HW 6

Evan Ott
UT EID: eao466

October 16, 2016

1 Proximal operators

(A)

$$\begin{aligned}
f(x) \approx \hat{f}(x; x_0) &= f(x_0) + (x - x_0)^{(t)} \nabla f(x_0) \\
\text{prox}_\gamma \hat{f}(x) &= \arg\min_z \left[\hat{f}(z) + \frac{1}{2\gamma} \|z - x\|_2^2 \right] \\
&= \arg\min_z \left[f(x_0) + (z - x_0)^{(t)} \nabla f(x_0) + \frac{1}{2\gamma} \|z - x\|_2^2 \right] \\
0 &= \frac{\partial}{\partial z} \left[f(x_0) + (z - x_0)^{(t)} \nabla f(x_0) + \frac{1}{2\gamma} \|z - x\|_2^2 \right] \\
&= 0 + \nabla f(x_0) + \frac{1}{2\gamma}(2z^* - 2x) = \nabla f(x_0) + \frac{1}{\gamma}(z^* - x) \\
\text{prox}_\gamma \hat{f}(x) = z^* &= x - \gamma \nabla f(x_0)
\end{aligned}$$

which is indeed the gradient-descent step for $f(x)$ of size γ starting at x_0.

(B)

$$\begin{aligned}
l(x) &= \frac{1}{2} x^\top P x - q^\top x + r \\
\text{prox}_{1/\gamma} l(x) &= \arg\min_z \arg\min_z \left[\hat{l}(z) + \frac{\gamma}{2} \|z - x\|_2^2 \right] \\
&= \arg\min_z \left[\frac{1}{2} z^\top P z - q^\top z + r + \frac{\gamma}{2} \|z - x\|_2^2 \right] \\
0 &= \frac{\partial}{\partial z} \left[\frac{1}{2} z^\top P z - q^\top z + r + \frac{\gamma}{2} \|z - x\|_2^2 \right] \\
&= P z^* - q + \gamma(z^* - x) \\
\gamma x + q &= (P + \gamma I) z^* \\
\text{prox}_{1/\gamma} l(x) = z^* &= (P + \gamma I)^{-1} (\gamma x + q)
\end{aligned}$$

assuming $(P + \gamma I)^{-1}$ exists.

If we have $y|x \sim N(Ax, \Omega^{-1})$ with y having n rows, then

$$\begin{aligned}
L(y|x) &= \frac{1}{\sqrt{2\pi}^n} |\Omega|^{-1/2} \exp\left[-\frac{1}{2} (y - Ax)^\top \Omega^{-1} (y - Ax) \right] \\
n(y|x) = -\log L(y|x) &= \frac{1}{2} \log |\Omega| + \frac{n}{2} \log(2\pi) + \frac{1}{2} (y - Ax)^\top \Omega^{-1} (y - Ax) \\
&= \frac{1}{2} y^\top \Omega^{-1} y - (Ax)^\top \Omega^{-1} y + \frac{1}{2} (Ax)^\top \Omega^{-1} Ax + \frac{1}{2} \log |\Omega| + \frac{n}{2} \log(2\pi)
\end{aligned}$$

So $P = \Omega^{-1}$, $q = \Omega^{-1} Ax$ (because $\Omega = \Omega^\top$ since it is a covariance matrix), and $r = \frac{1}{2}(Ax)^\top \Omega^{-1} Ax + \frac{1}{2} \log |\Omega| + \frac{n}{2} \log(2\pi)$.

(C)

$$
\begin{aligned}
\phi(x) &= \tau\|x\|_1 \\
\operatorname{prox}_\gamma \phi(x) &= \arg\min_z \left[\phi(z) + \frac{1}{2\gamma}\|z-x\|_2^2\right] \\
&= \arg\min_z \left[\tau\|z\|_1 + \frac{1}{2\gamma}\|z-x\|_2^2\right] \\
&= \arg\min_z \left[\tau\sum_{i=1}^{n}(|z_i|) + \frac{1}{2\gamma}\sum_{i=1}^{n}\left((z_i-x_i)^2\right)\right] \\
&= \arg\min_z \left[\sum_{i=1}^{n}\frac{1}{2\gamma}(z_i-x_i)^2 + \tau|z_i|\right] \\
&= \arg\min_z \left[\sum_{i=1}^{n}\frac{1}{2}(z_i-x_i)^2 + \tau\gamma|z_i|\right]
\end{aligned}
$$

The term being minimized for each component z_i is exactly $S_{\tau\gamma}(x_i)$ from the notation last week, and there are no interaction terms between the z_i and z_j for $i \neq j$, so

$$
\left(\operatorname{prox}_\gamma \phi(x)\right)_i = S_{\tau\gamma}(x_i)
$$