $\varepsilon > 0$ and polynomial dependency on κ. What is the running time?*

1.1 Part 1

Lemma 2.6 from [1], implies that for any $x, y \in \mathbb{R}^n$

$$
\begin{aligned}
f(y) &= f(x) + \int_0^1 \langle \nabla f(x + t(y-x)), y - x \rangle \, dt \\
&= f(x) + \int_0^1 \left\langle \nabla f(x) + \left(\int_0^t t(y-x)^\top \nabla^2 f(x + rt(y-x)) dr \right), y - x \right\rangle dt \\
&\qquad \text{(Using Lemma 2.6 to expand } \nabla f(x + t(y-x))) \\
&= f(x) + \int_0^1 \langle \nabla f(x), y - x \rangle \, dt + \int_0^1 \int_0^1 t(y-x)^\top \nabla^2 f(x + rt(y-x))(y-x) dr dt \\
&= f(x) + \langle \nabla f(x), y - x \rangle + \int_0^1 \int_0^1 t(y-x)^\top \nabla^2 f(x + rt(y-x))(y-x) dr dt. \qquad (1)
\end{aligned}
$$

Moreover, since $mI \preceq \nabla^2 f(x) \preceq MI$, for any vector $z \in \mathbb{R}^n$,

$$
m \|z\|_2^2 \le z^\top \nabla^2 f(x) z \le M \|z\|_2^2 .
$$

Combining these inequalities with Equation (1), implies that

$$
\begin{aligned}
f(y) &\le f(x) + \langle \nabla f(x), y - x \rangle + \int_0^1 \int_0^1 tM \|y - x\|_2^2 \, dr dt \\
&\le f(x) + \langle \nabla f(x), y - x \rangle + \frac{M}{2} \|y - x\|_2^2 , \qquad (2) \\
f(y) &\ge f(x) + \langle \nabla f(x), y - x \rangle + \int_0^1 \int_0^1 tm \|y - x\|_2^2 \, dr dt \\
&\ge f(x) + \langle \nabla f(x), y - x \rangle + \frac{m}{2} \|y - x\|_2^2 . \qquad (3)
\end{aligned}
$$

The required results follow by rearranging the above inequalities.

1.2 Part 2

By the definition of $y^\star$, for all $y \in \mathbb{R}^n$, $y^\star \le f(y)$. Since $y^\star \le f(y)$ and Equation (2) holds, for all $x, y \in \mathbb{R}^n$

$$
y^\star \le f(y) \le f(x) + \langle \nabla f(x), x - y \rangle + \frac{M}{2} \|y - x\|_2^2 .
$$

Further, as the above inequality holds for all $y \in \mathbb{R}^n$, it implies that

$$
\begin{aligned}
y^\star &\le \inf_{y \in \mathbb{R}^n} f(x) + \langle \nabla f(x), x - y \rangle + \frac{M}{2} \|y - x\|_2^2 \\
&= \inf_{z \in \mathbb{R}^n : \, \|z\| = 1} \; \inf_{t \in \mathbb{R}} \; f(x) + \frac{M}{2} t^2 + t \nabla f(x)^\top z \qquad \text{(Substituting } y = x + tz \text{ for some } t \in \mathbb{R} \text{ and } z \in \mathbb{R}^n) \\
&= \inf_{z \in \mathbb{R}^n : \, \|z\| = 1} \; f(x) - \frac{1}{2M} \langle \nabla f(x), z \rangle^2 \\
&= f(x) - \frac{1}{2M} \|\nabla f(x)\|_2^2 .
\end{aligned}
$$

Minimizing the RHS in Equation (3), implies that for all $x, y \in \mathbb{R}^n$

$$
\begin{aligned}
f(y) &\ge \inf_{w \in \mathbb{R}^n} f(x) + \langle \nabla f(x), w - x \rangle + \frac{m}{2} \|w - x\|_2^2 \\
&= \inf_{z \in \mathbb{R}^n : \, \|z\|_2 = 1} \inf_{t \in \mathbb{R}} f(x) + t \langle \nabla f(x), z \rangle + \frac{m}{2} t^2 \qquad \text{(Substituting } y = x + tz \text{ for some } t \in \mathbb{R} \text{ and } z \in \mathbb{R}^n) \\
&= \inf_{z \in \mathbb{R}^n : \, \|z\|_2 = 1} f(x) - \frac{2}{m} \langle \nabla f(x), z \rangle^2 \\
&= f(x) - \frac{2}{m} \|\nabla f(x)\|_2^2 .
\end{aligned}
$$

1.3 Part 3

Let $y = x_t - \alpha \nabla f(x_t)$ for some $\alpha \in \mathbb{R}$. Substituting $x = x_t$ in Equation (2), implies that:

$$\begin{aligned} f(y) &\leq f(x_t) + \langle \nabla f(x_t), y - x_t \rangle + \frac{M}{2} \|y - x_t\|_2^2 \\ &\leq f(x_t) + \langle \nabla f(x_t), -\alpha \nabla f(x_t) \rangle + \frac{M}{2} \|\alpha \nabla f(x_t)\|_2^2 \\ &\leq f(x_t) - \alpha \|\nabla f(x_t)\|_2^2 + \frac{M\alpha^2}{2} \|\nabla f(x_t)\|_2^2 . \end{aligned}$$

Since x_{t+1} is defined such that $f(x_{t+1}) = \inf_{\alpha \in \mathbb{R}} f(x_t - \alpha_t \nabla f(x_t))$, for any $\alpha \in \mathbb{R}$

$$f(x_{t+1}) \leq f(y) \leq f(x_t) - \alpha \|\nabla f(x_t)\|_2^2 + \frac{M\alpha^2}{2} \|\nabla f(x_t)\|_2^2 .$$

In particular, minimizing the RHS over α implies that

$$\begin{aligned} f(x_{t+1}) &\leq \inf_{\alpha \in \mathbb{R}} f(x_t) - \alpha \|\nabla f(x_t)\|_2^2 + \frac{M\alpha^2}{2} \|\nabla f(x_t)\|_2^2 \\ &\leq f(x_t) - \frac{1}{2M} \|\nabla f(x_t)\|_2^2 . \end{aligned} \tag{4}$$

1.4 Part 4

Subtracting $y^\star$ from both sides of Equation (4), implies that for all $t = 0, 1, \ldots$

$$\begin{aligned} f(x_{t+1}) - y^\star &\leq f(x_t) - y^\star - \frac{1}{2M} \|\nabla f(x_t)\|_2^2 \\ &\leq f(x_t) - y^\star - \frac{m}{M} (y^\star - f(x_t)) \\ &= \left(1 - \frac{m}{M}\right) \cdot (f(x_t) - y^\star) . \end{aligned}$$

Chaining the above inequality for all $t = 0, 1, \ldots, T-1$, implies that

$$f(x_T) - y^\star \leq \left(1 - \frac{m}{M}\right)^T \cdot (f(x_0) - y^\star) .$$

Recall that we require that $f(x_T) - y^\star \leq \varepsilon$, or equivalently that

$$f(x_T) - y^\star \leq \left(1 - \frac{m}{M}\right)^{T-1} (f(x_0) - y^\star) \leq \varepsilon \quad \iff \quad T \geq \log\left(\frac{\varepsilon}{f(x_0) - y^\star}\right) \cdot \left(\log\left(1 - \frac{m}{M}\right)\right)^{-1} + 1.$$

Thus, $\Theta\left(\frac{M}{m} \cdot \log\left(\frac{f(x_0) - y^\star}{\varepsilon}\right)\right)$ iterations suffice.

1.5 Part 5

Notice that the solution to $Ax = b$ is also the optimal solution to the optimization problem

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \left\langle A, xx^\top \right\rangle - b^\top x.$$

This holds because $\frac{1}{2} \left\langle A, xx^\top \right\rangle - b^\top x$ is convex (as A is PD) and its gradient is zero iff $Ax = b$. Define

$$f(x) \coloneqq x^\top \frac{A}{2} x - b^\top x.$$

One can observe that $\nabla^2 f(x) = A \succ 0$. Hence, it follows that $\nabla^2 f(x) = A \preceq \lambda_n(A)$ and that $\nabla^2 f(x) = A \succeq \lambda_1(A)$. We initialize the algorithm developed in the previous parts as follows:

- $m = \lambda_1(A)$ and $M = \lambda_n(A)$
- The gradient of f is $\nabla f(x) = -b + Ax$
- $f(x_t - \alpha \nabla f(x_t))$ is a convex function in α. Minimizing it over $\alpha \in \mathbb{R}$ and using the definition of α_t, we get that α_t (as a function of x_t) is

$$\alpha_t := \frac{(b - Ax_t)^\top (b - Ax_t)}{(b - Ax_t)^\top A(b - Ax_t)}$$

- We can initialize the starting point as $x_0 = 0$. This implies that $f(x_0) - y^\star \leq \frac{1}{2} bA^{-1}b$.

Thus,

$$\begin{aligned} T &= \log\left(\frac{2\varepsilon}{b^\top A^{-1} b}\right) \cdot \left(\log\left(1 - \frac{1}{\kappa}\right)\right)^{-1} + 1 \\ &= O\left(\kappa \log\left(\frac{2\varepsilon}{b^\top A^{-1} b}\right)\right). && \text{(Using that } \log\left(1 - \tfrac{1}{\kappa}\right) = -O\left(\tfrac{1}{\kappa}\right)\text{)} \end{aligned}$$

In every step, it takes $O(n^2)$ arithmetic operations in total to compute $m, M, \nabla f(x_t), \alpha_t$ and x_{t+1}, and hence, the amount of arithmetic steps are

$$O\left(n^2 \kappa \log \frac{2\varepsilon}{b^\top A^{-1} b}\right).$$

2 Problem 2

Problem 2.1. *Let $G = (V, E)$ be an undirected graph with n vertices and m edges. Let $B \in \mathbb{R}^{n \times m}$ be the vertex-edge incidence matrix of G. Assume that G is connected and let $\Pi := B^\top (BB^\top)^+ B$. Prove that, given a vector $g \in \mathbb{R}^m$, if we let x_g denote the projection of g on the subspace $K := \{x \in \mathbb{R}^m : Bx = 0\}$, then it holds that*

$$x_g = g - \Pi g.$$

Recall that the projection of a point $h \in \mathbb{R}^n$ on a closed convex nonempty set $K \subseteq \mathbb{R}^n$ is defined as the unique point $x_h \in K$ that minimizes the Euclidean distance to h:

$$x_h := \operatorname*{argmin}_{y \in K} \|y - h\|_2. \tag{5}$$

First, observe that for any $h \in \mathbb{R}^m$ x_h is feasible for the instance Equation (5): To see this observe that

$$\begin{aligned} Bx_h &= Bh - B\Pi h \\ &= Bh - BB^\top \left(BB^\top\right)^+ Bh && \text{(Using that } \Pi = B^\top \left(BB^\top\right)^+ B\text{)} \\ &= Bh - Bh \\ &= 0. \end{aligned}$$

Thus, $x_h \in K$ and, hence, feasible for Equation (5).

Next, observe K is a linear space, and hence, the projection of h on K is the unique point $x_h \in K$ such that $h - x_h$ is orthogonal to all points $z \in K$. Further, observe that K is the null space of B, and hence, K is also the null space of $B^\top B$ (exercise: prove this). Furthermore, since K is the null space of $B^\top B$, it is the span of the eigenvectors corresponding to the 0 eigenvalue of $B^\top B$. Since $B^\top B$ is the Laplacian of a connected graph, the all-ones vector is the unique eigenvector corresponding to the eigenvalue 0 (here, connectedness is required to guarantee uniqueness). Since $z \in K$ must be parallel to the all-ones vector, to check if $h - x_h$ is orthogonal to z, it suffices to show that $h - x_h$ is orthogonal to $1 \in \mathbb{R}^m$. This follows because

$$\begin{aligned}
\langle 1, h - x_h \rangle &= \langle 1, \Pi h \rangle && \text{(Using that } x_h = h - \Pi h\text{)} \\
&= 1^\top B^\top \left(BB^\top\right)^+ Bh && \text{(Using that } \Pi = B^\top \left(BB^\top\right)^+ B\text{)} \\
&= 0^\top \left(BB^\top\right)^+ Bh && \text{(Using that } B1 = 0\text{)} \\
&= 0.
\end{aligned}$$

References

[1] Nisheeth K. Vishnoi. *Algorithms for Convex Optimization.* Cambridge University Press, 2021.