Compressed Sensing

May 15, 2020

1 Preliminaries

We define useful mappings used in convex optimization, namely the proximal mapping

$$\text{Prox}_f^\lambda(y) = \arg\min_x \left(f(x) + \frac{1}{2\lambda}||x - y|| \right)$$

2 Overview

Short overwiew of matrix completion:

$$\min ||X||_* \quad s.t. \quad AX = B$$

We need to find the min rank X such that $AX = B$

3 Douglas- Rachford Splitting

In the noiseless case we can use DRS to solve this numerically via a fixed-point iteration:

$$z^{(k+1)} = z^{(k)} + \text{Prox}_g^\gamma \left(2 \, \text{Prox}_f^\gamma(z^{(k)}) - z^{(k)} \right) - \text{Prox}_g^\gamma(z^{(k)})$$

To do this, we express our objective as a sum of two functions $f + g$ where $f(X) = \delta_C(X)$ with $C = \{X|AX = B\}$ and $g(X) = ||X||_*$. We then obtain

$$\begin{aligned} \text{Prox}_f^\gamma(x) = \Pi_C(x) &= x + A^+(b - Ax) \\ &= x + A^T(AA^T)^{-1}(b - Ax) \\ &= x + A^T(b - Ax) \end{aligned}$$

To express the proximal mapping of the nuclear norm, we define the singular value thresholding operator $\mathcal{D}_\gamma$. For a matrix Y with the SVD $Y = U\Sigma V^T =$

$U\,\mathrm{diag}(\sigma(Y))V^T$ it is expressed by $\mathcal{D}_\gamma(Y) = U\,\mathrm{diag}(\sigma(Y) - \lambda)_+ V^T$. Thus we get

$$\begin{aligned}\mathrm{Prox}_g^\gamma(x) &= U(\sigma(x) - \gamma)_+ V^T \\ &= \mathcal{D}_\gamma(x)\end{aligned}$$

4 FISTA

In the noisy case we are trying to minimize the following objective (in the Lagrangian formulation):

$$\min_X \lambda||X||_* + \frac{1}{2}||AX - B||_2^2 \tag{4.1}$$

For some $\lambda > 0$. The FISTA algorithm is formulated with respect to the following objective:

$$\min_X f(x) + g(x)$$

where we assume f and g to be sufficiently smooth, i.e. $f \in C^{1,1}(\mathbb{R}^n)$, which means

$$\exists L(f) : ||\nabla f(x) - \nabla f(y)|| \leq L(f)||x - y|| \forall x, y \in \mathbb{R}^n$$

In our case we have $f(x) = \frac{1}{2}||AX - B||_2^2 = \frac{1}{2}\langle AX - B, AX - B\rangle$ and $g(x) = \lambda||X||_*$ .By simple calculation we get

$$\nabla f(X) = A^T AX - A^T B = A^T(AX - B) \tag{4.2}$$

$$L(f) = 2\lambda_{\max}(A^T A) \tag{4.3}$$

Algorithm 1 FISTA with constant step size

Input Lipschitz- constant $L(f)$ of ∇f, $y_1 = x_0 \in \mathbb{R}^n, t_1 = 1$

1: **for** $k = 1, ...$ **do**

2: $x_k = p_L(y_k)$

3: $t_{k+1} = \frac{1+\sqrt{1+4*t_k^2}}{2}$

4: $y_{k+1} = x_k + \frac{t_k - 1}{t_{k+1}}(x_k - x_{k-1})$

In this algorithm we def.

$$
\begin{aligned}
Q_L(x,y) &= f(y) + \langle x-y, \nabla f(y)\rangle + \frac{L(f)}{2}||x-y||^2 + g(x) \\
p_L(y) &= \arg\min_x Q_L(x,y) \\
&= \arg\min_x \Big(g(x) + \frac{L(f)}{2}\Big\| x - \Big(y - \frac{1}{L}\nabla f(y)\Big)\Big\|^2 \Big) \\
&= \arg\min_x \Big(\frac{g(x)}{\lambda} + \frac{L(f)}{2\lambda}\Big\| x - \Big(y - \frac{1}{L}\nabla f(y)\Big)\Big\|^2 \Big) \\
&= \mathrm{Prox}_{g/\lambda}^{\lambda/L}\Big(y - \frac{1}{L}\nabla f(y)\Big)
\end{aligned}
$$

Since we are trying to find the proximal mapping for the nuclear norm, we can apply the thresholding operator just like in our iteration step for DRS. We get

$$
\begin{aligned}
\mathrm{Prox}_{g/\lambda}^{\lambda/L}\Big(y - \frac{1}{L}\nabla f(y)\Big) &= U\,\mathrm{diag}\Big(\sigma\Big(y - \frac{1}{L}\nabla f(y)\Big) - \frac{\lambda}{L}\Big)_+ V^T \\
&= \mathcal{D}_{\lambda/L}\Big(y - \frac{1}{L}\nabla f(y)\Big) \\
&= \mathcal{D}_{\lambda/L}\Big(y - \frac{1}{L}A^T(Ay - B)\Big)
\end{aligned}
$$

I'm not sure if this is correct