Iteratively Reweighted Least-Squares solutions for non-linear reconstruction

Matthieu Guerquin-Kern

April 5, 2009


Introduction

Consider the problem of recovering some real original data $c_0$ from noisy real measurements:

$$m = Ec_0 + b.$$  

(1)

The $M \times N$ matrix real $E$ represents the linear forward model and $b$ 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 $E$ can be rectangular).

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

$$\tilde{c} = \arg \min_c \|m - Ec\|_2^2 + \lambda \|Rc\|_p^p.$$  

(2)

The matrix $R$ is often chosen as a differential operator: finite differences, wavelet analysis. The parameter $\lambda$ 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: $\tilde{c} = (E^HE + \lambda R^HR)^+E^Hm$.

Non-linear solutions with $p \leq 2$ 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

$$\tilde{x} = \arg \min_x \|m - Lx\|_2^2 + \lambda \|x\|_p^p,$$  

(3)

with $L = ER^{-1}$ and $\tilde{c} = R^{-1}\tilde{x}$.

In the sequel, we define the cost function $\tilde{c}(x) = \|m - Lx\|_2^2 + \lambda \|x\|_p^p$.

1 Quadratic upper-bound

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

1. $\Omega$ is a quadratic term,
2. Same values at point $x^*$: $\Omega(x^*) = \tilde{c}(x^*)$,
3. Same gradients at point $x^*$: $\nabla_x \Omega(x^*) = \nabla_x \tilde{c}(x^*)$,
4. $\Omega$ upper bounds $\tilde{c}$: $\forall x \neq x^* \Omega(x) > \tilde{c}(x)$

Considering constraint 1, the data fidelity term $\|m - Lx\|_2^2$ is kept. Only the regularization term $\|x\|_p^p$ needs a local quadratic approximation. To begin with, remark that:

$$\|x^*\|_p^p = \|x^*\|_D^2,$$  

(4)
Proof. The next estimate

Thus, we define \( \Omega(x) = \|m - Lx\|_2^2 + \lambda \left( \alpha \|x\|_D^2 + \beta \right) \).

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

We have

\[
\Omega(x) = \|m - Lx\|_2^2 + \frac{\lambda p}{2} \|x\|_D^2 + \lambda \left( 1 - \frac{p}{2} \right) \|x^*\|_D^p.
\]

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

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

\[
\Omega^*(x) - \mathcal{C}(x) = \frac{\lambda p}{2} \|x\|_D^2 + \lambda \left( 1 - \frac{p}{2} \right) \|x^*\|_D^p - \lambda \|x\|_D^p
\]

\[
= \lambda \sum_i \left[ \left( \frac{p}{2} b_i \right)^{p/2 - 1} a_i + \left( 1 - \frac{p}{2} \right) b_i^{p/2} - (a_i)^{p/2} \right]
\]

If \( p < 2 \) then \( (a)_{2/p} < \frac{p}{2} (b)^{p/2 - 1} a + (1 - \frac{p}{2}) (b)^{p/2} \) \( \forall a \neq b \in \mathbb{R}^+ \). Indeed, if \( p < 2 \) the function \( f : \mathbb{R}^+ \to \mathbb{R}^+ \) such that \( f(a) = (a)^{p/2} \) is strictly 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 \( \Omega^*(x) > \mathcal{C}(x) \) \( \forall x \neq x^* \).

2 Iterative minimization

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

The next estimate \( x^{(n+1)} \) is defined as the minimizer of \( \Omega^{(n)} \). As the latter function is quadratic, we get the linear solution:

\[ x^{(n+1)} = (L^H L + \frac{\lambda p}{2} D^{(n)})^\dagger L^H m. \]

The algorithm is defined as follows:

- Initialize \( x^{(0)} \).
- While convergence is not reached
  - \( D^{(n)} = \text{diag} (|x^{(n)}|^{p-2}) \),
  - \( x^{(n+1)} = (L^H L + \frac{\lambda p}{2} D^{(n)})^\dagger L^H m. \)
- End while

**Proposition 2.** If \( p > 1 \), the solution \( \tilde{x} \) is a fixed-point of the algorithm.

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

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

*Proof.* By definition of \( x^{(n+1)} \), we have \( \Omega^{(n)}(x^{(n+1)}) \leq \Omega^{(n)}(x^{(n)}) \). From constraint 2, \( \Omega^{(n)}(x^{(n)}) = \mathcal{C}(x^{(n)}) \). From constraint 4, we get \( \mathcal{C}(x^{(n+1)}) < \Omega^{(n)}(x^{(n+1)}) \).

Under the assumptions of Propositions 4 and 3 the series \( x^{(n)} \) is proven to converge to \( \tilde{x} \). 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 consider the particular case where the regularization operator is an orthonormal wavelet transform, i.e. $R = W$ and $W^{-1} = W^H$ ($W$ is a unitary matrix).

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

$$\left( E^H + \frac{\lambda p}{2} W^H DW \right)^{-1} E^H$$  \hspace{1cm} (9)

The algorithm can also be presented in the image domain:

- Initialize $c^{[0]}$.
- While convergence is not reached
  - $D^{(n)} = \text{diag} (|Wc^{(n)}|^p)$,
  - $c^{(n+1)} = (E^H + \frac{\lambda p}{2} W^{-1} D^{(n)} W) \frac{1}{E^H} m$.
- End while