THE UNIVERSITY OF TEXAS
AT AUSTIN

EE381V LARGE SCALE OPTIMIZATION

Problem Set 3

Edited by LaTeX

Department of Computer Science

STUDENT
Jimmy Lin
xl5224

COURSE COORDINATOR
Sujay Sanghavi

UNIQUE NUMBER
17350

RELEASE DATE
October 08, 2014

DUE DATE
October 16, 2014

TIME SPENT
5 hours

October 12, 2014

Table of Contents

List of Figures

Part I
Written Problems

1 Gradient descent with diminising step size

2 Gradient descent and non-convexity

3 Jacobi Method

Prove that, for a convex continuously differentiable f, and a step size $\alpha = 1/n$ where n is the number of coordinates, the next iterates of the Jacobi method produces a lower function value than x, provided x does not already minimize the function.

Proof. Let $x_i^* = (x_1, ..., x_{i-1}, \bar{x}_i, x_{i+1}, ..., x_n)$. Then we attempt to represent x^+ as convex combination of n points $(x_i^*,\ i = 1, ..., n)$.

$$
\begin{aligned}
x^+ &= x + \alpha(\bar{x} - x) && (1)\\
&= x + \frac{1}{n}(\bar{x} - x) && (2)\\
&= (1 - \frac{1}{n})x + \frac{1}{n}\bar{x} && (3)\\
&= (\frac{n-1}{n})x + \frac{1}{n}\bar{x} && (4)\\
&= \left(\frac{n-1}{n}x_1 + \frac{1}{n}\bar{x}_1, \frac{n-1}{n}x_2 + \frac{1}{n}\bar{x}_2, ..., \frac{n-1}{n}x_n + \frac{1}{n}\bar{x}_n\right) && (5)\\
&= \sum_{i=1}^{n} \frac{1}{n} x_i^* && (6)
\end{aligned}
$$

which presents us the convex combination. Then

$$
\begin{aligned}
f(x^+) &\le f(\sum_{i=1}^{n} \frac{1}{n} x_i^*) && && (7)\\
&\le \sum_{i=1}^{n} \frac{1}{n} f(x_i^*) && f \text{ is convex} && (8)\\
&\le \sum_{i=1}^{n} \frac{1}{n} f(x) && \forall i, f(x_i^*) \le f(x) && (9)\\
&= f(x) && && (10)
\end{aligned}
$$

where equality holds when $\forall i,\ x_i = \bar{x}_i$ $(x^+ = x)$, that is, point x does already minimize the function. And in other cases, the next iterate of the Jacobi method produces a lower function value than x. $\square$

4 Step size in Newton

(a) Values of t obtain global convergence

(b) Reason that convergence is not quadratic

5 Composite functions

(a) Run gradient descent on f and g

Show that the entire sequence of iterates will then be the same.

Proof. The gradient descent direction for $f(x)$ and $g(x)$ are respectively

$$\Delta x_{f(x)} = \nabla_x f(x) \tag{11}$$
$$\Delta x_{g(x)} = \nabla_x g(x) = \nabla_x \phi(f(x)) = \nabla_{f(x)} \phi(f(x)) \nabla_x f(x) \tag{12}$$

Apply direction to update rule, we have

$$x^+_{(f)} = x + t^*_{(f)} \nabla_x f(x) \tag{13}$$
$$x^+_{(g)} = x + t^*_{(g)} \nabla_{f(x)} \phi(f(x)) \nabla_x f(x) \tag{14}$$

where the optimal step size for $f(x)$ is

$$t^*_{(f)} = \arg\min_t f(x + t\nabla_x f(x)) \tag{15}$$

and the optimal step size for $g(x)$ is

$$t^*_{(g)} = \arg\min_t g(x + t\nabla_{f(x)} \phi(f(x)) \nabla_x f(x)) \tag{16}$$
$$= \arg\min_t \phi\big(f(x + t\nabla_{f(x)} \phi(f(x)) \nabla_x f(x))\big) \tag{17}$$
$$= \arg\min_t f(x + t\nabla_{f(x)} \phi(f(x)) \nabla_x f(x)) \qquad \phi(\cdot) \text{ is increasing function} \tag{18}$$

Now we observe that both step size $t^*_{(f)}$ and $t^*_{(g)}$ can be seen as exact line search of $f(\cdot)$ on point x towards direction $\nabla_x f(x)$. (Note that $\phi(f(x))$ is a scalar.) But two step size has different scale due to the existence of $\phi(f(x))$ on (18). Hence, we have

$$t^*_{(f)} = t^*_{(g)} \nabla_{f(x)} \phi(f(x)) \tag{19}$$

Thus,

$$x^+_{(f)} = x + t^*_{(f)} \nabla_x f(x) \tag{20}$$
$$= x + t^*_{(g)} \nabla_{f(x)} \phi(f(x)) \nabla_x f(x) \tag{21}$$
$$= x^+_{(g)} \tag{22}$$

which indicates that one iteration of gradient descent method on $f(\cdot)$ and $g(\cdot)$ starting with the same point x (arbitrarily) will go to the same point ($x^+_{(f)} = x^+_{(g)}$). Recursively apply this derivation, it is proved that the entire sequence of iterates will then be the same. □

(b) Run Newton method on f and g:

Show that

Proof. □