Supplementary Material:
Ab-initio Contrast Estimation and Denoising of Cryo-EM Images

Yunpeng Shi*[1] and Amit Singer[1,2]

[1]Program in Applied and Computational Mathematics, Princeton University
[2]Department of Mathematics, Princeton University

This supplementary material provides the results of applying our methods for contrast estimation and image denoising for two experimental datasets. Overall, the results for these two datasets are similar to the results with the experimental dataset reported in the main body text.

1 Results for EMPIAR-10005

We compare our algorithms using an experimental dataset of the TRPV1 ion channel, whose ID is EMPIAR-10005 [1]. Its estimated 3-D volume is also available on EMDB with ID EMD-5778 [1]. The dataset contains 35645 picked particle images of size 256×256 with pixel size of 1.2156 Å from 935 defocus groups.

We estimate the covariance using all images, and pick 19 defocus groups (0th, 50th, 100th, 150th, ..., 900th) to estimate the contrast of individual images and then denoise the selected images. The background subtraction, whitening, Fourier-Bessel expansion and covariance estimation takes 4 hours. It takes 1.5 seconds for SDP covariance refinement and less than 1 second for the GS one. Applying Wiener filtering to the 19 defocus groups takes 5 minutes. The time for contrast estimation from the Fourier-Bessel coefficients is less than one second.

We first present a box plot of both oracle contrasts (top subplot) and the contrasts estimated by CWF-GS (bottom subplot) for each of the 19 defocus groups in Figure 1. From left to right in each subfigure, the defocus values are sorted in ascending order, ranging from 0.9889 μm to 2.1976 μm. In each box plot, the 5 horizontal lines, from top to bottom, respectively correspond to max value, 75% quantile, median, 25% quantile, and min value. Similar to EMPIAR-10028, the box plots from the two subfigures are similar, and both subfigures show large contrast variation within each micrograph. However, from the box plots we do not see a strong correlation between the defocus values and the image contrasts, unlike the results for EMPIAR-10028.

Next, we present the scatter plot between the estimated contrasts and the oracle contrast. It is clear from Figure 2 that our estimates have much better correlation with the oracle.

*Corresponding author, yunpengs@princeton.edu

In particular, CWF-GS performs slightly better than CWF-SDP. The scatter plots of both methods tend to have slope smaller than 1, due to the low SNR of this dataset.

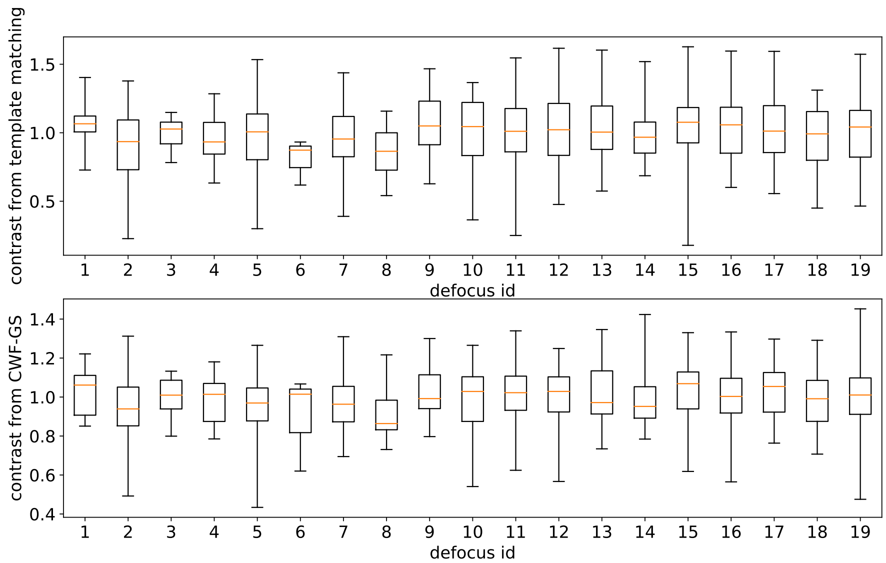

Figure 1: Box plot of the oracle contrasts (top) and our estimated contrasts (bottom) in 19 defocus groups of the dataset EMPIAR-10005. The defocus values are sorted in ascending order, ranging from 0.9889 μm to 2.1976 μm.

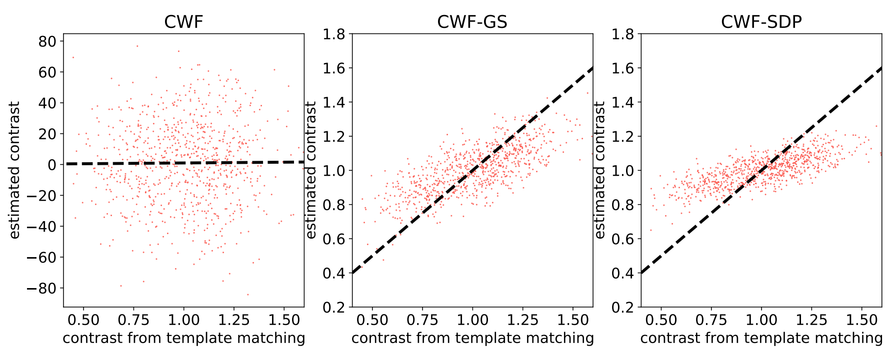

Figure 2: Scatter plot of the estimated contrast v.s. the oracle contrast for three defocus groups in the dataset EMPIAR-10005. The dashed lines correspond to the function $y = x$.

We next compare the image denoising performance by CWF and the ones with our refined covariance matrix. From Figure 3, we observe similar dark rings as in the example of EMPIAR-10028 for denoised images by CWF. This issue is largely mitigated by applying our covariance refinement methods.

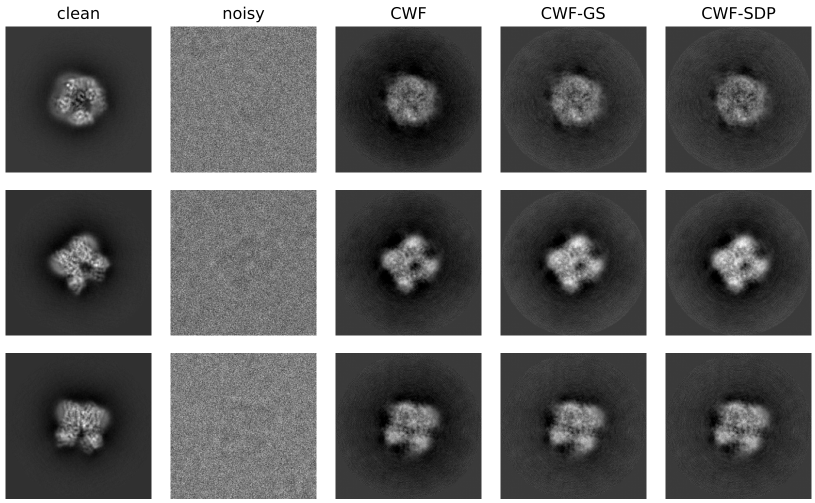

Figure 3: Denoising results of EMPIAR-10005.

At last, to quantitatively compare the denoising results of different methods, we compute the FRCs between the denoised images and their aligned clean templates (Figure 4). For each method, we compute the average FRC between the denoised images and the clean templates over 817 images from 19 defocus groups. Again, the FRCs of CWF-based methods are much higher than that of the naïve phase flipping method, and our methods achieve much better FRC than other methods at the first few radial frequencies due to the superior contrast estimation.

2 Results for EMPIAR-10073

We test the algorithms on the dataset of the yeast U4/U6.U5 tri-snRNP. The picked particles are downloadable from EMPIAR with ID EMPIAR-10073 [2]. Its 3-D reconstruction can be found on EMDB as EMD-8012 [2]. It consists of 138899 motion corrected and picked particle images of size 380×380 with 1.45 Å pixel size, from 2340 defocus groups.

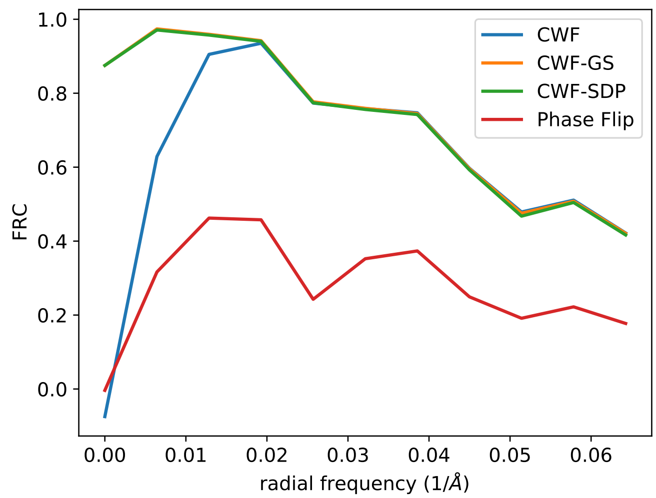

Figure 4: The average FRC between denoised images and the aligned clean templates over 817 images from EMPIAR-10005.

We estimate the covariance using all images, and pick 21 defocus groups (0th, 100th, 200th, ..., 2000th) to estimate the contrast of individual images and then denoise the selected images. The background subtraction, whitening, Fourier-Bessel expansion and covariance estimation takes 20 hours. In the current implementation of CWF the running time is proportional to the number of defocus groups, which is the reason why this step takes longer for this dataset. It takes 7 seconds for SDP covariance refinement and less than 1 second for the GS one. Applying Wiener filtering to the 21 defocus groups takes 9 minutes. The time for contrast estimation from the Fourier-Bessel coefficients is less than one second.

We first present a box plot of both oracle contrasts (top subplot) and contrasts estimated by CWF-GS (bottom subplot) for each of the 21 defocus groups in Figure 5. From left to right in each subfigure, the defocus values are sorted in ascending order, ranging from 0.4609 μm to 2.4008 μm. Similar to the previous example, we do not observe a clear correlation between defocus values and image contrast, and the contrast variation within each micrograph is large.

Next, we present the scatter plot of the estimated contrasts v.s. the oracle contrast. From Figure 6, our estimates have better correlation with the oracle. However, due to the low SNR of EMPIAR-10073, our methods perform worse than in previous examples.

We next compare the image denoising performance by CWF with our refined covariance matrix. From Figure 7, the denoised images by our methods have less dark areas, comparing to that of CWF, which suggests better CTF correction by our methods.

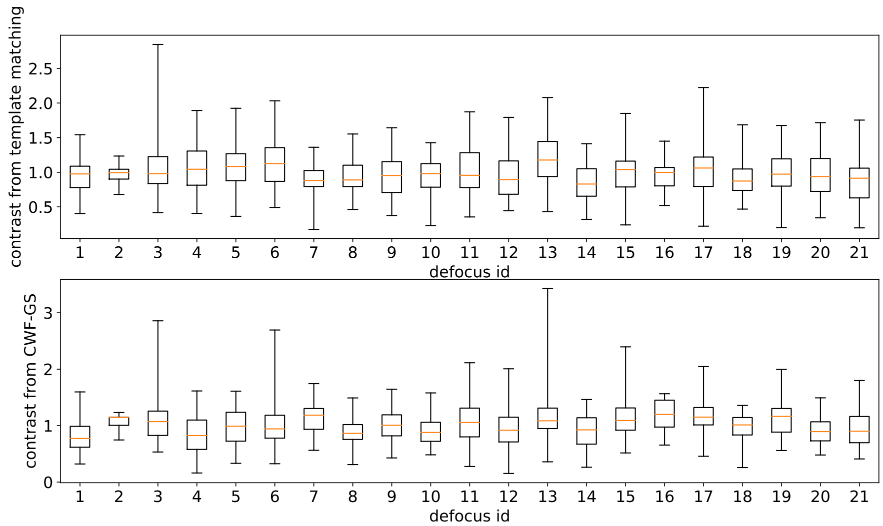

Figure 5: Box plot of the oracle contrasts (top) and our estimated contrasts (bottom) in 21 defocus groups of the dataset EMPIAR-10073. The defocus values are sorted in ascending order, ranging from 0.4609 μm to 2.4008 μm.

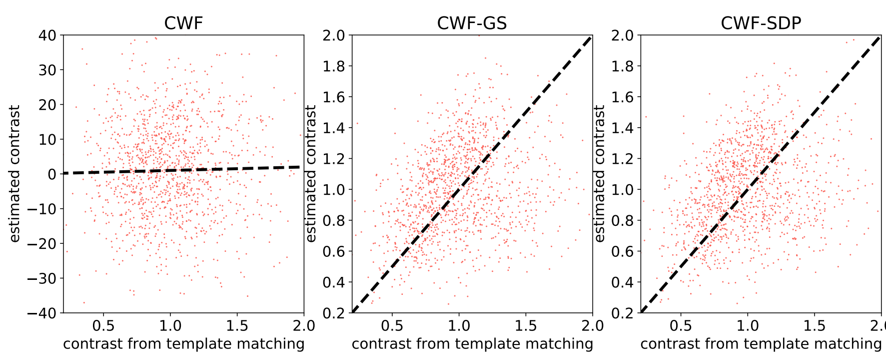

Figure 6: Scatter plot of the estimated contrast v.s. the oracle contrast for three defocus groups in the dataset EMPIAR-10073. The dashed lines correspond to the function $y = x$.

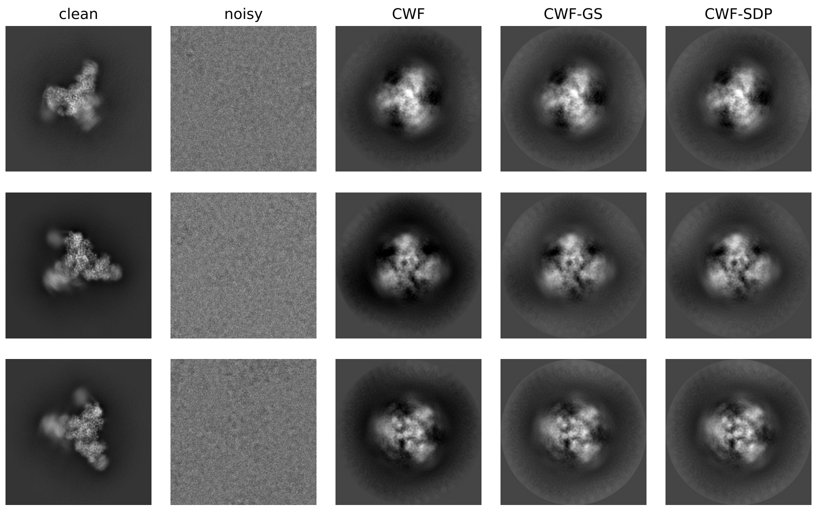

Figure 7: Denoising results of EMPIAR-10073.

At last, we present the FRCs between the denoised images and their aligned clean templates. For each method, we compute the average FRC between the denoised images and the clean templates over 1206 images from 21 defocus groups. Similar to the previous datasets, the denoised images by our methods have higher FRCs at the first few frequencies, compared to those of CWF.

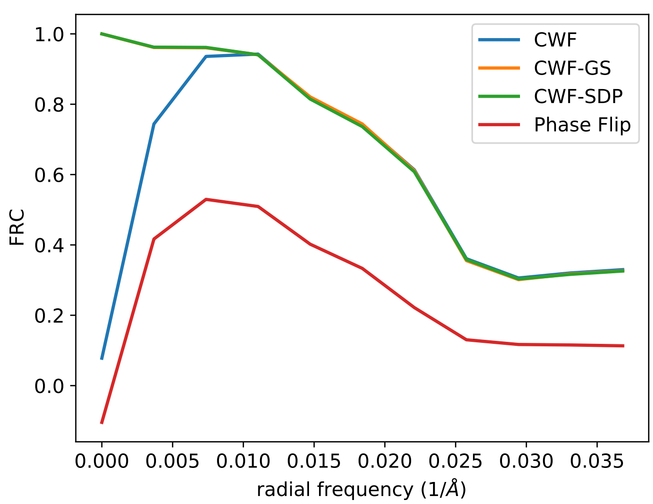

Figure 8: The average FRC between denoised images and the aligned clean templates over 1206 images from EMPIAR-10073.

References

[1] M. Liao, E. Cao, D. Julius, and Y. Cheng. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature*, 504(7478):107–112, 2013.

[2] T. H. D. Nguyen, W. P. Galej, X.-c. Bai, C. Oubridge, A. J. Newman, S. H. Scheres, and K. Nagai. Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature*, 530(7590):298–302, 2016.