A FAST ALGORITHM FOR STRUCTURED LOW-RANK MATRIX RECOVERY WITH APPLICATIONS TO UNDERSAMPLED MRI RECONSTRUCTION

Greg Ongie⋆ and Mathews Jacob†

⋆ Department of Mathematics, University of Iowa, IA, USA.
† Department of Electrical and Computer Engineering, University of Iowa, IA, USA.

ABSTRACT

Structured low-rank matrix priors are emerging as powerful alternatives to traditional image recovery methods such as total variation (TV) and wavelet regularization. The main challenge in applying these schemes to large-scale problems is the computational complexity and memory demand resulting from a lifting of the image to a high-dimensional dense matrix. We introduce a fast and memory efficient algorithm that exploits the convolutional structure of the lifted matrix to work in the original non-lifted domain, thus considerably reducing the complexity. Our experiments on the recovery of MR images from undersampled measurements show that the resulting algorithm provides improved reconstructions over TV regularization with comparable computation time.

Index Terms— Structured Low-Rank Matrix Recovery, Annihilating Filter Method, Finite Rate of Innovation, Compressed Sensing, MRI.

1. INTRODUCTION

Recovering image data from sparse and noisy measurements is a central problem in many biomedical imaging problems, including MRI, microscopy, and tomography. The traditional approach is to regularize the recovery using smoothness or sparsity priors. Recently, several researchers have introduced powerful structured low-rank matrix priors, which are emerging as more efficient alternatives to conventional regularization penalties [1–7]. These algorithms rely on annihilation relations between the Fourier transform coefficients of the image induced by sparsity and structure in the images. For example, we have recently shown that the Fourier coefficients of piecewise constant signals whose discontinuities are localized to band-limited curves satisfy these annihilation conditions [6, 8], and that their recovery translates to a structured matrix completion problem [7]. Since the above model can exploit the smoothness of the discontinuities along with sparsity, their use results in a significant improvement in reconstruction quality over traditional methods such as total variation (TV) minimization.

Despite the improvement in performance, the main challenge associated with the above structured low-rank algorithms is the dramatic increase in computational complexity over current methods such as TV minimization. The direct extension of these algorithms to multidimensional applications (e.g. dynamic MRI) will be prohibitively slow. Specifically, these schemes involve the recovery of a large-scale structured matrix (e.g. Hankel, Toeplitz), whose combined dimensions are two to three orders of magnitude higher than those of the image. Typical algorithmic approaches require a singular value decomposition (SVD) of the dense large-scale matrix at each iteration. Several strategies have been introduced to minimize or avoid SVD's. For example, [2, 4] rely on a low-rank assumption to replace the full SVD's with more efficient truncated SVD's or matrix inversions. However, strict low-rank approximations in this case can lead to slow to convergence, sensitivity to local minima, and suboptimal results compared to their unapproximated counterparts.

We introduce a novel, fast algorithm for the above class of problems that does not require overly strict low-rank approximations. Its distinguishing aspect is the direct exploitation of the convolution structure of the Hankel/Toeplitz matrices; none of the current algorithms fully exploit this structure. This enables us to evaluate the intermediate steps efficiently using fast Fourier transforms in the non-lifted domain, resulting in an algorithm with significant reductions in memory demand and computational complexity. Our approach is based on the iterative reweighted least squares (IRLS) algorithm for low-rank matrix completion [9, 10], which we adapt to the structured matrix setting. Our experiments show that the IRLS strategy significantly reduces the number of SVD's that have to be computed. Further, we show the subproblems of the algorithm simplify to (1) the robust estimation an annihilating filter, and (2) a least squares linear prediction step; this may be viewed as an iterative version of our two-step recovery scheme in [6].

2. BACKGROUND

To motivate the matrix recovery approach to MRI reconstruction, consider as a simple image model the sparse linear com-

bination of Dirac delta functions in spatial domain:

$$\rho(\mathbf{r}) = \sum_{k=1}^{K} c_k \, \delta(\mathbf{r} - \mathbf{r}_k), \ \ \forall \mathbf{r} \in [0,1]^2. \tag{1}$$

It can be shown that $\rho(\mathbf{r})\mu(\mathbf{r}) = 0$, where $\mu(\mathbf{r})$ is a trigonometric polynomial of degree K with zeros at locations $\{\mathbf{r}_k\}_{k=1}^{K}$. This spatial domain relation translates to the classical finite-rate-of-innovation annihilation conditions [11]

$$\hat{\rho}[\mathbf{k}] * \hat{\mu}[\mathbf{k}] = 0, \tag{2}$$

where $\widehat{\rho}$ and $\widehat{\mu}$ denote the Fourier coefficients of ρ and μ, respectively. The annihilation relationship can be expressed as a matrix relation

$$\mathcal{T}(\hat{\rho}) \, \hat{\boldsymbol{\mu}} = 0, \tag{3}$$

where $\mathcal{T}(\hat{\rho})$ is a Toeplitz matrix built from the entries of $\hat{\rho}$ and $\hat{\boldsymbol{\mu}}$ is the finite vector of Fourier coefficients of μ. Note that any polynomial $\eta(\mathbf{r}) = \mu(\mathbf{r})\gamma(\mathbf{r})$ will also satisfy the above annihilation condition, i.e. $\mathcal{T}(\hat{\rho})\hat{\boldsymbol{\eta}} = 0$. This implies that $\mathcal{T}(\hat{\rho})$ is a low-rank matrix. When we only have partial samples of $\hat{\rho}[\mathbf{k}]$ for $\mathbf{k} \in \Gamma$, we can use the low-rank property of $\mathcal{T}(\hat{\rho})$ to recover $\hat{\rho}$ as the solution to the following matrix completion problem:

$$\min_{\hat{\eta}} \ \text{rank}[\mathcal{T}(\hat{\eta})] \quad \text{such that} \quad \hat{\eta}[\mathbf{k}] = \hat{\rho}[\mathbf{k}], \ \ \forall \mathbf{k} \in \Gamma \tag{4}$$

We have shown that piecewise constant signals in two dimensions also satisfy an annihilation condition similar to (1), provided the signal discontinuities are localized to the zeroset of a smooth band-limited function $\mu(\mathbf{r})$; see [6,7] for details. In [7] we proposed to recover the Fourier coefficients of the piecewise smooth image $\rho(\mathbf{r})$ from undersampled measurements using a lifting formulation similar to (4), where instead we take $\mathcal{T}(\hat{\rho})$ to be the matrix obtained by stacking block Toeplitz matrices built from the Fourier coefficients of the partial derivatives of ρ. See also [2–5] for similar matrix lifting formulations based alternative image priors.

3. PROPOSED ALGORITHM FOR STRUCTURED LOW-RANK MATRIX RECOVERY

We derive a fast algorithm for the structured low-rank matrix recovery problem:

$$\min_{\mathbf{x}} \ \text{rank}[\mathcal{T}(\mathbf{x})] \quad s.t. \quad \mathcal{P}_\Gamma(\mathbf{x}) = \mathbf{b}, \tag{5}$$

where $\mathbf{x} \in \mathbb{C}^m$ represents a vectorized array of the k-space data to be recovered, $\mathcal{T}(\mathbf{x})$ is a structured lifting of $\mathbf{x}$ to a matrix in $\mathbb{C}^{M \times N}$, and $\mathbf{b}$ are the known k-space samples. In this work, we focus on the case where $\mathcal{T}(\mathbf{x})$ is a Toeplitz-like matrix encoding a convolution relationship of the form

$$\mathcal{T}(\mathbf{x})\mathbf{h} = \mathcal{P}_\Omega(\mathbf{Mx} * \mathbf{h}). \tag{6}$$

Here $\mathbf{M}$ is any linear transformation in k-space, and $\mathbf{Mx} * \mathbf{h}$ represents 2-D discrete convolution of $\mathbf{Mx}$ with an FIR filter $\mathbf{h}$ of fixed size, and $\mathcal{P}_\Omega$ is projection onto a finite set Ω of valid indices in k-space. The model (6) is general enough to incorporate the structured matrix liftings proposed in [3,4,7] with appropriate choices of $\mathbf{M}$ and Ω. For example, the piecewise constant image model of the authors in [7], $\mathbf{M}$ is multiplication by Fourier derivatives $-j2\pi k_x$ or $-j2\pi k_y$ where $\mathbf{k} = (k_x, k_y)$ is any k-space location; we call this the *gradient weighted* lifting.

We pose the recovery of $\mathbf{x}^\star$ from its noisy and undersampled measurements as

$$\mathbf{x}^\star = \arg\min_{\mathbf{x}} \|\mathcal{T}(\mathbf{x})\|_p + \frac{\lambda}{2}\|\mathcal{P}_\Gamma(\mathbf{x}) - \mathbf{b}\|_2^2, \tag{7}$$

where λ is a regularization parameter, and $\|\cdot\|_p$ denotes the family of Shatten p-norms, $0 \leq p \leq 1$, which is defined for an arbitrary matrix $\mathbf{X}$ by $\|\mathbf{X}\|_p = \frac{1}{p}Tr[(\mathbf{X}^*\mathbf{X})^{\frac{p}{2}}] = \frac{1}{p}\sum_i \sigma_i^p$, where σ_i are the singular values of $\mathbf{X}$. Additionally, we define the penalty $\|\mathbf{X}\|_0 = \sum_i \log(\sigma_i)$, which can be viewed as the limiting case of $\|\mathbf{X}\|_p$ when $p \to 0$. Note that $\|\cdot\|_1$ corresponds to the convex nuclear norm penalty, but the penalty $\|\cdot\|_p$ is non-convex for $0 \leq p < 1$.

We adapt an iteratively reweighted least squares (IRLS) approach to minimizing (7), originally proposed in [9,10]. The IRLS approach is motivated by the observation

$$\|\mathbf{X}\|_p = Tr[(\mathbf{X}^*\mathbf{X}) \underbrace{(\mathbf{X}^*\mathbf{X})^{\frac{p}{2}-1}}_{\mathbf{H}}] = \|\mathbf{X}\mathbf{H}^{\frac{1}{2}}\|_F^2,$$

supposing $\mathbf{X}$ has no zero singular values. This suggests an iterative algorithm that alternates between updating a weight matrix $\mathbf{H}$ and solving a weighted least squares problem. In particular, substituting $\mathbf{X} = \mathcal{T}(\mathbf{x})$ in the IRLS-p algorithm presented in [10] gives following iterative scheme for solving (7): initialize $\mathbf{x}_0 = \mathcal{P}_\Gamma^*\mathbf{b}$ and at the nth iteration compute

$$\mathbf{H} = [\mathcal{T}(\mathbf{x}_{n-1})^*\mathcal{T}(\mathbf{x}_{n-1}) + \epsilon_n\mathbf{I}]^{\frac{p}{2}-1} \tag{8}$$

$$\mathbf{x}_n = \arg\min_{\mathbf{x}} \|\mathcal{T}(\mathbf{x})\mathbf{H}^{\frac{1}{2}}\|_F^2 + \frac{\lambda}{2}\|\mathcal{P}_\Gamma(\mathbf{x}) - \mathbf{b}\|_2^2 \tag{9}$$

Here $\epsilon_n \to 0$ is a small parameter used to stabilize the inverse computed in (8).

We now show with a careful consideration of the structure (6) it is possible to simplify steps (8) and (9) to give a fast algorithm. In particular we show the iterates (8) and (9) can be interpreted as alternating between the estimation of an annihilating filter in (8) and enforcing an annihilation constraint in (9). This interpretation leads naturally to a novel approximation which greatly reduces the computational complexity. We call the proposed IRLS approach with these simplifications the *Generic Iteratively Reweighted Annihilating Filter* (GIRAF) algorithm.

4. SIMPLIFICATIONS

4.1. Step one: Least-squares annihilation

First, we focus on simplifying the least-squares problem (9). Expanding the Frobenius norm in terms of the columns $\mathbf{h}_i$ of $\mathbf{H}^{\frac{1}{2}}$, (9) is equivalent to

$$\min_{\mathbf{x}} \sum_{i=1}^{N} \|\mathcal{T}(\mathbf{x})\mathbf{h}_i\|_2^2 + \lambda\|\mathcal{P}_{\Gamma}(\mathbf{x}) - \mathbf{b}\|_2^2. \quad (10)$$

From the structure assumption (6) and the associativity of convolution, we have

$$\mathcal{T}(\mathbf{x})\mathbf{h}_i = \mathcal{P}_{\Omega}(\mathbf{Mx} * \mathbf{h}_i) = \mathcal{P}_{\Omega}(\mathbf{h}_i * \mathbf{Mx}) = \mathbf{PC}_i\mathbf{Mx} \quad (11)$$

for all $i = 1, \ldots, N$, where $\mathbf{C}_i$ is linear convolution by the filter $\mathbf{h}_i$, and $\mathbf{P}$ is a matrix representation of $\mathcal{P}_{\Omega}$. Because of the projection, we can assume $\mathbf{C}_i$ is a *circular* convolution on a sufficiently large grid, i.e. $\mathbf{C}_i$ is (block) circulant. Therefore we may write $\mathbf{C}_i = \mathbf{FS}_i\mathbf{F}^*$ where $\mathbf{F}$ is the 2-D DFT and $\mathbf{S}_i$ is a diagonal matrix representing pointwise multiplication by the inverse DFT of the zero-padded filter $\mathbf{h}_i$. From (11), computing the gradient of the objective in (10) and setting it equal to zero yields the linear system:

$$\underbrace{\left(\sum_{i=1}^{N}(\mathbf{M}^*\mathbf{C}_i^*\mathbf{P}^*\mathbf{PC}_i\mathbf{M}) + \lambda\mathcal{P}_{\Gamma}^*\mathcal{P}_{\Gamma}\right)}_{\mathbf{R}}\mathbf{x} = \lambda P_{\Gamma}^*\mathbf{b} \quad (12)$$

Observe that solving (12) with CG iterations will be computationally costly: forming matrix-vector products with $\mathbf{R}$ requires $4N$ FFTs, where N is the product of the filter dimensions. However, if the projection set Ω is large relative to the filter size, the projection operator $\mathbf{P}^*\mathbf{P}$ will be close to identity. Hence, we propose making the approximation $\mathbf{P}^*\mathbf{P} \approx \mathbf{I}$, so that an approximation of $\mathbf{R}$ is

$$\widetilde{\mathbf{R}} = \mathbf{M}^*\mathbf{F}\left(\sum_{i=1}^{N}\mathbf{S}_i^*\mathbf{S}_i\right)\mathbf{F}^*\mathbf{M} + \lambda\mathcal{P}_{\Gamma}^*\mathcal{P}_{\Gamma} \quad (13)$$

To form matrix-vector products with $\widetilde{\mathbf{R}}$ we only need 2 FFTs after pre-computing the diagonal matrix $\sum_i \mathbf{S}_i^*\mathbf{S}_i$. We propose solving (12) with $\widetilde{\mathbf{R}}$ in place of $\mathbf{R}$. In our experiments we show this approximation introduces only negligible errors into the iterations. Finally, note that in spatial domain the entries of $\sum_i \mathbf{S}_i^*\mathbf{S}_i$ corresponds to pointwise multiplication with gridded samples of the sum-of-squares function

$$\overline{\mu}(\mathbf{r}) = \sum_{i=1}^{N}|\mu_i(\mathbf{r})|^2, \quad \forall(\mathbf{r}) \in [0, 1]^2 \quad (14)$$

where $\mu_i(\mathbf{r})$ is the trigonometric polynomial given by the inverse Fourier transform of $\mathbf{h}_i$. This shows our approximation effectively condenses the collection of linearly independent annihilating filters $\mathbf{h}_i$ into a single spatial domain annihilating function (14).

	15×15 filter (SNR=13.8dB)			31×31 filter (SNR=16.1dB)		
Algorithm	# iter	total	svd cost/iter	# iter	total	svd cost/iter
SVT	8	110s	15.4s	12	790 s	63.4 s
GIRAF	6	20s	0.03s	7	44 s	0.3 s

Table 1: Convergence comparisons of singular value thresholding (SVT) and the proposed algorithm (GIRAF) for structured low-rank matrix recovery via nuclear norm minimization ((7) with $p = 1$). Convergence was judged as $MSE = \|\mathbf{x}^{\star} - \mathbf{x}_n\|^2/\|\mathbf{x}^{\star}\|^2 < 10^{-4}$ where $\mathbf{x}_n$ is the current iterate and $\mathbf{x}^{\star}$ is the solution obtained from 50 iterations of SVT.

4.2. Step two: Annihilating filter update

Now we show how to efficiently compute the weight matrix update (8). According to the previous section, the update (9) becomes a least squares annihilation subject to a filter specified by the columns of $\mathbf{H}^{\frac{1}{2}} = [\mathbf{h}_1, ..., \mathbf{h}_N]$. Fix an iterate $\mathbf{x} = \mathbf{x}_n$, and let $(\mathbf{V}, \mathbf{\Lambda})$ be the eigen-decomposition of $\mathcal{T}(\mathbf{x})^*\mathcal{T}(\mathbf{x})$, i.e. $\mathbf{V}$ is a orthogonal basis of eigenvectors $\mathbf{v}_i$ and $\mathbf{\Lambda}$ is diagonal matrix of the associated eigenvalues λ_i satisfying $\mathcal{T}(\mathbf{x}) = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^*$. Then (8) reduces to

$$\mathbf{H} = [\mathbf{V}(\mathbf{\Lambda} + \epsilon\mathbf{I})\mathbf{V}^*]^{\frac{p}{2}-1} = \mathbf{V}(\mathbf{\Lambda} + \epsilon\mathbf{I})^{\frac{p}{2}-1}\mathbf{V}^*.$$

Therefore, one choice of the matrix square root $\mathbf{H}^{\frac{1}{2}}$ is

$$\mathbf{H}^{\frac{1}{2}} = \mathbf{V}(\mathbf{\Lambda} + \epsilon\mathbf{I})^{\frac{p}{4}-\frac{1}{2}} = [\alpha_1^{\frac{1}{2}}\mathbf{v}_1, \ldots, \alpha_N^{\frac{1}{2}}\mathbf{v}_N].$$

where $\alpha_i = (\lambda_i + \epsilon)^{\frac{p}{2}-1}$. However, with our proposed approximation, step one of the algorithm only needs gridded samples of the sum-of-squares function (14), which in this case can be efficiently computed using N FFTs as
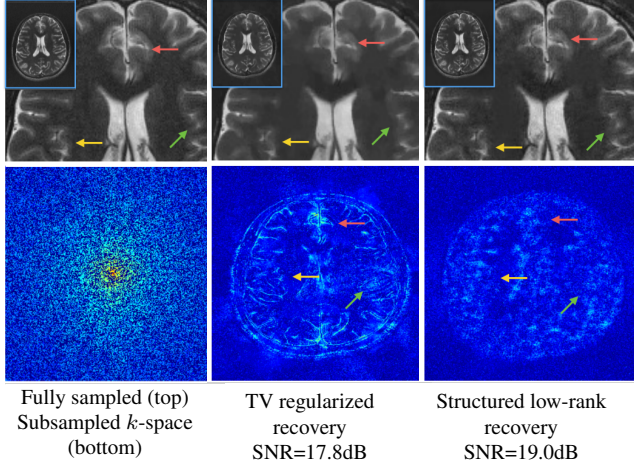
$$\overline{\mu}(\mathbf{r}) = \sum_{i=1}^{N}\alpha_i|\gamma_i(\mathbf{r})|^2 \quad (15)$$

where γ_i is the inverse Fourier transform of the filter $\mathbf{v}_i$. Note that the weight α_i only when $\mathbf{v}_i$ is close to the null space of $\mathcal{T}(\mathbf{x})$, i.e. when $\mathbf{v}_i$ is an annihilating filter. Therefore, we can interpret $\overline{\mu}$ as a regularized annihilating function that combines information from the entire null space of $\mathcal{T}(\mathbf{x})$.

5. EXPERIMENTS

5.1. Speedup of proposed GIRAF algorithm

To evaluate the potential speedup of GIRAF, we compare the computation time required to solve the problem (7) in the case of nuclear norm minimization ($p = 1$) against the singular value thresholding (SVT) algorithm [12]; see Table 1. For simplicity we restrict our comparisons to the matrix lifting (6) with gradient weighting, and reconstruct a synthetic brain phantom onto a 255×255 k-space grid from uniform random noiseless k-space data at 1.5-fold undersampling, where we

Fully sampled (top)
Subsampled k-space
(bottom)

TV regularized
recovery
SNR=17.8dB

Structured low-rank
recovery
SNR=19.0dB

Fig. 1: Recovery of single-coil MR data from 2-fold acceleration using the proposed GIRAF algorithm for structured low-rank completion. Error images shown below.
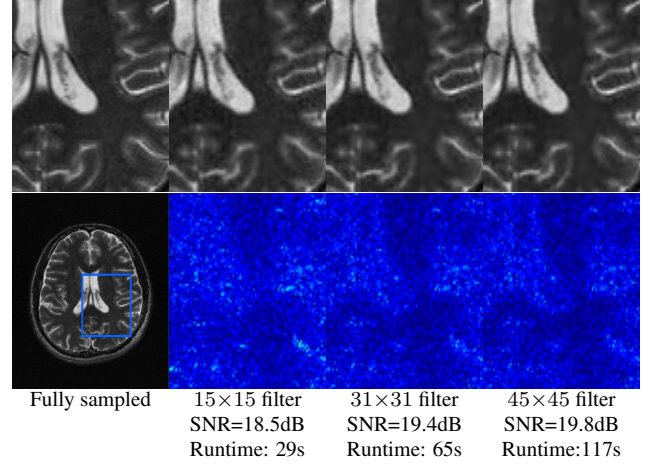

Fully sampled

15×15 filter
SNR=18.5dB
Runtime: 29s

31×31 filter
SNR=19.4dB
Runtime: 65s

45×45 filter
SNR=19.8dB
Runtime:117s

Fig. 2: Effect of filter size on reconstruction; zoomed for detail. Error images shown below.

fixed the regularization parameter to $\lambda = 10^{-10}$. Both algorithms were implemented in MATLAB on CPU for fair comparison. Observe the GIRAF algorithm is able to reach the same convergence tolerance as SVT with fewer iterations, at significantly less cost per iteration, and shows improved scalability with filter size. This is beacuse GIRAF only needs an $N \times N$ eigen-decomposition per iteration and not an $M \times N$ SVD, $M >> N$, as in the SVT algorithm.

5.2. Reconstruction from gradient weighting

In Fig. (1) we demonstrate the recovery of real MR data (225×225 sampling grid with 4 receiver coils, coil-compressed to a single channel) from 2-fold uniform random undersampling, using the gradient weighted lifting scheme and the proposed GIRAF algorithm (non-convex penalty $p = 0$, filter size 45×45). Despite the comparatively large filter size, the runtime was 132s with the proposed algorithm. We observe the recovery improves significantly over a standard TV reconstruction. In Fig. (2) we demonstrate the effect of filter-size on the recovery of the same dataset from 5-fold accelerated variable density random sampling. The reconstruction has noise-like artifacts which diminish in appearance as the filter size grows, justifying the use of larger filters enabled by the GIRAF algorithm.

6. DISCUSSION AND CONCLUSION

We introduced a novel algorithm with general applicability to structured low-rank matrix completion problems arising in MR image reconstruction. The method is computationally faster than previous approaches, and does not require any strict low-rank assumptions. This enables extensions of these methods to a much wider class of problems.

7. REFERENCES

[1] P. J. Shin, P. E. Larson, M. A. Ohliger, M. Elad, J. M. Pauly, D. B. Vigneron, and M. Lustig, "Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion," *Magnetic Resonance in Medicine*, 2013.

[2] J. P. Haldar, "Low-rank modeling of local k-space neighborhoods (loraks) for constrained mri." *Medical Imaging, IEEE Transactions on*, vol. 33, no. 3, pp. 668–681, 2014.

[3] ——, "Low-rank modeling of local k-space neighborhoods: from phase and support constraints to structured sparsity," in *SPIE Optical Engineering+ Applications*, 2015.

[4] K. H. Jin, D. Lee, and J. C. Ye, "A general framework for compressed sensing and parallel MRI using annihilating filter based low-rank hankel matrix," *arXiv preprint arXiv:1504.00532*, 2015.

[5] ——, "A novel k-space annihilating filter method for unification between compressed sensing and parallel mri," in *International Symposium on Biomedical Imaging (ISBI)*, April 2015.

[6] G. Ongie and M. Jacob, "Super-resolution MRI using finite rate of innovation curves," *IEEE International Symposium on Biomedical Imaging (ISBI)*, April 2015.

[7] ——, "Recovery of piecewise smooth images from few fourier samples," in *Sampling Theory and Applications (SampTA)*, May 2015, pp. 543–547.

[8] H. Pan, T. Blu, and P. L. Dragotti, "Sampling curves with finite rate of innovation," *Signal Processing, IEEE Transactions on*, vol. 62, no. 2, 2014.

[9] M. Fornasier, H. Rauhut, and R. Ward, "Low-rank matrix recovery via iteratively reweighted least squares minimization," *SIAM Journal on Optimization*, vol. 21, no. 4, pp. 1614–1640, 2011.

[10] K. Mohan and M. Fazel, "Iterative reweighted algorithms for matrix rank minimization," *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 3441–3473, 2012.

[11] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *Signal Processing, IEEE Transactions on*, vol. 50, no. 6, pp. 1417–1428, 2002.

[12] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.