FAST IMAGE RECONSTRUCTION ALGORITHMS COMBINING HALF-QUADRATIC REGULARIZATION AND PRECONDITIONING

Mila Nikolova

Département Traitement du Signal et des Images
ENST
46, rue Barrault, 75634 Paris Cédex 13, France

*Michael Ng**

Department of Mathematics
The University of Hong Kong
Pokfulam Road, Hong Kong

ABSTRACT

In this paper, we focus on image deconvolution and image reconstruction problems where a sought image is recovered from degraded observed data. The solution is defined to be the minimizer of an objective function combining a data-fidelity term and a edge-preserving, convex regularization term. Our objective is to speed up the calculation of the solution in a wide range of situations. To this end, we propose a method applying pertinent preconditioning to an adapted half-quadratic equivalent form of the objective function. The optimal solution is then found using an alternating minimization (AM) scheme. We focus specifically on Huber regularization. We exhibit the possibility get very fast calculations while preserving the edges in the solution. Preliminary numerical results are reported to illustrate the effectiveness of our method.

1. INTRODUCTION

This work addresses a wide class of image reconstruction situations where a sought image $\hat{x} \in R^p$ is recovered from degraded data $y \in R^q$ by minimizing an objective function $J : R^p \to R$ combining a data-fidelity term and a regularization term:

$$\hat{x} = \min_{x \in R^p} J(x) \tag{1}$$

$$J(x) = \|Ax - y\|^2 + \beta \sum_{i=1}^{r} \phi(d_i^T x). \tag{2}$$

In the expression above, $A \in R^{q \times p}$ represents the observation system and $\beta > 0$ is a parameter. The regularization term involves a smooth convex function $\phi : R \to R$, applied to a set of linear transforms of the image $d_i^T x$ which are typically first or second-order differences between neighboring pixels. Such reconstruction methods are well-known to allow the obtention of high quality image estimates $\hat{x}$ if $\phi(t)$ approaches affine function when $|t| \to \infty$ [4, 1, 6].

However, their use in different image reconstruction and image restoration applications can be practically limited by the numerical cost needed for the calculation of the estimate. The latter is usually calculated using iterative descent algorithms, based on the gradient of J. The calculation speed is mainly limited by the fact the gradient of J is nonlinear and by usually bad the conditioning of the objective function at each iteration. As an alternative, some authors used coordinate-wise minimization schemes [5]. The ambition of this paper is to propose a really fast algorithm for the calculation of $\hat{x}$ by using a half-quadratic equivalent form of J combined with pertinent preconditioning.

2. TWO FORMS OF HALF-QUADRATIC REGULARIZATION

Since [11] and [12], numerous algorithms have been proposed where $\hat{x}$ is calculated by minimizing an augmented objective function $G : R^p \times R^r \mapsto R$ which involves an auxiliary variable $s \in R^r$:

$$G(x, s) = \|Ax - y\|^2 + \beta \sum_{i=1}^{r} Q(d_i^T x, s_i) + \beta \sum_{i=1}^{r} \psi(s_i)$$

where for every $s_i \in R$ the function $Q(., s_i) : R \mapsto R$ is quadratic. The equivalence with (1-2) is ensured by the requirement that for every $x \in R^p$, $J(x) = \min_{s \in R^r} G(x, s)$. Based on the "continuous-valued" line variables introduced in [11], numerous authors considered quadratic terms of the form [2, 6, 8, 13]:

$$Q(d_i^T x, s_i) = (d_i^T x)^2 s_i. \tag{3}$$

A slightly different form for Q has been proposed in [12] to perform stochastic optimization for a specific concave regularization:

$$Q(d_i^T x, s_i) = (d_i^T x - s_i)^2. \tag{4}$$

*Research supported in part by Hong Kong Research Grants Council Grant Nos. HKU 7147/99P and HKU 7132/00P, and HKU CRCG Grant No. 10203501.

0-7803-6725-1/01/$10.00 ©2001 IEEE

Several authors generalized this approach [7, 1].

The advantage of such an equivalent formulation is twofold: for every s fixed, the gradient of $x \mapsto G(x,s)$ is an affine function whereas the non-quadratic part of G, that is $s \mapsto G(x,s)$, is separable in s and hence easy to minimize in a parallel way. Intuitively, such a scheme should speed up calculations. However, the numerical effects relevant to each formulation have never been examined.

Recall that the numerical efficiency of the minimization of G is tightly connected with the conditioning of its Hessian at each iteration. Let D denotes the $p \times r$ matrix yielding $[Dx]_i = d_i x$ for all i and x and diag(s) is a diagonal matrix with diagonal elements s_i. We observe that under (3), the Hessian of $G(.,s)$ is

$$2A^TA + \beta \text{diag}(s) D^T D. \tag{5}$$

The conditioning of $G(.,s)$ clearly depends on the value of s and this is an important drawback since the entries of s can take very different values. In comparison, the Hessian of $G(.,s)$ under (4) reads

$$H := 2A^TA + \beta D^T D, \tag{6}$$

hence it is independent of s. This is a useful advantage which pushes us to focus on (4).

3. PROPOSED ALGORITHM

The function G is classically constructed by using the theory of convex conjugate functions [16]. Under some technical assumptions [7, 1] ensuring that the maximum in (7) is well defined (e.g. that the function $t \mapsto ct^2/2 - \phi(t)$ is convex and coercive for some $c > 0$), it is found that the following function

$$\psi(s) = \max_{t \in R} \left[\phi(t) - \frac{1}{2} \left(\frac{s}{\sqrt{c}} - \sqrt{c}t \right)^2 \right] \tag{7}$$

is convex and for every t we have

$$\phi(t) = \min_{s \in R} \left[\psi(s) + \frac{1}{2} \left(\frac{s}{\sqrt{c}} - \sqrt{c}t \right)^2 \right]. \tag{8}$$

The above equality suggests to consider the following augmented criterion:

$$G(x,s) = \|Ax - y\|^2 + \beta \sum_{i=1}^{r} \left[\frac{1}{2} (d_i^T x - s_i)^2 + \psi(s_i) \right]$$

In our context, we ensure that ψ is convex, so G is convex function with respect to (x,s). Therefore, with an initial guess $(x^{(0)}, s^{(0)})$ for (x,s), we can minimize $G(x,s)$ by first solving $G(x^{(0)}, s^{(1)}) \equiv \min_s G(x^{(0)}, \cdot)$ and then $G(x^{(1)}, s^{(1)}) \equiv \min_x G(\cdot, s^{(1)})$. We develop an alternating minimization (AM) algorithm in which the function value $G(x^{(n)}, s^{(n)})$ always decreases as n increases. More precisely, the algorithm is stated as follows:

Assume we have $x^{(k)}$ and $s^{(k)}$:

- Find the entries of $s^{(k+1)}$ by solving

$$\min_{s_i^{(k+1)}} \psi(s_i^{(k+1)}) + (d_i^T x^{(k)} - s_i^{(k+1)})^2 \tag{9}$$

  for $i = 1, \ldots, r$. This minimization is very easy since it is performed for each i separately. Notice that the next step does not involve the minimum value $\psi(s_i^{(k+1)})$ but only its argument $s_i^{(k+1)}$.

- Solve for $x^{(k+1)}$

$$Hx^{(k+1)} = 2A^Ty + \sum_{i=1}^{r} s_i^{(k+1)} d_i. \tag{10}$$

The matrix in the left side above H is as given in (6) and it is constant, whereas the right-side is easily updated at each iteration. The costly stage in the algorithm is the inversion of H. To this end, we propose to use pertinent preconditioning of H in order to speed up the computation of $x^{(k+1)}$.

4. PRECONDITIONING OF G

We will especially focus on applications where the matrices A modelling the observation system are Toeplitz-like, [15]. In this paper, we will consider the preconditioned conjugate gradient method to solve the linear system in (10) with Toeplitz-like coefficient matrix. Optimal transformed based matrices are used to precondition Toeplitz-like matrices in conjugate gradient iterations [10]. Part of their motivation was to exploit the fast inversion of transform based matrices [15]. Numerical results suggest that the method converges very fast for a wide range of Toeplitz-like observation operators A.

5. APPLICATIONS TO HUBER REGULARIZATION AND COSINE PRECONDITIONERS

We now concentrate on Huber regularizations which are defined using

$$\phi(t) = \begin{cases} \frac{1}{2}t^2 & \text{if} \quad |t| \le \alpha \\ \alpha|t| - \frac{1}{2}\alpha^2 & \text{if} \quad |t| > \alpha \end{cases} \tag{11}$$

The threshold parameter α controls the size of the discontinuities modeled by the prior by providing a less severe edge penalty. We point out several advantages of such a regularization.

- It is among the best functions which are both convex and edge-preserving. Unlike total variation regularization (corresponding to $\phi(t) = |t|$) which yields a "blocky effect" [9, 14], the quadratic part near 0 in (11) allows smoothly varying regions in the image to be restored (see [14]).
- Although ϕ' is discontinuous at $\pm\alpha$, we show that the chance to get a minimizer $\hat{x}$ involving a difference such that $|d_i^T\hat{x}| = \alpha$ is null. Normally, the relevant J is $\mathcal{C}^\infty$ on a neighborhood of its minimizers.
- When A is singular, J is non-strictly convex and may give rise to non-strict minimizers. However, we show that meaningful restored images can almost never be non-strict minimizers.
- A crucial numerical advantage is that Huber function (11) gives rise to only linear and constant terms in the derivative of both J and G.

We give more details in the full paper.

By a proper choice of c in (7), we may get ψ which is differentiable. However, we prefer taking $\psi(s) = \alpha|s|$ which corresponds to $c = 1$. It is easy to check that (8) remains true. The minimization of $G(.,x)$ involved in (9) is particularly simple in this case since only $d_i^T x^{(k)}$ need to be computed:

$$s^{(k+1)} = \begin{cases} 0 & \text{if } |d_i^T x^{(k)}| \leq \alpha \\ |d_i^T x^{(k)} - \alpha| & \text{if } |d_i^T x^{(k)}| > \alpha \end{cases}$$

5.1. Experimental Results

In the experiments, we consider reconstructing high resolution images from multiple under-sampled, shifted, degraded frames with sub-pixel displacement errors [3].

In Figure 1 are displayed: (a) the original image, (b) the observed blurred and noisy version, (c) a reconstruction obtained using Huber function and (d) a reconstruction using Laplacian regularization. It is clear that the image obtained using the Huber regularization is more neat than that using the Laplacian regularization. We remark that the stopping criteria of the AM method and the preconditioned conjugate gradient method are $||x^{(k)} - x^{(k-1)}||_2 < 20$ and $||r^{(j)}||/||r^{(0)}|| < 10^{-6}$ respectively, where $r^{(j)}$ is the normal equations residual after j iterations.

Next we show the efficiency of our method. The number of AM iterations are listed in Tables 1 and 2. When α increases, the number of AM iterations decreases. However, when β changes, the numbers of AM iterations are almost the same for different α.

We also note that the numbers of iterations of using the preconditioned conjugate gradient (PCG) method for inverting the matrix in (6) are almost the same for each AM iteration. In Tables 3 and 4, the total numbers of CG and PCG iterations required to solve the linear systems in each AM step are listed. Here we used the optimal cosine transform based preconditioners for the image reconstruction problem in §5.1. We see from the tables that it takes significantly more iterations using the CG method that those using the PCG method. Next we compare the number of iterations required to solve the linear systems with the coefficient matrices in (5) and (6). We find that the number of iterations of using CG or PCG for solving the linear system with the coefficient matrix in (5) is in average more than 300. This number is much greater than those required for solving linear system with the coefficient matrix in (6). This demonstrates the effectiveness of our new formulation.

β	α	Iter.
1e-2	(5,10,20,40)	(20,14,9,5)
1e-3	(5,10,20,40)	(20,15,9,5)
1e-4	(5,10,20,40)	(28,16,9,5)

Table 1. Number of AM iterations for SNR=40dB.

β	α	Iter.
1e-2	(5,10,20,40)	(20,15,9,5)
1e-3	(5,10,20,40)	(17,13,9,5)
1e-4	(5,10,20,40)	(15,11,8,5)

Table 2. Number of AM iterations for SNR=50dB.

β	α	CG	PCG
1e-2	(5,10,20,40)	(480,336,216,120)	(120,84,54,30)
1e-3	(5,10,20,40)	(1180,885,531,295)	(140,105,63,35)
1e-4	(5,10,20,40)	(4172,2384,1341,745)	(252,144,81,45)

Table 3. Number of CG iterations for solving the system with the coefficient matrix in (6) when SNR=40dB.

β	α	CG	PCG
1e-2	(5,10,20,40)	(480,360,216,120)	(120,84,54,30)
1e-3	(5,10,20,40)	(969,741,513,285)	(140,105,63,35)
1e-4	(5,10,20,40)	(1995,1463,1064,665)	(252,144,81,45)

Table 4. Number of CG iterations for solving the system with the coefficient matrix in (6) when SNR=50dB.

Finally, we give an another example as shown as in Figure 2. It takes 49 PCG iterations to reconstruct the image (e) using our new model. Again, we see from the figure that the error (rel. error = 0.04601) of using the Huber regularization (f) is less than that (relative error = 0.05019) using the Laplacian regularization (d).

Preliminary numerical results show that our method is quite efficient and effective by combining the new formulation of half-quadratic regularization and preconditioning. In the future work, we plan to make a detailed comparisons and extend to other regularization functionals.

6. REFERENCES

[1] G. Aubert and L. Vese, "A variational method in image recovery," *SIAM Journal on Numerical Analysis*, vol.

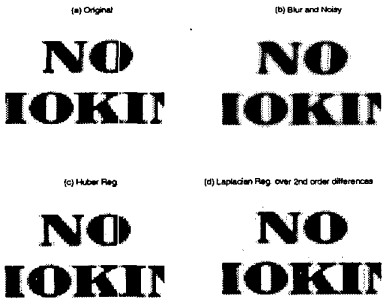

Fig. 1. Results of the first test image.

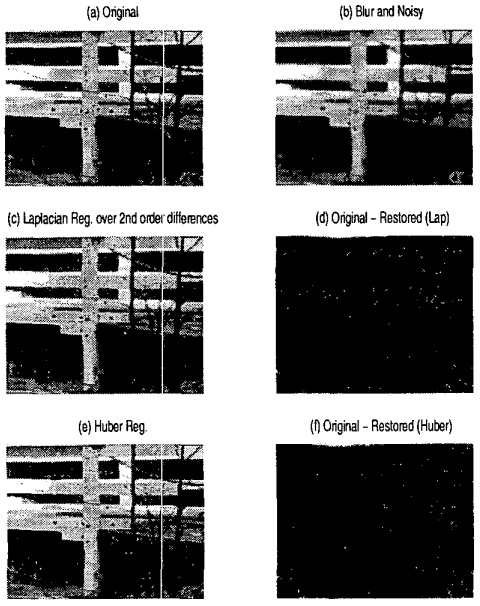

Fig. 2. Results of the second test image.

34, no. 5, pp. 1948–1979, 1997.

[2] Michael Black and Anand Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications to early vision," *International Journal of Computer Vision*, vol. 19, no. 1, pp. 57–91, 1996.

[3] N. Bose and K. Boo, "High-resolution Image Reconstruction with Multisensors," International Journal of Imaging Systems and Technology, vol 9, pp. 294–304, 1998.

[4] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Transactions on Image Processing*, vol. IP-2, no. 3, pp. 296–310, July 1993.

[5] Ch. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Transactions on Image Processing*, vol. 5, no. 3, pp. 480–492, Mar. 1996.

[6] P. Charbonnier, Laure Blanc-Féraud, Gilles Aubert, and Michel Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Transactions on Image Processing*, vol. 6, no. 2, pp. 298–311, Feb. 1997.

[7] L. D. Cohen, "Auxiliary variables and 2-step iterative algorithms in computer vision problems," *Journal of Mathematical Imaging and Vision*, vol. 6, pp. 59–83, 1996.

[8] A. H. Delanay and Y Bressler, "Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography," *IEEE Transactions on Image Processing*, vol. 7, pp. 204–221, Feb. 1998.

[9] D. Dobson and F. Santosa, "Recovery of blocky images from noisy and blurred data," *SIAM Journal on Applied Mathematics*, vol. 56, no. 4, pp. 1181–1199, Aug. 1996.

[10] R. Chan and M. Ng, "Conjugate gradient methods for Toeplitz systems," *SIAM Review*, vol 38, pp. 427–482, 1996.

[11] Donald Geman and Georges Reynolds, "Constrained restoration and recovery of discontinuities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-14, no. 3, pp. 367–383, Mar. 1992.

[12] Donald Geman and Chengda Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Transactions on Image Processing*, vol. IP-4, no. 7, pp. 932–946, July 1995.

[13] P. Kornprobst, R. Deriche, and G. Aubert, "Image sequence analysis via partial differential equations," *Journal of Mathematical Imaging and Vision*, vol. 11, no. 1, pp. 5–26, Oct. 1999.

[14] Mila Nikolova, "Local strong homogeneity of a regularized estimator," *SIAM Journal on Applied Mathematics*, vol. 61, no. 2, pp. 633–658, 2000.

[15] M. Ng, R. Chan and W. Tang, "A fast algorithm for deblurring models with Neumann boundary conditions," *SIAM J. Sci. Comput.*, vol 21 pp. 851–866, 1999.

[16] R. Tyrell Rockafellar, *Convex Analysis*, Princeton University Press, 1970.