Iteratively reweighted regularization methods for image reconstruction from non-uniform Fourier data

VICTOR CHURCHILL[1,*]

[1] Department of Mathematics, Dartmouth College
[*] Corresponding author: victor.a.churchill.gr@dartmouth.edu

Compiled March 12, 2018

In this paper, we apply two popular image reconstruction algorithms that use iteratively reweighted regularization to Fourier data acquired non-uniformly. We find that while both methods reconstruct a test image better than standard TV regularization using the same number of measurements, further improvements lie ahead.

OCIS codes: (140.3490) Lasers, distributed feedback; (060.2420) Fibers, polarization-maintaining;(060.3735) Fiber Bragg gratings.

http://dx.doi.org/10.1364/ao.XX.XXXXXX

1. INTRODUCTION

Since the development of compressed sensing (CS) [1, 2], total variation (TV) regularization [3], based on the ℓ_1 norm has been ubiquitous in image reconstruction. The general method is to reconstruct an image as the solution to the optimization problem

$$\arg\min_{\mathbf{f}} ||A\mathbf{f} - \hat{\mathbf{f}}||_2^2 + \lambda||D\mathbf{f}||_1 \quad (1)$$

where A is the forward model for the imaging system, $\hat{\mathbf{f}}$ is the collected data, λ is a regularization parameter weighting fidelity against regularization, and D is the matrix representation of the TV transform. The accuracy of (1) relies on two things. The first is an appropriate choice for λ, which is typically difficult and problem-dependent, [4]. The second is sparsity in the edge domain of the image. That is, the number of jumps in the image needs to be much lower than the total number of points.

In ideal cases, (1) has been shown to reconstruct images exactly using a sampling rate significantly below that prescribed by the Shannon-Nyquist sampling theorem, [1]. In many cases, however, the resulting reconstruction is not actually as sparse in the edge domain as desired. This could be due to noise, non-uniform sampling, the fact that the TV transform is not actually sparsifying with respect to a particular image, or that the ℓ_1 norm penalizes large magnitudes more heavily. When dealing with data from a real-world imaging system, there is likely a combination of these factors. As a result, the overall accuracy of the reconstruction is diminished.

One promising approach for correcting this issue is to use iteratively reweighted (IR) regularization, [5–11]. IR methods employ multiple runs of TV regularization based on a weighted ℓ_p norm, where typically $p \in \{1,2\}$. The main idea in this class of algorithms is to remove the magnitude dependence by penalizing non-zero magnitudes in the edge domain more fairly than is done by an unweighted ℓ_1 norm. These methods typically prescribe weights that are inversely proportional to the magnitude of each point in the edge domain. In this way, they end up regularizing less in areas with jumps and more in areas without jumps.

These IR methods have been shown to increase the accuracy of a reconstruction when data are acquired as uniform Fourier samples, [5]. That is, as in (1) the forward model is $A = F$, the discrete Fourier transform (DFT). In this paper we perform the straightforward extension of these algorithms to non-uniformly acquired Fourier data. This will require the implementation of a non-uniform discrete Fourier transform forward model, or more practically a non-unifrom fast Fourier transform (NFFT), [12–14]. This is an important addition, as data are frequently acquired non-uniformly in real-world imaging systems and non-uniform sampling creates additional reconstruction errors.

This paper will focus on applying the two most popular (by citation count) IR methods, those of [5] and [6], to non-uniform Fourier data image reconstruction. Comparisons of these two methods were also included in [6] and [10], but only for Gaussian forward models and one-dimensional sparse signals.

2. PROBLEM SET UP

Let $f : [-1,1]^2 \to \mathbb{R}$ be a piecewise smooth function. We can think of this as the underlying function that describes the image we want to reconstruct. Suppose we are given finitely many non-uniform Fourier samples of f,

$$\hat{f}(\lambda_{\mathbf{k}}) = \frac{1}{4}\int_{-1}^{1}\int_{-1}^{1} f(x,y)e^{-\pi i\lambda_{\mathbf{k},1}x}e^{-\pi i\lambda_{\mathbf{k},2}y}dxdy \quad (2)$$

where $\lambda_{\mathbf{k}} = \{(\lambda_{k_1},\lambda_{k_2}) : |k_1|,|k_2| \leq M\} \in \mathbb{R}^2$. Specifically, we look at the jittered sampling pattern for $\lambda_{\mathbf{k}}$ given by

$$\lambda_{\mathbf{k}} = \mathbf{k} - \left\lfloor \frac{2M+1}{2} \right\rfloor - 1 + \frac{1-2\xi_{\mathbf{k}}}{4} \quad (3)$$

where $\xi_{\mathbf{k}} \sim U([0,1])^2$. Figure 1 gives a visualization of (3). This sampling pattern simulates data theoretically acquired on a uniform grid with some error that sometimes occurs in real world measurement systems. Reconstructing using a standard

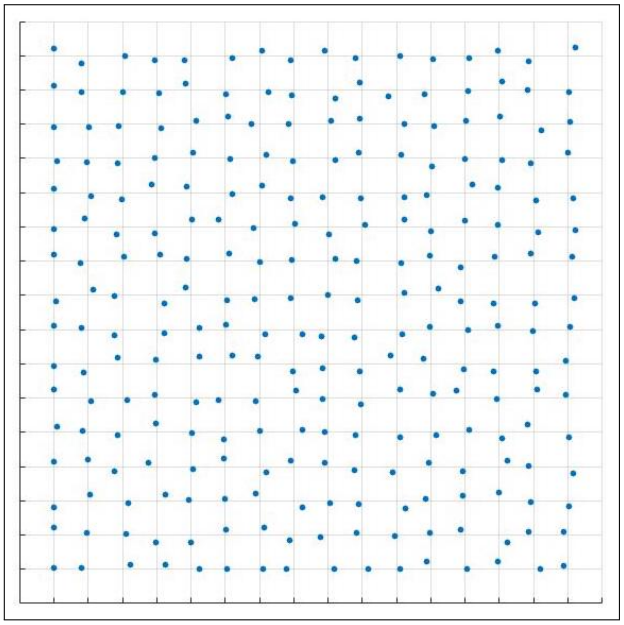

Fig. 1. Jittered non-uniform sampling pattern as in (3).

DFT as the forward model will not take into consideration these perturbations.

Let $\mathbf{f} = \{f(x_j, y_j) : j = 1, \ldots, N^2\}$ be a vector of the reconstruction points and $\hat{\mathbf{f}} = \{\hat{f}(\lambda_{k_1}, \lambda_{k_2}) : |k_1|, |k_2| \leq M\}$ be a vector of the Fourier modes. The standard TV reconstruction is the solution to the unconstrained optimization problem

$$\arg\min_{\mathbf{f}} ||F_N\mathbf{f} - \hat{\mathbf{f}}||_2^2 + \lambda||D\mathbf{f}||_1, \tag{4}$$

where F_N is the NDFT matrix. The top right box of Figure 2 shows an example of (4) implemented on the Modified Shepp-Logan phantom. Iteratively reweighted regularization methods seek to improve upon this result.

3. ITERATIVELY REWEIGHTED SCHEMES

As explained in [1], reconstructing an image via solving

$$\arg\min_{\mathbf{f}} ||F_N\mathbf{f} - \hat{\mathbf{f}}||_2^2 + \lambda||D\mathbf{f}||_0, \tag{5}$$

where $||\cdot||_0$ counts non-zero values, promotes the most sparsity in the edge domain of the image. However, this combinatorial problem is NP-hard. The ℓ_1 norm as in (4) acts as a convex surrogate for the ℓ_0 term, making the problem easier to solve. But, it is not as sparsity-encouraging. Naturally, this begs the question of whether there are better surrogates that generate solvable optimization problems. We will look at two different approaches that both use IR schemes to approximate nonconvex functions that are closer surrogates to the ℓ_0 term.

Both methods employ multiple runs of TV regularization based on a weighted ℓ_1 or ℓ_2 norm. The main idea behind these methods is that large weights can be used to discourage non-zero entries in the edge domain, while small weights can be used to encourage non-zero entries. Hence they penalize non-zero edge domain magnitudes more fairly than the magnitude-dependent ℓ_1 regularization. To achieve this, weights inversely proportional to the true edge domain magnitudes are used. In this way, IR methods encourage sparsity in the edge domain by regularizing less in areas with jumps and more in areas without jumps.

A. Iteratively reweighted ℓ_1 (IRℓ_1)

In lieu of the ℓ_0 or ℓ_1 regularization terms, the approach of [5] is to regularize using the log-sum function, a concave penalty function that more closely resembles the ℓ_0 norm. The log-sum penalty function has the potential to be much more sparsity-encouraging than the ℓ_1 norm, and is a closer surrogate to the ℓ_0 norm. See Figure 3 in [5] for a heuristic explanation of this. They want to solve

$$\arg\min_{\mathbf{f}} ||F_N\mathbf{f} - \hat{\mathbf{f}}||_2^2 + \sum_{1\leq i,j\leq n} \log(|(D\mathbf{f})_{i,j}| + \epsilon) \tag{6}$$

where ϵ is a parameter to stay in the domain of log. However, since the log-sum function is nonconvex, (6) is difficult minimize. So the authors propose to approximate it with a series of weighted ℓ_1 regularized minimizations. Details for their IR algorithm are in Algorithm 1.

A similar algorithm from [8] uses the same framework with support estimates and constant weights. Another support-driven ℓ_1-based approach is in [9]. Zhao and Li in [**?**] provide a unified framework for reweighted ℓ_1 minimization.

Algorithm 1. IRℓ_1

1: Set $\ell = 0$ and $w_{i,j}^{(\ell)} = 1$ for $1 \leq i, j \leq N$, where N is the number of grid points in each dimension.
2: Solve the weighted regularization minimization problem

$$\mathbf{f}^{(\ell)} = \arg\min_{\mathbf{f}} ||F_N\mathbf{f} - \hat{\mathbf{f}}||_2^2 + \lambda \sum_{1\leq i,j\leq 2J} w_{i,j}^{(\ell)} |(D\mathbf{f})_{i,j}|.$$

3: Update the weights. For each (i, j), $1 \leq i, j \leq N$,

$$w_{i,j}^{(\ell+1)} = \frac{1}{|(D\mathbf{f}^{(\ell)})_{i,j}| + \epsilon}$$

To avoid division by zero and normalize the weights, $\epsilon > 0$ is a small constant. The authors claim the algorithm is fairly robust to the choice of ϵ.
4: Terminate on convergence or when ℓ attains a specified maximum number of iterations ℓ_{max}. Otherwise, increment ℓ and go to step 2.

B. Iteratively reweighted least squares (IRLS)

The method from [6] ends up using a similar surrogate for the ℓ_9 term to the one above, but takes a different approach to finding it. The authors attempt to use iteratively reweighted ℓ_2 regularization, or iteratively reweighted least squares, to solve the nonconvex problem

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} ||F_N\mathbf{f} - \hat{\mathbf{f}}||_2^2 + \lambda||D\mathbf{f}||_p^p \tag{7}$$

for $p \in [0, 1)$. The best results occur using $p = 0$, which uncoincidentally, similar to Algorithm 1, corresponds to regularizing with the concave log-sum penalty function with a squared argument

$$\sum_{1\leq i,j\leq N} \log((D\mathbf{f})_{i,j}^2 + \epsilon). \tag{8}$$

Algorithm 2 details the procedure. A basic comparison in [6] shows that Algorithm 2 slightly outperforms Algorithm 1 in recovering a sparse signal from random measurements when an appropriate adaptive choice of ϵ is applied. In addition to

accuracy, another benefit of the ℓ_2 approach is computational simplicity. Each iteration of Algorithm 1 requires solving an ℓ_1 minimization problem. This quickly becomes computationally expensive. As for IRLS, even in the case that the inversion in step 2 is expensive, this inner problem can be solved using a conjugate gradient method, which is much more efficient than methods used to solve ℓ_1 regularized problems.

Further work has been done on convergence of iteratively reweighted least squares methods in [7].

Algorithm 2. IRLS

1: Set $\ell = 0$ and $w_{i,j}^{(\ell)} = 1$ for $1 \leq i,j \leq N$, where N is the number of grid points in each dimension.
2: Solve the weighted regularization minimization problem

$$\mathbf{f}^{(\ell)} = \left(F_N^T F_N + \lambda (W^{(\ell)} D)^T W^{(\ell)} D\right)^{-1} F_N^T \hat{\mathbf{f}}$$

where $W^{(\ell)} = \text{diag}(w^{(\ell)})$.
3: Update the weights. For each (i,j), $1 \leq i,j \leq N$,

$$w_{i,j}^{(\ell+1)} = \frac{1}{(D\mathbf{f}^{(\ell)})_{i,j}^2 + \epsilon}$$

where ϵ is a parameter similar to that in Algorithm 1.
4: Terminate on convergence or when ℓ attains a specified maximum number of iterations ℓ_{max}. Otherwise, increment ℓ and go to step 2.

4. NUMERICAL RESULTS

Our test image is the Modified Shepp-Logan phantom of size 256×256 pixels. See top left box of Figure 2. In the frequency domain, we acquire a 50×50 grid of jittered (non-uniform) Fourier samples. This is few enough so that the standard TV reconstruction is flawed, which allows us to test the improvement of each method.

For the implementation, for Algorithm 1 we perform each ℓ_1 minimization using the Split Bregman method [15]. For Algorithm 2, each ℓ_2 minimization is performed using conjugate gradient descent [16]. See code for more. In terms of speed, IRLS objectively runs faster than IRℓ_1. This is completely expected since in general ℓ_2 regularized minimizations are much easier to solve then ℓ_1 regularized minimizations. However, we did not consider accelerated algorithms that have been proposed for IRℓ_1, [17].

For the regularization parameter, since we did not consider additive noise, we use a low weight $\lambda = 10^{-5}$ for all algorithms. For the choice of the parameter ϵ, [5] recommends that a choice of ϵ slightly smaller than the expected nonzero magnitudes of $D\mathbf{f}$ provides the stability necessary to correct for inaccurate coefficient estimates while still improving upon the unweighted TV algorithm. On the other hand, [6] gives evidence that changing ϵ in each iteration yields superior results for the problem of sparse signal recovery. However, the implementation suggested does not make sense for a use-case with much higher error. Therefore, we did not consider an adaptive ϵ parameter, only fixed. In particular, some basic testing gave the best results for $\epsilon = 10^{-1}$ for IRℓ_1 and $\epsilon = 10^{-4}$ for IRLS.

For the main result, Figure 2 shows the reconstructions. In the eyeball norm we see that both IRℓ_1 and IRLS improve upon the standard TV reconstruction. However, in both cases we still see some spurious oscillations as well as the over-smoothing of edges. It follows that we see minimal gains in relative error. The relative error for the TV minimization was $||\mathbf{f}^* - \mathbf{f}||_2 / ||\mathbf{f}||_2 = 0.3994$. For IR$\ell_1$, the relative error was 0.3629, and for IRLS, 0.3705.

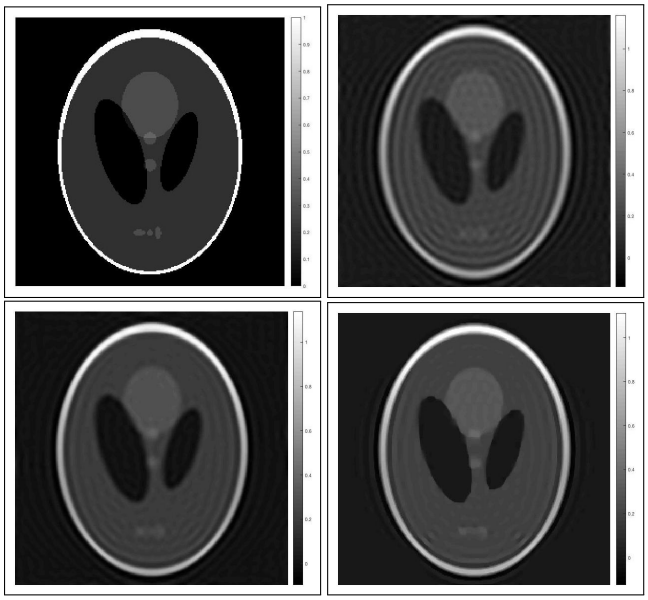

Fig. 2. Reconstruction comparison of 256×256 pixel Modified Shepp-Logan phantom from 50×50 jittered Fourier modes. Top left is the original while the others are reconstructions via (top right) TV regularization, (bottom left) IRℓ_1, and (bottom right) IRLS.

5. CONCLUSION AND FUTURE WORK

This paper provides some evidence that IR methods are superior to single-run methods when reconstructing images with sparse edges from non-uniform Fourier data. Both IRℓ_1 and IRLS can recover an image better than standard TV reconstruction using the same number of measurements. However, further testing is needed when complex Gaussian noise is added to the data as well as when different sampling patterns are used.

There are also some drawbacks to IR methods. Run time is increased for these methods, since they require multiple TV minimization steps. In addition, they introduce another uncertainty

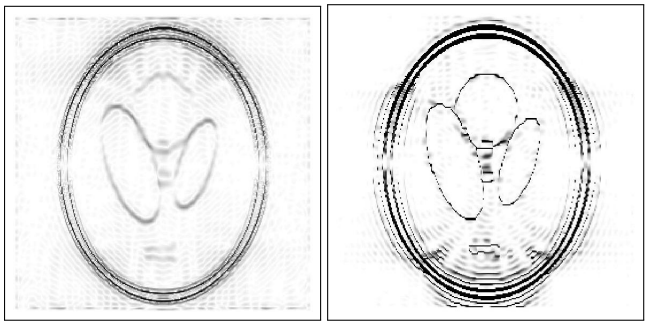

Fig. 3. Comparison of x-direction IRℓ_1 (left) and IRLS (right) weighting matrices after 5 iterations. Black indicates minimum weight while white indicates maximum weight.

in the parameter ϵ, upon which the success of these algorithms depends. Lastly and chiefly, the accuracy of these methods is not terribly better than standard TV reconstruction. It seems that one source of error is from the difficulty both of these methods have with weighting points between 0 and $\frac{1}{\epsilon}$. Since there is no smooth variation away from jumps in our test image, many of these values falsely indicate small jumps. Troublingly, Figure 3 shows both methods generate many of these false jump identifications, leading to overall less accurate reconstructions.

Future work focused on a better way to assign weights the regularization term has promise to increase the accuracy and speed of these methods. In particular, the authors of [5] claim that without prior information about non-zero elements in the edge domain of an image, the procedure for selecting these weights is iterative in nature. However, the weighting schemes adopted by these methods are likely not most direct way to penalize non-zero locations in the edge domain when the problem starts from Fourier data. In particular, there are robust methods available for edge detection from Fourier data [18]. Performing a pre-processing edge detection and applying a zero weight to a point where an edge is detected and a high constant weight to points where no edge is detected may yield better accuracy and could also improve run time since it would only require a single optimization step. Further study is certainly in this direction is required.

REFERENCES

1. E. J. Candès, J. Romberg, and T. Tao, IEEE Transactions on information theory **52**, 489 (2006).
2. D. L. Donoho, IEEE Transactions on information theory **52**, 1289 (2006).
3. L. I. Rudin, S. Osher, and E. Fatemi, Phys. D: Nonlinear Phenom. **60**, 259 (1992).
4. S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, Multiscale Model. & Simul. **4**, 460 (2005).
5. E. J. Candes, M. B. Wakin, and S. P. Boyd, J. Fourier analysis applications **14**, 877 (2008).
6. R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in "Acoustics, speech and signal processing, 2008. ICASSP 2008. IEEE international conference on," (IEEE, 2008), pp. 3869–3872.
7. I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, Commun. on Pure Appl. Math. **63**, 1 (2010).
8. H. Mansour and Ö. Yilmaz, "Support driven reweighted ℓ_1 minimization," in "Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on," (IEEE, 2012), pp. 3309–3312.
9. Y. Wang and W. Yin, SIAM J. on Imaging Sci. **3**, 462 (2010).
10. D. Wipf and S. Nagarajan, IEEE J. Sel. Top. Signal Process. **4**, 317 (2010).
11. I. F. Gorodnitsky and B. D. Rao, IEEE Transactions on signal processing **45**, 600 (1997).
12. A. Dutt and V. Rokhlin, SIAM J. on Sci. computing **14**, 1368 (1993).
13. L. Greengard and J.-Y. Lee, SIAM review **46**, 443 (2004).
14. J.-Y. Lee and L. Greengard, J. Comput. Phys. **206**, 1 (2005).
15. T. Goldstein and S. Osher, SIAM journal on imaging sciences **2**, 323 (2009).
16. G. H. Golub and C. F. Van Loan, *Matrix computations*, vol. 3 (JHU Press, 2012).
17. M. S. Asif and J. Romberg, IEEE Transactions on Signal Process. **61**, 5905 (2013).
18. A. Gelb and G. Song, J. Sci. Comput. **71**, 737 (2017).