Iteratively Reweighted Least-Squares solutions for non-linear reconstruction

Matthieu Guerquin-Kern

April 5, 2009

As a main reference: positive form of the half-quadratic minimization in Optimisation course of Mila Nikolova. First works on half-quadratic minmization: D. Geman and C. Yang, **Nonlinear image recovery with half-quadratic regularization**, *IEEE Transactions on Image Processing*, IP-4 (1995), pp. 932–946.

Introduction

Consider the problem of recovering some real original data $\mathbf{c}_0$ from noisy real measurements:

$$\mathbf{m} = \mathbf{E}\mathbf{c}_0 + \mathbf{b}. \tag{1}$$

The $M \times N$ matrix real $\mathbf{E}$ represents the linear forward model and $\mathbf{b}$ is the vector representing both model mismatch and noise.

The original data and the measurements do not necessarily have the same size (*i.e.* the matrix $\mathbf{E}$ can be rectangular).

A popular way to define the solution $\widetilde{\mathbf{c}}$ of this inverse problem is:

$$\widetilde{\mathbf{c}} = \arg\min_{\mathbf{c}} \|\mathbf{m} - \mathbf{E}\mathbf{c}\|_{\ell^2}^2 + \lambda \|\mathbf{R}\mathbf{c}\|_{\ell^p}^p . \tag{2}$$

The matrix $\mathbf{R}$ is often chosen as a differential operator: finite differences, wavelet analysis. The parameter λ balances the effect of the two terms: the fidelity to the data and to the *a priori* .

Quadratic regularization corresponds to $p = 2$ and leads to the linear solution: $\widetilde{\mathbf{c}} = (\mathbf{E}^H\mathbf{E} + \lambda\mathbf{R}^H\mathbf{R})^\dagger \mathbf{E}^H\mathbf{m}$.

Non-linear solutions with $p \leqslant 2$ are often preferred when the reconstruction problem is severely ill-conditioned.

Here we study the Iteratively Reweighted Least-Squares (IRLS) method that solve such non-quadratic minimization problem.

For starters, let us rewrite the problem as

$$\widetilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{m} - \mathbf{L}\mathbf{x}\|_{\ell^2}^2 + \lambda \|\mathbf{x}\|_{\ell^p}^p , \tag{3}$$

with $\mathbf{L} = \mathbf{E}\mathbf{R}^{-1}$ and $\widetilde{\mathbf{c}} = \mathbf{R}^{-1}\widetilde{\mathbf{x}}$.

In the sequel, we define the cost function $\mathcal{C}(\mathbf{x}) = \|\mathbf{m} - \mathbf{L}\mathbf{x}\|_{\ell^2}^2 + \lambda \|\mathbf{x}\|_{\ell^p}^p$.

1 Quadratic upper-bound

The goal of this section is to define a quadratic upper-bound $\mathcal{Q}$ of $\mathcal{C}$, adapted to a point $\mathbf{x}^*$. Here are the constraints we want to impose:

1. $\mathcal{Q}$ is a quadratic term,
2. Same values at point $\mathbf{x}^*$: $\mathcal{Q}(\mathbf{x}^*) = \mathcal{C}(\mathbf{x}^*)$,
3. Same gradients at point $\mathbf{x}^*$: $\nabla_{\mathbf{x}}\mathcal{Q}(\mathbf{x}^*) = \nabla_{\mathbf{x}}\mathcal{C}(\mathbf{x}^*)$,
4. $\mathcal{Q}$ upper bounds $\mathcal{C}$: $\forall \mathbf{x} \neq \mathbf{x}^* \quad \mathcal{Q}(\mathbf{x}) > \mathcal{C}(\mathbf{x})$

Considering constraint 1, the data fidelity term $\|\mathbf{m} - \mathbf{L}\mathbf{x}\|_{\ell^2}^2$ is kept. Only the regularization term $\|\mathbf{x}\|_{\ell^p}^p$ needs a local quadratic approximation. To begin with, remark that:

$$\|\mathbf{x}^*\|_{\ell^p}^p = \|\mathbf{x}^*\|_{\mathbf{D}}^2 , \tag{4}$$

where $\|\mathbf{x}\|_{\mathbf{M}} = \langle \mathbf{x}, \mathbf{M}\mathbf{x} \rangle$ denotes a weighted ℓ^2-norm and $\mathbf{D}$ is a diagonal positive-definite matrix whose elements are $d_i = |x_i^*|^{p-2}$.

Thus, we define $\mathcal{Q}$ as:

$$\mathcal{Q}(\mathbf{x}) = \|\mathbf{m} - \mathbf{L}\mathbf{x}\|_{\ell^2}^2 + \lambda \left(\alpha \|\mathbf{x}\|_{\mathbf{D}}^2 + \beta \right). \tag{5}$$

The conditions 2 and 3 impose $\alpha = \frac{p}{2}$ and $\beta = (1 - \frac{p}{2}) \|\mathbf{x}^*\|_{\ell^p}^p$.

We have

$$\mathcal{Q}(\mathbf{x}) = \|\mathbf{m} - \mathbf{L}\mathbf{x}\|_{\ell^2}^2 + \frac{\lambda p}{2} \|\mathbf{x}\|_{\mathbf{D}}^2 + \lambda \left(1 - \frac{p}{2}\right) \|\mathbf{x}^*\|_{\ell^p}^p. \tag{6}$$

Proposition 1. *If $p < 2$, the function $\mathcal{Q}$, as defined in Eq. 6, is an upper-bound of $\mathcal{C}$. They join only at point $\mathbf{x} = \mathbf{x}^*$.*

Proof. Let note $a_i = x_i^2$ and $b_i = (x_i^*)^2$. We can write:

$$\begin{aligned} \mathcal{Q}^*(\mathbf{x}) - \mathcal{C}(\mathbf{x}) &= \frac{\lambda p}{2} \|\mathbf{x}\|_{\mathbf{D}}^2 + \lambda \left(1 - \frac{p}{2}\right) \|\mathbf{x}^*\|_{\ell^p}^p - \lambda \|\mathbf{x}\|_{\ell^p}^p & (7) \\ &= \lambda \sum_i \left[\frac{p}{2} (b_i)^{p/2-1} a_i + (1 - \frac{p}{2}) (b_i)^{p/2} - (a_i)^{p/2} \right] & (8) \end{aligned}$$

If $p < 2$ then $(a)^{p/2} < \frac{p}{2}(b)^{p/2-1}a + (1 - \frac{p}{2})(b)^{p/2} \quad \forall a \neq b \in \mathbb{R}^+$. Indeed, if $p < 2$ the function $f : \mathbb{R}^+ \to \mathbb{R}^+$ such that $f(a) = (a)^{p/2}$ is stricly concave and upper-bounded by its tangent at every point $b \in \mathbb{R}^+$: $g(a) = f(b) + f'(b).(a - b) = \frac{p}{2}(b)^{p/2-1}a + (1 - \frac{p}{2})(b)^{p/2}$.

As a consequence, if $p < 2$ we have $\mathcal{Q}^*(\mathbf{x}) > \mathcal{C}(\mathbf{x}) \quad \forall \mathbf{x} \neq \mathbf{x}^*$. □

2 Iterative minimization

In the previous section, we upper-bounded the cost function $\mathcal{C}$ with a well-suited function $\mathcal{Q}$ that mimics its local behaviour at a point $\mathbf{x}^*$. The idea is now to iteratively update the upper-bound $\mathcal{Q}^{(n)}$ with the weights $\mathbf{D}^{(n)}$ depending on the current estimate $\mathbf{x}^{(n)}$.

The next estimate $\mathbf{x}^{(n+1)}$ is defined as the minimizer of $\mathcal{Q}^{(n)}$. As the latter function is quadratic, we get the linear solution: $\mathbf{x}^{(n+1)} = (\mathbf{L}^H\mathbf{L} + \frac{\lambda p}{2}\mathbf{D}^{(n)})^\dagger \mathbf{L}^H\mathbf{m}$.

The algorithm is defined as follows:

- Initialize $\mathbf{x}^{(0)}$.

  While convergence is not reached
- $\mathbf{D}^{(n)} = \text{diag}\left(|\mathbf{x}^{(n)}|^{p-2}\right)$,
- $\mathbf{x}^{(n+1)} = (\mathbf{L}^H\mathbf{L} + \frac{\lambda p}{2}\mathbf{D}^{(n)})^\dagger \mathbf{L}^H\mathbf{m}$.

  End while

Proposition 2. *If $p > 1$, the solution $\widetilde{\mathbf{x}}$ is a fixed-point of the algorithm.*

Proof. If $p > 1$, $\mathcal{C}$ is differentiable and admits a unique minimizer $\widetilde{\mathbf{x}}$. We have the following property: $\nabla_{\mathbf{x}}\mathcal{C}(\widetilde{\mathbf{x}}) = \mathbf{0}$. This yields $2\mathbf{L}^H\mathbf{L}\widetilde{\mathbf{x}} - 2\mathbf{L}^H\mathbf{m} + p\mathbf{D}^*\widetilde{\mathbf{x}} = \mathbf{0}$, with $\mathbf{D}^* = \text{diag}\left(|\widetilde{\mathbf{x}}|^{p-2}\right)$. The latter rewrites $\widetilde{\mathbf{x}} = (\mathbf{L}^H\mathbf{L} + \frac{\lambda p}{2}\mathbf{D}^*)^\dagger \mathbf{L}^H\mathbf{m}$. □

Proposition 3. *If $p < 2$, as long as $\mathbf{x}^{(n+1)} \neq \mathbf{x}^{(n)}$, we have $\mathcal{C}(\mathbf{x}^{(n+1)}) < \mathcal{C}(\mathbf{x}^{(n)})$.*

Proof. By definition of $\mathbf{x}^{(n+1)}$, we have $\mathcal{Q}^{(n)}(\mathbf{x}^{(n+1)}) \leqslant \mathcal{Q}^{(n)}(\mathbf{x}^{(n)})$. From constraint 2, $\mathcal{Q}^{(n)}(\mathbf{x}^{(n)}) = \mathcal{C}(\mathbf{x}^{(n)})$. From constraint 4, we get $\mathcal{C}(\mathbf{x}^{(n+1)}) < \mathcal{Q}^{(n)}(\mathbf{x}^{(n+1)})$. □

Under the assumptions of Propositions 4 and 3 the serie $\mathbf{x}^{(n)}$ is proven to converge to $\widetilde{\mathbf{x}}$. Consequently, if $1 < p < 2$ the IRLS method solves the minimization problem (3).

3 IRLS for wavelet regularization

Many natural-looking images, in particular the ones obtained through MRI scanners, have sparse approximations in the wavelet domain. It is then logical to try to exploit this sparsity *a priori* to make the inverse problem better conditioned.

In this Section, we condider the particular case where the regularization operator is an orthonormal wavelet transform, *i.e.* $\mathbf{R} = \boldsymbol{W}$ and $\boldsymbol{W}^{-1} = \boldsymbol{W}^{\mathrm{H}}$ ($\boldsymbol{W}$ is a unitary matrix).

Proposition 4. *If the matrix* $\mathbf{D}$ *is invertible, the image reconstruction matrix* $\mathbf{W}^{\mathrm{H}}\left(\mathbf{L}^{\mathrm{H}}\mathbf{L} + \frac{\lambda p}{2}\mathbf{D}\right)^{-1}\mathbf{L}^{\mathrm{H}}$ *presented in Section 2 can be rewritten:*

$$\left(\mathbf{E}^{\mathrm{H}}\mathbf{E} + \frac{\lambda p}{2}\boldsymbol{W}^{\mathrm{H}}\mathbf{D}\boldsymbol{W}\right)^{-1}\mathbf{E}^{\mathrm{H}} \tag{9}$$

The algorithm can also be presented in the image domain:

- Initialize $\mathbf{c}^{(0)}$.

  While convergence is not reached

- $\mathbf{D}^{(n)} = \mathrm{diag}\left(|\boldsymbol{W}\mathbf{c}^{(n)}|^{p-2}\right)$,

- $\mathbf{c}^{(n+1)} = \left(\mathbf{E}^{\mathrm{H}}\mathbf{E} + \frac{\lambda p}{2}\boldsymbol{W}^{-1}\mathbf{D}^{(n)}\boldsymbol{W}\right)^{\dagger}\mathbf{E}^{\mathrm{H}}\mathbf{m}$.

  End while