SPARSE IMAGE RESTORATION USING ITERATED LINEAR EXPANSION OF THRESHOLDS

Hanjie Pan , Thierry Blu

Electronic Engineering
The Chinese University of Hong Kong, Hong Kong
Emails: {hjpan, tblu}@ee.cuhk.edu.hk

ABSTRACT

We focus on image restoration that consists in regularizing a quadratic data-fidelity term with the standard ℓ_1 sparse-enforcing norm. We propose a novel algorithmic approach to solve this optimization problem. Our idea amounts to approximating the result of the restoration as a linear sum of basic thresholds (e.g. soft-thresholds) weighted by unknown coefficients. The few coefficients of this expansion are obtained by minimizing the equivalent low-dimensional ℓ_1-norm regularized objective function, which can be solved efficiently with standard convex optimization techniques, e.g. *iterative reweighted least square* (IRLS). By iterating this process, we claim that we reach the global minimum of the objective function. Experimentally we discover that very few iterations are required before we reach the convergence.

***Index Terms*—** Image deconvolution, Iterative Shrinkage Threshold (IST), sparsity, thresholding, Linear Expansion of Thresholds (LET)

1. INTRODUCTION

1.1. Problem formulation

Consider the standard image restoration problem: given the degraded measurements $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, find a good estimation of the original signal $\mathbf{x}$. Here $\mathbf{H}$ models certain linear transformations between the original image $\mathbf{x}$ and measurements $\mathbf{y}$. Specifically, $\mathbf{H}$ can be the convolution matrix for deblurring problems, Radon transform in tomography reconstructions, or missing partial pixels for inpainting problems, etc. And $\mathbf{n}$ is the additive noise whose energy is known or can be robustly estimated with median filtering [1]. In many cases, $\mathbf{H}$ is ill-conditioned, which precludes the possibility to apply inverse filtering to the measurements. One way to overcome ill-conditioning difficulty is to consider the minimization of the following objective function:

$$J(\mathbf{c}) = \|\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{c}\|_2^2 + \lambda\|\mathbf{c}\|_1 \quad (1)$$

This work was supported in part by a grant from RGC (No.: CUHK410110) of HK SAR.

where $\mathbf{x} = \mathbf{W}\mathbf{c}$. Here $\|\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{c}\|_2^2$ is a data-fitting term and $\|\mathbf{c}\|_1 = \sum_i |c_i|$ is a ℓ_1-norm regularization. And $\mathbf{W}$ is some transformation that maps the signal to the transformation domain, e.g. wavelet. The underlying reasoning to use ℓ_1-norm regularization is that natural images often have sparse representations in some transformation domains, typically wavelet [2, 3, 4, 5, 6], or curvelet [7] and ℓ_1 minimization tends to enforce sparsity of the reconstructed signal. Image restoration is known to be efficient when images are processed in transformed domains where natural images have sparse representations. Specifically, in wavelet-based image restoration algorithms, $\mathbf{W}$ is the wavelet synthesis matrix and $\mathbf{c}$ is the corresponding wavelet coefficients.

1.2. Previous algorithms

The optimization problem (1) is closely related to *basis pursuit* [8] criterion and *least absolute shrinkage and selection operator* [9] (LASSO), which has been known to statistics community for decades. Computational complexity in most image processing problems prevents the usefulness of standard convex optimization algorithms, such as interior point methods. Other approaches that can readily handle large scale problems can be adopted, including gradient methods [10, 11], fix-point continuation methods [12], gradient projection methods [5], and iterative shrinkage threshold (IST) methods [6, 3].

IST is probably one of the most popular methods because of its simplicity. It requires one matrix multiplication of $\mathbf{H}$ and $\mathbf{H}^{\mathrm{T}}$ followed by the shrinkage step:

$$\mathbf{c}^{(n+1)} = \theta_{\lambda\tau/2}\Big(\mathbf{c}^{(n)} - \tau\mathbf{W}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\big(\mathbf{H}\mathbf{W}\mathbf{c}^{(n)} - \mathbf{y}\big)\Big) \quad (2)$$

where $\theta_\alpha(t) = \mathrm{sign}(t)\max(|t| - \alpha, 0)$ is the soft-threshold function and τ is the IST step size. It can be derived with *expectation maximization* (EM) [2] or a more generalized *majorization minimization* (MM) approach [4]. However, it is recognized that IST is a slow-converging algorithm [13, 14]. Recent efforts to speed up IST led to the emergence of algorithms like TwIST [15], FISTA [13], and NESTA [10, 11]. They all can be categorized as gradient-based methods that take into account previous two or more iterates results.

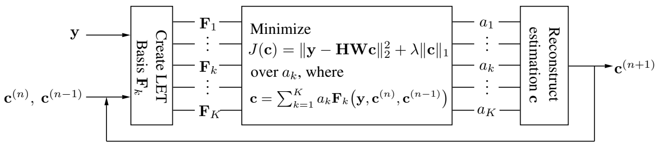

Fig. 1. Schematical view of LET deconvolution. LET coefficients are obtained by minimizaing the objective function value (1). The reconstructed signal is then feedback to the next iterate.

Another class of algorithms, which is related to *iterated reweighted least square* (IRLS), has attracted significant amount of attentions, e.g. iterative reweighted shrinkage [4], which exhibits much faster convergence rate compared than the IST algorithm.

2. ITERATIVE LET RESTORATION

From our previous experience in image denoising, we know that it is possible to have a good approximation of the original signal by decomposing the denoising process into a linear combination of elementary processes (or linear expansion of thresholds, i.e. LET) [16, 17]. Here we adopt the same principle by representing the wavelet coefficients $\mathbf{c}$ that minimizes (1) with a set of basis thresholding functions weighted by unknown coefficients:

$$\mathbf{c} = \mathbf{F}(\mathbf{y}) = \sum_{k=1}^{K} a_k \mathbf{F}_k(\mathbf{y})$$

where $\mathbf{F}_k(\mathbf{y})$ for $k = 1, \dots, K$ are LET bases and a_k's are coefficients to be determined. Therefore, our goal is to find the optimal function (or more specifically the a_k's for a given set of $\mathbf{F}_k(\mathbf{y})$) that maps the measurements $\mathbf{y}$ to the reconstructed signal $\mathbf{c}$ *iteratively*. The optimality is in the sense that the reconstructed signal gives minimal objective function values in (1):

$$\mathbf{c}^{(n+1)} = \sum_{k=1}^{K} a_k \mathbf{F}_k\big(\mathbf{y}, \mathbf{c}^{(n)}, \mathbf{c}^{(n-1)}\big) \tag{3}$$

where $a_k = \arg\min_{a_k} J\Big(\sum_{k=1}^K a_k \mathbf{F}_k\big(\mathbf{y}, \mathbf{c}^{(n)}, \mathbf{c}^{(n-1)}\big)\Big)$. At each iteration, the LET bases depend on both the measurements $\mathbf{y}$, current and previous iterate $\mathbf{c}^{(n)}$, $\mathbf{c}^{(n-1)}$ (Fig. 1). We include the previous iterate for faster convergence, as in FISTA [13] and TwIST [15]. If we substitute the LET representation (3) of the reconstructed signal $\mathbf{c}$ into (1), the resultant objective function is till convex in a_k's but with much smaller dimension. The minimization can be easily achieved with standard convex optimization methods and leads to $\hat{\mathbf{c}} = \mathbf{c}^{(\infty)}$. We can then reconstruct the solution $\hat{\mathbf{x}} = \mathbf{W}\hat{\mathbf{c}}$.

2.1. LET bases

In principle, we have the freedom to include arbitrary LET bases in (3). However, for simplicity (and also for the sake of convergence: see below), we only focus on one specific set of LET expansions, which follows naturally from the IST algorithm. A good choice of LET restoration bases $\mathbf{F}_k$ are of the form:

$$\Theta_1 = \mathbf{c}^{(n-1)}$$
$$\Theta_2 = \mathbf{c}^{(n)}$$
$$\Theta_3 = \theta_{\lambda\tau/2}\Big(\mathbf{c}^{(n)} - \tau \mathbf{W}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}(\mathbf{HW}\mathbf{c}^{(n)} - y)\Big) \tag{4}$$
$$\Theta_4 = (\mathbf{W}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{HW} + \mu\mathbf{I})^{-1}\big(\mathbf{c}^{(n)} - \Theta_3\big) \tag{5}$$

Notice that the third basis (4) is exactly the same as the shrinkage step in IST, where λ is the regularization weight and τ is the step-size. The fourth basis (5) can be treated as the result obtained with optimization techniques that involves the regularized Hessian of the data-fitting term in $J(\mathbf{c})$. Here instead of using the exact gradient, $\mathbf{c} - \Theta_3$ is used, which behaves similarly to a gradient for $J(\mathbf{c})$. With the LET approach, we can prove that global convergence is guaranteed provided one of the LET bases is a soft-threshold (4), *without any limitation on the value of* τ. Due to lack of space, the proof of this claim is left for a future publication. Here instead of choosing μ empirically for best performance, we use several Θ_4 with different but fixed μ's for the LET bases. The LET coefficients balance the influence of each basis via the minimization of objective function $J(\mathbf{c})$.

2.2. Computing LET coefficients

At each iteration, the LET coefficients are obtained by minimizing the objective function $J(\mathbf{c})$. To this aim, we use *iterated reweighted least square* (IRLS), which is found to be efficient to solve ℓ_1 regularized minimization [4, 19]. It is generally difficult to implement IRLS for image restoration problems because of the need to solve a very large linear system of equations [4, 15, 20]. But we do not have the same issue in LET restoration approach. Indeed, if we introduce (3) into (1), then we only need to solve a small linear system of equations at each iteration: $\mathbf{M}^{(n)}\mathbf{a} = \mathbf{b}$ where $\mathbf{a}$ and $\mathbf{b}$ are $K \times 1$ vectors and $\mathbf{M}^{(n)}$ is a $K \times K$ matrix, given by:

$$\mathbf{M}^{(n)} = (\mathbf{HWF}_k)^{\mathrm{T}}(\mathbf{HWF}_l) + \frac{1}{2}\lambda\mathbf{F}_k^{\mathrm{T}}\mathbf{D}^{(n)}\mathbf{F}_l$$
$$\mathbf{b} = (\mathbf{HWF}_k)^{\mathrm{T}}\mathbf{y}$$

Here $\mathbf{D}^{(n)}$ is a diagonal matrix with $\big[D^{(n)}\big]_{i,i} = \frac{1}{|c_i^{(n)}|}$. Observe that the dimensionality is reduced dramatically from the size of image NM to the degree of freedom of LET bases K for an $N \times M$ image ($K \ll NM$).

3. SIMULATION RESULTS AND DISCUSSIONS

In this section, we illustrate the iterative LET (*i*-LET) restoration approach with wavelet deconvolution problems. Several convolution kernels are used in experiments, including

- type 1: 9×9 uniform blur
- type 2: $h_{i,j} = 1/(1 + i^2 + j^2)$ for $i, j = -7, \dots, 7$
- type 3: $h_{i,j} = [1, 4, 6, 4, 1]^{\mathrm{T}}[1, 4, 6, 4, 1]/256$

The additive noise level is varied with blurred signal-to-noise-ratio (BSNR) ranging from 40dB to 10dB. In all experiments, we use sym8 decimated wavelet transform (DWT) with three levels and the test image is 256×256 *cameraman*. The ℓ_1 regularization weights λ in (1) are chosen consistently for all cases based on the noise level σ^2 and the wavelet coefficients $\mathbf{c}$ of the original image: $\lambda = \frac{\sigma^2 NM}{\|\mathbf{c}\|_1}$ for an $N \times M$ image. We choose τ based on the regularization weight λ in (4) with[1] $\tau = 150/\lambda$, and three different μ's in (5) with $\mu_1 = 10^{-2}\lambda$, $\mu_2 = 10^{-1}\lambda$, and $\mu_3 = \lambda$. Therefore, we have six LET bases in total. Note that these parameter settings do not need to be finely adjusted: the optimization of the LET coefficients automatically performs the finer tuning.

We compare the proposed *i*-LET deconvolution method with the-state-of-art algorithms: FISTA [13] and SALSA [18] in terms of iterations required to reach the same objective function value. Since FISTA is an extension to Nesterov's algorithms and it has been reported to outperform NESTA [21], we do not include the results of NESTA here. Table 1 summarizes the results obtained from experiments for deconvolution problems with decimated wavelet transform. The objective function values at convergence are obtained by running FISTA with 1000 iterations yielding $\mathbf{c}_{\mathrm{FISTA}}^{(1000)}$ and the number of iterations required for other algorithms to reach $1.001 \times J\big(\mathbf{c}_{\mathrm{FISTA}}^{(1000)}\big)$ are compared. When the algorithm does not reach the value after running 1000 iterations, we report it as "does not converge" (DNC). In all cases, the LET deconvolution outperforms SALSA and FISTA in terms of iteration numbers required. In our current implementation, one *i*-LET iteration takes 2.7 times the execution time per FISTA iteration. This means that our approach is already faster than FISTA and SALSA in a number of practical cases. In addition, we may expect further gain by a more careful choice of LET bases and a more efficient implementation. Experiments with other images than *cameraman* with the same settings have led to similar results (not shown here).

The *i*-LET approach is also capable of handling restoration problems when redundant transforms are used, which are known to provide better reconstruction quality. Here we use 4-level undecimated Haar wavelet transform. Unfortunately, FISTA and SALSA do not perform well, maybe due to the default parameter settings in the current software distributions. Then the *i*-LET approach outperforms both of them significantly. We include one example when the image is blurred with type 2 convolution kernel with 30dB BSNR. The regularization parameter λ is chosen in a way such that the basis-pursuit criterion is met, i.e. $\|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|_2^2 = NM\sigma^2$ for an $N \times M$ image. The restoration result as well as the evolution of objective function values are shown in Fig. 2. The deconvolution result obtained with DWT under the same experiment conditions is also shown for comparison.

[1]Notice again that we do not need to limit the value of τ, contrary to what is required for the IST.

Table 1. Iteration # required to reach convergence

BSNR	40	35	30	25	20	15	10
Method	*cameraman* 256×256, type 1 blur						
FISTA	282	187	116	60	33	18	8
SALSA	29	52	78	166	DNC	DNC	DNC
i-LET	**22**	**21**	**19**	**14**	**11**	**9**	**5**
Method	*cameraman* 256×256, type 2 blur						
FISTA	182	110	69	33	21	11	6
SALSA	19	35	71	168	DNC	DNC	DNC
i-LET	**16**	**13**	**11**	**8**	**6**	**5**	**3**
Method	*cameraman* 256×256, type 3 blur						
FISTA	341	199	120	55	14	8	4
SALSA	46	74	158	423	DNC	DNC	DNC
i-LET	**38**	**37**	**32**	**23**	**9**	**6**	**5**

4. CONCLUSIONS

In this paper, we have proposed a novel framework to solve the ℓ_1 image restoration problem by expanding the restoration process onto a basis of elementary processes, whose individual effect is weighted by coefficients — LET coefficients. Results obtained from standard deconvolution benchmark experiments show that the LET scheme has significant speed improvements under various experiment settings consistently.

Compared with conventional algorithms, several advantages of the LET approach are notable: 1) Dimensionality of the problem is reduced dramatically from the size of images to the degree of freedom of LET bases K. 2) The linear relationships between the reconstruction $\hat{\mathbf{c}}$ and LET coefficients a_k simplify the original optimization problem. For a given merit criterion, e.g. minimizing the objective function (1), the solution can be obtained by solving a much smaller linear system of equations. 3) The LET scheme is flexible and extensible. Even though we exemplify the LET reconstruction approach with a particular set of LET bases for image deconvolution problems, the LET framework does not specify what kind of bases should be included. The algorithm can always take the best of each LET basis via the adjustment of the linear coefficients a_k.

5. REFERENCES

[1] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the american*

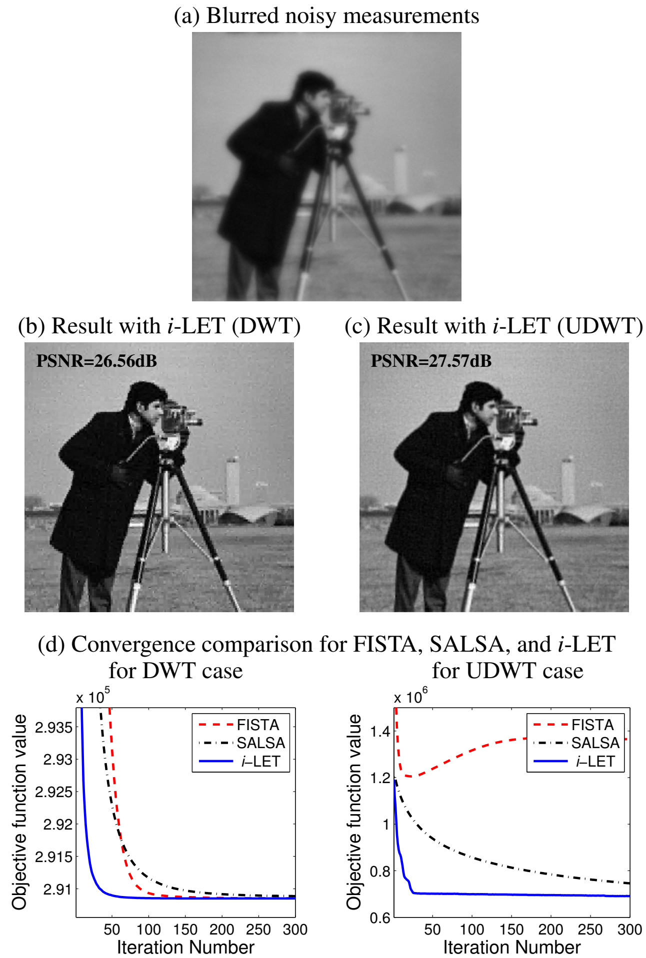

Fig. 2. Blurred measurements and deconvolution results obtained with *i*-LET for type 2 blur and 30dB BSNR with Haar decimated wavelet transform (DWT) and redundant wavelet transform (UDWT).

statistical association, vol. 90, no. 432, 1995.

[2] M. A. T. Figueiredo and R. D. Nowak, "An em algorithm for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 12, pp. 906–916, August 2003.

[3] C. Vonesch and M. Unser, "A fast iterative thresholding algorithm for wavelet-regularized deconvolution," in *Proceedings of the SPIE Conference on Mathematical Imaging: Wavelet XII*, San Diego CA, USA, August 26-29, 2007, vol. 6701, pp. 67010D–1–67010D–5.

[4] M. A. T. Figueiredo, J. M. Bioucas-Dias, and R. D. Nowak, "Majorization-minimization algorithms for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2980 –2991, December 2007.

[5] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, pp. 586–597, Dec. 2007.

[6] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *ArXiv Mathematics e-prints*, July 2003.

[7] J. L. Starck, D. L. Donoho, and E. J. Candès, "Astronomical image representation by the curvelet transform," *Astronomy Astrophysics*, vol. 398, pp. 785–800, Feb. 2003.

[8] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[9] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[10] Y. Nesterov, "Smooth minimization of non-smooth functions," *Mathematical Programming*, vol. 103, no. 1, pp. 127–152, 2005.

[11] Y. Nesterov, *Gradient methods for minimizing composite objective function*, CORE, 2007.

[12] E. T. Hale, W. Yin, and Y. Zhang, "A fixed-point continuation method for l1-regularized minimization with applications to compressed sensing," *CAAM TR07-07, Rice University*, 2007.

[13] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[14] K. Bredies and D. A. Lorenz, "Linear convergence of iterative soft-thresholding," *ArXiv e-prints*, Sept. 2007.

[15] J. M. Bioucas-Dias and M. A. T. Figueiredo, "A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Transactions on Image Processing*, vol. 16, pp. 2992–3004, Dec. 2007.

[16] F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 593–606, 2007.

[17] T. Blu and F. Luisier, "The SURE-LET approach to image denoising," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2778–2786, 2007.

[18] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2345–2356, 2010.

[19] D. Wipf and S. Nagarajan, "Iterative reweighted l1 and l2 methods for finding sparse solutions," *SPARS09, Rennes, France*, 2009.

[20] J. M. Bioucas-Dias, "Bayesian wavelet-based image deconvolution: a gem algorithm exploiting a class of heavy-tailed priors," *IEEE Transactions on Image Processing*, vol. 15, pp. 937–951, Apr. 2006.

[21] S. Becker, J. Bobin, and E. Candès, "NESTA: A fast and accurate first-order method for sparse recovery," *Arxiv preprint arXiv*, vol. 904, 2009.