A PROXIMAL METHOD FOR INVERSE PROBLEMS IN IMAGE PROCESSING.

Pierre Weiss † *and Laure Blanc-Féraud* ‡

† Institute for Computational Mathematics - Center for Mathematic Imaging Vision
Hong Kong Baptist University, Kowloon Tong, Hong Kong
email: pierre.armand.weiss@gmail.com
‡ Projet Ariana - CNRS/INRIA/UNSA
INRIA Sophia-Antipolis, 2004 route des Lucioles, 06902 Sophia-Antipolis - FRANCE
email: laure.blanc_feraud@sophia.inria.fr

ABSTRACT

In this paper, we present a new algorithm to solve some inverse problems coming from the field of image processing. The models we study consist in minimizing a regularizing, convex criterion under a convex and compact set. The main idea of our scheme consists in solving the underlying variational inequality with a proximal method rather than the initial convex problem. Using recent results of A. Nemirovski [13], we show that the scheme converges at least as $O\left(\frac{1}{k}\right)$ (where k is the iteration counter). This is in some sense an optimal rate of convergence. Finally, we compare this approach to some others on a problem of image cartoon+texture decomposition.

1. INTRODUCTION

Recently, the interest for convex problems in signal processing increased a lot. Several reasons might explain that trend. One of them is that new theorems (like those related to compressed sensing [4]) provide a very nice theory to support their use. Furthermore, some impressive results were obtained in image reconstruction, using combinations of harmonic and convex analysis (see e.g. [9]). Finally, it seems natural to optimize some criteria in order to carry out a given task. Unfortunately, very few optimization problems can be solved (globally) in reasonable computing times. This is not the case for convex problems: they can be considered as solvable under very mild assumptions. Indeed, the solutions of most convex problems can be obtained with an accuracy ε in less than $O\left(\frac{n^q}{\varepsilon^p}\right)$ arithmetic operations, where n is the problem's dimension and p and q are positive numbers [2].

All those reasons make convex problems very attractive in image processing. However, due to the huge dimensions involved, solving those problems fast and *accurately* is still critical and new effective methods are still required for many applications. In this paper, we provide a first order scheme which has many interesting theoretical features. One of them is its "optimality" which will be described later. The scheme, known as *extra-gradient* method, actually dates back from 1976 [10] and seems to have been ignored until now. We show its efficiency on a problem of image cartoon+texture decomposition.

1.1 Notations

The following notations will be used throughout the paper. X and Y denote subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively and $Z = X \times Y$. Those sets are endowed with the usual scalar products $\langle\cdot,\cdot\rangle_X$, $\langle\cdot,\cdot\rangle_Y$ and $\langle\cdot,\cdot\rangle_Z$ respectively. $||\cdot||_p$ denotes the standard l^p-norm on either spaces. Let A be a linear transform. Its complex conjugate (or transpose in the real case) is denoted A^*. $|||A||| = \max_{x,\ ||x||_2\leq 1} (||Ax||_2)$ denotes the highest singular value of A. Finally $\Pi_Z(z^0) = \arg\min_{z\in Z} \left(||z-z^0||_2\right)$ is the Euclidean projector on Z.

1.2 Problem statement

In this paper, we consider the following class of problems:

$$\min_{x\in X} (\Psi(x)) \tag{1}$$

where X is a convex, *compact* subset of $\mathbb{R}^n$, Ψ is a convex function which can be written under a conjugate form:

$$\Psi(x) = \max_{y\in Y} (\langle Ax, y\rangle_Y - \phi(y)) \tag{2}$$

where Y is a convex, *compact* subset of $\mathbb{R}^m$, $A : X \to Y$ is a linear transform, $\langle\cdot,\cdot\rangle_Y$ is the usual scalar product and ϕ is a closed convex function with Lipschitz gradient:

$$||\nabla\phi(y_1) - \nabla\phi(y_2)||_2 \leq L_\phi ||y_1 - y_2||_2,\ \forall (y_1,y_2) \in Y \times Y. \tag{3}$$

We further make the strong assumption that we can compute projections on X and Y exactly.

1.3 Applications

This class of problems contains a number of interesting applications. Let us give some examples:

1. If we set $A = \nabla$ (the discrete gradient), $\phi = 0$, $Y = \{y \in \mathbb{R}^m, ||y||_\infty \leq 1\}$, then $\Psi(x)$ corresponds to the discrete total variation. Giving different definitions to the set X allows to achieve tasks like denoising of Gaussian, uniform or impulse noise, compression noise reduction, texture+geometry decompositions,... A detailed review of those applications is given in [15].
2. If we change Y to an l^2 or l^1-ball, then problem (1) becomes a Tikhonov regularization or an infinite-Laplacian problem. These are interesting when the signals to be restored are smooth or have bounded derivatives.
3. If we keep the same setting as previously, only changing A to be a wavelet transform, then Ψ corresponds to a Besov semi-norm [6] which is known as a good regularizer.

4. This formalism also allows to do compressed sensing in a particular case. Let B be a frame (i.e. a linear mapping such that $c||x||_2^2 \leq ||Bx||_2^2 \leq C||x||_2^2$, $\forall x \in \mathbb{R}^n$ with $C \geq c > 0$). Then the set $X = \{x \in \mathbb{R}^n, ||Bx - y^0||_2^2 \leq \alpha\}$ is bounded and projectors can be computed using linear algebra. Thus we can solve problems of the form:

$$\min_{x,||Bx-y^0||_2^2\leq\alpha} (||x||_1) \quad \text{or} \quad \min_{x,||Bx-y^0||_2^2\leq\alpha} (||\nabla x||_1) \tag{4}$$

where y^0 is the datum to be processed. Those are a particular cases of the “compressed sensing” problem.
5. Until now we considered only the case $\phi = 0$. For instance, if we choose $\phi(y) = \frac{1}{2}||y||_2^2$ and $Y = \{y \in \mathbb{R}^m, ||y||_\infty \leq 1\}$, then Ψ becomes the Huber regularization of the l^1-norm, which is used in many applications.

1.4 Algorithm features

In this paper, we show that the *extra-gradient* method can be used successfully to solve the above cited problems. This scheme was proposed as early as 1976 in [10] and used broadly since then in different areas. Strangely, it seems that it has never been used yet in image processing, while it is considered as one of the most efficient first order methods [3]. Recently, A. Nemirovski studied this scheme from a theoretical point of view and got many interesting theoretical results including convergence rates [13]. Let us summarize some of them:

- It converges to the set of minimizers [10].
- It converges with a rate of convergence better than $\Psi(x^k) - \Psi(x^*) \leq O(1)\frac{n}{k}$ where $\{x^k\}$ are the iterates, x^* is a solution of the problem [13] and n is the problem's dimension. Furthermore, one can provide a bound on the term $O(1)$.
- Its rate of convergence is “optimal”, in the sense that no first order method can converge faster (up to a multiplicative factor) on the class of problems considered [13].
- It requires only one parameter: the number of iterations or the precision.
- At each iteration it is possible to evaluate the duality gap. Thus the scheme has a reliable stopping condition.

Those features are quite unusual in the image processing literature.

2. ALGORITHM'S DESCRIPTION

The idea of the proposed scheme consists in solving the variational inequality associated with (1) rather than tackling its minimization directly.

Let us show how (1) can be rewritten as a variational inequality. The fact that X and Y are bounded and that ϕ is convex is sufficient to ensure the following central equality (see e.g. [7, p.176]):

$$\min_{x\in X}\left(\underbrace{\max_{y\in Y}\left(\langle Ax,y\rangle_Y - \phi(y)\right)}_{\Psi(x)}\right) \tag{5}$$

$$= \max_{y\in Y}\left(\underbrace{\min_{x\in X}\left(\langle x,A^*y\rangle_X\right) - \phi(y)}_{\Psi^*(y)}\right). \tag{6}$$

Moreover, the two problems (5) and (6) have the same set of solutions. Note that the solution of (1) is not unique in general. x is what we will call a *primal* variable and y is a *dual* variable. The scheme we propose can be seen as primal-dual in the sense that both x and y are updated recursively and will converge to the set of minimizers.

Let x^* and y^* denote a primal and a dual solution respectively. The optimality condition for x^* writes (see e.g. [7, p.36]):

$$\langle A^*y, x - x^*\rangle_X \geq 0,\ \forall x \in X. \tag{7}$$

The optimality condition for y^* writes:

$$\langle Ax - \nabla\phi(y^*), y - y^*\rangle_Y \leq 0,\ \forall y \in Y. \tag{8}$$

Now, let us define the following mapping:

$$\begin{array}{rccl} F: & X\times Y & \rightarrow & \mathbb{R}^n\times\mathbb{R}^m \\ & z=(x,y) & \mapsto & \begin{bmatrix} A^*y \\ -Ax+\nabla\phi(y)\end{bmatrix}. \end{array} \tag{9}$$

In view of (7) and (8), finding a solution of (5) is equivalent to solving the following problem:

$$\text{find } z^* = (x^*,y^*) \text{ s.t. } \langle F(z^*), z - z^*\rangle_Z \geq 0,\ \forall z \in Z, \tag{10}$$

and this is exactly what we have been naming *variational inequality* associated with (1).

There exists a huge body of literature that deals with methods for solving variational inequalities, as they are more general than convex problems (see e.g. [8] for a recent reference). In this paper, we concentrate only on a simple version of the extra-gradient method. This method can be used as soon as F is a monotonous and Lipschitz operator [10].

Let us show that those conditions are satisfied. Monotonicity is simply due to the fact that ϕ is convex. Thus:

$$\langle F(z_1) - F(z_2), z_1 - z_2\rangle_Z \geq 0,\ \forall(z_1,z_2) \in Z\times Z \tag{11}$$

Moreover F is Lipschitz continuous:

$$||F(z_1) - F(z_2)||_2 \leq L||z_1 - z_2||_2, \tag{12}$$

with $L \leq \sqrt{2(|||A|||^2 + L_\phi^2)}$ and $L \leq |||A|||$ in the case $L_\phi = 0$. Indeed, denoting $z_1 = (x_1,y_1)$ and $z_2 = (x_2,y_2)$, we have:

$$\begin{aligned} & ||F(z_1) - F(z_2)||_2^2 \\ = \ & ||A^*y_1 - A^*y_2||_2^2 + ||Ax_1 - Ax_2 + \nabla\phi(y_1) - \nabla\phi(y_2)||_2^2 \\ \leq \ & |||A|||^2(||y_1 - y_2||_2^2 + 2||x_1 - x_2||_2^2) + 2L_\phi^2||y_1 - y_2||_2^2 \\ \leq \ & (2|||A|||^2 + 2L_\phi^2)||z_1 - z_2||_2^2. \end{aligned}$$

Thus, all the requirements are satisfied in order to use an extra-gradient method.

This method writes simply as follows:

Algorithm 1 Extra-gradient method

Choose a number of iterations N.
Set a point $z^0 = (x^0, y^0)$ (as close as possible to z^*).
for $k = 1$ to N **do**
$w^k = \Pi_Z(z^{k-1} - \gamma F(z^{k-1}))$
$z^k = \Pi_Z(z^{k-1} - \gamma F(w^k))$
end for

and z^k converges to the set of minimizers as long as $\gamma \leq \frac{1}{L}$ [10] where L is defined in (12). In [13], the author shows an additional result:

Proposition 1 (Convergence rate)
Let $\gamma = \frac{1}{\sqrt{2}L}$ *and* $\bar{z}^k = \frac{1}{k}\sum_{i=1}^{k} z^k$. *This sequence satisfies:*

$$\Psi(\bar{z}^k) - \Psi(z^*) \leq \frac{\sqrt{2}L\,\mathrm{diam}(Z)^2}{k}, \tag{13}$$

where $\mathrm{diam}(Z) = \max_{(z^1,z^2)\in Z\times Z} \left(||z^1 - z^2||_2\right)$.

This convergence rate is neatly sublinear and might look bad. However, we would like to point out that we are considering non differentiable and non elliptic problems, for which there can exist no better rates. Result (13) is somehow optimal: A. Nemirovski [12] shows that some instances of (1) cannot be solved faster than $O\left(\frac{1}{k}\right)$ using only the mapping F. So that any improvements of this method will only result in lowering the multiplicative constants of the convergence rates, but not their asymptotic behavior.

Finally, let us precise that this method can be modified in order to give a reliable stopping criteria. In view of equality (6), we have:

$$\forall (x,y) \in X \times Y,\ \ \Psi^*(y) \leq \Psi^*(y^*) = \Psi(x^*) \leq \Psi(x). \tag{14}$$

Thus the quantity:

$$\Delta(x,y) = \Psi(x) - \Psi^*(y) \tag{15}$$

satisfies $\Delta(x,y) \geq \Psi(x) - \Psi(x^*)$. This quantity is called *duality gap*. It is a reliable measure of convergence. It is thus interesting to modify scheme (1) as follows:

Algorithm 2 Extra-gradient method (with stopping criterion)

Choose a precision ε.
Set a point $z^0 = (x^0, y^0)$ (as close as possible to z^*).
while $\Delta(x^k, y^k) > \varepsilon$ **do**
$w^k = \Pi_Z(z^{k-1} - \gamma F(z^{k-1}))$
$z^k = \Pi_Z(z^{k-1} - \gamma F(w^k))$
end while

Note that in view of Proposition 1, it might seem interesting to change the condition in the **while** statement by $\Delta(\bar{x}^k, \bar{y}^k) > \varepsilon$. However, we observed that $\bar{z}^k$ converged less rapidly than z^k on the tested problems.

3. NUMERICAL RESULTS

In this section we provide some preliminary comparisons of the scheme's efficiency. We consider a problem of image cartoon+texture decomposition proposed initially in [11]. The main idea is to decompose the image into a cartoon part which has a low total variation (i.e. has little oscillations) and a texture part which has a low "G-norm". The G-norm of an oscillatory function is low. This model should thus separate piecewise smooth and oscillatory patterns.

3.1 The problem considered

Let x^0 denote the image to be decomposed. The problem of decomposition consists in solving:

$$\min_{x, ||x||_\infty \leq \alpha} \left(||\nabla(x^0 - \mathrm{div}(x))||_1\right), \tag{16}$$

where ∇ and div are the discrete gradient and divergence. This model was proposed in [11] and further investigated in [1]. Reformulation (16) was proposed in [15]. We look for a decomposition of the image x^0 into two components u and v. u is the cartoon component defined as $u = x^0 - \mathrm{div}(x^*)$ where x^* is a solution of (16). v is the texture component defined as $v = \mathrm{div}(x^*)$. We thus have $x^0 = u + v$. We refer the reader to [11, 1, 15] for more insight on those models.

This problem can be stated in our formalism. It suffices to choose:

- $X = \{x \in \mathbb{R}^n, ||x||_\infty \leq \alpha\}$ (n is twice the number of pixels).
- $A = -\nabla\mathrm{div}$. Thus $m = n$ and using the discretization of ∇ and div proposed in [5], we get $|||A||| = 8$.
- $\phi(y) = -\langle \nabla x^0, y\rangle_Y$ (thus $L_\phi = 0$).
- $Y = \{y \in \mathbb{R}^m, ||y||_\infty \leq 1\}$.

With those choices, we have $\Psi(x) = ||\nabla(x^0 - \mathrm{div}(x))||_1$ and $\Psi^*(y) = -\alpha||\nabla\mathrm{div}(y)||_1 + \langle \nabla x^0, y\rangle_Y$. The theoretical worst case convergence rate can be computed and is:

$$\Psi(\bar{z}^k) - \Psi(z^*) \leq \frac{8n(\alpha^2+1)}{k}. \tag{17}$$

Note that *it grows linearly* with the problem's dimension n. This is the best expectable behavior. This result indicates that whatever the size of x^0, the same number of iterations will be required to get a desired accuracy. It is indeed reasonable to require a precision which depends linearly on n (if we double the number of pixels, it is natural to double the required accuracy).

3.2 Methods chosen for comparisons

In this paper we will consider and compare only first order approaches. They are known to have bad asymptotic convergence rates, but their low cost iterations, their low memory requirements and their good behavior at the origin makes them very attractive to get approximate solutions fast (which is often sufficient in image processing). Second order methods like second order cone programming are accurate, but seem to be unusable for large images. Until now, the combinatorial approaches proposed in the literature seem to be adapted only to narrow classes of imaging problems.

In this paper, we compare 4 first order methods. We will not describe them due to space restriction.

- The first one is the proposed approach. We tried different choices of step-sizes (including variable step-sizes) but for this problem, the one presented is seemingly the best.
- The second one is the one proposed by J-F. Aujol et al in [1].
- The third one is a projected subgradient descent with optimal step. It is unusable in practice, because the optimum in (16) must be known.
- The fourth one is the one proposed in [15]. It is based on results by Y. Nesterov [14]. This scheme also has an "optimal" $O\left(\frac{1}{k}\right)$ rate of convergence, but it is somehow more difficult to use because the number of iterations must be chosen depending on a regularization parameter.

For comparisons, we evaluate the duality gap $\Delta(x^k, y^k)$ which is the only reliable criterion available. It can be computed only in the proposed approach and in the approach [15]. We also evaluate the cost function for the subgradient descent. Unfortunately, those criteria cannot be evaluated in the approach of [1]. Indeed this approach does not ensure that the constraint $x \in X$ is satisfied. We will only provide comparisons on the resulting images (see Figure 2).

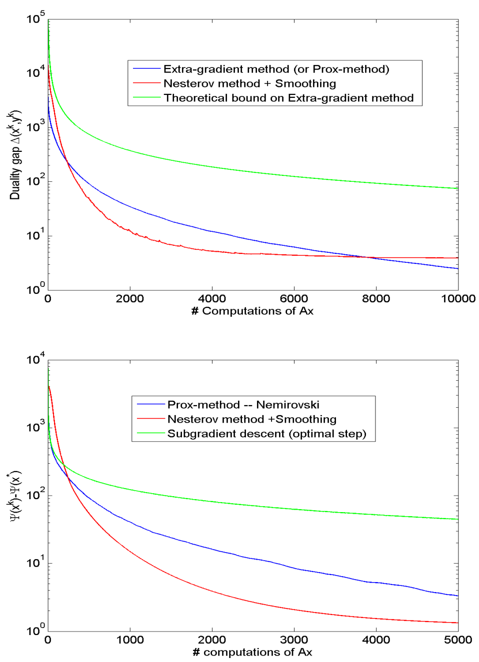

Figure 1: Comparisons of some schemes efficiency.

According to these curves, we see that the prox-method proposed in this paper and the one proposed in [15] are the most efficient. The number of iterations required to obtain satisfactory solutions is quite high (of order 2000). We could not obtain satisfactory solutions using the approach described in [1] nor with the subgradient descent. Those methods give rough approximations very fast, but do not seem to allow to get precise solutions. Figure (2) illustrates this fact. Multiple experiments led us to the conclusion that the proposed approach is completely reliable and easy to use (the user only has to provide the precision which can be made dimension and data independent). We think that those results are very encouraging. Further work will include the use of more advanced techniques. To finish, we give an example of image decomposition using another model. It simply consists in replacing $||x||_\infty$ by $||x||_1$ in (16). This new model allows the extraction of oscillations with larger amplitude (see e.g. Figure (2)).

REFERENCES

[1] J.-F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle. Image Decomposition into a Bounded Variation Component and an Oscillating Component. *Journal of Mathematical Imaging and Vision*, 22:71–88, 2005.

[2] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization*. SIAM, MPS-SIAM Series on optimization, 2001.

[3] H. Binsheng, X. Yuan, and J. J.Z. Zhang. Comparison of two kinds of prediction-correction methods for monotone variational inequalities. *Computational Optimization and Applications*, 27, 2004.

[4] E. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Inf. Theory*, 2006.

[5] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.*, 20(1-2):89–97, 2004.

[6] Antonin Chambolle, Ronald Devore, Nam yong Lee, and Bradley J. Lucier. Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Trans. Image Processing*, 7:319–335, 1998.

[7] I. Ekeland and R. Temam. *Convex analysis and variational problems. Studies in Mathematics and its Applications*. American Elsevier Publishing Company, 1976.

[8] F. Facchinei and J-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer, 2003.

[9] M.J. Fadili, J-L. Starck, and F. Murtagh. Inpainting and zooming using sparse representations. *The Computer Journal*, 2007.

[10] G. Korpelevich. The extragradient method for finding saddle points and other problems. *Ekonomica i Matematicheskie Metody*, 12:747,756, 1976.

[11] Y. Meyer. Oscillating patterns in image processing and in some nonlinear evolution equations. *The Fifteenth Dean Jaqueline B. Lewis Memorial Lectures*, 2001.

[12] A. Nemirovski. Information-based complexity of linear operator equations. *Journal of Complexity*, 8:153–175, 1992.

[13] A. Nemirovski. Prox-method with rate of convergence o(1/t) for variational inequalities with lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, pages 229–251, 2004.

[14] Y. Nesterov. Smooth minimization of non-smooth functions. *Math. Program.*, 103(1):127–152, 2005.

[15] P. Weiss, L. Blanc-Féraud, and G. Aubert. Efficient schemes for total variation minimization under constraints in image processing. *SIAM journal on Scientific Computing*, 31(3):2047–2080, 2009.

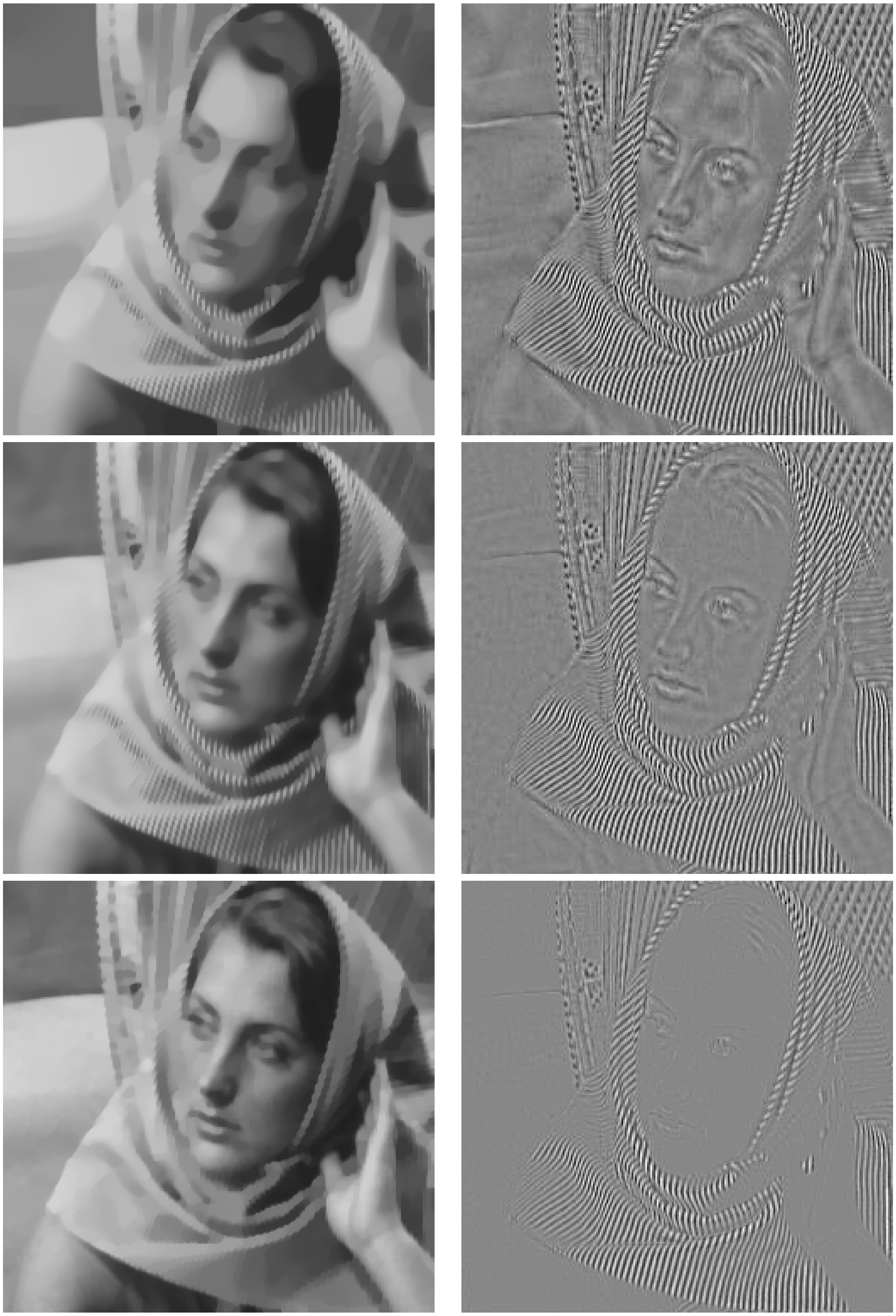

Figure 2: Left: Cartoon part – Right: texture part. From top to bottom: solution R^1 using our algorithm (10 minutes) – solution R^2 using the approach in [1] (10 minutes) – result replacing $||x||_\infty$ by $||x||_1$ in (16). (Note: the total variation of R^2 is higher than that of R^1 and the texture component of R^2 does not satisfy the constraint. R^1 is thus more accurate than R^2.)