MSc. Data Science & Artificial Intelligence

Inverse Problems in Image Processing

Faisal JAYOUSI, Laure BLANC-FERAUD & Luca CALATRONI

Assignment 1

Author: Joris LIMONIER

joris.limonier@gmail.com

Due: February 19, 2023

Contents

1 Exercise 1

Let f be given by:

$$f(x) = \frac{1}{2}\|Ax - y\|_2^2 \tag{1}$$

we want to compute the gradient of f. We have for a given direction $v \in \mathbb{R}^n$:

$$\begin{aligned}
\nabla_v f(x) &= \lim_{\varepsilon \to 0} \frac{f(x + \varepsilon v) - f(x)}{\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{\|A(x + \varepsilon v) - y\|^2 - \|Ax - y\|^2}{2\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{\|(Ax - y) + A\varepsilon v\|^2 - \|Ax - y\|^2}{2\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{\|Ax - y\|^2 + 2\varepsilon < Ax - y, Av > + \varepsilon^2\|Av\|^2 - \|Ax - y\|^2}{2\varepsilon} \\
&= \lim_{\varepsilon \to 0} (Ax - y)^T Av + \underbrace{\frac{\varepsilon}{2}\|Av\|^2}_{\to 0} \\
&= \left\langle A^T(Ax - y), v \right\rangle,
\end{aligned}$$

We see that we obtain a scalar product with v on one side, as we wished. The other side of the scalar product (*i.e.* $A^T(Ax - y)$) corresponds to the gradient of f in the direction v. In other words, this is the change that will occur when we take a small step in the direction of v.

2 Exercise 2

2.1 Proof of the Lipschitz continuity of the gradient

Recall that a function g : is said to be L-Lipschitz continuous if $\forall\ x_1, x_2 \in \mathbb{R}^n$:

$$\|g(x_1) - g(x_2)\| \leq L\|x_1 - x_2\|. \tag{2}$$

In particular for $g \equiv \nabla f$, we have:

$$\begin{aligned}
\|\nabla f(x_1) - \nabla f(x_2)\| &= \|A^T(Ax_1 - y) - A^T(Ax_2 - y)\| \\
&= \|A^T A(x_1 - x_2)\| \\
&\leq \|A^T A\| \cdot \|x_1 - x_2\|
\end{aligned}$$

We therefore have that that ∇f is L-Lipschitz continuous, with $L := \|A^T A\|$.

2.2 Computation of the Lipschitz constant

Note that $\|A^T A\|$ is the matrix norm of $A^T A$, which is the largest singular value of $A^T A$. Let us call $\sigma_{max}(M)$ the largest singular value of a given matrix M. Recall that the singular value decomposition of M is given by:

$$M = U\Sigma V^T \tag{3}$$

where U and V are orthogonal matrices and Σ is a diagonal matrix with the singular values of M on the diagonal, which we assume to be sorted in decreasing order (without loss of generality, thanks to potential reordering of the rows, columns of U, V respectively). We have:

$$\begin{aligned}
\|A^T A\| &= \sigma_{max}(A^T A) \\
&= \sigma_{max}(U\Sigma \underbrace{V^T V}_{=Id} \Sigma U^T) \\
&= \sigma_{max}(U\Sigma^2 U^T) \\
&= \sigma_{max}(\Sigma^2) \\
&= \sigma_{max}(\Sigma)^2 \\
&= \sigma_{max}(U\Sigma V^T)^2 \\
&= \|A\|^2
\end{aligned}$$

Thus we have that the Lipschitz constant L is given by:

$$L = \|A^T A\| = \|A\|^2.$$

3 Exercise 3

We must split the problem into two subproblems: $x = 0$ and $x \neq 0$.
We will first consider the case $x = 0$. In this case, we have that the subdifferential of f at a point x is given by:

$$\partial f(x) := \{c \in \mathbb{R} \mid \forall\ y \in \mathbb{R}, \quad f(y) \geq f(x) + \langle c, y - x \rangle\} \tag{4}$$

We compute the following:

$$\begin{aligned}
\partial f(0) &= \{c \in \mathbb{R} \mid \forall\ y \in \mathbb{R}, \quad f(y) \geq f(0) + \langle c, y - 0 \rangle\} \\
&= \{c \in \mathbb{R} \mid \forall\ y \in \mathbb{R}, \quad |y| \geq \langle c, y \rangle\} \\
&= \{c \in \mathbb{R} \mid \forall\ y \in \mathbb{R}, \quad |y| \geq cy\} \\
&= \begin{cases} \{c \in \mathbb{R} \mid \forall\ y \in \mathbb{R}, \quad y \geq cy\}, & y \geq 0 \\ \{c \in \mathbb{R} \mid \forall\ y \in \mathbb{R}, \quad -y \geq cy\}, & y < 0 \end{cases} \\
&= \begin{cases} \{c \in \mathbb{R} \mid \forall\ y \in \mathbb{R}, \quad c \leq 1\}, & y \geq 0 \\ \{c \in \mathbb{R} \mid \forall\ y \in \mathbb{R}, \quad c \geq -1\}, & y < 0 \end{cases} \\
&= \{c \in \mathbb{R} \mid \forall\ y \in \mathbb{R}, \quad c \in [-1, 1]\} \\
&= [-1, 1]
\end{aligned}$$

Now for the case $x \neq 0$, we have

$$x \neq 0 \implies \partial f(x) = \{\mathrm{sign}(x)\}$$

4 Exercise 4

We want to compute the subdifferential of the following function:

$$F(x) := \frac{1}{2}\|Ax - y\|^2 + \lambda \|x\|_1^2 \tag{5}$$

We define F_1 and F_2 as follows:

$$\begin{aligned} F_1(x) &:= \frac{1}{2}\|Ax - y\|^2 \\ F_2(x) &:= \lambda\|x\|_1^2 \end{aligned}$$

and we note that $F(x) = F_1(x) + F_2(x)$.
We also note that F_1 is the function seen in exercise 1, while F_2 is somewhat of a generalization of the function seen in exercise 3 to the n-dimensional case (we will expand on that later). Furthermore, by proposition 1, we have that $\partial F(x) = \partial F_1(x) + F_2(x)$.
Let us first compute the subdifferential of F_1. By exercise 1, we know that F_1 is differentiable and we have:

$$\begin{aligned} \partial F_1(x) &= \{\nabla F_1(x)\} \\ &= \left\{A^T(Ax - y)\right\} \end{aligned}$$

Now let us compute the subdifferential of F_2. By proposition 2, we have:

$$\begin{aligned} \partial F_2(x) &= \partial\left(\lambda\|x\|_1\right) \\ &= \lambda\partial\|x\|_1 \end{aligned}$$

We notice that $\|x\|_1$ is the sum of the absolute values of the components of x, which is a convex separable function. We can therefore apply proposition 3 to compute the subdifferential of $\|x\|_1$:

$$\begin{aligned} \partial\|x\|_1 &= \partial\sum_{i=1}^{n}|x_i| \\ \implies \partial F_2(x) &= \{\lambda(p_1, \ldots, p_n) \mid \forall i = 1, \ldots, n,\ p_i \in \partial|x_i|\} \end{aligned}$$

with $\partial|x_i|$ as defined in exercise 3.
Combining the results of the previous two computations, we have:

$$\begin{aligned} \partial F(x) &= \partial F_1(x) + \partial F_2(x) \\ &= \left\{A^T(Ax - y)\right\} + \{\lambda(p_1, \ldots, p_n) \mid \forall i = 1, \ldots, n,\ p_i \in \partial|x_i|\} \\ &= \left\{A^T(Ax - y) + \lambda(p_1, \ldots, p_n) \mid \forall i = 1, \ldots, n,\ p_i \in \partial|x_i|\right\} \end{aligned}$$