Optimization Hw3

Toby Harvey

November 20, 2019

Problem 1
Since $f(x)$ is differential everywhere but 0, we have that the subgradient is:

$$\begin{aligned}\nabla f(x) = \frac{\partial}{\partial x_i} \|x\|_2 = \frac{\partial}{\partial x_i}(x_1^2 + x_2^2 + ... + x_n^2)^{\frac{1}{2}}, \qquad \forall x_i \\ = \frac{1}{2(x_1^2 + x_2^2 + ... + x_n^2)^{\frac{1}{2}}}(2x_i) \qquad \forall x_i \\ = \frac{x}{\|x\|_2}\end{aligned}$$

If $x = 0$, $f(x)$ is not differentiable, but we can still show that the subgradient is any y st. $\|y\|_2 \leq 1$
Let y be any y such that $\|y\| < 1$, then we have:

$y \cdot (x - 0) = y \cdot (x) \leq \|y\| \|x\| \leq \|x\| = \|x\| - \|0\|$

Where the first inequality is cauchy-schwarz, and the secondy is by hypothesis that $\|y\| < 1$. Then have found what we wanted since we started with abritrary $\|y\| \leq 1$ and showed it was a subgradient.

Problem 2

If $f_1(\tilde{x}) > f_2(\tilde{x})$ at $\tilde{x}$, then $f(\tilde{x}) = f_1(\tilde{x})$, since $f_1(\tilde{x})$ is differentable, we have that the subgraident is $\nabla f_1(x)$. If $f_2(\tilde{x}) > f_1(\tilde{x})$ at $\tilde{x}$, then $f(\tilde{x}) = f_2(\tilde{x})$, since $f_2(\tilde{x})$ is differentable, we have that the subgraident is $\nabla f_2(x)$.
If $f_1(\tilde{x}) = f_2(\tilde{x})$, then we have both:

$$\nabla f_1(\tilde{x}) \cdot (x - \tilde{x}) \leq f_1(x) - f_1(\tilde{x}), \qquad \forall x \tag{1}$$
$$\nabla f_2(\tilde{x}) \cdot (x - \tilde{x}) \leq f_2(x) - f_2(\tilde{x}), \qquad \forall x \tag{2}$$

Multiplying (1) through by t, and and (2) through by $(1-t)$, where $0 \leq t \leq 1$, since $f(x)$ is convex, and adding these two equations gives:

$$t\nabla f_1(\tilde{x}) \cdot (x - \tilde{x})) + (1-t)\nabla f_2(\tilde{x}) \cdot (x - \tilde{x}) \leq t(f_1(x) - f_1(\tilde{x})) + (t-1)(f_2(x) - f_2(\tilde{x}))$$
$$\implies (t\nabla f_1(\tilde{x}) + (1-t)\nabla f_2(\tilde{x})) \cdot (x - \tilde{x}) \leq (tf_1(x) + (1-t)f_2(x)) - (tf_1(\tilde{x}) - (1-t)f_2(\tilde{x}))$$

Since we have that $f_1(\tilde{x}) = f_2(\tilde{x})$ the second term on the RHS is not a line but a point and therefore equals $f(\tilde{x})$ (plugging in $f(\tilde{x})$, this works algebrically too). In the first term on the RHS, we know that both $f_1(x)$ and $f_2(x)$ are less then or equal to $f(x)$ therefore the entire first term is less then or equal to $f(x)$.Taking these facts into account we have:

$$(t\nabla f_1(\tilde{x}) + (1-t)\nabla f_2(\tilde{x})) \cdot (x - \tilde{x}) \leq f(x) - f(\tilde{x})$$

Which says that the line segment between the gradients is the sub differential.

Problem 3

(1) Differentating $f(x)$ we get $-\frac{1}{2\sqrt{x}}$ which doesn't exist at 0, this is ok. So by way of contradition let $f(x)$ be subdifferentiable at $x = 0$, then we have:

$$g \cdot (x - 0) \leq -x^{\frac{1}{2}} - 0 \qquad \forall x$$
$$\implies g \leq -\frac{x^{\frac{1}{2}}}{x} \qquad \forall x$$

Using L'Hôpital's rule on this we get $-\frac{1}{2\sqrt{x}}$ which diverges as $x \to 0$. Therefore $g \to \infty$, which is a contradiction. Also we can note that the leftside limit does exist, and we need both limits to exist for the subdifferential to exist.

(2) By way of contradiction, Assume that $f(x)$ is subdifferentiable at $x = 0$. Then there exists least a g such that:

$$g \cdot (x - 0) \leq f(x) - 1 \qquad \forall x$$

Letting $x > 0$ we have:

$$g \leq -\frac{1}{x} \qquad \forall x$$

So with arbitrarly small x, $g \to \infty$, which is a contradiction.

Problem 4
From the text on the convergence of the subgradient method we have:

$$\begin{aligned}
\left\|x^+ - x^*\right\|_2^2 &= \|x - \alpha g - x^*\|_2^2 \\
&= \|x - x^*\|_2^2 - 2\alpha g(x - x^*) + a^2 \|g\|_2^2 \\
&\leq \|x - x^*\|_2^2 - 2\alpha(f(x) - fx^*)) + a^2 \|g\|_2^2
\end{aligned}$$

Using the given stepsize to show the convergence we have:

$$\begin{aligned}
\left\|x^+ - x^*\right\|_2^2 &\leq \|x - x^*\|_2^2 - 2\alpha(f(x) - fx^*)) + a^2 \|g\|_2^2 \\
&= \|x - x^*\|_2^2 - 2\left(\frac{2(f(x) - f(x^*))}{\|g\|_2^2}\right)(f(x) - fx^*)) + \left(\frac{2(f(x) - f(x^*))}{\|g\|_2^2}\right)^2 \|g\|_2^2 \\
&= \|x - x^*\|_2^2 - 4\left(\frac{(f(x) - f(x^*))}{\|g\|_2^2}\right)(f(x) - fx^*)) + 4\left(\frac{(f(x) - f(x^*))^2}{\|g\|_2^2}\right) \\
&= \|x - x^*\|_2^2
\end{aligned}$$

Taking the square root shows the result.

Problem 5
If we decompose over S, since the objective function is convex and each subproblem we are minimizing is convex, we can minimize the orginal objective (i.e. even though we are dual decomposing we should still be able to find x^*, since each sub problem is convex). The lagrangian is:

$$\begin{aligned}
L(x_{s,u}, \alpha_s, \beta_u, \gamma_{s,u}) &= \sum_{s,u} a_{s,u} x_{s,u} + \sum_{s,u} b_{s,u}((x_{s,u} + 1)\ln(x_{s,u} + 1) - x_{s,u}) \\
&+ \sum_s \alpha_s(c_s - \sum_u x_{s,u}) + \sum_u \beta_u(c_u - \sum_s x_{s,u}) - \sum_{s,u} \gamma_{s,u} x_{s,u}
\end{aligned}$$

If we decompose over S for each $s \in S$ we get a subproblem which is constrained by the seperable constraints and minimizes over that single s:

$$\begin{aligned} \min \qquad & \sum_u a_{s,u} x_{s,u} + \sum_u b_{s,u}((x_{s,u}+1)\ln(x_{s,u}+1) - x_{s,u}) - \sum_u \gamma_{s,u} x_{s,u} \\ \text{st.} \qquad & c_s - \sum_u x_{s,u} \leq 0 \\ & c_u - x_{s,u} \leq 0 \qquad \forall u \end{aligned}$$

We can use $\gamma_{s,u}$ to couple the subproblems together. Using gradient decent we have an update rule for $\gamma_{s,u}$ after all sub problems have been solved each round. The gradient with respect to $\gamma_{s,u}$ is $-x_{s,u}$. Since we are maximizing we need $x_{s,u}$. The full algorithim looks like:

1. choose arbitrary $\gamma_{s,u}$
2. Solve each of the subproblems for optimal fixed s
3. update $\gamma_{s,u}$ with gradient decent: $\gamma_{s,u}^{k+1} = \gamma_{s,u}^{k} + t x_{s,u}$, and repeat