Toward single particle reconstruction without particle picking: Breaking the detection limit

Tamir Bendory, Nicolas Boumal, William Leeb, Eitan Levin and Amit Singer

July 25, 2018

Abstract

Here comes the abstract

1 Introduction

[Revise–Cryo–electron microscopy (cryo–EM) is an innovative technology for single particle reconstruction (SPR) of macromolecules.] In a cryo–EM experiment, biological samples are rapidly frozen in a thin layer of vitreous ice. Within the ice, the molecules are randomly oriented and positioned. The microscope produces a 2-D tomographic image of the samples embedded in the ice, called a *micrograph*. Each micrograph contains tomographic projections of the samples at unknown locations and under unknown viewing directions. The goal is to construct 3-D models of the molecules from the micrographs.

The signal to noise ratio (SNR) of the projections in the micrographs is a function of two dominating factors. On the one hand, the SNR is a function of the electron dose. To keep radiation damage within acceptable bounds, the dose must be kept low, which leads to high noise levels. On the other hand, the SNR is a function of the molecule size. The smaller the molecules, the fewer detected electrons carry information about them.

All contemporary methods in the field split the reconstruction procedure into several stages. The first stage consists in extracting the various particle projections from the micrographs. This is called *particle picking*. Later stages aim to construct a 3-D model of the molecule from these projections. The quality of the reconstruction eventually hinges on the quality of the particle picking stage. Figure 1 illustrates how particle picking becomes increasingly challenging as the SNR degrades.

Crucially, it can be shown that reliable detection of individual particles is impossible below a certain critical SNR. This fact has been recognized early on by the cryo-EM community. In particular, in an influential paper from 1995, Henderson [8] investigates the following questions:

> *For the purposes of this review, I would like to ask the question: what is the smallest size of free-standing molecule whose structure can in principle be determined by phase-contrast electron microscopy? Given what has already been demonstrated in published work, this reduces to the question: what is the smallest size of molecule for which it is possible to determine from images of unstained molecules the five*

parameters needed to define accurately its orientation (three parameters) and position (two parameters) so that averaging can be performed?

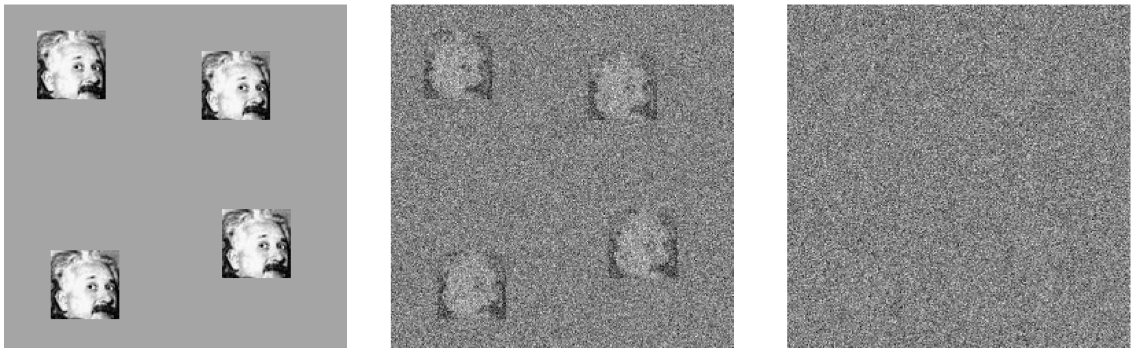
In that paper and in others that followed (e.g., [5]), it was established that particle picking is impossible for molecules below a certain weight (below ∼50 kDa). The same was mentioned by Joachim Frank in his Nobel prize lecture "*Using the ribosome as an example, it became clear from the formula we obtained that the single-particle approach to structure research was indeed feasible for molecules of sufficient size*" [4]. As a result, it is impossible to reconstruct such small molecules by any of the existing computational pipelines for single particle analysis in cryo–EM, as the particles themselves cannot be picked from the micrographs. This has motivated recent technical advances in the field, including the use of Volta phase plates [10, 13] and scaffolding cages [14].

Despite this progress, detecting small molecules in the micrographs remains a challenge. We note that nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography are well suited to reconstruct small molecules. Yet, cryo–EM has a lot to offer even for molecules with already known structures obtained via NMR spectroscopy or X-ray crystallography, because these methods have limited ability to distinguish conformational variability. [Need a ref for this claim.]

In this paper, we argue that there is a gap between the two questions in the quoted excerpt above, and that one may be able to exploit it to design better reconstruction algorithms. Specifically, the impossibility of particle picking does not necessarily imply impossibility of particle reconstruction. Indeed, the aim is only to reconstruct the molecule: estimating the locations of the particles in the micrograph is merely a helpful intermediate stage when it can be done. Our main message is that the limits particle picking imposes on molecule size do not translate into limits on particle reconstruction.

As a proof of concept, we study two simplified models. In first model, an unknown image appears numerous times at unknown locations in several micrographs, each affected by additive Gaussian noise—see Figure 1 for an illustration. The goal is to estimate the planted image. The task is interesting in particular when the SNR is low enough that particle picking cannot be done reliably. This problem is interesting of its own as it appears in other scientific applications, such as spike sorting [12], passive radar [6] and system identification [15]. We study it in more details in a companion paper [2]. An extension to multiple planted images is analyzed in [2]. In the second model, we consider the 3-D reconstruction problem as it appears in cryo-EM, while neglecting important aspects as discussed in detail in TKTK. [Here we need to elaborate a bit.] This is the main result of this paper. [We probably need to switch the order of this paragraph.]

In order to recover the 3-D volume, we use autocorrelation analysis. In a nutshell, we relate the autocorrelation functions of the micrographs to the autocorrelation functions of the volume. For any noise level, these autocorrelations can be estimated to any desired accuracy, provided the projections are well separated and we acquire enough of them. Importantly, there is no need to detect the projections. The autocorrelations of the micrographs are straightforward to compute and require only one pass over the data. These directly yield estimates for the autocorrelations of the volume. To estimate the volume itself from its estimated autocorrelations, we solve a nonlinear inverse problem via least-squares. We use similar technique for the simpler model of planted image as illustrated in Figure 2. As a side note, we mention that expectation-maximization (EM)—a popular framework in SPR—is



(a) $\sigma = 0$ (b) $\sigma = 0.5$ (c) $\sigma = 3$

Figure 1: Example of micrographs of size 250×250 with additive white Gaussian noise of variance σ^2 for increasing values of σ. Each micrograph contains the same four occurrences of a 50×50 image of Einstein. In panel (c), the noise level is such that it is very challenging to locate the occurrences of the planted image. In fact, it can be shown that at low SNR, reliable detection of individual image occurrences is impossible, even if the true image is known. By analogy to cryo-EM, this depicts a scenario where particle picking cannot be done. [Do we want to replace with a cryo-EM figure?]

intractable for this problem; see [2] for the details.

Another interesting feature of the described approach pertains to model bias, whose importance in cryo-EM was stressed by a number of authors [17, 19, 9, 18]. In the classical "Einstein from noise" experiment, multiple realizations of pure noise are aligned to a picture of Einstein using cross-correlation and then averaged. In [17], it was shown that the averaged noise rapidly becomes remarkably similar to the Einstein template. In the context of cryo–EM, this experiment exemplifies how prior assumptions about the particles may influence the reconstructed structure. This model bias is common to all particle picking methods based on template matching. In our approach, no templates are required, thus significantly reducing concerns about model bias. [To add reference to our example.]

2 Results

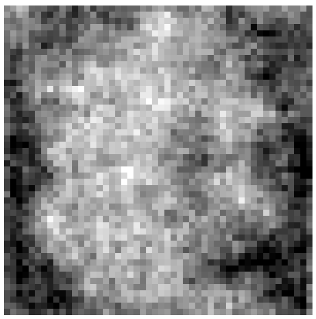
We start this section by considering the simplified model of planted image. This experiment aims to recover a 2-D image from an increasing number of micrographs with high noise, similar to the rightmost panel of Figure 1. This is done using moments of second order, as these are sufficient to recover a 2-D image up to elementary symmetries. As outlined below, we find that it is indeed possible to recover accurate estimates of the ground truth signals from the highly corrupted micrographs, without particle picking. Furthermore, we find that the quality of estimation increases with the amount of data collected, despite the fact that particle picking remains challenging. Then, we show that the same holds true for the 3-D reconstruction setup. The Methods section provides additional details.

In the first experiment, we estimated Einstein's image of size 50×50 and mean zero from a growing number of micrographs, each of size 4096×4096 pixels. A micrograph contains, on average, 700 occurrences of the target image at random locations. The latter are chosen so that two occurrences are always separated by at least 49 pixels in each direction. [it is not clear at this point why the separation is needed. Maybe comment about this, or at least mention that the separation condition is discussed later on.] Thus, about 10% of each micrograph contains signal. The micrographs are contaminated with additive white Gaussian noise with standard deviation $\sigma = 3$ (this corresponds to $\mathsf{SNR} = 1/20$). This high noise level is illustrated in the right panel of Figure 1. In this first experiment, we assume knowledge of σ and of the total number of signal occurrences across all micrographs.[Need to mention that in the 3-D reconstruction we do not make these assumptions.]

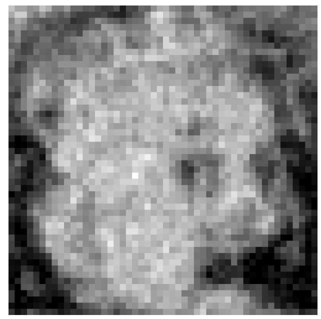
We compute the average autocorrelation of the micrographs (equivalently, the average of their power spectra). This is a particularly simple computation. In the methods section, we show how, owing to separation of the occurrences, a determined portion of the averaged autocorrelation allows to estimate the power spectrum of the unknown image itself. Mathematically, it is easy to show that the quality of this estimate improves steadily as the amount of data grows, regardless of noise level. Then, to estimate the target image, we resort to a standard phase retrieval algorithm called relaxed-reflect-reflect (RRR) [3]. RRR is initialized far away from the ground truth, and it iterates to produce the estimate, up to a reflection ambiguity.

Figure 2 shows several estimated images for a growing number of micrographs, and a movie is available in [supplementary material]. Figure 3 presents the normalized recovery error as a function of the amount of data available. Error is measured as the ratio of the root mean square error (RMSE) to the norm of the ground truth (square root of the sum of squared pixel intensities.) This is computed after fixing elementary symmetries (see Methods.) As evidenced by these figures, the ground truth image can be estimated increasingly well from increasingly many micrographs, without particle picking.

[Here the cryo-EM setup shows up]



(a) $P = 512$



(b) $P = 512 \times 10$

(c) $P = 512 \times 10^2$

(d) $P = 512 \times 10^3$

Figure 2: Recovery of Einstein from micrographs at noise level $\sigma = 3$ (see Figure 1(c)). Averaged autocorrelations of the micrographs allow to estimate the power spectrum of the target image. This does not require particle picking. A phase retrieval algorithm (RRR) produces the estimates here shown, initialized with an image of the physicist Isaac Newton. Estimates are obtained from P micrographs (growing across panels), each containing 700 image occurrences on average. [To add: redo the figures according to Amit's comments]

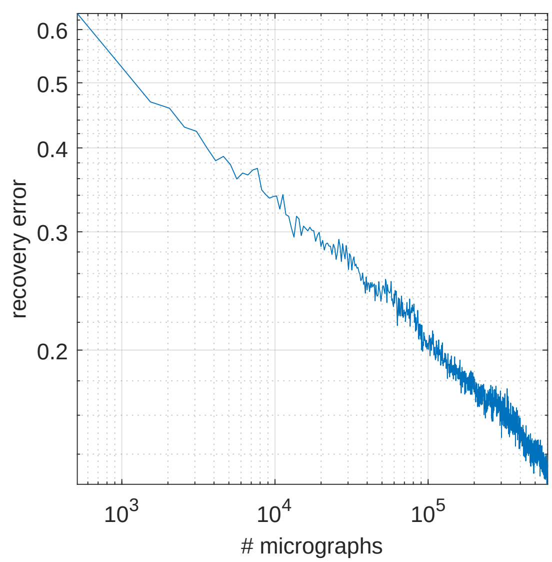


Figure 3: Relative root mean square error of the estimate of Einstein's image as a function of the number of observed micrographs (logarithmic scale along both axes.)

3 Discussion

[I suggest starting the Discussion section with a summary statement highlighting the significance of the work. Something along the lines (this is not polished, just to give you a rough idea): "In this paper we illustrated that it is possible to determine the 3-D molecular structure in single particle cryo-EM directly from the micrographs without performing particle picking. Our work implies that it is possible to reconstruct arbitrarily small molecules, in particular, molecules that are too small to being detected and located in micrographs. Ultimately, this work significantly increases the range of molecules to which cryo-EM can be successfully applied."

We should also incorporate Alberto's comment that our technique also allows to use much lower defocus values. Lower defocus means lower contrast, but will maintain higher frequency information. From that perspective, we may be able to get high resolution reconstructions from fewer micrographs, just because we would be using lower defocus.]

In the simplified model for cryo-EM we examined, we showed it is possible to estimate the 3-D structure of small volumes. Our strategy is to compute autocorrelation functions of the data and to relate these statistics to the unknown parameters. Recovering the parameters from the statistics reduces to solving a set of polynomial equations. Crucially, the outlined approach involves no particle picking, hence a fortiori no viewing direction estimation or conformation clustering. As a result, it may not be limited to large molecules in the same way that particle picking approaches are. Concerns for model bias would also greatly be reduced.

Of course, we recognize that significant challenges lay ahead for the implementation of the proposed approach to 3-D reconstruction directly from the micrographs. We discuss a few now.

One possible concern is that the numerical experiments conducted here suggest a large amount of data may be necessary.[1] Recent trends in high-throughput cryo-EM technology [?] give hope that this may be a lesser concern in the long term. Still, large amounts of data also imply large amounts of computations. On this front, we note that computing autocorrelations of low orders can be done efficiently on CPUs and GPUs, and in parallel across micrographs. It can even be done in streaming mode, as only one look at each micrograph is necessary. The output of this data processing stage is a succinct summary in the form of autocorrelation estimates: its size is a function of the resolution, not a function of the number of observed micrographs. Subsequent steps, which involve solving the system of polynomial equations, scale only in the size of that summary. Of course, an important question then is whether the equations can be solved meaningfully in practice. The proof-of-concept experiments above suggest they might.

Beyond data acquisition and computational challenges, there are modeling issues to consider. As stated, our approach relies on two core assumptions that are not necessarily verified in SPR experiments. First, we assume an additive white noise model, while in practice the noise may be correlated or signal dependent. To address this point, it may be necessary to investigate better noise models and to extend the autocorrelation analysis accordingly. [another issue is that micrographs also contain contaminants such as carbon, ice crystals, etc..

[1] Whether or not this large amount of data would be necessary for any method to succeed given the unfavorable SNR is an interesting research question.

] Second, we assume that any two signal occurrences are sufficiently separated, and we use this assumption to derive algebraic relations between autocorrelations of the micrographs and autocorrelations of the target signals. Perhaps this separation could be induced by careful experimental design [?]. Alternatively, if the signals are not well separated, one can introduce new parameters which encode the distribution of the spacing between occurrences. Here as well, relations between autocorrelation functions of the data and of the signals can be derived.

We also admit several aspects of SPR experiments that were neglected in this paper, including power spectrum estimation, Contrast Transfer Function (CTF) correction and the non-uniformity of the viewing directions. All this aspects must be taken into account so the method could be applied on real data. We hope to take care of this issues in a future research, as well extending this method for the possibility to estimate several volumes simultaneously.

[Where and how do we cite Kam? Fred?]

4 Methods

4.1 2-D experiment (we need a name)

To explain the method, we begin by deriving the algebraic relation between the autocorrelation functions of the micrographs and the autocorrelation functions of the planted image.

Let x be the sought image and let $y \in \mathbb{R}^{N\times N}$ be the observed micrograph (notice that it is equivalent to think of the data as being one long micrograph or multiple smaller micrographs concatenated into one.) The forward model (or "image" formation model) is as follows. An unknown binary signal indicates (with 1's) the starting positions of all occurrences of x in y, so that, with additive white Gaussian noise:

$$y = x * s + \varepsilon, \qquad\qquad \varepsilon \sim \mathcal{N}(0, \sigma^2 I_N), \tag{4.1}$$

where $*$ denotes linear convolution. Let $\mathbf{i} := (i_1, i_2)$ and similarly for $\mathbf{j}$. The binary signals obey the following property: [Rewrite in terms of δ functions for consistency with 3D]

$$\text{If } s[\mathbf{i}] = 1 \text{ then } s[\mathbf{j}] = 0 \text{ for all} \|\mathbf{i} - \mathbf{j}\|_\infty \leq 2L - 1. \tag{4.2}$$

In words: the starting positions of any two occurrences must be separated by at least $2L-1$ positions in each direction, so that their end points are necessarily separated by at least $L-1$ signal-free entries in the micrograph.

From y, we aim to recover x. In this simplified model, we assume to know the number of occurrences of the signal. In contrast, particle picking is the task of estimating the binary signal s, which cannot be performed reliably if σ is large (that is, at low SNR.)

Aperiodic autocorrelation functions For the purpose of this experiment, we need only the mean of the signal and its second-order autocorrelation defined for an arbitrary signal $z \in \mathbb{R}^{m\times m}$ by [to verify]

$$a_z^2[\ell_1, \ell_2] = \frac{1}{m^2} \sum_{i_1=\max\{0,-\ell_1\}}^{m-1+\min\{0,-\ell_1\}} \sum_{i_2=\max\{0,-\ell_2\}}^{m-1+\min\{0,-\ell_2\}} z[i_1, i_2] z[i_1 + \ell_1, i_2 + \ell_2]. \tag{4.3}$$

It can be shown that in the limit of $N \to \infty$, the mean of the micrograph is equal to the mean of the signal x, scaled by the scalar

$$\gamma = \frac{ML^2}{N^2}, \tag{4.4}$$

which denotes the density of x in y (that is, the fraction of entries of y occupied by occurrences of x.) The spacing constraint (4.2) imposes $\gamma \leq \frac{L}{2L-1} \approx 1/2$. Similarly, the second-order autocorrelations are related through

$$\lim_{N\to\infty} a_y^2[\ell_1, \ell_2] = \gamma a_x^2[\ell_1, \ell_2] + \sigma^2 \delta[\ell_1, \ell_2].$$

These relations, and their extension to multiple planted images, are proven and discussed in [2].

Numerical experiment with 2-D image. For the 2-D experiment shown in Figures 2 and 3, we generate P micrographs of size 4096×4096 pixels. In each micrograph, we place Einstein's image (of zero mean) of size 50×50 in random locations, while preserving the separation condition (4.2). This is done by randomly selecting 4000 placements in the micrograph, one at a time with an accept/reject rule based on the separation constraint and locations picked so far. On average, 700 images are placed in each micrograph. Then, i.i.d. Gaussian noise with standard deviation $\sigma = 3$ is added, inducing an SNR of approximately 1/20. An example of a micrograph's excerpt is presented in the right panel of Figure 1.

In this experiment, we assume we know the noise level σ and the total number of occurrences of the target image across all micrographs. In stark contrast with the 1-D setup, the second-order autocorrelation determines almost any target image uniquely, up to reflection through the origin [7] (see also [1] for a review). This is because the second-order autocorrelations correspond to the Fourier magnitudes of the signal through the 2-D Fourier transform. Therefore, we estimate the signal's Fourier magnitudes (or power spectrum) from the Fourier magnitudes of the micrographs, at the cost of one 2-D fast Fourier transform (FFT) per micrograph. These can be computed highly efficiently and in parallel.

To recover the target image from the estimated power spectrum, we use a standard phase retrieval algorithm called relaxed-reflect-reflect (RRR). This algorithm iterates the map

$$z \leftarrow z + \beta(P_2(2P_1(z) - z) - P_1(z))$$

on an image z of size $2L \times 2L$. We set the parameter β to 1. The map is designed so that, if the estimated power spectrum is exact, then fixed points contain Einstein's image in the upper-left corner of size $L \times L$, possibly reflected through its origin, and zeros elsewhere. The operator $P_2(z)$ combines the Fourier phases of the current estimation z with the estimated Fourier magnitudes. The operator $P_1(z)$ zeros out all entries of z outside the $L \times L$ upper-left corner.

In order to compare the performance in multiple cases and at different noise levels, the algorithm is stopped after a fixed number of iterations (1000) and the iterate with the smallest error compared to the ground truth (up to the reflection ambiguity) is chosen as the solution. While this cannot be done in practice (since we do not have access to the ground truth to determine which iterate is best), this procedure enables us to compare a large number of instances in different noise environments. [Note the last two sentences!]

4.2 3-D experiment

[Add similarities to 2-D]

[Proposal: Use W for size of projection, M for size of micrograph]

Let ϕ be the Coulomb potential representing the molecule we wish to recover, and denote it 3-D Fourier transform by $\widehat{\phi}$. In this paper, we assume that $\widehat{\phi}$ is given as a finite expansion of the form

$$\widehat{\phi}(k, \theta, \varphi) = \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} \sum_{s=1}^{S(\ell)} a_{\ell,m,s} Y_\ell^m(\theta, \varphi) j_{\ell,s}(k), \tag{4.5}$$

where Y_ℓ^m are complex spherical harmonics and $j_{\ell,s}$ are normalized spherical Bessel functions, see Appendix TKTK for definitions. [Mention somewhere that since ϕ is real-valued, we recover only coefficients with $m \geq 0$.] Given $\omega \in \mathrm{SO}(3)$, denote by I_ω the projection obtained from viewing direction ω, and denote by $\widehat{I_\omega}$ its 2-D Fourier transform. By the Fourier projection-slice theorem [16, p. 11], the 2D Fourier transform of I_ω is given by [Move to appendix?]

$$\widehat{I_\omega}(k, \varphi) = \sum_{\ell,m,m',s} a_{\ell,m,s} D^\ell_{m',m}(\omega) Y_\ell^{m'}\left(\frac{\pi}{2}, \varphi\right) j_{\ell,s}(k), \tag{4.6}$$

where $D^\ell_{m',m}(\omega)$ is a Wigner-D matrix rotating spherical harmonics.

Let $\mathcal{I} \in \mathbb{R}^{M\times M}$ denote a micrograph. For this paper, we assume that the micrograph consists of shifted copies of projections and perturbed by additive white Gaussian noise:

$$\mathcal{I} = \sum_{j=1}^{N} I_{\omega_j} * \delta_{\mathbf{s}_j} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I), \tag{4.7}$$

where the viewing directions ω_j are uniformly distributed over SO(3), $\mathbf{s}_j$ denotes the location of the center of the jth projection in the micrograph, and we impose a separation condition similarly to Eq. [2D]:

$$||\mathbf{s}_j - \mathbf{s}_i||_\infty \geq 2L. \tag{4.8}$$

[Say something about nonuniformity in practice? Remember that biologists will jump at this...] Define the kth autocorrelation of $\mathcal{I}$ as

$$m_k[\mathcal{I}](\Delta\mathbf{i}_1, \ldots, \Delta\mathbf{i}_{k-1}) = \frac{1}{M^2} \sum_{i,j=1}^{M} \mathcal{I}(\mathbf{i})\mathcal{I}(\mathbf{i} + \Delta\mathbf{i}_1) \cdots \mathcal{I}(\mathbf{i} + \Delta\mathbf{i}_{k-1}), \tag{4.9}$$

where $\mathbf{i} = (i, j)$, $\Delta\mathbf{i} = (\Delta i, \Delta j)$, we require $||\Delta\mathbf{i}||_\infty \leq 2L - 1$, and we set $\mathcal{I}(i, j) = 0$ if either $i, j \notin \{1, \ldots, M\}$. In our procedure, we compute the first three autocorrelations, called the mean, power spectrum, and bispectrum, of the micrographs. [This definition and terminology should come earlier]. Because of the separation condition Eq. (4.8), the autocorrelations of the micrograph are related to those of the projections by

$$m_k[\mathcal{I}] \to \gamma \langle m_k[I_\omega] \rangle_\omega + b_k, \tag{4.10}$$

where $\langle\cdot\rangle$ denotes averaging, b_k is a bias term, and we take the limit as the number of projections N, and hence the dimensions of the micrograph M, tend to ∞. The mean is unbiased so $b_1 = 0$, the power spectrum has bias b_2 depending only on σ^2, the variance of the noise, and the bias of the bispectrum b_3 depends additionally on the mean of ϕ, which can be accurately estimated from the mean of the micrograph. We can therefore estimate the quantities $\gamma\langle m_k[I_\omega]\rangle_\omega$ directly from the micrograph.

Since we assume that the viewing directions ω_j are uniformly distributed over SO(3), we would like to average over all in-plane rotations of the projections in the computation of the autocorrelations. To do so, we follow [11, 20] and expand the second and third autocorrelations of the micrograph in Prolate Spheroidal Wave Functions (PSWFs). It can then be shown that the expansion coefficients $\mathfrak{m}_k$ of the autocorrelations in PSWFs are related to the expansion coefficients of the volume by

$$\mathfrak{m}_1\{a_{\ell,m,s}\} = \frac{L}{\sqrt{4\pi}} \sum_s a_{0,0,s} j_{0,s}(0) \tag{4.11}$$

$$\mathfrak{m}_2\{a_{\ell,m,s}\}(q) = \frac{\sqrt{2\pi}}{4\pi} \sum_{\substack{\ell,m \\ s_1,s_2}} C_2^{(q)}(\ell, s_1, s_2) a_{\ell,m,s_1} \overline{a_{\ell,m,s_2}} \tag{4.12}$$

$$\mathfrak{m}_3\{a_{\ell,m,s}\}(k, q_1, q_2) = \sum_{\substack{\ell_1,m_1,s_1 \\ \ell_2,m_2,s_2 \\ s_3}} \sum_{\ell_3=|\ell_1-\ell_2|}^{\min(L,\ell_1+\ell_2)} C_3^{(k,q_1,q_2)}(\ell_1, m_1, s_1; \ell_2, m_2, s_2; \ell_3, s_3) a_{\ell_1,m_1,s_1} a_{\ell_2,m_2,s_2} \overline{a_{\ell_3,m_1+m_2,s_3}}, \tag{4.13}$$

see Appendix TKTK for the definitions of C_2, C_3.

To recover the expansion coefficients $\{a_{\ell,m,s}\}$ from the autocorrelation coefficients estimated from the micrograph $\mathfrak{m}_k[\mathcal{I}]$, we minimize the weighted least-squares problem

$$\min_{\{a_{\ell,m,s}\}} \sum_{k=1}^{3} \frac{\left|\left|\mathfrak{m}_k\{a_{\ell,m,s}\} - \mathfrak{m}_k[\mathcal{I}]\right|\right|_F^2}{||\mathfrak{m}_k[\mathcal{I}]||_F^2}. \tag{4.14}$$

Acknowledgment

Let's thank them all

References

[1] Tamir Bendory, Robert Beinert, and Yonina C Eldar. Fourier phase retrieval: Uniqueness and algorithms. In *Compressed Sensing and its Applications*, pages 55–91. Springer, 2017.

[2] Tamir Bendory, Nicolas Boumal, William Leeb, Eitan Levin, and Amit Singer. Estimation below the detection limit with application to cryo-em. *preprint*, 2018.

[3] Veit Elser. Matrix product constraints by projection methods. *Journal of Global Optimization*, 68(2):329–355, 2017.

[4] Joachim Frank. Single-particle reconstruction of biological molecules-story in a sample (nobel lecture). *Angewandte Chemie International Edition*, 2018.

[5] Robert M Glaeser. Electron crystallography: present excitement, a nod to the past, anticipating the future. *Journal of structural biology*, 128(1):3–14, 1999.

[6] Sandeep Gogineni, Pawan Setlur, Muralidhar Rangaswamy, and Raj Rao Nadakuditi. Passive radar detection with noisy reference channel using principal subspace similarity. *IEEE Transactions on Aerospace and Electronic Systems*, 2017.

[7] MHMH Hayes. The reconstruction of a multidimensional sequence from the phase or magnitude of its fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 30(2):140–154, 1982.

[8] Richard Henderson. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quarterly reviews of biophysics*, 28(2):171–193, 1995.

[9] Richard Henderson. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences*, 110(45):18037–18041, 2013.

[10] Maryam Khoshouei, Mazdak Radjainia, Wolfgang Baumeister, and Radostin Danev. Cryo-EM structure of haemoglobin at 3.2 å determined with the Volta phase plate. *Nature communications*, 8:16099, 2017.

[11] Boris Landa and Yoel Shkolnisky. Steerable principal components for space-frequency localized images. *SIAM journal on imaging sciences*, 10(2):508–534, 2017.

[12] Michael S Lewicki. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computation in Neural Systems*, 9(4):R53–R78, 1998.

[13] Yi-Lynn Liang, Maryam Khoshouei, Mazdak Radjainia, Yan Zhang, Alisa Glukhova, Jeffrey Tarrasch, David M Thal, Sebastian GB Furness, George Christopoulos, Thomas Coudrat, et al. Phase-plate cryo-EM structure of a class B GPCR–G-protein complex. *Nature*, 546(7656):118, 2017.

[14] Yuxi Liu, Shane Gonen, Tamir Gonen, and Todd O. Yeates. Near-atomic cryo-EM imaging of a small protein displayed on a designed scaffolding system. *Proceedings of the National Academy of Sciences*, 2018.

[15] Lennart Ljung. System identification. In *Signal analysis and prediction*, pages 163–173. Springer, 1998.

[16] Frank Natterer. *The mathematics of computerized tomography*, volume 32. SIAM, 1986.

[17] Maxim Shatsky, Richard J Hall, Steven E Brenner, and Robert M Glaeser. A method for the alignment of heterogeneous macromolecules from electron microscopy. *Journal of structural biology*, 166(1):67–78, 2009.

[18] Marin van Heel. Finding trimeric HIV-1 envelope glycoproteins in random noise. *Proceedings of the National Academy of Sciences*, 110(45):E4175–E4177, 2013.

[19] Marin van Heel, Michael Schatz, and Elena Orlova. Correlation functions revisited. *Ultramicroscopy*, 46(1–4):307–316, 1992.

[20] Zhizhen Zhao, Yoel Shkolnisky, and Amit Singer. Fast steerable principal component analysis. *IEEE transactions on computational imaging*, 2(1):1–12, 2016.

A Moments derivation

[To rewrite for the discrete case]

We consider the moments of $\mathcal{I}$ in the limit $N \to \infty$. To take this limit, we shall assume that $N = \Omega(M^2)$, and that $\gamma = \lim_{N\to\infty} \frac{N}{M^2} \in (0,1)$ is a constant. Because of the separation condition Eq. (4.8), if $\mathbf{i}$ is in the support of the jth projection, then $\mathbf{i} + \Delta\mathbf{i}$ is either in the support of the same projection or outside the support of any projection. Formally, $\mathbf{i}$ is in the support of the jth projection if and only if $\mathbf{0} \le \mathbf{i} - \mathbf{s}_j < (L, L)^T$ where the inequalities apply to each coordinate. Then because of the separation condition,

note that if $\mathbf{r} \in S_i$ for some i, then $\mathbf{r} + \Delta\mathbf{r} \in S_i \cup Z$ whenever $||\Delta\mathbf{r}|| \le 2L$ by the separation assumption $d > 2L$. Since $\mathbb{R}^2 = Z \sqcup \bigsqcup_{i=1}^N S_i$, and $\mathcal{I}(\mathbf{r}) = 0$ if $\mathbf{r} \in Z$, we have for $||\Delta\mathbf{r}|| \le 2L$

$$
\begin{aligned}
m_2[\mathcal{I}](\Delta\mathbf{r}) &= \frac{1}{M^2} \int_{\mathbb{R}^2} \mathcal{I}(\mathbf{r})\mathcal{I}(\mathbf{r} + \Delta\mathbf{r})\, d\mathbf{r} \\
&= \frac{1}{M^2} \int_{Z} \mathcal{I}(\mathbf{r})\mathcal{I}(\mathbf{r} + \Delta\mathbf{r})\, d\mathbf{r} + \frac{1}{M^2} \sum_{i=1}^{N} \int_{S_i} \mathcal{I}(\mathbf{r})\mathcal{I}(\mathbf{r} + \Delta\mathbf{r})\, d\mathbf{r} \\
&= \frac{1}{M^2} \sum_{i=1}^{N} \int_{\mathbb{R}^2} I_{\omega_i}(\mathbf{r}) I_{\omega_i}(\mathbf{r} + \Delta\mathbf{r})\, d\mathbf{r} \\
&= \frac{N}{M^2} \cdot \frac{1}{N} \sum_{i=1}^{N} L^2 m_2[I_{\omega_i}](\Delta\mathbf{r}) \\
&\to \gamma \int_{\mathrm{SO}(3)} L^2 m_2[I_\omega](\Delta\mathbf{r})\, d\omega.
\end{aligned}
$$

If we set $m_2(\Delta\mathbf{r}) = 0$ for $||\Delta\mathbf{r}|| > 2L$, the above equality holds for all $\Delta\mathbf{r}$, and taking its Fourier transform (with respect to $\Delta\mathbf{r}$) and interchanging the Fourier integral with the integral over SO(3) (which can *always* be done), we get

$$
\widehat{m_2[\mathcal{I}]}(\mathbf{k}) \to \gamma \int_{\mathrm{SO}(3)} L^2 \widehat{m_2[I_\omega]}(\mathbf{k})\, d\omega = \gamma \langle |\widehat{I}_\omega(\mathbf{k})|^2 \rangle_\omega,
$$

where $\langle \cdot \rangle_\omega$ denotes average over the orientations, so over SO(3).

Similarly, for the third moment if $\mathbf{r} \in S_i$ for some i then $\mathbf{r} + \Delta\mathbf{r}_1, \mathbf{r} + \Delta\mathbf{r}_2 \in S_i \cup Z$, and hence we again get for $||\Delta\mathbf{r}_1|| \leq 2L$ and $||\Delta\mathbf{r}_2|| \leq 2L$ that

$$\begin{aligned}
m_3[\mathcal{I}](\Delta\mathbf{r}_1, \Delta\mathbf{r}_2) &= \frac{1}{M^2}\int_{\mathbb{R}^2} \mathcal{I}(\mathbf{r})\mathcal{I}(\mathbf{r}+\Delta\mathbf{r}_1)\mathcal{I}(\mathbf{r}+\Delta\mathbf{r}_2)\, d\mathbf{r} \\
&= \frac{1}{M^2}\int_{Z} \mathcal{I}(\mathbf{r})\mathcal{I}(\mathbf{r}+\Delta\mathbf{r}_1)\mathcal{I}(\mathbf{r}+\Delta\mathbf{r}_2)\, d\mathbf{r} + \frac{1}{M^2}\sum_{i=1}^{N}\int_{S_i} \mathcal{I}(\mathbf{r})\mathcal{I}(\mathbf{r}+\Delta\mathbf{r}_1)\mathcal{I}(\mathbf{r}+\Delta\mathbf{r}_2)\, d\mathbf{r} \\
&= \frac{1}{M^2}\sum_{i=1}^{N}\int_{\mathbb{R}^2} I_{\omega_i}(\mathbf{r})I_{\omega_i}(\mathbf{r}+\Delta\mathbf{r}_1)I_{\omega_i}(\mathbf{r}+\Delta\mathbf{r}_2)\, d\mathbf{r} \\
&= \frac{N}{M^2}\cdot\frac{1}{N}\sum_{i=1}^{N} L^2 m_3[I_{\omega_i}](\Delta\mathbf{r}_1, \Delta\mathbf{r}_2) \\
&\to \gamma\int_{\mathrm{SO}(3)} L^2 m_3[I_\omega](\Delta\mathbf{r}_1, \Delta\mathbf{r}_2)\, d\omega.
\end{aligned}$$

Again, setting $m_3(\Delta\mathbf{r}_1, \Delta\mathbf{r}_2) = 0$ if either $||\Delta\mathbf{r}_1|| > 2L$ or $||\Delta\mathbf{r}_2|| > 2L$, we also get

$$\widehat{m_3[\mathcal{I}]}(\mathbf{k}_1, \mathbf{k}_2) = \gamma\langle \widehat{I}_\omega(\mathbf{k}_1)\widehat{I}_\omega(\mathbf{k}_2)\overline{\widehat{I}_\omega(\mathbf{k}_1+\mathbf{k}_2)}\rangle_\omega.$$

B Steering

The first two moments are cheap to compute and store - m_1 is a scalar and m_2 is radially symmetric (in the limit $N \to \infty$), so it suffices to compute a single ray of it. The third moment however is infeasible to store - it has $(2L-1)^4$ entries, and in modern microscopes we can easily have $L \geq 300$. In addition, the above scheme does not average over in-plane rotations, which can improve estimation in the precense of noise (and accelerate convergence to the population moments without noise). We therefore propose the following steering procedure:

The third moment $m_3[\mathcal{I}](\Delta\mathbf{r}_1, \Delta\mathbf{r}_2)$ as defined above is compactly supported (at least numerically) with respect to each of its variables in both real and Fourier space, with support radii $2L$ and $1/2$ in real and Fourier space, respectively (see above formulas for m_3 in real and Fourier space). Therefore, we can expand it in a suitable product basis that is compactly supported in both real and Fourier space, e.g. Prolate Spheroidal Wave Functions (PSWFs) or Fourier-Bessel (FB):

$$m_3[\mathcal{I}](\Delta\mathbf{r}_1, \Delta\mathbf{r}_2) = \sum_{k_1,k_2=-\infty}^{\infty}\sum_{q_1,q_2=1}^{\infty} \mathfrak{m}_3(k_1,q_1;k_2,q_2)\psi_{k_1,q_1}(\Delta\mathbf{r}_1)\overline{\psi_{k_2,q_2}(\Delta\mathbf{r}_2)},$$

where we conjugate the second set of basis functions for convenience, that will become apparent below (of course, the set of conjugates is also a basis, as $\overline{\psi_{k,q}} = \psi_{-k,q}$ for PSWFs and $\overline{\psi_{k,q}} = (-1)^k\psi_{-k,q}$ for FB). We shall assume that the basis $\{\psi_{k,q}(\mathbf{r})\}$ is orthonormal, so $\int_{\mathbb{R}^2}\psi_{k_1,q_1}(\mathbf{r})\overline{\psi_{k_2,q_2}}(\mathbf{r}) = \delta_{k_1,k_2}\delta_{q_1,q_2}$, valid for both PSWFs and FB. Then, the coefficients are given by

$$\begin{aligned}
\mathfrak{m}_3(k_1,q_1;k_2,q_2) &= \int_{||\Delta\mathbf{r}_1||,||\Delta\mathbf{r}_2||\leq 2L} m_3[\mathcal{I}](\Delta\mathbf{r}_1, \Delta\mathbf{r}_2)\overline{\psi_{k_1,q_1}}(\Delta\mathbf{r}_1)\psi_{k_2,q_2}(\Delta\mathbf{r}_2) \\
&= \int_{\mathbf{r}} \mathcal{I}(\mathbf{r})\left(\int_{||\Delta\mathbf{r}_1||\leq 2L} \mathcal{I}(\mathbf{r}+\Delta\mathbf{r}_1)\overline{\psi_{k_1,q_1}}(\Delta\mathbf{r}_1)\right)\left(\int_{||\Delta\mathbf{r}_2||\leq 2L} \mathcal{I}(\mathbf{r}+\Delta\mathbf{r}_2)\psi_{k_2,q_2}(\Delta\mathbf{r}_2)\right).
\end{aligned}$$

Note that if we expand $\mathcal{I}(\mathbf{r}+\Delta\mathbf{r})$ in $\{\psi_{k,q}(\Delta\mathbf{r})\}$ for fixed $\mathbf{r}$ and $||\Delta\mathbf{r}|| \leq 2L$, so

$$\mathcal{I}(\mathbf{r}+\Delta\mathbf{r}) = \sum_{k,q} a_{k,q}(\mathbf{r})\psi_{k,q}(\Delta\mathbf{r}), \quad a_{k,q}(\mathbf{r}) = \int_{||\Delta\mathbf{r}||\leq 2L} \mathcal{I}(\mathbf{r}+\Delta\mathbf{r})\overline{\psi_{k,q}}(\Delta\mathbf{r}),$$

and use the fact that $\mathcal{I}$ is real, our expression becomes

$$\mathfrak{m}_3(k_1,q_1;k_2,q_2) = \int_{\mathbf{r}} \mathcal{I}(\mathbf{r})a_{k_1,q_1}(\mathbf{r})\overline{a_{k_2,q_2}}(\mathbf{r}).$$

We further assume that all in-plane rotations of the micrograph and its reflection, or equivalently, all in-plane rotations of each such disc of radius $2L$ and its reflection, are present in our dataset. Noting that rotations and reflections commute with the Fourier transform and using the derivation of [Zhao, Landa], the expansion of the disc about $\mathbf{r}$ by an angle α is given by

$$\mathcal{I}^{\alpha,+}(\mathbf{r}+\Delta\mathbf{r}) = \sum_{k,q} a_{k,q}e^{-ik\alpha}\psi_{k,q}(\Delta\mathbf{r}),$$

and the expansion of the reflection of that disc rotated by an angle α by

$$\mathcal{I}^{\alpha,-}(\mathbf{r}+\Delta\mathbf{r}) = \sum_{k,q} \overline{a_{k,q}}e^{-ik\alpha}\psi_{k,q}(\Delta\mathbf{r}),$$

where we used the fact that for real-valued images we have $a_{-k,q} = \overline{a_{k,q}}$ for both PSWFs and FB. We thus have

$$\begin{aligned}
\mathfrak{m}_3(k_1,q_1;k_2,q_2) &= \int_{\mathbf{r}} \mathcal{I}(\mathbf{r})\left(\frac{1}{4\pi}\int_0^{2\pi} \left[a_{k_1,q_1}(\mathbf{r})\overline{a_{k_2,q_2}}(\mathbf{r}) + \overline{a_{k_1,q_1}}(\mathbf{r})a_{k_2,q_2}(\mathbf{r})\right] e^{-i(k_1-k_2)\alpha}d\alpha\right), \\
&= \delta_{k_1,k_2}\int_{\mathbf{r}} \mathcal{I}(\mathbf{r})\Re\{a_{k_1,q_1}(\mathbf{r})\overline{a_{k_2,q_2}}(\mathbf{r})\} \\
&= \Re\left\{\delta_{k_1,k_2}\int_{\mathbf{r}} \mathcal{I}(\mathbf{r})a_{k_1,q_1}(\mathbf{r})\overline{a_{k_2,q_2}}(\mathbf{r})\right\},
\end{aligned}$$

Thus, the bispectrum in our steerable basis $\mathfrak{m}_3(k_1,q_1;k_2,q_2)$ is block-diagonal, and is effectively a 3-tensor.

Similarly, the power spectrum is compactly supported in space (with support $2L-1$) and bandlimited as in Fourier space it is the average squared magnitude of the Fourier transform of the projections, each of which is supposedly bandlimited. Therefore, we may expand it as

$$m_2[\mathcal{I}](\Delta\mathbf{r}) = \sum_{k,q} \mathfrak{m}_2(k,q)\psi_{k,q}(\Delta\mathbf{r}),$$

and obtain the expansion coefficients as

$$\begin{aligned}
\mathfrak{m}_2(k,q) &= \int_{\Delta\mathbf{r}} m_2[\mathcal{I}](\Delta\mathbf{r})\overline{\psi_{k,q}(\Delta\mathbf{r})} \\
&= \int_{\mathbf{r}} \mathcal{I}(\mathbf{r})\int_{\Delta\mathbf{r}} \mathcal{I}(\mathbf{r}+\Delta\mathbf{r})\overline{\psi_{k,q}(\Delta\mathbf{r})} \\
&= \int_{\mathbf{r}} \mathcal{I}(\mathbf{r})a_{k,q}(\mathbf{r}).
\end{aligned}$$

Taking all rotations of $\mathcal{I}$ and its reflection, we get

$$\begin{aligned}\mathfrak{m}_2(k,q) &= \int_{\mathbf{r}} \mathcal{I}(\mathbf{r}) \left(\frac{1}{4\pi} \int_0^{2\pi} [a_{k,q}(\mathbf{r}) + \overline{a_{k,q}}(\mathbf{r})] e^{-ik\alpha}\, d\alpha \right) \\ &= \delta_{k,0} \int_{\mathbf{r}} \mathcal{I}(\mathbf{r}) a_{0,q}(\mathbf{r}),\end{aligned}$$

where in the last equality we dropped the real part since both $\mathcal{I}(\mathbf{r})$ and $\psi_{0,q}$ are real valued, and hence $a_{0,q}(\mathbf{r}) = \int_{\Delta\mathbf{r}} \mathcal{I}(\mathbf{r}+\Delta\mathbf{r})\psi_{0,q}(\Delta\mathbf{r})$ is real as well. Thus, the average power spectrum is effectively a vector.

C Connection to volume

We derive a relation between the steered bispectrum derived in Sect. B to the volume, expanded in a suitable basis. Specifically, expand the Fourier-transformed volume as

$$\widehat{V}(c\mathbf{k}) = \sum_{\ell,m,s} a_{\ell,m,s} j_{\ell,s}(k) Y_{\ell,m}(\mathbf{k}/k),$$

for $k \leq 1$ and zero otherwise, where c is the assumed bandlimit, $Y_{\ell,m}$ may be taken to be either real or complex, and $j_{\ell,s}(r)$ is some radial basis function (in practice, either spherical bessel or the radial part of the 3D PSWFs as in [Lederman]). We work here with the complex spherical harmonics, given by

$$Y_{\ell,m}(\theta,\varphi) = \sqrt{\frac{2\ell+1}{4\pi} \cdot \frac{(\ell-m)!}{(\ell+m)!}} P_\ell^m(\cos\theta) e^{im\varphi},$$

where P_ℓ^m are the associated Legendre polynomials with the Condon-Shortley phase, and $j_{\ell,s}$ are normalized spherical Bessel functions given by

$$j_{\ell,s}(k) = \frac{4}{|j_{\ell+1}(u_{\ell,s})|} j_\ell(2u_{\ell,s}k),$$

where j_ℓ is the spherical Bessel function of order ℓ, $u_{\ell,s}$ is the sth positive zero of j_ℓ, and the radius satisfies $k \in [0, 1/2]$ in Fourier space since $1/2$ is the Nyquist frequency [ISBI].

Then

$$\widehat{I}_\omega(ck,\theta) = \sum_{\ell,m,m',s} a_{\ell,m,s} Y_{\ell,m'}(\pi/2, 0) D^\ell_{m',m}(\omega) j_{\ell,s}(k) e^{im'\theta},$$

whenever $k \leq 1$. Express this in 2D PSWFs so

$$\widehat{I}_\omega(ck,\theta) = \sum_{N,n} b_{N,n} \psi_{N,n}(k,\theta),$$

where in papers the 2D PSWFs are defined as

$$\psi_{N,n}(k,\theta) = \frac{1}{\sqrt{2\pi}} R_{N,n}(k) e^{ik\theta},$$

but in Boris' code the expansion is actually performed with respect to

$$\widetilde{\psi}_{N,n}(k,\theta) = \frac{1}{2\sqrt{2\pi}}\alpha_{N,n}R_{N,n}(k)e^{ik\theta},$$

where $\alpha_{N,n}$ are the eigenvalues associated with $\psi_{N,n}$ - see Sect. **??** below.

We then get

$$\begin{aligned}
b_{N,n} &= \frac{1}{\sqrt{2\pi}}\int_0^{2\pi}\int_0^1 \widehat{I}_\omega(ck,\theta)R_{N,n}(k)e^{-iN\theta}k\,dk\,d\theta, \\
&= \sum_{\ell,m,m',s} a_{\ell,m,s}[\sqrt{2\pi}Y_{\ell,m'}(\pi/2,0)]D^\ell_{m',m}(\omega)\left(\int_0^1 j_{\ell,s}(k)R_{N,n}(k)k\,dk\right)\left(\frac{1}{2\pi}\int_0^{2\pi} e^{i(m'-N)\theta}\,d\theta\right), \\
&= \sum_{\ell\geq|N|}\sum_{m,s} a_{\ell,m,s}D^\ell_{N,m}(\omega)\beta_{\ell,s;N,n},
\end{aligned}$$

where

$$\beta_{\ell,s;N,n} = \begin{cases} \sqrt{2\pi}Y_{\ell,N}(\pi/2,0)\int_0^1 j_{\ell,s}(k)R_{N,n}(k)k\,dk, & \ell\geq|N|, \\ 0, & \ell<|N| \end{cases},$$

can be precomputed.

Then, back in real space,

$$\begin{aligned}
I_\omega(r,\varphi) &= \sum_{N,n}\widehat{\alpha}_{N,n}b_{N,n}\psi_{N,n}(r,\varphi), \\
&= \sum_{\ell=0}^{L}\sum_{N,m=-\ell}^{\ell}\sum_{n=0}^{n_{\max}(N)}\sum_{s=1}^{S(\ell)} a_{\ell,m,s}\widehat{\beta}_{\ell,s;N,n}D^\ell_{N,m}(\omega)\psi_{N,n}(r,\varphi).
\end{aligned}$$

where $\alpha_{N,n}$ is the eigenvalue corresponding to the (N,n)th PSWF, $\widehat{\alpha}_{N,n} = (c/2\pi)^2\alpha_{N,n}$, and $\widehat{\beta}_{\ell,s;N,n} = \widehat{\alpha}_{N,n}\beta_{\ell,s;N,n}$. In this notation, we assumed I_ω has bandlimit c and is concentrated in the unit ball in real space. We then consider the product

$$\begin{aligned}
m_3(\Delta\mathbf{r}_1,\Delta\mathbf{r}_2) &= \int_{\mathbf{r}}\langle I_\omega(\mathbf{r})I_\omega(\mathbf{r}+\Delta\mathbf{r}_1)\overline{I_\omega(\mathbf{r}+\Delta\mathbf{r}_2)}\rangle_\omega, \\
&= \sum_{\substack{N_1,n_1\\N_2,n_2\\N_3,n_3}}\langle b_{N_1,n_1}b_{N_2,n_2}\overline{b_{N_3,n_3}}\rangle_\omega\int_{\mathbf{r}}\psi_{N_1,n_1}(\mathbf{r})\psi_{N_2,n_2}(\mathbf{r}+\Delta\mathbf{r}_1)\overline{\psi_{N_3,n_3}(\mathbf{r}+\Delta\mathbf{r}_2)}.
\end{aligned}$$

Now,

$$\begin{aligned}
\langle b_{N_1,n_1}b_{N_2,n_2}\overline{b_{N_3,n_3}}\rangle_\omega = &\sum_{\substack{\ell_1,m_1,s_1\\\ell_2,m_2,s_2\\\ell_3,m_3,s_3}} a_{\ell_1,m_1,s_1}a_{\ell_2,m_2,s_2}\overline{a_{\ell_3,m_3,s_3}}\langle D^{\ell_1}_{N_1,m_1}(\omega)D^{\ell_2}_{N_2,m_2}\overline{D^{\ell_3}_{N_3,m_3}}\rangle_\omega \\
&\times\beta_{\ell_1,s_1;N_1,n_1}\beta_{\ell_2,s_2;N_2,n_2}\overline{\beta_{\ell_3,s_3;N_3,n_3}},
\end{aligned}$$

and [Tamir's note on Kam's bispectrum]

$$\langle D^{\ell_1}_{N_1,m_1}(\omega)D^{\ell_2}_{N_2,m_2}\overline{D^{\ell_3}_{N_3,m_3}}\rangle_\omega = (-1)^{N_3+m_3}\begin{pmatrix}\ell_1 & \ell_2 & \ell_3\\ N_1 & N_2 & -N_3\end{pmatrix}\begin{pmatrix}\ell_1 & \ell_2 & \ell_3\\ m_1 & m_2 & -m_3\end{pmatrix},$$

and since $\begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = 0$ unless $m_1+m_2+m_3=0$ and $|\ell_1-\ell_2| \le \ell_3 \le \ell_1+\ell_2$, we conclude that

$$\begin{aligned}\langle b_{N_1,n_1} b_{N_2,n_2} \overline{b_{N_3,n_3}} \rangle_\omega = \delta_{N_3,N_1+N_2} & \sum_{\substack{\ell_1,m_1,s_1 \\ \ell_2,m_2,s_2 \\ s_3}} \sum_{\ell_3=|\ell_1-\ell_2|}^{\min(L,\ell_1+\ell_2)} a_{\ell_1,m_1,s_1} a_{\ell_2,m_2,s_2} \overline{a_{\ell_3,m_1+m2,s_3}} \\ & \times (-1)^{N_1+N_2+m_1+m_2} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ N_1 & N_2 & -N_1-N_2 \end{pmatrix} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & -m_1-m_2 \end{pmatrix} \\ & \times \beta_{\ell_1,s_1;N1,n_1} \beta_{\ell_2,s_2;N_2,n_2} \overline{\beta_{\ell_3,s_3;N_1+N2,n_3}}.\end{aligned}$$

Finally, we expand

$$m_3(\Delta\mathbf{r}_1, \Delta\mathbf{r}_2) = \sum_{k,q_1,q_2} \mathfrak{m}_3(k,q_1,q_2) \psi_{k,q_1}(\Delta\mathbf{r}_1) \overline{\psi_{k,q_2}(\Delta\mathbf{r}_2)},$$

where we only include the block-diagonal terms in the expansion. Defining

$$\Psi_{\ell,N,s}(\mathbf{r}) = \sum_{n=0}^{n_{\max}(N)} \beta_{\ell,s;N,n} \psi_{N,n}(\mathbf{r}),$$

the final formula reads

$$\begin{aligned}\mathfrak{m}_3(k,q_1,q_2) = & \sum_{\substack{\ell_1,m_1,s_1 \\ \ell_2,m_2,s_2 \\ s_3}} \sum_{\ell_3=|\ell_1-\ell_2|}^{\min(L,\ell_1+\ell_2)} a_{\ell_1,m_1,s_1} a_{\ell_2,m_2,s_2} \overline{a_{\ell_3,m_1+m_2,s_3}} \\ & \times (-1)^{m_1+m_2} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & -m_1-m_2 \end{pmatrix} \\ & \times \sum_{N_1=-\ell_1}^{\ell_1} \sum_{N_2=-\ell_2}^{\ell_2} (-1)^{N_1+N_2} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ N_1 & N_2 & -N_1-N2 \end{pmatrix} \int_{\mathbf{r}} \Psi_{\ell_1,N_1,s_1}(\mathbf{r}) \rho^{(k,q_1)}_{\ell_2,N_2,s_2}(\mathbf{r}) \overline{\rho^{(k,q_2)}_{\ell_3,N_1+N_2,s_3}(\mathbf{r})},\end{aligned}$$

where

$$\rho^{(k,q)}_{\ell,N,s} = \int_{\Delta\mathbf{r}} \Psi_{\ell,N,s}(\mathbf{r}+\Delta\mathbf{r}) \overline{\psi_{k,q}(\Delta\mathbf{r})}.$$

In practice, the last line of the above expression for $\mathfrak{m}_3(k,q_1,q_2)$ is precomputed, and both the integration over $\mathbf{r}$ and over $\Delta\mathbf{r}$ is performed on the grid of the images in the dataset, to match the integration performed on the actual images.

C.1 The power spectrum

The power spectrum is easier to derive directly in Fourier space, to avoid integration of shifted prolates against centered ones. The average power spectrum in Fourier space can be derived from Kam's original formula [Kam, 1980; Eq. 10] by setting $\mathbf{k}_1 = \mathbf{k}_2$ to obtain

$$\langle |\widehat{I}_\omega(k,\theta)|^2 \rangle_\omega = \frac{1}{4\pi} \sum_{\ell,m} \left| \sum_s a_{\ell,m,s} j_{\ell,s}(k) \right|^2 = \frac{1}{4\pi} \sum_{\substack{\ell,m \\ s_1,s2}} a_{\ell,m,s_1} \overline{a_{\ell,m,s_2}} j_{\ell,s_1}(k) j_{\ell,s_2}(k),$$

where we used the fact that the normalized spherical Bessel functions $j_{\ell,s}$ are real. To expand the above in 2D PSWFs, we write

$$\langle|\widehat{I}_\omega(k,\theta)|^2\rangle_\omega = \sum_q \mathfrak{m}_2(q)\psi_{0,q}(k),$$

and conclude that

$$\mathfrak{m}_2(q) = \frac{\sqrt{2\pi}}{4\pi}\sum_{\substack{\ell,m\\ s_1,s_2}} a_{\ell,m,s_1}\overline{a_{\ell,m,s_2}}\int_0^1 j_{\ell,s_1}(k)j_{\ell,s_2}(k)R_{0,q}(k)k\,dk.$$

C.2 The mean

Since the Fourier transformed volume is given by

$$\widehat{V}(k,\theta,\varphi) = \sum_{\ell,m,s} a_{\ell,m,s}Y_{\ell,m}(\theta,\varphi)j_{\ell,s}(k),$$

since $j_{\ell,s}(0) = 0$ unless $\ell = 0$, and since $Y_{0,0}(\theta,\varphi) = \frac{1}{\sqrt{4\pi}}$, we conclude that

$$m_1[V] = \widehat{V}(\mathbf{0}) = \frac{1}{\sqrt{4\pi}}\sum_s a_{0,0,s}j_{0,s}(0).$$

$$C_2^{(q)} = \int_0^1 j_{\ell,s_1}(k)j_{\ell,s_2}(k)R_{0,q}(k)k\,dk$$

$$C_3^{(k,q_1,q_2)}\ell_1,m_1,s_1;\ell_2,m_2,s_2;\ell_3,s_3) := (-1)^{m_1+m_2}\begin{pmatrix}\ell_1 & \ell_2 & \ell_3\\ m_1 & m_2 & -m_1-m_2\end{pmatrix}$$
$$\times\sum_{N_1=-\ell_1}^{\ell_1}\sum_{N_2=-\ell_2}^{\ell_2}(-1)^{N_1+N_2}\begin{pmatrix}\ell_1 & \ell_2 & \ell_3\\ N_1 & N_2 & -N_1-N2\end{pmatrix}\int_{\mathbf{r}}\Psi_{\ell_1,N_1,s_1}(\mathbf{r})\rho^{(k,q_1)}_{\ell_2,N_2,s_2}(\mathbf{r})\overline{\rho^{(k,q_2)}_{\ell_3,N_1+N_2,s_3}(\mathbf{r})},$$

and

$$\rho^{(k,q)}_{\ell,N,s} = \int_{\Delta\mathbf{r}}\Psi_{\ell,N,s}(\mathbf{r}+\Delta\mathbf{r})\overline{\psi_{k,q}(\Delta\mathbf{r})}.$$

C.3 Implementation details

To implement the above formula, we precompute the quantities

$$W(\ell_1,\ell_2,\ell_3,m_1,m_2) = (-1)^{m_1+m_2}\begin{pmatrix}\ell_1 & \ell_2 & \ell_3\\ m_1 & m_2 & -m_1-m_2\end{pmatrix},$$

and

$$B^{(k,q_1,q_2)}(\ell_1,\ell_2,\ell_3,s_1,s_2,s_3) = \sum_{N_1=-\ell_1}^{\ell_1}\sum_{N_2=-\ell_2}^{\ell_2}W(\ell_1,\ell_2,\ell_3,N_1,N_2)\int_{\mathbf{r}}\Psi_{\ell_1,N_1,s_1}(\mathbf{r})\rho^{(k,q_1)}_{\ell_2,N_2,s_2}(\mathbf{r})\overline{\rho^{(k,q_2)}_{\ell_3,N_1+N_2,s_3}(\mathbf{r})},$$

so in each iteration of the optimization we compute

$$\mathfrak{m}_3(k,q_1,q_2) = \sum_{\substack{\ell_1,\ell_2,\ell_3\\ m_1,m_2\\ s_1,s_2,s_3}} W(\ell_1,\ell_2,\ell_3,m_1,m_2) B^{(k,q_1,q_2)}(\ell_1,\ell_2,\ell_3,s_1,s_2,s_3)\times$$
$$a_{\ell_1,m_1,s_1} a_{\ell_2,m_2,s_2} \overline{a_{\ell_3,m_1+m_2,s_3}}.$$

The Wigner $3j$ symbols are computed from the Racah formula

$$\begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & -m_3 \end{pmatrix} = \delta_{m_3,m_1+m_2}(-1)^{\ell_1-\ell_2+m_3}\sqrt{\Delta(\ell_1,\ell_2,\ell_3)}$$
$$\times \prod_{i=1}^{3}\sqrt{(\ell_i-m_i)!(\ell_i+m_i)!}\times\sum_t \frac{(-1)^t}{x(t)},$$

where

$$x(t) = t!(\ell_3-\ell_2+t+m_1)!(\ell_3-\ell_1+t-m_2)!(\ell_1+\ell_2-\ell_3+t)!(\ell_1-t-m_1)!(\ell_2-t+m_2)!,$$

the sum is over all integer t for which the arguments in the factorials in $x(t)$ are all nonnegative, and

$$\Delta(\ell_1,\ell_2,\ell_3) = \frac{(\ell_1+\ell_2-\ell_3)!(\ell_1-\ell_2+\ell_3)!(-\ell_1+\ell_2+\ell_3)!}{(\ell_1+\ell_2+\ell_3+1)!}.$$

D Noisy micrograph moments

We now suppose that the micrograph is perturbed by additive white Gaussian noise, so we observe $\widetilde{\mathcal{I}} = \mathcal{I} + \xi$ where $\xi \overset{\text{iid}}{\sim} \mathcal{N}(0,\sigma^2 I)$. We proceed to derive $\lim_{N\to\infty} m_2[\widetilde{\mathcal{I}}],\ m_3[\widetilde{\mathcal{I}}]$. For simplicity of notation, we shall use vectorized indices $\mathbf{i} = (i,j)$.

For the power spectrum:

$$\begin{aligned} m_2[\mathcal{I}+\xi](\Delta\mathbf{i}) &= \frac{1}{M^2}\sum_{i,j=1}^{M}\widetilde{\mathcal{I}}(\mathbf{i})\widetilde{\mathcal{I}}(\mathbf{i}+\Delta\mathbf{i}) \\ &= \frac{1}{M^2}\sum_{i,j=1}^{M}\mathcal{I}(\mathbf{i})\mathcal{I}(\mathbf{i}+\Delta\mathbf{i}) + \frac{1}{M^2}\sum_{i,j=1}^{M}\mathcal{I}(\mathbf{i})\xi(\mathbf{i}+\Delta\mathbf{i}) \\ &+ \frac{1}{M^2}\sum_{i,j=1}^{M}\xi(\mathbf{i})\mathcal{I}(\mathbf{i}+\Delta\mathbf{i}) + \frac{1}{M^2}\sum_{i,j=1}^{M}\xi(\mathbf{i})\xi(\mathbf{i}+\Delta\mathbf{i}). \end{aligned}$$

Considering the terms one-by-one, the first term is independent of the noise, and as shown above converges to $\gamma\langle m_2[I_\omega]\rangle_\omega$ as $N\to\infty$. Denoting the center of the instance of I_ω in $\mathcal{I}$ by $\mathbf{s}_\omega$, the second term satisfies

$$\begin{aligned} \frac{1}{M^2}\sum_{\mathbf{i}}\mathcal{I}(\mathbf{i})\xi(\mathbf{i}+\Delta\mathbf{i}) &= \frac{NL^2}{M^2}\cdot\frac{1}{NL^2}\sum_{\omega,\mathbf{i}} I_\omega(\mathbf{i})\xi_\omega(\mathbf{i}+\Delta\mathbf{i}) \\ &\to \gamma\mathbb{E}[\xi]\mathbb{E}[I_\omega] = 0, \end{aligned}$$

where $\xi_\omega(\mathbf{i}) = \xi(\mathbf{i} + \mathbf{s}_\omega)$ and $\mathbb{E}[I_\omega]$ is proportional to the mean of the volume. A similar argument applied to the third term shows that it also vanishes as $N \to \infty$. For the fourth term, if $\Delta\mathbf{i} \neq \mathbf{0}$ then since the noise is zero mean and i.i.d. this term vanishes. If $\Delta\mathbf{i} = \mathbf{0}$ then

$$\frac{1}{m^2}\sum_{i,j=1}^{m} \xi(\mathbf{i})^2 \to \sigma^2.$$

Thus, we conclude

$$m_2[\mathcal{I} + \xi](\Delta\mathbf{i}) \to \gamma\langle m_2[I_\omega](\Delta\mathbf{i})\rangle_\omega + \sigma^2\delta(\Delta\mathbf{i}).$$

For the third moments, we get 8 terms:

$$\begin{aligned} m_3[\mathcal{I} + \xi](\Delta\mathbf{i}_1, \Delta\mathbf{i}_2) = {} & \underbrace{\frac{1}{M^2}\sum_{\mathbf{i}} \mathcal{I}(\mathbf{i})\mathcal{I}(\mathbf{i} + \Delta\mathbf{i}_1)\mathcal{I}(\mathbf{i} + \Delta\mathbf{i}_2)}_{(1)} + \underbrace{\frac{1}{M^2}\sum_{\mathbf{i}} \xi(\mathbf{i})\xi(\mathbf{i} + \Delta\mathbf{i}_1)\xi(\mathbf{i} + \Delta\mathbf{i}_2)}_{(2)} \\ & + \underbrace{\frac{1}{M^2}\sum_{\mathbf{i}} \mathcal{I}(\mathbf{i})\xi(\mathbf{i} + \Delta\mathbf{i}_1)\mathcal{I}(\mathbf{i} + \Delta\mathbf{i}_2)}_{(3)} + \underbrace{\frac{1}{M^2}\sum_{\mathbf{i}} \mathcal{I}(\mathbf{i})\mathcal{I}(\mathbf{i} + \Delta\mathbf{i}_1)\xi(\mathbf{i} + \Delta\mathbf{i}_2)}_{(4)} \\ & + \underbrace{\frac{1}{M^2}\sum_{\mathbf{i}} \xi(\mathbf{i})\mathcal{I}(\mathbf{i} + \Delta\mathbf{i}_1)\mathcal{I}(\mathbf{i} + \Delta\mathbf{i}_2)}_{(5)} + \underbrace{\frac{1}{M^2}\sum_{\mathbf{i}} \mathcal{I}(\mathbf{i})\xi(\mathbf{i} + \Delta\mathbf{i}_1)\xi(\mathbf{i} + \Delta\mathbf{i}_2)}_{(6)} \\ & + \underbrace{\frac{1}{M^2}\sum_{\mathbf{i}} \xi(\mathbf{i})\xi(\mathbf{i} + \Delta\mathbf{i}_1)\mathcal{I}(\mathbf{i} + \Delta\mathbf{i}_2)}_{(7)} + \underbrace{\frac{1}{M^2}\sum_{\mathbf{i}} \xi(\mathbf{i})\mathcal{I}(\mathbf{i} + \Delta\mathbf{i}_1)\xi(\mathbf{i} + \Delta\mathbf{i}_2)}_{(8)}. \end{aligned}$$

We address these terms one by one:

- Term (1) is $m_3[\mathcal{I}]$, shown above to converge to $\gamma\langle m_3[I_\omega]\rangle_\omega$.
- Term (2) is $m_3[\xi]$, the bispectrum of pure noise, which vanishes as shown above.
- Terms (3)-(5) depend linearly on the noise and hence converge to zero.
- For term (6), if $\Delta\mathbf{i}_1 \neq \Delta\mathbf{i}_2$ the term vanishes as then $\xi(\mathbf{i} + \Delta\mathbf{i}_1)$ and $\xi(\mathbf{i} + \Delta\mathbf{i}_2)$ are independent. If $\Delta\mathbf{i}_1 = \Delta\mathbf{i}_2$ the term becomes

  $$\frac{1}{M^2}\sum_{\mathbf{i}} \mathcal{I}(\mathbf{i})\xi(\mathbf{i} + \Delta\mathbf{i}_1)^2 = \frac{NL^2}{M^2} \cdot \frac{1}{NL^2}\sum_{\omega,\mathbf{i}} I_\omega(\mathbf{i})\xi_\omega(\mathbf{i} + \Delta\mathbf{i})^2 \to \gamma\sigma^2\mathbb{E}[I_\omega] = \gamma\sigma^2 L m_1[V],$$

  where $m_1[V]$ is the mean of the volume (and hence also the mean of each projection I_ω).
- For term (7), if $\Delta\mathbf{i}_1 \neq \mathbf{0}$ then once again this term vanishes, whereas if $\Delta\mathbf{i}_2 = \mathbf{0}$ then it becomes

  $$\frac{1}{M^2}\sum_{\mathbf{i}} \xi(\mathbf{i})^2\mathcal{I}(\mathbf{i} + \Delta\mathbf{i}_2) = \frac{NL^2}{M^2} \cdot \frac{1}{NL^2}\sum_{\omega,\mathbf{i}} \xi_\omega(\mathbf{i})^2 I_\omega(\mathbf{i} + \Delta\mathbf{i}_2) \to \gamma\sigma^2 L m_1[V].$$

  Similarly, term (8) vanishes if $\Delta\mathbf{i}_2 \neq \mathbf{0}$ and converges to $\gamma\sigma^2 L m_1[V]$ otherwise.

Thus, we conclude that

$$m_3[\mathcal{I}+\xi](\Delta\mathbf{i}_1,\Delta\mathbf{i}_2) \to \gamma\langle m_3[I_\omega](\Delta\mathbf{i}_1,\Delta\mathbf{i}_2)\rangle_\omega + \gamma\sigma^2 L m_1[V]\Big(\delta(\Delta\mathbf{i}_1-\Delta\mathbf{i}_2)+\delta(\Delta\mathbf{i}_1)+\delta(\Delta\mathbf{i}_2)\Big).$$

Note that in practice, we have $m_1[\mathcal{I}+\xi] \to m_1[\mathcal{I}] \approx \gamma L m_1[V]$ since the noise has zero mean, so we do not need prior knowledge of γ to effectively debias the bispectrum.

D.1 Expansion in PSWFs

In practice, we compute the moments of our noisy micrograph in a product basis of 2D prolates, so we need to derive the expansion coefficients of the above bias terms in these functions for debiasing.

To this end, writing (in the continuous case again)

$$\delta(\Delta\mathbf{r}_1-\Delta\mathbf{r}_2) = \sum_{k,q_1,q_2} \mathfrak{d}(k,q_1,q_2)\psi_{k,q_1}(\Delta\mathbf{r}_1)\overline{\psi_{k,q_2}(\Delta\mathbf{r}_2)},$$

we get

$$\begin{aligned}
\mathfrak{d}(k,q_1,q_2) &= \int_{\Delta\mathbf{r}_1,\Delta\mathbf{r}_2} \delta(\Delta\mathbf{r}_1-\Delta\mathbf{r}_2)\overline{\psi_{k,q_1}(\Delta\mathbf{r}_1)}\psi_{k,q_2}(\Delta\mathbf{r}_2) \\
&= \int_{\Delta\mathbf{r}_2} \overline{\psi_{k,q_1}(\Delta\mathbf{r}_2)}\psi_{k,q_2}(\Delta\mathbf{r}_2) \\
&= \delta_{q_1,q_2}.
\end{aligned}$$

Similarly, writing

$$\delta(\Delta\mathbf{r}_1) = \sum_{k,q_1,q_2} \mathfrak{d}^{(1)}(k,q_1,q_2)\psi_{k,q_1}(\Delta\mathbf{r}_1)\overline{\psi_{k,q_2}(\Delta\mathbf{r}_2)},$$

we get

$$\begin{aligned}
\mathfrak{d}^{(1)}(k,q_1,q_2) &= \int_{\Delta\mathbf{r}_1,\Delta\mathbf{r}_2} \delta(\Delta\mathbf{r}_1)\overline{\psi_{k,q_1}(\Delta\mathbf{r}_1)}\psi_{k,q_2}(\Delta\mathbf{r}_2) \\
&= \overline{\psi_{k,q_1}(\mathbf{0})}\int_{\Delta\mathbf{r}_2} \psi_{k,q_2}(\Delta\mathbf{r}_2) \\
&= \delta_{k,0}R_{0,q_1}(0)\int_0^1 R_{0,q_2}(r)r\,dr \\
&= \delta_{k,0}\frac{\alpha_{0,q_2}}{2\pi}R_{0,q_1}(0)R_{0,q_2}(0),
\end{aligned}$$

where we used the fact that $R_{0,q}$ satisfies

$$\frac{\alpha_{0,q}}{2\pi}R_{0,q}(r) = \int_0^1 R_{0,q}(\rho)J_0(cr\rho)\rho\,d\rho,$$

where J_k is the Bessel function of the first kind, and that $J_0(0)=1$. A similar derivation applies to $\delta(\Delta\mathbf{r}_2)$.

Thus, in terms of the prolate expansion coefficients, our bias formulas become

$$\widetilde{\mathfrak{m}}_3(k,q_1,q_2) = \mathfrak{m}_3(k,q_1,q_2) + \sigma^2 m_1[\widetilde{\mathcal{I}}]\left[\delta_{q_1,q_2} + \delta_{k,0}\frac{1}{2\pi}(\alpha_{0,q_1}+\alpha_{0,q_2})R_{0,q_1}(0)R_{0,q_2}(0)\right].$$

Similarly, for the power spectrum we get

$$\widetilde{\mathfrak{m}}_2(q) = \mathfrak{m}_2(q) + \sigma^2 \frac{1}{\sqrt{2\pi}} R_{0,q}(0).$$