FAST ALGORITHMS FOR NONCONVEX COMPRESSIVE SENSING: MRI RECONSTRUCTION FROM VERY FEW DATA

Rick Chartrand

Los Alamos National Laboratory

ABSTRACT

Compressive sensing is the reconstruction of sparse images or signals from very few samples, by means of solving a tractable optimization problem. In the context of MRI, this can allow reconstruction from many fewer k-space samples, thereby reducing scanning time. Previous work has shown that nonconvex optimization reduces still further the number of samples required for reconstruction, while still being tractable. In this work, we extend recent Fourier-based algorithms for convex optimization to the nonconvex setting, and obtain methods that combine the reconstruction abilities of previous nonconvex approaches with the computational speed of state-of-the-art convex methods.

Index Terms— Magnetic resonance imaging, image reconstruction, compressive sensing, nonconvex optimization.

1. INTRODUCTION

1.1. Compressive Sensing and MRI

Results of Candès et al. [1] and Donoho [2] demonstrated that sparse images can be reconstructed from fewer linear measurements than previously thought possible, in what is now known as *compressive sensing* (among other related terms). The results take advantage of the sparsity inherent in real-world images: the ubiquity of image compression points to the existence of transforms converting a digital image into one having relatively few values significantly different from zero. Also, the success of total-variation regularization for image restoration indicates that most images can be well approximated by those having a sparse gradient.

The approach has been to solve the following convex optimization problem:

$$\min_u \|\Psi u\|_1, \quad \text{subject to} \quad \Phi u = b. \tag{1}$$

Here, b consists of the samples Φx where x is an image in 2- or 3-D, Φ is the measurement or sampling operator, and Ψ is a sparsifying operator, such as a wavelet transform or discrete gradient. Then, provided Φ and Ψ are sufficiently *incoherent* (roughly that rows of Φ can't be sparsely expressed in terms of columns of Ψ, and vice versa; see [3]), the solution to (1) will be exactly x, even when the linear system is severely underdetermined. When the measurements are noisy, it is best to relax the equality constraint and solve the following instead:

$$\min_u \|\Psi u\|_1 + (\mu/2)\|\Phi u - b\|_2^2, \tag{2}$$

with μ a regularization parameter whose choice is dependent on the noise level.

The application of compressive sensing to MRI reconstruction was implicit in [1], which demonstrated perfect reconstruction of the Shepp-Logan phantom from Fourier-space samples along 22 radial lines, amounting to 9% of k-space. Explicit application was undertaken by Lustig et al. [4]. Both phantoms and anatomical images were reconstructed well with as little as 20% sampling, using patterns such as spirals and randomly-positioned lines. (The increase over the sampling of [1] is typical for real images, which are less sparse than synthetic phantoms.) They obtained improved results by having both a wavelet transform and a discrete gradient in the objective.

1.2. Nonconvex Compressive Sensing

Numerical results in [5] showed that one can reduce the sampling of Fourier space needed for reconstruction by replacing the ℓ^1 norm in (1) with the ℓ^p quasi-norm, where $0 < p < 1$:

$$\min_u \|\Psi u\|_p^p, \quad \text{subject to} \quad \Phi u = b. \tag{3}$$

This gives a nonconvex optimization problem, but one that in practice appears to be solvable using simple algorithms [5, 6]. Theoretical work [7, 8] has justified these observations, with [8] also showing that the nonconvex approach increases robustness to noise and image nonsparsity. An iteratively-reweighted least-squares (IRLS) approach to ℓ^p minimization was previously taken by Rao and Kreutz-Delgado [9], but with unimpressive results due to getting trapped in local minima. With both gradient descent [5] and IRLS [6], the key to avoiding local minima is the successive regularization of the ℓ^p objective, in a manner reminiscent of the graduated nonconvexity approach of Blake and Zisserman [10]. In recovering sparse signals perfectly, these algorithms are able to find the global minimum of the ℓ^0 analog of (1), though it has yet to be proven that the global minimum of (3) itself is obtained.

We gratefully acknowledge support of this work by the U.S. Department of Energy through the LANL/LDRD Program.

In the MRI context, for the same Shepp-Logan phantom synthetic example of [1] it was shown in [5] that the number of radial lines could be reduced from 22 to 10 (or 3.8% of k-space). Trzasko and Manduca [11] reproduced this result with a faster algorithm, while also considering real MRI data examples. Their approach uses an objective function that approaches the ℓ^0 norm asymptotically, which creates a graduated nonconvexity that appears to allow global convergence. They obtain reconstructions with fewer samples than in [4], and show that reconstruction quality is improved over ℓ^1 minimization. They obtain reconstructions of 256×256 images in 1–3 minutes in MATLAB on a 3 GHz desktop computer.

New algorithms for MRI reconstruction via ℓ^1 minimization that are much faster than their predecessors have been developed by Ma et al. [12] and Goldstein and Osher [13]. They both are able to replace iterative linear solvers with Fourier-domain computations, with substantial time savings. In this work we generalize these approaches to the nonconvex setting of ℓ^p minimization with $p < 1$. As before we find that this reduces the number of k-space samples needed for reconstruction of a given quality. The resulting algorithms are much faster than any existing algorithms for nonconvex compressive sensing. The computation time is dominated by a few FFT computations per iteration, resulting in desirable scaling to large images that will make 3-D reconstructions much more feasible than before.

2. ALGORITHM

Now we present our algorithm approach, beginning with a generalization of an operator-splitting algorithm of Ma et al. [12], which is equivalent to the case of $p = 1$ in what follows. Our derivation generalizes that of the related algorithm of Yang et al. [14]. We present the 2-D version, with extension to 3-D being straightforward. For the time being fix $\beta > 0$ and any real p, and define a regularized ℓ^p objective by

$$\varphi(t) = \begin{cases} \gamma|t|^2 & \text{if } |t| \le \alpha \\ |t|^p/p - \delta & \text{if } |t| > \alpha. \end{cases} \quad (4)$$

Here t can be a scalar or 2-D vector (in which case $|t|$ is the length of t), the parameters γ and δ are chosen to make φ a C^1 function, and α will be chosen shortly. For convenience we adopt the unusual convention that when $p = 0$, $|t|^p/p$ will mean $\log|t|$.

Next, we show that there is a function ψ satisfying

$$\varphi(t) = \min_s \{\psi(s) + (\beta/2)\|s - t\|_2^2\}. \quad (5)$$

Yang et al. [14] exploit convex duality and calculate for $p = 1$ that $\psi(s) = |s|$ (in our notation). In our case, ϕ is not convex, but by taking $\alpha = \beta^{1/(p-2)}$ (or larger), the auxiliary function $f(t) = |t|^2/2 - (1/\beta)\varphi(t)$ is convex. Letting $g = f^*$ be the convex conjugate (or Legendre-Fenchel transform) of f, the convexity and continuity of f ensures that $f = g^* = f^{**}$ as well. Simple manipulations then show that (5) is satisfied by $\psi(s) = \beta(g(s) - |s|^2/2)$. Moreover, the minimizer of (5) will be given by $s^* = \nabla f(t)$ [15, p. 476], which can be shown to be given by what we call the *p-shrinkage* operator S_α^p:

$$S_\alpha^p(t) = \max\{|t| - \alpha|t|^{p-1}, 0\}t/|t|. \quad (6)$$

This generalizes the shrinkage operator used in several compressive sensing algorithms as well as soft wavelet thresholding. It will not be necessary to compute ψ explicitly. This can be done for special values of p such as 0 and $1/2$, which can then be used to derive stopping conditions from optimality conditions as in [14]. We will pursue this in future work.

We now incorporate the above into our algorithm. As in [4] we use an objective with both a discrete gradient Du and an orthogonal wavelet transform Ψu:

$$\min_u \sum_i \big(\varphi((Du)_i) + \lambda\varphi((\Psi u)_i)\big) + (\mu/2)\|\Phi u - b\|_2^2. \quad (7)$$

We find it sufficient to use $\lambda = 1$ (or 0 for gradient-sparse images). Now we apply the splitting and obtain

$$\min_{u,v,w} \sum_i \big(\psi(v_i) + \lambda\psi(w_i)\big) + (\beta_D/2)\|v - Du\|_2^2 + (\beta_W/2)\|w - \Psi u\|_2^2 + (\mu/2)\|\Phi u - b\|_2^2, \quad (8)$$

where i ranges over all pixels of the image. Note that each v_i is vector-valued while the u_i and w_i are scalars. For fixed β_D, β_W, and μ, we solve this iteratively by solving for each of u, v, and w in turn while holding the other two fixed. The benefit of the splitting approach is that each of the three subproblems is simple and computationally fast to solve. The v and w subproblems are both separable, involving the p-shrinkage operator (6) applied to each pixel separately, making the computation vectorized and parallelizable. The solution to the u subproblem can be computed directly using FFTs and an FWT, avoiding the need for an expensive linear solver.

By letting $\beta_D, \beta_W \to \infty$, v is forced to approach Du and w to approach Ψu, making the solution to (8) approach that of (7). Ma et al. use a continuation approach, starting with small β values and then increasing them geometrically, using the solution at each stage to initialize the next stage. Instead, we adopt the Bregman iteration approach of Goldstein and Osher [13]. Bregman iterations were first used in image restoration [16]; the basic approach is that one can enforce a data constraint more and more tightly by adding the residual back to the data at each iteration. In image denoising, this amounts to adding what is ostensibly the "noise" back to the noisy image, a counterintuitive approach that is nonetheless effective and theoretically justified. This approach was applied by Goldstein and Osher to the above in the case of $p = 1$, and we find for $p < 1$ that it gives better performance than continuation. The resulting algorithm is as follows. Let $\mathcal{F}$ denote the 2-D discrete Fourier transform, and K a projection operator onto the k-space locations being sampled, so that $\Phi = K\mathcal{F}$.

Provided we use periodic boundary conditions, we can regard discrete derivatives as circular convolution with two-element kernels, so that there is d such that $Du = \mathcal{F}^{-1} d \mathcal{F} u$. We also need Bregman iterates, vector-valued (at each pixel) b_D and scalar-valued b_W.

Input: k-space data b, k-space locations projection K, parameters $\mu, \lambda, \beta_D, \beta_W$
Precompute: Fourier-domain denominator
$G = \mu K + \beta_D |d|^2 + \beta_W$
Initialize: $u^0 = \mathcal{F}^{-1} b$, $v^0 = b_D^0 = \mathbf{0}$, $w^0 = b_W^0 = 0$, $b^0 = b$
for *number of Bregman iterations* **do**
  for *number of inner iterations* **do**
    $u^{n+1} = \mathcal{F}^{-1}\big(\big[\mu b^m + \mathcal{F}\big(\beta_D D^T (v^n - b_D^n) + \beta_W W^T (w^n - b_W^n)\big)\big]/G\big)$
    $v^{n+1} = S^p_{1/\beta_D}(Du^{n+1} + b_D^n)$
    $w^{n+1} = S^p_{1/\beta_W}(\Psi u^{n+1} + b_W^n)$
    $b_D^{n+1} = b_D^n + Du^{n+1} - v^{n+1}$
    $b_W^{n+1} = b_W^n + \Psi u^{n+1} - w^{n+1}$
  end
  $b^{m+1} = b^m + b - K\mathcal{F} u^{n+1}$
end
Output: Reconstructed image u

Algorithm 1: Fast nonconvex MRI reconstruction

Note that the variables are not reinitialized at the close of each inner loop, so that n would not be reset. We use manually determined numbers of iterations, deferring more sophisticated stopping criteria to later work.

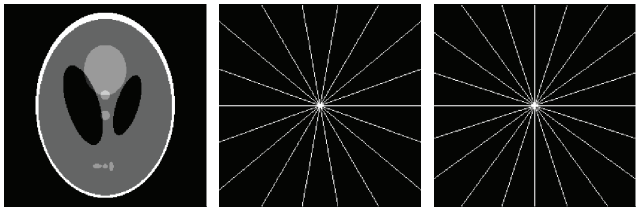
We obtain a second algorithm by replacing the p-shrink operator with a weighted 1-shrink operator $S^{c_i}_{\alpha}(t) = \max\{|t| - \alpha c_i, 0\} t/|t|$. The approach is to compute the weights as $c_i = |(Du)_i|$ or $|(Wu)_i|$, but only updated once each inner loop has completed. This decreases the iteration time slightly, while slightly increasing the number of iterations needed; which algorithm performs better depends on the particular test.

3. EXPERIMENTS

All experiments were done in MATLAB on a 1.2 GHz laptop with 3 GB of memory. We begin with the 256×256 Shepp-Logan phantom. Because the phantom has a very sparse gradient, we do not use the wavelet regularization, and let $\lambda = \beta_W = 0$. We set $\mu = 10^5$ and $\beta_D = 1$. An early coding mistake led to the accidental use of $p = -1/2$, which allowed the phantom to be recovered exactly from 9 radial lines of samples, or 3.5% of k-space. This is fewer than the previous best of 10 radial lines, first done in [5]. Reconstruction was unsuccessful with nonnegative p. While using $p < 1$ makes sense from the perspective of trying to approximate the ℓ^0 norm, we do not understand why $p < 0$ should improve performance, but after this discovery we consistently found this in all of our tests. (Rao and Kreutz-Delgado [9] considered $p < 0$, with no benefit.) Using 40 inner iterations, after 32 outer (Bregman) iterations, the reconstruction quality was 51.0 dB, taking 66.5 s. The reconstruction reaches 200 dB after 217 outer iterations and 647 s, at which point the worst pixel error was 6.58×10^{-10}.

	$p = 1$	$p = 1/2$	$p = 0$	$p = -1/2$
phantom	6.8 dB	50.5 dB	50.3 dB	50.0 dB
uterus	13.0 dB	13.6 dB	13.8 dB	13.9 dB
phantom	11.5 s	17.5 s	13.0 s	13.0 s
uterus	346 s	465 s	438 s	469 s

Table 1. Reconstruction SNR and CPU time for the 256×256 Shepp-Logan phantom (10 radial lines, 3.8% sampling) and th 1024×1024 uterus image (random phase, 15% sampling). The uterus SNRs are misleadingly low due to the noise in the original image and the denoising effect of the regularization.



(a) Shepp-Logan (b) 9 radial lines (c) 10 radial lines

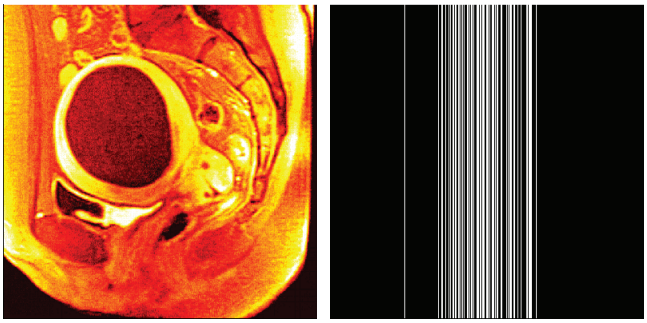
Fig. 1. Shepp-Logan phantom and radial sampling. Reconstructions are visually identical to the original, so not shown.

Next we consider 10 radial lines (3.8% sampling), the previous benchmark. In this case we get better performance from the reweighted version of the algorithm. For $p < 1$, reconstructions to 50 dB take as little as 13.0 s; more details comparing different values of p are in Table 1. This is about 10 times faster than reported in [11], with better scaling to larger images expected. Reconstruction fails utterly with $p = 1$.

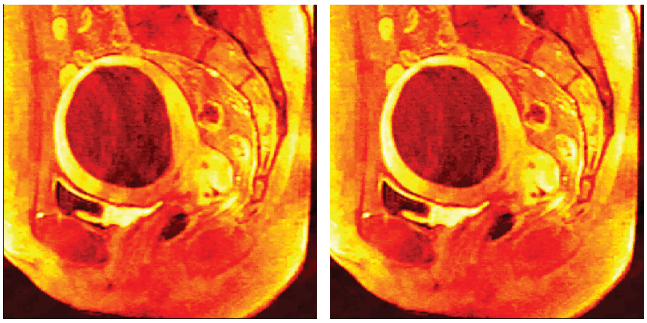
We conclude with simulated k-space samples of the noisy 1024×1024 image in Figure 2, from [17]. We use Gaussian-density random phase encoding (cf. [4]) with standard deviation 100 columns, and 15% sampling. We use Algorithm 1 with 40 inner iterations and 5 outer iterations. We set $\mu = 10$, $\lambda = 1$, $\beta_D = 1$, and $\beta_W = 10$. Results with $p \in \{1, 1/2, 0, -1/2\}$ are in Table 1. In Figure 2 we display the poorest and best results, namely $p = 1$ and $p = -1/2$. In the $p = 1$ image we see more pronounced aliasing effects. Running times of 6–8 minutes roughly fit the $N \log N$ scaling to be expected, though direct comparison with the Shepp-Logan phantom and radial sampling is inappropriate.

4. CONCLUSIONS

We presented an algorithm that can reconstruct MR images from few k-space samples in much less time than previously possible. This makes use of previous operator splitting methods and the Bregman iteration method, as well as a novel p-



(a) noisy image (b) k-space samples



(c) $p = 1$ (d) $p = -1/2$

Fig. 2. Synthetic k-space samples are generated for a noisy image, using Gaussian random phase encoding. Reconstruction with $p = 1$ shows greater aliasing than with $p = -1/2$.

shrinkage operator. Further investigation is needed into the effect of sampling patterns and sparsifying transforms. The development of sparser representations, perhaps customized to each class of medical images, will lead to better reconstruction fidelity from very few k-space measurements. It will also widen the gap between convex and nonconvex compressive sensing, as the very sparse phantom examples suggest.

5. REFERENCES

[1] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, 2006.

[3] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, pp. 969–985, 2007.

[4] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, pp. 1182–1195, 2007.

[5] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, pp. 707–710, 2007.

[6] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2008.

[7] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 035020, pp. 1–14, 2008.

[8] R. Saab, R. Chartrand, and Özgür Yilmaz, "Stable sparse approximations via nonconvex optimization," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2008.

[9] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Process.*, vol. 47, pp. 187–200, 1999.

[10] A. Blake and A. Zisserman, "Visual reconstruction and the GNC algorithm," in *Parallel Architectures and Computer Vision*, pp. 33–48, 1988.

[11] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic ℓ_0-minimization," *IEEE Trans. Med. Imaging*, vol. 28, pp. 106–121, 2009.

[12] S. Ma, W. Yin, Y. Zhang, and A. Chakraborty, "An efficient algorithm for compressed MR imaging using total variation and wavelets," in *IEEE Conference on Computer Vision and Pattern Recognition, 2008*, pp. 1–8, 2008.

[13] T. Goldstein and S. Osher, "The split Bregman method for L1 regularized problems," CAM Report 08-29, UCLA, 2008.

[14] J. Yang, W. Yin, Y. Zhang, and Y. Wang, "A fast algorithm for edge-preserving variational multichannel image restoration." To appear in SIAM J. Imaging Sciences.

[15] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*. Berlin: Springer-Verlag, 2004.

[16] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An iterative regularization method for total variation-based image restoration," *Multiscale Model. Simul.*, vol. 4, pp. 460–489, 2005.

[17] T. J. Kroencke, A. Gauruder-Burmester, C. N. Enzweiler, M. Taupitz, and B. Hamm, "Disintegration and stepwise expulsion of a large uterine leiomyoma with restoration of the uterine architecture after successful uterine fibroid embolization: Case report," *Hum. Reprod.*, vol. 18, pp. 863–865, 2003.