Problem Set 1

Faisal Jayousi

Problem 1

1. Given $y \in \mathbb{R}^n$ and a linear operator $A : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, compute the **gradient** of the n-dimensional function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ defined as
$$f(x) = \frac{1}{2}\|Ax - y\|_2^2. \tag{1}$$
2. Compute the gradient of $g\colon x \mapsto \|x\|_2^2$.
3. Let $\lambda > 0$. Deduce $\nabla(f + \lambda g)(x)$

Solution

1. **Method 1** Let $x, v \in \mathrm{dom}(f) = \mathbb{R}^n$. The gradient of f at x if it exists, $\nabla f(x)$, is such that
$$\nabla_v f(x) = \langle \nabla f(x), v \rangle$$
where $\nabla_v f(x)$ is the directional derivative of f at x along the **unit** vector v. It follows that
$$\begin{aligned}
\nabla_v f(x) &= \lim_{h \to 0} \frac{f(x + hv) - f(x)}{h} \\
&= \lim_{h \to 0} \frac{||A(x + hv) - y||^2 - ||Ax - y||^2}{2h} \\
&= \lim_{h \to 0} \frac{||Ax - y||^2 + 2h\langle Ax - y, Av \rangle + h^2||Av||^2 - ||Ax - y||^2}{2h} \\
&= \lim_{h \to 0} (Ax - y)^T Av + \frac{h}{2}||Av||^2 \\
&= \langle A^T(Ax - y), v \rangle,
\end{aligned}$$
We conclude that
$$\nabla f(x) = A^T(Ax - y)$$
Method 2 Using matrix calculus. We have
$$\begin{aligned}
\nabla f(x) &= \frac{\partial}{\partial x}\Big(\frac{1}{2}\|y\|^2 + \frac{1}{2}\|Ax\|^2 - \langle Ax, y \rangle\Big) \\
&= \frac{1}{2}\frac{\partial}{\partial x} x^T A^T Ax - \frac{\partial}{\partial x} y^T Ax
\end{aligned}$$

Let $G = (g_{ij})_{1\le i,j\le n} := A^T A$. For all $k \in \{1,\dots,n\}$ we have

$$\begin{aligned}
\frac{\partial}{\partial x_k} x^T A^T A x &= \frac{\partial}{\partial x_k} \sum_{i,j} g_{ij} x_i x_j \\
&= \sum_j g_{kj} x_j + \sum_i g_{ik} x_i \\
&= [Gx + G^T x]_k \\
&= 2[A^T A x]_k
\end{aligned}$$

Similarly,

$$\begin{aligned}
\frac{\partial}{\partial x_k} y^T A x &= \frac{\partial}{\partial x_k} \sum_{i,j} a_{ij} y_i x_j \\
&= \sum_i a_{ik} y_i \\
&= [A^T y]_k
\end{aligned}$$

We can now conclude

$$\nabla f(x) = A^T A x - A^T y = A^T(Ax - y)$$

2.

$$\begin{aligned}
\frac{g(x+hv) - g(x)}{h} &= \frac{h^2\|v\|^2 + 2hx^T v}{h} \\
&= h\|v\|^2 + 2x^T v \longrightarrow 2x^T v = \langle 2x, v\rangle
\end{aligned}$$

It is clear that $\nabla g = 2\,\mathrm{Id}$.

3. The gradient ∇ is a linear differential operator, i.e.,

$$\nabla(f + \lambda g)(x) = (\nabla f + \lambda \nabla g)(x)$$

It follows that

$$\begin{aligned}
(\nabla f + \lambda \nabla g)(x) &= \nabla f(x) + \lambda \nabla g(x) \\
&= A^T(Ax - y) + 2\lambda x
\end{aligned}$$