Shrinkage/Thresholding Iterative Methods

- Nonquadratic regularizers
  - Total Variation
  - $l_p$-norm
  - Wavelet orthogonal/redundant representations
  - Sparse regression
- Majorization Minimization revisited
- IST- Iterative Shrinkage Thresholding Methods
- TwIST-Two step IST
Linear inverse problems - LIPs

\[ \hat{f} = \arg \min_f L(f) \]

\[ L : \mathbb{R}^m \rightarrow \mathbb{R} = [-\infty, +\infty] \]

\[ L(f) = \frac{1}{2} \| g - Af \|^2 + \alpha \Phi(f), \]

\( \Phi(f) \) is non-quadratic and possibly non-smooth
References


More references


Majorization minorization (MM) framework

Let \( \hat{x} = \arg \min_x L(x) \)

Majorization Minorization algorithm: \( \hat{x}^{(t+1)} = \arg \min_x Q(x|\hat{x}^{(t)}) \)

\( Q(x|\hat{\theta}^{(t)}) \geq L(x) \) ....with equality if and only if \( x = \hat{x}^{(t)} \)

Easy to prove monotonicity:

\[ L(\hat{x}^{(t+1)}) \leq L(\hat{x}^{(t)}) \]

Notes: \( Q(x|\hat{x}^{(t)}) \) should be easy to maximize

EM is an algorithm of this type.
MM algorithms for LIPs

\[ L(f) = \frac{1}{2} \| g - Af \|^2 + \alpha \Phi(f) \]

**IST Class:** Majorize \[ \| g - Af \|^2 \]

**IRS Class:** Majorize \[ \Phi(f) \]

**IST/IRS:** Majorize \[ \| g - Af \|^2 \] and \[ \Phi(f) \]
MM algorithms: IST class

\[ L(f) = \frac{1}{2} \Vert g - Af \Vert^2 + \alpha \Phi(f) \]

\[ L_d(f) = L_d(f') + (f - f')^* \nabla L(f') + (f - f')^* A^* A(f - f') \]

Assume that \[ \| A^* A \|_2 \leq 1 \]

\[ L_d(f) \leq L_d(f') + (f - f')^* \nabla L(f') + (f - f')^* (f - f') \]

\[ = \frac{1}{2} \| f - y \| + c^{te} \]

\[ y = f + A^* (g - Af') \]
MM algorithms: IST class

\[ L(f) = \frac{1}{2} \| g - Af \|^2 + \alpha \Phi(f) \]

\[ \| A^* A \|_2 \leq 1 \]

Majorizer:

\[ Q(f; f_k) = \frac{1}{2} \| f - y_k \|^2 + \alpha \Phi(f) \]

\[ y_k = f + A^*(g - Af_k) \]

Let:

\[ \psi_\alpha(y) = \arg \min_f \frac{1}{2} \| f - y \|^2 + \alpha \Phi(f) \]

### IST Algorithm

\[ f_{k+1} = \psi_\alpha(f + A^*(g - Af_k)) \]
**MM algorithms: IST class**

**Overrelaxed IST Algorithm**

\[ f_{k+1} = (1 - \beta) + \beta \psi_\alpha(f + A^*(g - Af_k)) \]

Convergence: [Combettes and V. Wajs, 2004]

- \( \Phi(f) \) is convex
- \( \|A\|^2_2 < 2 \)
- the set of minimizers, \( G \), of \( L(f) = \frac{1}{2} \|g - Af\|^2 + \alpha \Phi(f) \) is non-empty
- \( \beta \in ]0,1] \)
- Then \( f_k \) converges to a point in \( G \)
Denoising with convex regularizers

\[ \Phi(\cdot) \text{ convex } \Rightarrow \left\{ \frac{1}{2} \|\cdot - y\|^2 + \alpha \Phi(\cdot) \right\} \text{ strictly convex} \]

\[ \psi_\alpha(y) = \arg\min_f \frac{1}{2} \|f - y\|^2 + \alpha \Phi(f) \]

\text{Denoising function also known as the Moreou proximal mapping}

Classes of convex regularizers:

1- homogeneous (TV, lp-norm (p>1))
2- p power of an lp norm
1-Homogeneous regularizers

\[ \Phi(\cdot) \text{ convex and } \Phi(tf) = t\Phi(f) \text{ for } t \geq 0 \]

Then

\[ \psi_{\alpha}(y) = y - P_{\alpha C}(y) \]

where \( C \in \mathbb{R}^m \) is a closed convex set

\[ C = \{ f \in \mathbb{R}^m | \langle f, u \rangle \leq \Phi(u), \forall u \in \mathbb{R}^m \} \]

and \( P_A : \mathbb{R}^m \rightarrow \mathbb{R}^m \) denotes the orthogonal projection on the convex set \( A \in \mathbb{R}^m \)
Total variation regularization

Total variation [S. Osher, L. Rudin, and E. Fatemi, 1992]

\[ \Phi_{iTV}(f) = \sum_i \sqrt{(\Delta_i^h f)^2 + (\Delta_i^v f)^2} \]
\[ \Phi_{niTV}(f) = \sum_i |\Delta_i^h f| + |\Delta_i^v f| \]

\[ \Delta_i^h f = f_{i,j} - f_{i,j-1} \quad \Delta_i^v f = f_{i,j} - f_{i-1,j} \]

\( \Phi \) is convex (although not strictly) and 1-homogeneous

\[ \Phi(tf) = t\Phi(f), \text{ for } t \geq 0 \]

Total variation is a discontinuity-preserving regularizer

\( f_1, f_2 \) have the same TV
Total variation regularization

\[ \Phi(\cdot) \text{ convex and } \Phi(tf) = t\Phi(f) \text{ for } t \geq 0 \]

Then

\[ \Psi_\alpha(y) = y - P_\alpha C(y) \]

\[ C = \left\{ \text{div } p \mid p \in \mathbb{R}^m \times \mathbb{R}^m, |p_{i,j}| \leq 1, \forall i, j = 1, 2, \ldots \right\} \]

[Chambolle, 2004]

**Theorem 3.1.** Let \( \tau \leq 1/8 \). Then, \( \lambda \text{div } p^n \) converges to \( \pi_{\lambda K}(g) \) as \( n \to \infty \).

\[
p_{i,j}^{n+1} = \frac{p_{i,j}^n + \tau(\nabla(\text{div } p^n - g/\alpha))_{i,j}}{1 + \tau |(\nabla(\text{div } p^n - g/\alpha))_{i,j}|}
\]
Total variation denoising
Total variation deconvolution

2000 IST iterations !!!
Weighted $l_p$-norms

$$\Phi_{l_p^w}(f) = \|f\|_{p,w} - \left(\sum_i w_i |f_i|^p\right)^{1/p}$$

where $w_i > 0$.

$p \geq 1 \Rightarrow \Phi$ is convex (although not strictly) and 1-homogeneous.

There is no closed form expression for $\psi_\alpha$ except for some particular cases.

$p = 1 \Rightarrow \Phi$, is 1-homogeneous and $\psi_\alpha$ is decoupled.

Thus

$$\psi_\alpha(y) = (\psi_\alpha(y_1), \ldots, \psi_\alpha(y_N))$$

$$\psi_\alpha(y_i) = y_i - P_\alpha C_i(y)$$

$$C_i = \{ f_i \in \mathbb{R}^m | \langle f_i, u_i \rangle \leq |u_i|, \forall u \in \mathbb{R} \}$$
Soft thresholding: $p=1$

\[
\Phi_{\ell^1}(f) = \|f\|_1 = \sum_i |f_i|
\]

$p = 1 \Rightarrow \Phi$, is 1-homogeneous and $\Psi_\alpha$ is decoubled

Thus

\[
\Psi_\alpha(y) = (\Psi_\alpha(y_1), \ldots, \Psi_\alpha(y_N))
\]

\[
\Psi_\alpha(y_i) = y_i - P_{\alpha C_i}(y)
\]

\[C_i = \{f_i \in \mathbb{R} \mid \langle f_i, u_i \rangle \leq |u_i|, \forall u \in \mathbb{R}\} = [-1, 1]
\]

\[
P_{\alpha C_i}(y) = \begin{cases} 
\alpha & y \geq \alpha \\
-\alpha & y \leq \alpha \\
y & |y| < \alpha.
\end{cases}
\]

\[
\Psi_\alpha(y) = y - P_{\alpha C}(y) = \begin{cases} 
y - \alpha & y \geq \alpha \\
y + \alpha & y \leq \alpha \\
0 & |y| < \alpha.
\end{cases}
\]
Soft thresholding: \( p=1 \)

\[
\Phi_{\ell^1}(f) = \|f\|_1 = \sum_i |f_i| 
\]

\[
\psi_{\alpha}(y) = y - P_{\alpha C}(y) = \begin{cases} 
   y - \alpha & y \geq \alpha \\
   y + \alpha & y \leq \alpha \\
   0 & |y| < \alpha.
\end{cases}
\]

\( \equiv \text{soft}(y, \alpha) \)

![Diagram showing soft thresholding with lines at \(-\alpha\), \(\alpha\), and \(0\) and a dotted line for \(\alpha\) and \(-\alpha\)]
Soft thresholding: $p=1$

\[ \Phi_{\ell^1}(f) = \|f\|_1 = \sum_i |f_i| \]

\[ \Psi_{\alpha}(y) = y - P_{\alpha\mathcal{C}}(y) = \begin{cases} 
  y - \alpha & y \geq \alpha \\
  y + \alpha & y \leq \alpha \\
  0 & |y| < \alpha.
\end{cases} \equiv \text{soft}(y, \alpha) \]
Another way to look at it:

\[ L(f) = \frac{1}{2} \| f - y \|^2 + \alpha \Phi(f) \]

Since \( L \) is convex: the point \( \hat{f} \) is a global minimum of \( L \) iif

\[ 0 \in \partial L(\hat{f}) \]

where \( \partial L(f) \) is the subdifferential of \( L \) at \( f \)

\[ \partial L(f') = \{ u | L(f) - L(f') \geq \langle u, f - f' \rangle \} \]
Example: Wavelet-based restoration

\[ \theta = W f \]

\[ \text{Wavelet basis} \]
\[ \text{Wavelet coefficients} \]

\[ g, h \text{ – quadrature mirror filters} \]

Approximation coefficients (g-low pass filter)

Detail coefficients (h – high pass filter)

DWT, Harr, J=2
Example: Wavelet-based restoration