FAST WAVELET-BASED SINGLE-PARTICLE RECONSTRUCTION IN CRYO-EM

Cédric Vonesch[1], **Lanhui Wang**[1], **Yoel Shkolnisky**[2], and **Amit Singer**[1]

[1]Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ, USA

[2]Department of Applied Mathematics, Tel Aviv University, Tel Aviv, Israel

Abstract

This paper presents a novel algorithm for the 3D tomographic inversion problem that arises in single-particle electron cryo-microscopy (Cryo-EM). It is based on two key components: 1) a variational formulation that promotes sparsity in the wavelet domain and 2) the Toeplitz structure of the combined projection/back-projection operator. The first idea has proven to be very effective for the recovery of piecewise-smooth signals, which is confirmed by our numerical experiments. The second idea allows for a computationally efficient implementation of the reconstruction procedure, using only one circulant convolution per iteration.

Index Terms

Inverse problem; electron cryo-microscopy (Cryo-EM); single particle reconstruction; projection-slice theorem; 3D; wavelets; sparsity; one-norm regularization; Toeplitz structure; non-uniform FFT

1. INTRODUCTION

Together with X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, electron cryo-microscopy (Cryo-EM) is one of the key techniques for determining the structure of biological macromolecules. Cryo-EM can achieve resolutions in the nanometer range and can be used for imaging a variety of subcellular components [1]. While X-ray crystallography can provide sub-atomic resolutions, Cryo-EM has the advantage of a simpler specimen preparation, as it does not require crystallization. It is also more suitable for imaging large molecules than NMR.

One of the main challenges of Cryo-EM lies in the high radiation sensitivity of biological samples. Even when using a low-dose mode, imaging a sample ultimately leads to its destruction. The maximum tolerance is on the order of 5 to 10 electrons per square Ångström, leading to signal-to-noise ratios (SNR) that are sometimes well below unity [2].

3D cryo-EM imaging techniques can be divided into two main categories: 1) electron tomography, where the specimen is physically tilted along a single axis and imaged several times, and 2) "single-particle" reconstruction, where the sample contains multiple copies of the same molecule under different orientations, which are imaged only once. The first category has the advantage that the projection angles are known, thus allowing for standard tomographic inversion procedures. For the second category, the orientations of the molecules are usually unknown. While this requires a broader computational framework for the reconstruction, it potentially allows for improved SNRs (through data averaging) and higher resolutions (thanks to the lower radiation and absence of a "missing-frequency wedge").

The main steps of a full single-particle reconstruction procedure can be summarized as follows.

1. Preprocessing: the individual molecules are detected and extracted from the acquired data.

2. Class averaging: images with high correlation are averaged together to improve their SNR.

3. Orientation estimation: this is done either via finding common lines in the frequency domain (based on the projection-slice theorem stated below) or via reprojection (based on a correlation measurement with templates generated from a previous reconstruction). This step may also include a shift estimation.

4. Reconstruction: a 3D image is generated by a tomographic inversion algorithm.

5. Iterative refinement: the whole procedure is repeated from step 3, using the outcome of the most recent reconstruction to improve the results.

The diversity of pattern-recognition and image-processing problems that need to be solved at each of these steps constitutes very fertile grounds for algorithm development. Here we will focus on the reconstruction step.

The reconstruction algorithms that are most commonly found in the literature—see [3] for a recent review—are either based on a direct inversion in the frequency-domain (filtered back-projection, gridding) or on a quadratic variational formulation. In particular, least squares and regularized least squares formulations are still very much in use, typically leading to linear reconstruction procedures such as the Landweber algorithm or the Kaczmarz (ART) method.

The present work has two main aspects. First, we use a non-quadratic regularization that favors estimates with a sparse wavelet expansion. Wavelet-domain sparsity has proven to be successful paradigm for a wide range of other inverse problems—see e.g. [4] for a recent application to optical microscopy. It leads to a non-linear variant of the Landweber algorithm that incorporates a wavelet-domain thresholding operation at every iteration.

Second, we take advantange of the Toeplitz structure of the iteration operator, which allows for a computationally effective implementation of the algorithm. Furthermore, we precompute recurring quantities using non-uniform FFTs [5]. The Toeplitz property appears in [5] and has been used e.g. for 2D magnetic-resonance imaging [6, 7, 8], but to our knowledge it is not commonly exploited in the context of 3D Cryo-EM.

2. STRUCTURE OF THE IMAGING MODEL

2.1. Fourier-transform conventions

We will use the following definition for the Fourier transform of a D-dimensional function $f(\mathbf{x}) = f(x_1, \ldots, x_D)$:

$$\hat{f}(\omega_1, \ldots, \omega_D) = \hat{f}(\omega) = \int_{\mathbb{R}^D} f(\mathbf{x}) \, e^{-i\omega^T \mathbf{x}} \, \mathrm{d}\mathbf{x}.$$

Given an even integer $N > 0$, we define $S_N = \{-N/2, \ldots, N/2 - 1\}$. In this paper we will consider D-dimensional discrete signals indexed in S_N^D. To distinguish them from functions, we will use square brackets, such as in $f[\mathbf{n}] = f[n_1, \ldots, n_D]$. With these conventions, the Discrete Fourier Transform (DFT) is defined as

$$\hat{f}[k_1, \ldots, k_D] = \hat{f}[\mathbf{k}] = \sum_{\mathbf{n} \in S_N^D} f[\mathbf{n}] \, e^{-i 2\pi \mathbf{k}^T \mathbf{n}/N}.$$

2.2. The projection-slice theorem

In the context of electron microscopy, a molecule is characterized by its Coulomb potential, a function of the three spatial coordinates $f(\mathbf{x}) = f(x_1, x_2, x_3)$. A simple model for Cryo-EM imaging is a projection along the axial direction of the microscope, say x_3:

$$\mathscr{P}f(x_1, x_2) = \int_{\mathbb{R}} f(x_1, x_2, x_3) \, \mathrm{d}x_3.$$

In practice we observe a large number of projections of the same molecule under different (random) orientations. Each orientation is associated with a certain rotation matrix $\mathbf{R}$ and its corresponding rotation operator $\mathscr{R}f(\mathbf{x}) = f(\mathbf{R}\mathbf{x})$. With the above definition of the Fourier transform, we have:

Property 1 (Projection-slice theorem).

$$\widehat{\mathscr{P}\mathscr{R}}f(\omega_1, \omega_2) = \widehat{\mathscr{R}f}(\omega_1, \omega_2, 0)$$

In the sequel we will index the rotation operators and matrices corresponding to the different orientations of the molecule by $p \in \{1, \ldots, P\}$.

2.3. Discretization

Let us assume that $f(\mathbf{x})$ is supported in the open ball $B(r) = \{\mathbf{x} \in \mathbb{R}^3 \mid \|\mathbf{x}\| < r\}$, and that we measure the samples

$$g_p[\mathbf{n}] = \mathscr{P}\mathscr{R}_p f(2r\mathbf{n}/N) \text{ for } \mathbf{n} \in S_N^2.$$

From these measurements, our goal will be to reconstruct the samples

$$f[\mathbf{n}] = f(2r\mathbf{n}/N) \text{ for } \mathbf{n} \in S_N^3.$$

When N is sufficiently large, the continuous Fourier transforms in the projection-slice theorem can be approximated using Riemann sums. This leads to the following model (which is in terms of the DFTs of the measurements).

Property 2 (Discretized imaging model).

For $\mathbf{k} = [k_1, k_2] \in S_N^2$, define $\omega_{p,\mathbf{k}} = 2\pi \mathbf{R}_p [k_1 \ k_2 \ 0]^T/N$. Then $\hat{g} = Af + \hat{b}$, where the operator A is defined as

$$(Af)_p[\mathbf{k}] = \frac{r}{N} \sum_{\mathbf{n} \in S_N^3} f[\mathbf{n}] \, e^{-i\omega_{p,\mathbf{k}}^T \mathbf{n}.}$$

and $\hat{b}_p[\mathbf{k}]$ represents measurement and discretization errors.

In the remainder of this paper, the symbol f will refer to the discrete signal $f[\mathbf{n}]$.

3. WAVELET-REGULARIZED RECONSTRUCTION

Standard variational approaches to single-particle reconstruction are based on the minimization of a cost functional of the form

$$\mathscr{C}(f) = \|\hat{g} - Af\|_{\ell_2}^2 + \lambda\mathscr{Q}(f),$$

where $\mathscr{Q}(f)$ is a quadratic regularization term. A typical choice is the squared ℓ_2 norm of a discretized Laplacian applied to f, which favors smooth solutions as λ increases.

In this paper we will set $\mathscr{Q}(f) = \|Wf\|_{\ell_1}$, where W represents a wavelet decomposition operator (for simplicity, we will assume that the wavelet basis is orthonormal). The key idea is that, compared to a standard ℓ_2 norm, the ℓ_1 norm behaves as follows: small coefficients are more penalized and large coefficients are less penalized. Thus we favor solutions whose energy is concentrated in a small number of large wavelet coefficients.

3.1. The thresholded Landweber algorithm

Using results from the theory of non-smooth convex optimization [9], it can be shown that f is a minimizer if and only if it is a fixed point of the non-linear operator

$$U\{f\} = W^* T_{\lambda\tau/2}\{W(f - \tau A^*Af + \tau A^*\hat{g})\}.$$

Here τ is a step-size parameter, A^* is the adjoint of A and $T_{\lambda\tau/2}$ denotes the a soft-thresholding operation (see e.g. [4]). This leads to a non-linear iterative reconstruction procedure defined by $f^{(k+1)} = U\{f^{(k)}\}$. Note that without thresholding ($\lambda = 0$), one recovers the standard Landweber iteration.

An important observation is that the computation of $h = A^*\hat{g}$ can be done once and for all in the beginning and thus the iteration essentially boils down to the application of the symmetrized operator A^*A. The adjoint of A can be expressed as

$$h[\mathbf{n}] = A^*\hat{g}[\mathbf{n}] = \frac{r}{N}\sum_{p,\mathbf{k}}\hat{g}_p[\mathbf{k}]\, e^{i\omega_{p,\mathbf{k}}^T \mathbf{n}}. \tag{1}$$

For the operator A^*A, we will use the following result:

Property 3 (Toeplitz structure of the symmetrized operator).

$$A^*Af[\mathbf{n}] = \sum_{\mathbf{k}\in S_N^3} f[\mathbf{k}]c[\mathbf{n} - \mathbf{k}],$$

where, for $\mathbf{n} \in S_{2N}^3$,

$$c[\mathbf{n}] = \frac{r^2}{N^2} \sum_{p,\mathbf{k}} e^{i\omega_{p,\mathbf{k}}^T \mathbf{n}}. \tag{2}$$

This property allows for a fast implementation of the symmetrized operator: once the convolution kernel $c[\mathbf{n}]$ has been precomputed, it is sufficient to extend $f[\mathbf{n}]$ to the domain S_{2N}^3 by zero-padding, to perform a circular convolution with $c[\mathbf{n}]$, and to restrict the result back to S_N^3. The circulant convolution can be computed efficiently in the frequency domain using the FFT algorithm.

3.2. Back-projection and kernel computation

It remains to explain how to compute the back-projection $h[\mathbf{n}]$ and the kernel $c[\mathbf{n}]$ efficiently. Formula (1), as well as its particular case (2), is not a standard DFT because the frequencies $\omega_{p,\mathbf{k}}$ do not necessarily lie on the grid $\frac{2\pi}{N} S_N^3$. We thus use the (type-1) non-uniform FFT algorithm [5], which is based on the following steps:

1. interpolate the frequency coefficients on a (usually finer-resolution) uniform grid (Cost: $\mathcal{O}N^2P$);

2. apply a uniform FFT algorithm (Cost: $\mathcal{O}N^3 \log N$);

3. compensate for the effect of the interpolation procedure by a coefficient-wise division (Cost: $\mathcal{O}N$).

In practice the interpolation step often has the dominant computational cost. It requires choosing a window function whose spatial expression $w(\mathbf{x})$ and Fourier transform $\hat{w}(\omega)$ are known analytically (typically a Gaussian). To compute (1) we can then define

$$\hat{a}(\omega) \sum_{p,\mathbf{k}} \hat{g}_p[\mathbf{k}] \hat{w}(\omega - \omega_{p,\mathbf{k}}),$$

whose inverse Fourier transform is

$$\frac{1}{(2\pi)^D} \int_{\mathbb{R}^D} \hat{a}(\omega) \, e^{i\omega^T \mathbf{n}} d\omega = w(\mathbf{x}) \sum_{p,\mathbf{k}} \hat{g}_p[\mathbf{k}] \, e^{i\omega_{p,\mathbf{k}}^T \mathbf{x}}.$$

In particular, for $\mathbf{x} = \mathbf{n} \in S^D$, we can write that

$$\frac{1}{(2\pi)^D} \int_{[-\pi,\pi]^D} \tilde{a}(\omega) \, e^{i\omega^T \mathbf{n}} d\omega = w(\mathbf{n}) h[\mathbf{n}],$$

where $\tilde{a}(\omega) = \sum_{\mathbf{k} \in \mathbb{Z}^D} \hat{a}(\omega - 2\pi \mathbf{k})$.

Replacing the last integral by a Riemann sum leads to

$$h[\mathbf{n}] \approx \frac{1}{w(\mathbf{n}) M^D} \sum_{\mathbf{m} \in S_M^3} \tilde{a}(2\pi \mathbf{m}/M) \, e^{i2\pi \mathbf{m}^T \mathbf{n}/M},$$

where M is a (small) multiple of N that can be freely chosen. The above sum is an inverse DFT of the coefficients $\tilde{a}[\mathbf{m}] = \tilde{a}(2\pi\mathbf{m}/M)$, which are obtained through a non-uniform interpolation procedure (with periodic boundary conditions).

4. NUMERICAL EXPERIMENTS

We now present the results of two numerical experiments on synthetic data of size $N^3 = 64^3$. To simulate measurements, the forward imaging model was implemented using a type-2 non-uniform FFT algorithm (see [5]). We used $P = 500$ projections and Gaussian white noise of variance σ^2 was added to each of them. To have a relative measure of the noise level, we define the signal-to-noise ratio (SNR) as (Variance of the projections)/σ^2.

In the first experiment, we used a 3D modified Shepp-Logan phantom as the original data f. We compared the performance of the thresholded Landweber (TL) algorithm with a simple least-squares (LS) formulation obtained for $\lambda = 0$, as well as a regularized least-squares (RLS) formulation with a Laplacian regularization. The latter was implemented using the steepest-descent algorithm described by Penczek in [3]. The initial estimate $f^{(0)}$ was set to zero for all algorithms. We assigned a fixed budget of 50 iterations to each algorithm and kept the estimate $f^{(k)}$ for which the SNR improvement $20 \log_{10} (\|f^{(0)} - f\|/\|f^{(k)} - f\|)$ was the highest. For the RLS and TL algorithms, λ was adjusted so as to maximize this quantity. For TL, we used Haar wavelets with three decomposition levels, combined with the random-shift technique described in [4]. Table 1 lists the results for three different input SNRs. It is seen that the TL algorithm consistently yields better estimates.

In the second experiment, we used the 3D image of a ribosome as the original data, and the input SNR was set to 1/8. This time we used Symlet4 wavelets for regularization. Visually the result of the TL algorithm contains less background noise than the one obtained with RLS. This is illustrated in Fig. 1, which shows cross-sections through the different data volumes. The SNR improvements confirm this impression (RLS: 13.17 dB; TL: 14.26 dB). In Fig. 2, a plot of the Fourier shell correlation (FSC) for this experiment suggests that the TL algorithm achieves a slightly better resolution (see e.g. [2] for a discussion of the FSC).

5. CONCLUSION

We have presented an efficient implementation of the thresholded Landweber algorithm using the specific structure of the single-particle reconstruction problem in Cryo-EM. It requires only two initial non-uniform FFTs for computing the back-projection and the Toeplitz kernel associated with the symmetrized operator appearing in the iteration. The actual iterations are performed with standard FFTs.

The first results obtained here confirm the potential of wavelet regularization for 3D Cryo-EM imaging. The algorithmic refinements that we have used carry over to the case where the Contrast Transfer Function (CTF) of the system must be taken into account.
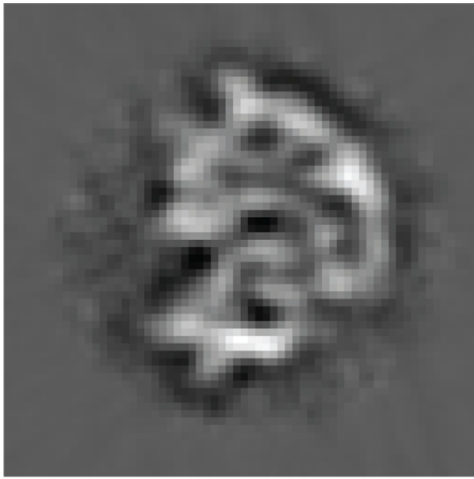
In the future we plan to perform more extensive comparisons with standard single-particle-reconstruction software packages. Ultimately the goal is to incorporate the algorithm into a full iterative-refinement loop. In this framework it is certainly worth investigating further acceleration methods such as preconditioning or the multilevel scheme described in [4].
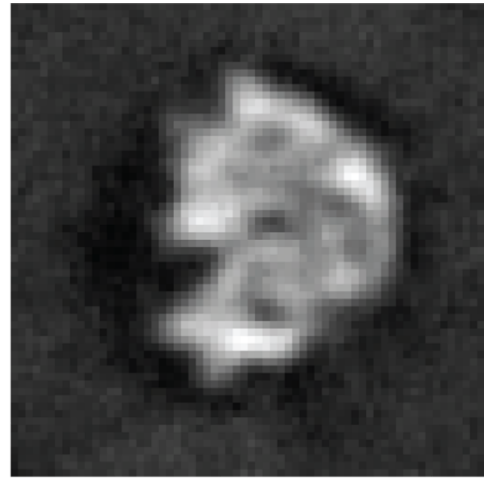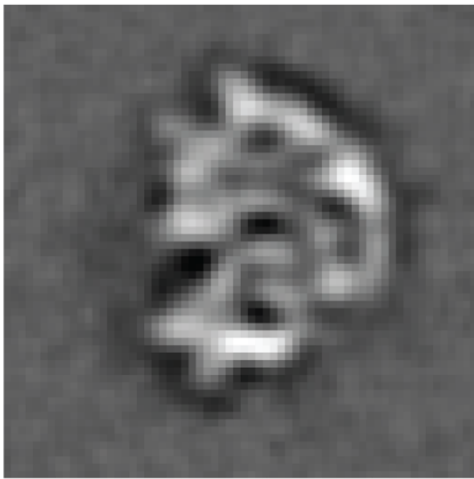
Acknowledgments

REFERENCES

1. Frank, J. Electron tomography: methods for three-dimensional visualization of structures in the cell. second edition. New York: Springer; 2006.

2. van Heel M, Gowen B, Matadeen R, Orlova EV, Finn R, Pape T, Cohen D, Stark H, Schmidt R, Schatz M, Patwardhan A. Single-particle electron cryo-microscopy: towards atomic resolution. Quarterly Reviews of Biophysics. 2000; vol. 33:307–369. [PubMed: 11233408]

3. Penczek, PA. Cryo-EM, Part B: 3-D Reconstruction, vol 482 of Methods in Enzymology, chapter Fundamentals of Three-Dimensional Reconstruction from Projections. Elsevier; 2010. p. 1-33.

4. Vonesch C, Unser M. A fast multilevel algorithm for wavelet-regularized image restoration. IEEE Transactions on Image Processing. 2009 Mar; vol. 18(no. 3):509–523. [PubMed: 19188124]

5. Dutt A, Rokhlin V. Fast Fourier transforms for nonequispaced data. SIAM Journal on Scientific Computing. 1993 Nov; vol. 14(no. 6):1368–1393.

6. Wajer, FTAW.; Pruessmann, KP. Major speedup of reconstruction for sensitivity encoding with arbitrary trajectories. Proceedings of ISMRM, 8th Annual Meeting; Glasgow. 2001. p. 767

7. Fessler JA, Lee S, Olafsson VT, Shi HR, Noll DC. Toeplitz-based iterative image reconstruction for MRI with correction for magnetic field inhomogeneity. IEEE Transactions on Signal Processing. 2005 Sep; vol. 53(no. 9):3393–3402.

8. Guerquin-Kern, M.; Van De Ville, D.; Vonesch, C.; Baritaux, J-C.; Pruessmann, KP.; Unser, M. Proceedings of ISBI. Boston: 2009 Jul. Wavelet-regularized reconstruction for rapid MRI; p. 193-196.

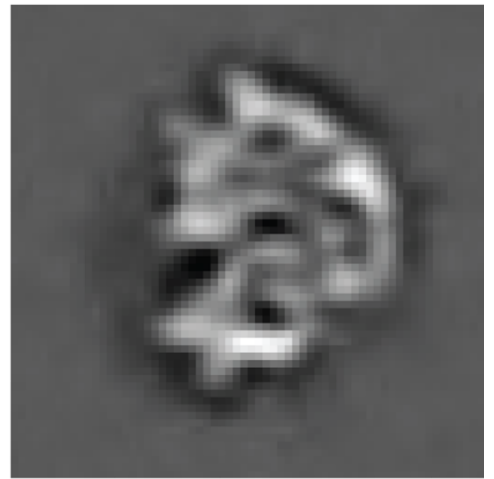9. Bertsekas, DP.; Nedic, A.; Ozdaglar, AE. Convex Analysis and Optimization. Athena Scientific; 2003 Apr.

(a) Original

(b) Back-projected

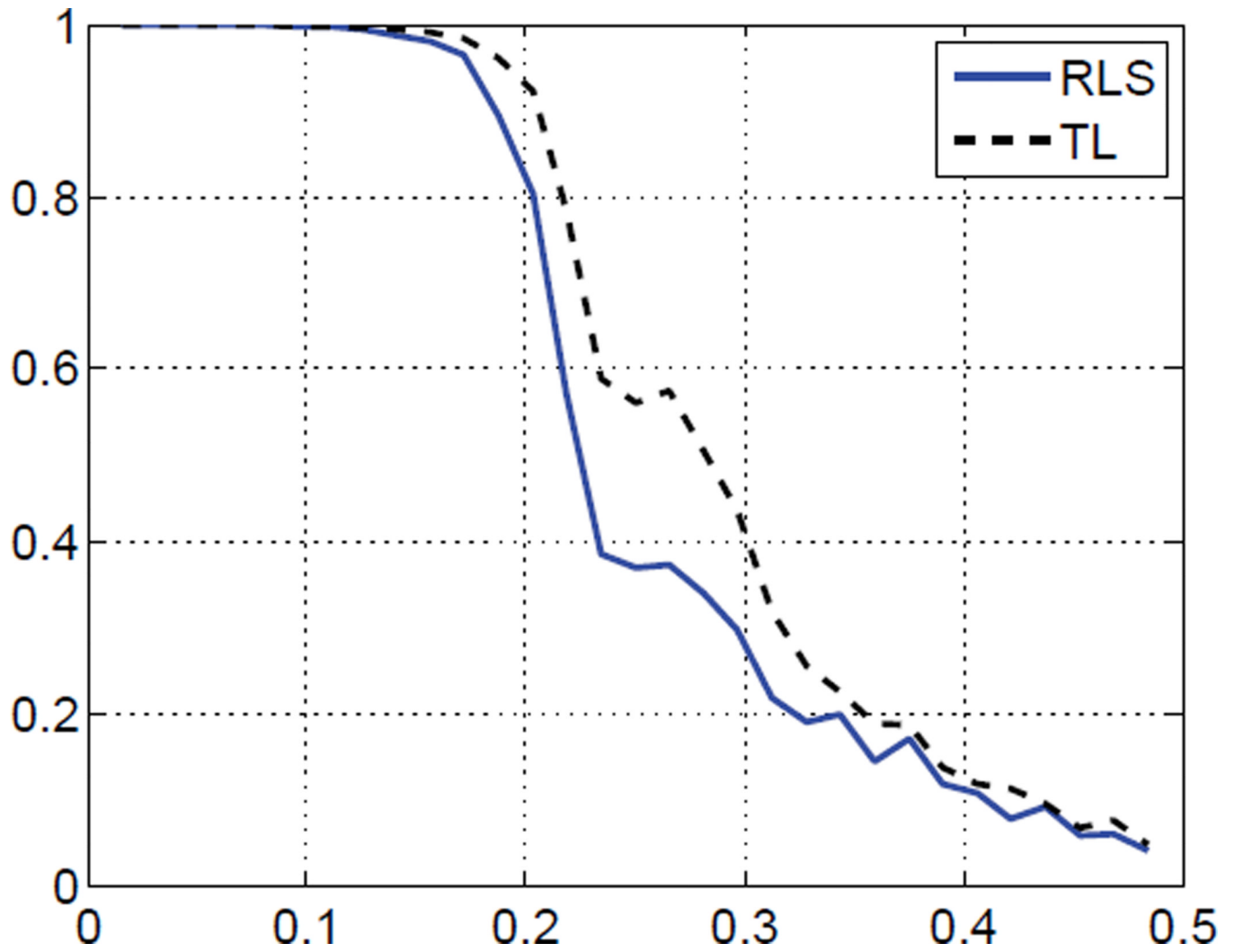(c) RLS

(d) TL

Fig. 1.
Results of Experiment 2.

Fig. 2.
Fourier-shell correlation for Experiment 2.

Table 1

SNR improvements for Experiment 1.

Input SNR	LS	RLS	TL
1/1	8.19 dB	8.42 dB	9.54 dB
1/8	4.74 dB	5.36 dB	7.27 dB
1/64	2.82 dB	3.66 dB	4.66 dB