TP3: Image reconstruction in X-ray tomography

Dimitri Bouche
dimitri.bouche@ensae.fr
Vincent Plassier
vincent.plassier@ens-paris-saclay.fr

March 2019

1 Part 2: Optimization problem

Question 2 We introduce the following extended notation for ψ: for $z \in \mathbb{R}^{2N}$, $\psi(z)$ denotes the vector z on which ψ has been applied coordinate-wise:

$$\psi(z) := (\psi(z_1), ..., \psi(z_N))$$

. We use the same convention for the first order derivative of ψ, ψ'

$$f(x) = \frac{1}{2}||Hx - y||^2 + \lambda \mathbf{1}_{2N}^T \psi(Gx)$$

Where $\mathbf{1}_{2N}^T$ denotes the vector of $\mathbb{R}^{2N}$ with all entries equal to 1.

$$\nabla f(x) = H^T(Hx - y) + \lambda G^T \psi'(Gx)$$

With :

$$\psi'(u) = \frac{u}{\delta\sqrt{\delta^2 + u^2}}$$

Question 3 In this question, we are interested in the lipschitz constant of the ∇f gradient. For this, we will use the triangular inequality to separately major the terms.
Take $x, z \in \mathbb{R}^N$, we know that

$$\|H^T(Hx - y) - H^T(Hz - y)\| \leq \|H^T H(x - z)\|$$

By property of the spectral norm:

$$\leq \|H^T H\| \|x - z\|$$

Now let's look at the second term. For this, we will look for a majoration of the second derivative of $\psi \in \mathcal{C}^2(\mathbb{R}^N)$ function. Let $u \in \mathbb{R}^N$, we have

$$\psi'(u) = \frac{u}{\delta\sqrt{\delta^2 + u^2}}$$

As well as

$$\psi''(u) = \frac{\delta}{(\delta^2 + u^2)^{3/2}} \leq \frac{1}{\delta^2}$$

By grouping the previous results, by triangular inequality, we obtain:

$$\|\nabla f(x) - \nabla f(z)\| \leq \|H^\star H(x - z)\| + \frac{\lambda}{\delta^2}\|G^T\| \left\| \left(\psi'\left([Gx]^{(i)}\right) - \psi'\left([Gz]^{(i)}\right)\right)_i \right\|$$

$$\overset{cf(\star)}{\leq} \left(\|H^\star H\| + \frac{\lambda}{\delta^2}\|G^T\|\|G\|\right)\|x - z\|$$

By definition of the spectral norm: $\|G^T\| = \|G\|$ and $\|H^T H\| = \|H\|^2$

$$\leq \left(\|H\|^2 + \frac{\lambda}{\delta^2}\|G\|^2\right)\|x - z\|$$

Now let's show the passage $(\star)$, indeed by inequality of the finite increments we have

$$\begin{aligned}\left\|\left(\psi'\left([Gx]^{(i)}\right)-\psi'\left([Gz]^{(i)}\right)\right)_i\right\|^2 &= \sum_{i=1}^{2N}\left|\psi'\left([Gx]^{(i)}\right)-\psi'\left([Gz]^{(i)}\right)\right|^2 \\ &\leq \|\psi''\|_\infty^2 \sum_{i=1}^{2N}\left|[G(x-z)]^{(i)}\right|^2 \\ &= \|\psi''\|_\infty^2 \|G(x-z)\|^2 \\ &\leq \|G\|^2\|x-z\|^2\end{aligned}$$

We have demonstrated that ∇f is L-Lipschitzian with $L := \|H\|^2 + \frac{\lambda}{\delta^2}\|G\|^2$.

2 Part 3: Optimization algorithms

2.1 Gradient descent

2.2 MM quadratic algorithm

Question 1 We split the majoration in two parts. The first part: $x \longmapsto \frac{1}{2}\|Hx-y\|_2^2$, is simple as its Hessian is H^TH which does not depend on x, thus we take H^TH for this part.

For the second, part, we apply slide 11 from the class. Admitting from slide 13 that ψ satisfy all the conditions. Since we are seeking the curvature term, we only take the second order term which is:

$$\frac{\psi'(|u|)}{|u|}(u-v)^2$$

We apply this $\forall i \in [\![1, 2N]\!]$ at $u = [Gx]^{(i)}$ and $v = [Gz]^{(i)}$

$$\frac{\psi'(|[Gx]^{(i)}|)}{|[Gx]^{(i)}|}([Gx]^{(i)} - [Gz]^{(i)})^2$$

Stacking all $1 \leq i \leq 2N$ terms in vector form we can write:

$$(x-z)^T G^T \mathbf{Diag}\left[\left(\frac{\psi'(|[Gx]^{(i)}|)}{|[Gx]^{(i)}|}\right)_{1\leq i\leq 2N}\right] G(x-z)$$

Which is exactly $\|x-z\|_{B(x)}^2$ with $B(x) := G^T\mathbf{Diag}\left[\left(\frac{\psi'(|[Gx]^{(i)}|)}{|[Gx]^{(i)}|}\right)_{1\leq i\leq 2N}\right] G$

Summing our two curvatures term (multiplying also $B(x)$ by λ) we get:

$$A(x) = H^TH + \lambda B(x) = H^TH + \lambda G^T \mathbf{Diag}\left[\left(\frac{\psi'(|[Gx]^{(i)}|)}{|[Gx]^{(i)}|}\right)_{1\leq i\leq 2N}\right] G$$

Important note on the implementation: The inversion of the linear operator with the conjugate gradient takes a lot of time in this setting if we want to get a decent precision on the inverse. However we found that setting a low maximum number of iteration for this inversion step still yielded a very good descent direction and took much less time. Meaning that we are using a very rough approximation of the inverse (we set a maximum of iteration of 50 for the conjugate gradient inverse, thus the conjugate gradient has not converged at all), however it still works very well.

2.3 3MG algorithm (MM memory gradient)

2.4 Block-coordinate MM quadratic algorithm

In this part, we focused in another acceleration strategy. This time, at each iteration we split the matrix $A(x)$ into blocks of length $J \in \mathbb{N}^*$. For every $x_n \in \mathbb{R}^N$, we iterate the algorithm at step n by choosing the block number $j_n := mod(n, J)$, then by updating

$$x_n^{(j_n)} = x_{n-1}^{(j_n)} - \theta_{n-1} A_{j_n}(x_{n-1})^{-1}\nabla f(x_{n-1})$$

In order to compared the differents algorithms, we choose $J = 4$ number of blocks.

2.5 Parallel MM quadratic algorithm

For every $x \in \mathbb{R}^N$, let $B(x) \in \mathbb{R}^{N\times N}$ a diagonal matrix such that $b^{(i)} = \mathcal{H}^{(i)T}\mathbf{1} + \lambda \mathcal{G}^{(i)T}\left(\frac{\psi'(Gx)}{Gx}\right)$. Let's prove that $A(x) \preceq B(x)$. To prove this result, we demonstrate for every $x \in \mathbb{R}^N$

$$x^T \mathbf{Diag}\left[\left(\mathcal{H}^{(i)T}\mathbf{1}\right)_{1\le i\le M}\right] x - x^T H^T H x \ge 0 \tag{1}$$

and

$$x^T \mathbf{Diag}\left[\left(\mathcal{G}^{(i)T}\left(\frac{\psi'(Gx)}{Gx}\right)\right)_{1\le i\le 2N}\right] x - x^T G^T \mathbf{Diag}\left[\left(\frac{\psi'(|[Gx]^{(i)}|)}{|[Gx]^{(i)}|}\right)_{1\le i\le 2N}\right] Gx \ge 0 \tag{2}$$

We start by demonstrating the inequality 1. By developping the sclar product, we get

$$\begin{aligned}
x^T \mathbf{Diag}\left[\left(\mathcal{H}^{(i)T}\mathbf{1}\right)_{1\le i\le M}\right] x - x^T H^T H x &= \sum_{m=1}^{M}\left[\sum_{i,j=1}^{N} x_i^2 |H^{(m,i)}||H^{(m,j)}| - x_i x_j H^{(m,i)} H^{(m,j)}\right] \\
&= \sum_{m=1}^{M}\left[\sum_{i=1}^{N} x_i^2 |H^{(m,i)}| \left(\sum_{j=1}^{N} |H^{(m,j)}|\right)\right] \\
&- \sum_{m=1}^{M}\left[\left(\sum_{i=1}^{N}\left(\mathrm{sign}(H^{(m,i)})\sqrt{|H^{(m,i)}|}x_i\right)\sqrt{|H^{(m,i)}|}\right)^2\right]
\end{aligned}$$

By Cauchy-Schwarz on the second term with :

$$v_1 = \mathrm{sign}(H^{(m,i)})\sqrt{|H^{(m,i)}|}x_i \text{ and } v_2 = \sqrt{|H^{(m,i)}|}$$

We finally get :

$$x^T \mathbf{Diag}\left[\left(\mathcal{H}^{(i)T}\mathbf{1}\right)_{1\le i\le M}\right] x - x^T H^T H x \ge 0$$

This shows the statement 1. Hence, we deduce that $H^T H \preceq \mathbf{Diag}\left[\left(\mathcal{H}^{(i)T}\mathbf{1}\right)_{1\le i\le M}\right]$.

Let's show in the same way inequality 2. We have

$$\begin{aligned}
&x^T \mathbf{Diag}\left[\left(\mathcal{G}^{(i)T}\left(\frac{\psi'(Gx)}{Gx}\right)\right)_{1\le i\le 2N}\right] x - x^T G^T \mathbf{Diag}\left[\left(\frac{\psi'([Gx]^{(i)})}{[Gx]^{(i)}}\right)_{1\le i\le 2N}\right] Gx \\
&= \sum_{n=1}^{2N}\left[\sum_{i,j=1}^{N} x_i^2 |G^{(n,i)}||G^{(n,j)}| \frac{\psi'([Gx]^{(i)})}{[Gx]^{(i)}} - x_i x_j G^{(n,i)} G^{(n,j)} \frac{\psi'([Gx]^{(i)})}{[Gx]^{(i)}}\right] \\
&= \sum_{m=1}^{M} \frac{\psi'([Gx]^{(i)})}{[Gx]^{(i)}}\left[\sum_{i=1}^{N} x_i^2 |G^{(m,i)}|\left(\sum_{j=1}^{N}|G^{(m,j)}|\right) - \left(\sum_{i=1}^{N}\left(\mathrm{sign}(G^{(m,i)})\sqrt{|G^{(m,i)}|}x_i\right)\sqrt{|G^{(m,i)}|}\right)^2\right]
\end{aligned}$$

Since $\frac{\psi'([Gx]^{(i)})}{[Gx]^{(i)}} \ge 0$ and by Cauchy-Schwarz on the second term with

$$v_1 = \mathrm{sign}(G^{(m,i)})\sqrt{|G^{(m,i)}|}x_i \text{ and } v_2 = \sqrt{|G^{(m,i)}|}$$

We finally get :

$$x^T \mathbf{Diag}\left[\left(\mathcal{G}^{(i)T}\left(\frac{\psi'(Gx)}{Gx}\right)\right)_{1\le i\le 2N}\right] x - x^T G^T \mathbf{Diag}\left[\left(\frac{\psi'([Gx]^{(i)})}{[Gx]^{(i)}}\right)_{1\le i\le 2N}\right] Gx \ge 0$$

Hence the expected result. In conclusion, we have shown that $A(x) \preceq B(x)$.

We obtain the following speed of convergence:

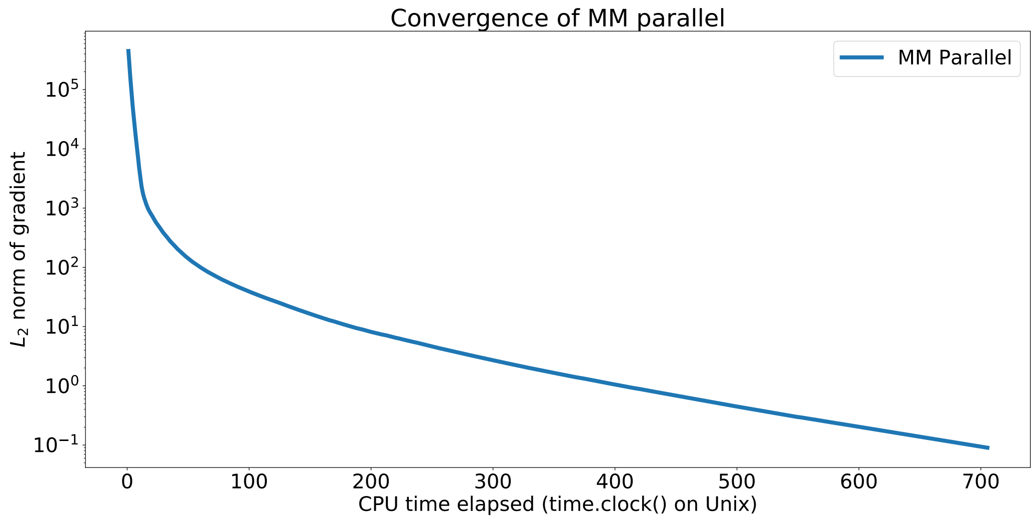

Figure 1: Convergence of MM Parallel

This takes more time to converge than the other methods as can be seen by comparing figure 1 with 2, however, we did not have any parallel architecture to run id on, so we cannot really appreciate the speed. However this may come from a tradeoff, we used a majoration $B(x)$ that is less precise than the one we used in all other MM methods ($A(x)$). As a consequence, we traded off a majoration function that is efficient with one that is less efficient, but that enables us to perform the update in parallel. Thus it seems normal that on a non parallel architecture, this should be slower.

2.6 Comparison of the methods

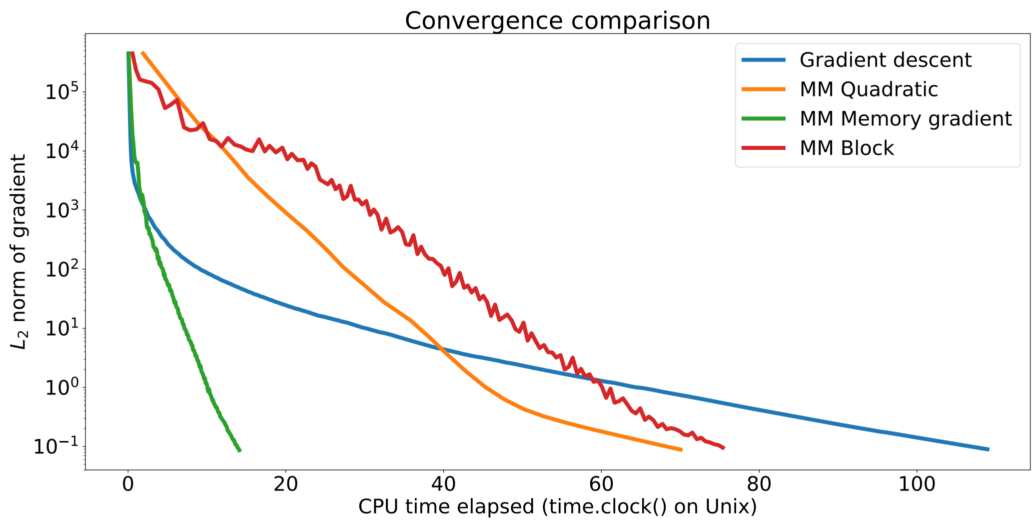

Figure 2: Comparison of the speed of convergence

Question 2 We have set $J = 4$ blocks for the block MM quadratic. The fastest of all algorithms to reach the precision is MM with Memory gradient subspace search. We also notice that with MM quadratic with blocks, the convergence is less smooth because of the partial updates that are performed at each time.

Question 3 Since 3MMG (MM Memory gradient) performed the better, we use it for this question.

Performing a small grid-search with $\lambda \in [0.01, 1.0]$ and $\delta \in [0.01, 0.5]$, we found that optimal parameters with respect to the signal to noise ratio on our grid are $\lambda = 1$ and $\delta = 0.01$. There is a compromise in δ. Increasing it yields more details in the image however this also allows more noise which is highly penalized by the Signal to Noise ratio. In figure 3 we can see on the left the reconstruction with parameters optimal (for SNR) on our grid, and on the right the reconstruction with the same λ but with a higher δ. We then see some noise but we have more details also on the image.

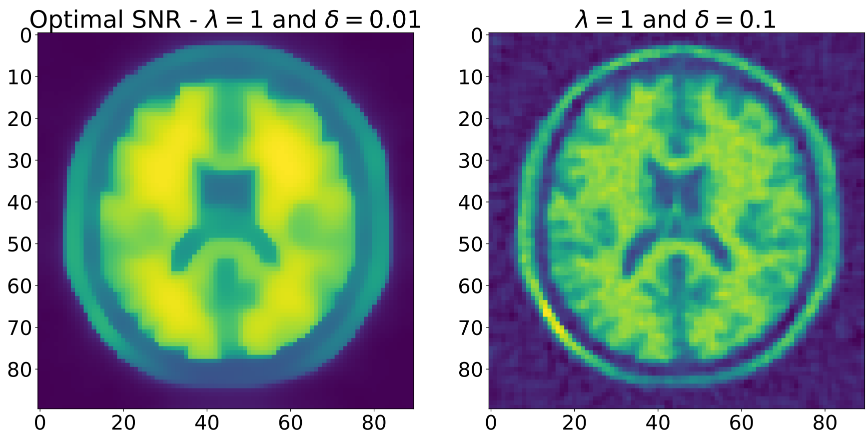

Figure 3: Reconstructed signals