凸优化第 11 周作业

1 预习

下节课讲授算法讲义剩余内容，没有小测，但请提前预习。

2 作业题

1. (Non-convex gradient descent) Consider minimizing a differentiable function f with $\operatorname{dom} f = \mathbb{R}^n$, whose gradient is L-Lipschitz continuous for a constant $L > 0$, i.e.,

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2, , \quad \forall x, y \in \mathbb{R}^n.$$

Consider the gradient descent method starting from x_0, with the updates

$$x_{k+1} = x_k - t\nabla f(x_k),$$

where $t \leq 1/L$. Notice that we **do not** assume that f is convex. Prove the following statements:

(a)
$$f(y) \leq f(x) + \nabla f(x)^\top (y - x) + \frac{L}{2}\|y - x\|_2^2, \quad \forall x, y \in \mathbb{R}^n.$$

(b)
$$f\left(x_{k+1}\right) \leq f\left(x_k\right) - \left(1 - \frac{Lt}{2}\right) t \left\|\nabla f\left(x_k\right)\right\|_2^2.$$

(c)
$$\left\|\nabla f\left(x_k\right)\right\|_2^2 \leq \frac{2}{t}\left(f\left(x_k\right) - f\left(x_{k+1}\right)\right).$$

(d)
$$\sum_{i=0}^{k} \left\|\nabla f\left(x_i\right)\right\|_2^2 \leq \frac{2}{t}\left(f\left(x_0\right) - f\left(x^*\right)\right).$$

(e)
$$\min_{i=0,\ldots,k} \left\|\nabla f\left(x_i\right)\right\|_2 \leq \sqrt{\frac{2}{t(k+1)}\left(f\left(x_0\right) - f\left(x^*\right)\right)}$$

As a result, we have proved that gradient descent reaches an ϵ-substationary point (i.e., $\|\nabla f(x)\|_2 \leq \epsilon$) in $O(1/\epsilon^2)$ iterations.

2. (Ill-conditioned linear equation) Consider solving the linear equation $Ax = b$, where A is an ill-conditioned matrix. This is equivalent to solving the optimization problem

$$\text{minimize } f(x) = \frac{1}{2}\|Ax - b\|_2^2.$$

Consider the proximal operator

$$\text{Prox}_{\alpha f}(v) = \arg\min_x \alpha f(x) + \frac{1}{2}\|x - v\|^2.$$

   (a) Derive the exact form of the iteration $x_{k+1} = \text{Prox}_{\alpha f}(x_k)$.

   (b) Implement the above iteration and solve the problem using the data `1A.csv` and `1b.csv` provided in the attachment. Choose α by yourself to make the iteration numerically stable and efficient. Stop when $f(x_k) \le 0.02$. Plot i) $\log(\|x_k - x^*\|_2)$ vs. $\log(k)$; ii) $\log(f(x_k))$ vs. $\log(k)$. Assume that the total number of iterations is T, then you may treat last iterate x_T as the true optimal solution x^*, and make the plot for $k = 0, 1, \ldots, T-1$.

3. (LASSO) Consider the LASSO problem:

$$\text{minimize } h(x) = \frac{1}{2}\|Ax - b\|_2^2 + \|x\|_1,$$

where A, b are provided in the attachments `2A.csv` and `2b.csv`.

   (a) Implement the sub-gradient method with the following iteration:

$$x_{k+1} = x_k - \alpha_k g_k,\ g_k \in \partial h(x),$$

   where $\alpha_k = c \cdot k^{-\beta}$, with $c = 0.01, \beta = 0.5$. Use the starting point x_0 provided in `2x0.csv`. Stop when $\|x_k - x_{k-1}\|_2^2 < 10^{-8}$. Plot i) $\log(\|x_{k+1} - x_k\|_2)$ vs. $\log(k)$; ii) $\log(\|x_k - x^*\|_2)$ vs. $\log(k)$; iii) $\log(h(x_k))$ vs. $\log(k)$. Assume that the total number of iterations is T, then you may treat last iterate x_T as the true optimal solution x^*, and make the plot for $k = 0, 1, \ldots, T-1$.

   (b) When A is full column rank, the function $f(x) = \frac{1}{2}\|Ax - b\|_2^2$ is strongly convex, i.e.,

$$f(y) \ge f(x) + \nabla f(x)^\top (y - x) + \frac{m}{2}\|y - x\|^2.$$

   With A, b provided in `2A.csv` and `2b.csv`, find the maximal m such that the above inequality holds. Repeat problem (a) with step size $\alpha_k = 1/(mk)$.

3 作业说明

1. 编程作业部分需要撰写报告，包含推导步骤和运行结果。报告提交电子版，和代码一起打包提交至网络学堂。
2. 编程语言不限，但不能使用现成的优化器。
3. **请大家务必在截止时间之前提交作业**，迟交一周以内的作业得分是卷面分的 50%，迟交超过一周的作业不得分。
4. 每次作业的满分是 25 分。