Total Variation Minimization with Separable Sensing Operator

Serge L. Shishkin, Hongcheng Wang, and Gregory S. Hagen

United Technologies Research Center, 411 Silver Ln, MS 129-15, East Hartford, CT 06108, USA

Abstract. Compressed Imaging is the theory that studies the problem of image recovery from an under-determined system of linear measurements. One of the most popular methods in this field is Total Variation (TV) Minimization, known for accuracy and computational efficiency. This paper applies a recently developed Separable Sensing Operator approach to TV Minimization, using the Split Bregman framework as the optimization approach. The internal cycle of the algorithm is performed by efficiently solving coupled Sylvester equations rather than by an iterative optimization procedure as it is done conventionally. Such an approach requires less computer memory and computational time than any other algorithm published to date. Numerical simulations show the improved — by an order of magnitude or more — time vs. image quality compared to two conventional algorithms.

1 Introduction

Compressed Imaging (CI) methods perform image recovery from a seemingly insufficient set of measurements [1]. This is a fast growing field with numerous applications in optical imaging, medical MRI, DNA biosensing, seismic exploration, etc. One promising but still open application for CI is the problem of on-line processing of "compressed" video sequences. The two main obstacles are that, first, even the fastest recovery algorithms are not fast enough for on-line processing of the video; and second, the memory requirements of the algorithms are too high for the relatively small processors typical for video devices.

A big step toward overcoming these limitations was done recently in [2,3] with the suggestion to use only separable — i.e. presentable in the form $A = R \otimes G$ — Sensing matrices (SM), where "$\otimes$" stands for the Kronecker product. In such a case, for any image F the equality holds:

$$A\,\mathbf{vect}(F) = \mathbf{vect}(RFG)\,, \tag{1}$$

where "**vect**" denotes a vectorization operator. For M measurements performed on the image of the size $n \times n$, the conventional approach — left hand side of (1) — demands $O(Mn^2)$ operations and computer memory while the right hand side of equation (1) requires $O(M^{1/2}n)$ operations and computer memory. This simplification fits into any CI method and yields significant reduction of computational time and memory almost independently of the method schematics.

A. Elmoataz et al. (Eds.): ICISP 2010, LNCS 6134, pp. 86–93, 2010.

However, the potential advantages of separable SMs are not limited to the simple consideration above. As will be shown in this paper, some CI algorithms can be customized according to specific qualities of the SM and this leads to further speed-up of image recovery.

In this paper, we consider an approach to CI recovery via Total Variation (TV) minimization. TV minimization was first introduced for image denoising in [4] and later was very successfully used for CI recovery in [5], [6] and other papers. A theoretical justification of the approach is presented in [7].

In TV framework, the image is recovered by solving the one of two problems:

$$\text{minimize } \|\nabla_x F\|_1 + \|\nabla_y F\|_1 \qquad \text{s.t. } \|A\,\mathbf{vect}(F) - y\| \le \varepsilon\,, \tag{2}$$

$$\text{minimize } \|\nabla_x F\|_1 + \|\nabla_y F\|_1 + \frac{\mu}{2}\|A\,\mathbf{vect}(F) - y\|_2^2\,. \tag{3}$$

where ε and μ are small constants, A is the $M \times N$ Sensing matrix, y is M-vector of measurements, $M < N$.

A variant of this approach called "Anisotropic TV minimization" involves solving one of the slightly modified problems, respectively:

$$\text{minimize } \sum_{i,j} \sqrt{|\nabla_x F_{i,j}|^2 + |\nabla_y F_{i,j}|^2} \qquad \text{s.t. } \|A\,\mathbf{vect}(F) - y\| \le \varepsilon\,, \tag{4}$$

$$\text{minimize } \sum_{i,j} \sqrt{|\nabla_x F_{i,j}|^2 + |\nabla_y F_{i,j}|^2} + \frac{\mu}{2}\|A\,\mathbf{vect}(F) - y\|_2^2\,. \tag{5}$$

Since the problems (2) and (4) can be solved by reduction to the form (3) and (5) respectively, either with iterative adjustment of the parameter μ or by introducing an auxiliary variable as in [9], we will consider only (3) and (5). With the assumption that the SM has separable structure $A = R \otimes G$, these two problems take the following form, respectively:

$$\text{minimize } \|\nabla_x F\|_1 + \|\nabla_y F\|_1 + \frac{\mu}{2}\|RFG - Y\|_2^2\,, \tag{6}$$

$$\text{minimize } \sum_{i,j} \sqrt{|\nabla_x F_{i,j}|^2 + |\nabla_y F_{i,j}|^2} + \frac{\mu}{2}\|RFG - Y\|_2^2\,, \tag{7}$$

where the matrices R and G of appropriate dimensions are fixed *a priori* and $y = \mathbf{vect}(Y)$. To enable CI recovery, the Sensing matrix $A = R \otimes G$ must satisfy certain conditions. We do not discuss them here; a reader is referred to [1], [3].

We propose that the problems (6) and (7) should be solved using a Split-Bregman approach [8], [9], with the important alternation: the internal cycle of minimization is replaced by solving two Sylvester equations. Such a modification is possible due to separable structure of the SM. This significantly reduces the computational load of the method and improves the accuracy.

The structure of this paper is as follows: Section 2 presents the fast modification of the Split-Bregman method to solve the problems (6) and (7). Results of numerical simulations are presented and discussed in Section 3. Conclusions are formulated in Section 4.

2 TV Minimization Algorithm

Let us start with problem (6). According to the Split-Bregman approach [8], [9], we introduce new auxiliary variables $V^k, W_k, D_x^k, B_x^k, D_y^k, B_y^k$ and replace the unconstrained minimization problem (6) by the equivalent problem

$$\text{minimize } \|D_x\|_1 + \|D_y\|_1 + \frac{\mu}{2}\|RFG - Y\|_2^2 \quad \text{such that } D_x = \nabla_x V,\ D_y = \nabla_y V,\ V = F\,. \quad (8)$$

Here we see the main idea of the proposed algorithm: "splitting" the main variable into two — F and V — so that the former is responsible for the term $\|RFG - Y\|_2^2$ while the latter is being l_1 regularized by TV minimization. The condition $F = V$ guarantees that this is, in fact, one variable; however such a split allows us to decompose the most computationally expensive step of the algorithm into two much simpler steps.

This approach looks similar to the version of the Split Bregman method proposed in [9] for solving problem (2) with $\varepsilon = 0$. However, the method presented here is developed for a slightly different problem (3) and additional differences will be noted in the development. The method presented here can easily be modified along the lines of [9] to address problems (2) or (4).

The solution of (8) can be found by unconstrained optimization:

$$\underset{F,V,D_x,D_y}{\text{minimize}}\ \|D_x\|_1 + \|D_y\|_1 + \frac{\lambda}{2}\|\nabla_x V - D_x\|_2^2 + \frac{\lambda}{2}\|\nabla_y V - D_y\|_2^2 + \frac{\nu}{2}\|V - F\|_2^2 + \frac{\mu}{2}\|RFG - Y\|_2^2 \quad (9)$$

with appropriate choice of the coefficients λ, ν, μ. Consistently applying the Split-Bregman procedure, we replace (9) by the sequence of minimization problems

$$\text{minimize } \|D_x^k\|_1 + \|D_y^k\|_1 + \frac{\lambda}{2}\|\nabla_x V^k - D_x^k + B_x^k\|_2^2 + \frac{\lambda}{2}\|\nabla_y V^k - D_y^k + B_y^k\|_2^2 + \frac{\nu}{2}\|V^k + W^k - F^k\|_2^2 + \frac{\mu}{2}\|RFG - Y\|_2^2\,, \quad (10)$$

where auxiliary variables B_x^k, B_y^k, and W^k are updated as follows:

$$W^{k+1} = V^{k+1} + W^k - F^{k+1}\,, \quad (11)$$

$$B_x^{k+1} = \nabla_x V^{k+1} + B_x^k - D_x^{k+1}\,, \qquad B_y^{k+1} = \nabla_y V^{k+1} + B_y^k - D_y^{k+1}\,. \quad (12)$$

The minimization (10) can be split into the following subtasks:

$$F^{k+1} = \arg\min_F \frac{\mu}{2}\|RFG - Y\|_2^2 + \frac{\nu}{2}\|F - V^k - W^k\|_2^2\,, \quad (13)$$

$$V^{k+1} = \arg\min_V \frac{\lambda}{2}\|D_x^k - \nabla_x V - B_x^k\|_2^2 + \frac{\lambda}{2}\|D_y^k - \nabla_y V - B_y^k\|_2^2 + \frac{\nu}{2}\|V - F^{k+1} + W^k\|_2^2\,, \quad (14)$$

$$D_x^{k+1} = \arg\min_D \|D\|_1 + \frac{\lambda}{2}\|D - \nabla_x V^{k+1} - B_x^k\|_2^2\,, \quad (15)$$

$$D_y^{k+1} = \arg\min_D \|D\|_1 + \frac{\lambda}{2}\|D - \nabla_y V^{k+1} - B_y^k\|_2^2\,. \quad (16)$$

Let us consider these subtasks one-by-one.

Differentiating (13), we obtain the necessary condition on F^{k+1}:

$$\mu(R^T R)F^{k+1}(GG^T) + \nu F^{k+1} = \mu R^T Y G^T + \nu V^k + \nu W^k \tag{17}$$

This is a classical Sylvester equation. To solve it, let us compute the eigen-decomposition of the matrices $(R^T R)$ and (GG^T), respectively:

$$R^T R = U_R^T L_R U_R, \qquad GG^T = U_G^T L_G U_G,$$

where U_R, U_G are orthogonal matrices, and L_R, L_G are diagonal ones. Denote

$$\widetilde{F^k} = U_R F^k U_G^T, \quad \widetilde{V^k} = U_R V^k U_G^T, \quad \widetilde{W^k} = U_R W^k U_G^T, \quad \widetilde{Y} = U_R R^T Y G^T U_G^T.$$

Multiplying (17) by U_R^T from left and by U_G from right, we obtain

$$\mu L_R \widetilde{F^{k+1}} L_G + \nu \widetilde{F^{k+1}} = \mu \widetilde{Y} + \nu \widetilde{V^k} + \nu \widetilde{W^k} \tag{18}$$

which is easily solvable on entry-by-entry basis, yielding:

$$\widetilde{F_{i,j}^{k+1}} = (\mu \widetilde{Y} + \nu \widetilde{V^k} + \nu \widetilde{W^k})_{i,j} / \big(\nu + \mu (L_R)_{ii} (L_G)_{jj}\big) \qquad i,j = 1,\dots,n \tag{19}$$

Now, let us denote by $\mathfrak{D}_x, \mathfrak{D}_y$ the matrices of differential operators ∇_x, ∇_y: $\nabla_x F = \mathfrak{D}_x F, \quad \nabla_y F = F \mathfrak{D}_y$ and compute the eigen-decomposition of the matrices $(\mathfrak{D}_x^T \mathfrak{D}_x)$ and $(\mathfrak{D}_y \mathfrak{D}_y^T)$, respectively:

$$\mathfrak{D}_x^T \mathfrak{D}_x = U_x^T L_x U_x, \qquad \mathfrak{D}_y \mathfrak{D}_y^T = U_y^T L_y U_y.$$

Following the same steps as in the analysis of (13), we obtain the solution of the problem (14), which is given by

$$\widehat{V_{i,j}^{k+1}} = \Big(U_x \big(\lambda \mathfrak{D}_x^T (D_x^k - B_x^k) + \lambda (D_y^k - B_y^k) \mathfrak{D}_y^T + \nu F^{k+1} - \nu W^k \big) U_y^T \Big)_{i,j} / l_{i,j} \tag{20}$$

where $l_{i,j} = \nu + \lambda (L_x)_{ii} + \lambda (L_y)_{jj}, \quad i,j = 1,\dots,n$. Note that U_x, U_y are Fourier matrices, and thus all matrix multiplications in (20) can be performed very efficiently by Fast Fourier transforms and by differentiation w.r.t. x, y.

In the "classical" Split Bregman scheme, the subproblems (13), (14) would be presented as one optimization problem, involving all the terms now distributed between them. However, solving such a problem would not be straightforward: either some iterative minimization algorithm has to be employed to find an accurate solution of the subproblem at each step (as in [8]), or some severe limitations (incompatible with TV minimization) on all the operators must be imposed (as in [9]). Our method, as was demonstrated above, allows the application of the Silvester method for solving such problems — thus it is much faster than any existing optimization procedure.

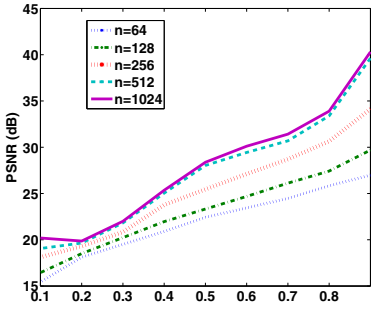

Fig. 1. Image PSNR vs. sampling rate r (the number of measurements $M = r \cdot n \cdot n$)

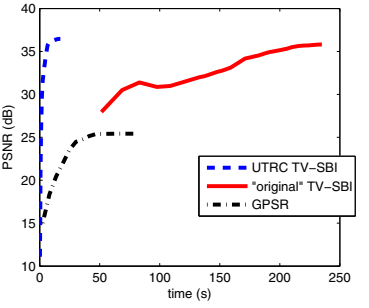

Fig. 2. Comparison of image quality vs. computational time for three algorithms

Now, let us consider the problems (15), (16). Since there is no coupling between the elements of D in either subproblem, we can solve them on an "element-wise" basis. It is straightforward to obtain the formula for the solution ([8], [9]):

$$(D_x^{k+1})_{i,j} = \mathbf{shrink}\big((\nabla_x V^{k+1} + B_x^k)_{i,j}, 1/\lambda\big), \qquad i,j = 1,\dots,n \quad (21)$$

$$(D_y^{k+1})_{i,j} = \mathbf{shrink}\big((\nabla_y V^{k+1} + B_y^k)_{i,j}, 1/\lambda\big), \qquad i,j = 1,\dots,n \quad (22)$$

$$\mathbf{shrink}(d,\gamma) = \operatorname{sign}(d)\max(|d| - \gamma, 0), \qquad d,\gamma \in \mathbb{R}$$

Assembling (11), (12)and (19)–(22) we obtain the formal description of the algorithm. Note that the matrices $U_R, U_G, U_x, U_y, L_R, L_G, L_x, L_y$ are independent of the particular image and can be computed in advance.

Solving problem (7) is quite similar to (6); the only difference is that the subproblems (15), (16) have to be replaced by the subproblem

$$[D_x^{k+1}, D_y^{k+1}] = \arg\min_{D_x, D_y} \sum_{i,j} \sqrt{(D_x)_{i,j}^2 + (D_y)_{i,j}^2} + \frac{\lambda}{2}\|D_x - \nabla_x V^{k+1} - B_x^k\|_2^2 + \frac{\lambda}{2}\|D_y - \nabla_y V^{k+1} - B_y^k\|_2^2 . \quad (23)$$

The solution of (23) can be expressed by

$$(D_x^{k+1})_{i,j} = \mathbf{shrk}\left((\nabla_x V^{k+1} + B_x^k)_{i,j}, (\nabla_y V^{k+1} + B_y^k)_{i,j}, 1/\lambda\right), \quad (24)$$

$$(D_y^{k+1})_{i,j} = \mathbf{shrk}\left((\nabla_y V^{k+1} + B_y^k)_{i,j}, (\nabla_x V^{k+1} + B_x^k)_{i,j}, 1/\lambda\right), \quad (25)$$

$$\mathbf{shrk}(d_1, d_2, \gamma) = d_1 (d_1^2 + d_2^2)^{-1/2} \max\left((d_1^2 + d_2^2)^{-1/2} - \gamma, 0\right), \qquad d_1, d_2, \gamma \in \mathbb{R}$$

Thus the algorithm for solving the anisotropic problem (7) is defined by (11), (12), (19), (20), (24), (25).

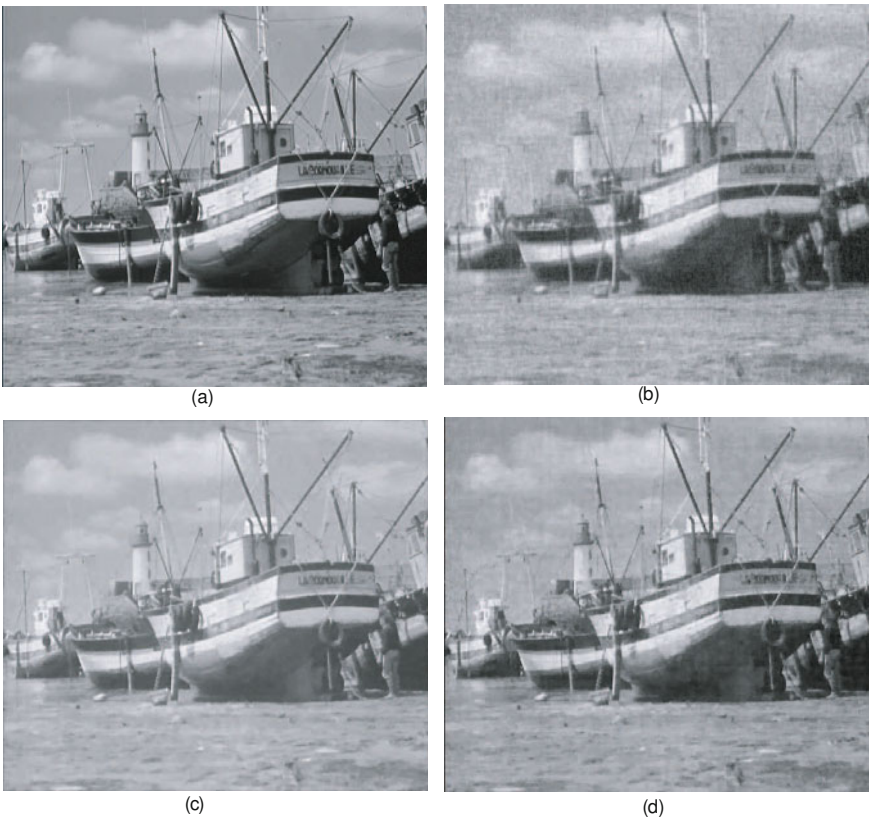

Fig. 3. Algorithm performance comparison. The sampling rate is $r = 0.3$, the image resolution is 512×512. (a) Original image. (b) Reconstruction by GPSR, PSNR = 25.3dB, time=29.9 seconds. (c) Reconstruction by original TV-SBI with separable matrix and PCG solver, PSNR = 30.1dB, time=134.9 seconds. (d) Reconstruction by using our "UTRC TV-SBI" algorithm, PSNR = 29.3dB, time=16.9 seconds.

3 Numerical Experiments

We have performed a set of numerical experiments to test the effectiveness of our proposed algorithm. The parameters of the method were fixed as follows: $\mu = 1, \lambda = 50, \nu = 5$. All the experiments were run using MATLAB code on a standard PC with an Intel Duo Core 2.66GHz CPU and 3GB of RAM.

We first tested the image quality (measured using PSNR) with respect to the CI sampling rate. The number of algorithm iterations is kept constant and thus the computational time varies just slightly. Typical results are shown in Fig. 1. As expected, as the number of measurements increases, the reconstructed image quality gets better. At the same time, the PSNR gets better as image resolution increases with the same sampling rate. This is because the relative image sparsity with respect to image resolution is smaller for larger images. Experiments on different images generated similar results.

Fig. 4. An example image of 16MP reconstructed by our algorithm

We then compared the performance of our proposed algorithm (named "UTRC TV-SBI") with the original TV-SBI algorithm with the PCG (Preconditioned Conjugate Gradient) solver [8]. This method was sped up by using separable sensing matrices as proposed in [2]. Also, we used for comparison one of the most efficient algorithms in the literature, — GPSR (Gradient Projection for Sparse Reconstruction) [10]. We did not compare the original TV-SBI algorithm [8] without separable sensing matrices because the computer runs out of memory even for moderate image sizes. For the same reason, we had to use GPSR with a sparse Sensing Matrix, as a full one would require 20Gb for storage. Fig. 2 presents the image quality vs. computational time results for the three algorithms. We used a sampling rate of $r = 0.3$. Each line in the plot was generated based on different convergence criteria of the algorithms. As the number of iterations increases, the final image quality increases and saturates at some PSNR. The line corresponding to "original" TV-SBI starts so high because the very first iteration of the method takes significant time due to slow convergence of PCG; as the reward, though, it produces the result with high PSNR. It can be easily seen that, with comparable image quality, our algorithm is by up to an order of magnitude faster than GSPR and up to two orders of magnitude faster then original TV-SBI algorithm with separable sensing matrices.

The final results of image reconstruction by each of the three algorithms are shown in Fig. 3. The image resolution is 512×512. The proposed algorithm yields image quality comparable with the original TV-SBI with separable matrices, but much better image quality than GPSR. Besides, the proposed algorithm is several times more efficient than the original TV-SBI and almost twice as efficient as GSPR. Experiments with different sampling rate generated similar results.

Finally, we have used our algorithm to recover the images as large as 16MPixel. Fig. 4 shows an example of such reconstruction with compression rate 20%, which took 1.5 hours the same computer platform described earlier in this paper. The sampling rate is $r = 0.3$. The result image has PSNR of 31.7dB.

4 Conclusions

This paper presents a new algorithm for Compressive Imaging built upon the approaches of Total Variation minimization and a Separable Sensing Matrix. Rather than mechanically combining these two ideas, the algorithm welds them together in a specific and computationally efficient way that allows speed-up of the image recovery by almost an order of magnitude. Numerical experiments show significant computational advantage of the proposed algorithm over such efficient methods as Split Bregman ([8], [2]) and GSPR ([10]).

Our algorithm also has the advantage of memory efficiency because it uses separable sensing matrices. With a standard PC, our algorithm can reconstruct up to 16 Mega-Pixel images. To the best of our knowledge, this is the largest image reported in the literature to date using a typical PC.

The authors thankfully acknowledge help and support kindly given to them by Dr. Alan Finn, Dr. Yiqing Lin and Dr. Antonio Vincitore at the United Technologies Research Center.

References

1. Donoho, D.: Compressed sensing. IEEE Trans. Inform. Theory 52(4), 1289–1306 (2006)
2. Rivenson, Y., Stern, A.: Compressed imaging with a separable sensing operator. IEEE Signal Processing Let. 16(6), 449–452 (2009)
3. Duarte, M., Baraniuk, R.: Kronecker Compressive Sensing. Submitted to IEEE Trans. Image Proc., 30 (2009)
4. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D 60, 259–268 (1992)
5. Osher, S., Burger, M., Goldfarb, D., Xu, J., Yin, W.: An iterative regularization method for total variation-based image restoration. Multiscale Modeling and Simulation 4, 460–489 (2005)
6. Ma, S., Yin, W., Zhang, Y., Chakraborty, A.: An efficient algorithm for compressed MR imaging using total variation and wavelets. In: IEEE Conf. on Computer Vision and Pattern Recognition 2008, pp. 1–8 (2008)
7. Han, W., Yu, H., Wang, G.: A General Total Variation Minimization Theorem for Compressed Sensing Based Interior Tomography. Int. J. of Biomedical Imaging 2009, Article ID 125871, 3 pages (2009), doi:10.1155/2009/125871
8. Goldstein, T., Osher, S.: The Split Bregman Method for L1 Regularized Problems. UCLA CAM Report, p. 21 (2008)
9. Plonka, G., Ma, J.: Curvelet-Wavelet Regularized Split Bregman Iteration for Compressed Sensing (2009) (preprint)
10. Figueiredo, M.A.T., Nowak, R.D., Wright, S.J.: Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. IEEE Journal of Selected Topics in Signal Processing: Special Issue on Convex Optimization Methods for Signal Processing 1(4), 586–598 (2007)