Noise limits in the assembly of diffraction data

Veit Elser

Department of Physics, Cornell University, Ithaca, NY 14853, USA

(Dated: January 9, 2014)

We obtain an information theoretic criterion for the feasibility of assembling diffraction signals from noisy tomographs when the positions of the tomographs within the signal are unknown. For shot-noise limited data, the minimum number of detected photons per tomograph for successful assembly is much smaller than previously believed necessary, growing only logarithmically with the number of resolution elements of the diffracting object. We also demonstrate assembly up to the information theoretic limit with a constraint-based algorithm.

Arbitrarily high levels of noise can be tolerated when an unlimited number of measurements are available and can be averaged to obtain the signal. A new challenge is introduced when the signal is interrogated tomographically, that is, by means of multiple sections of the signal. If the position of each tomograph within the signal is unknown, then each measurement must have at least the minimum information required to position the tomograph, for signal averaging to be feasible. We will refer to the signal processing demands posed by this scenario as *crypto-tomography*.

An instance of crypto-tomography is the assembly of diffraction data in the proposed x-ray free electron laser (XFEL) investigations of single molecules [1]. The goal of these experiments is to obtain the 3D structure of a molecule by algorithmically inverting to direct space the measurement of its continuous diffraction pattern. Each XFEL measurement provides information about one 2D tomograph (an Ewald sphere) of the 3D diffraction pattern. Since the molecular orientation in each measurement is random, and the total number of photons collected small, the data assembly problem will test the noise limits of crypto-tomography. To help assess the feasibility of these proposed experiments, we have investigated crypto-tomography in the case of a weak, shot-noise limited signal. Because our approach is information theoretic [2], the results apply to any algorithm that aims to assemble and average noisy diffraction data. We conclude with some results obtained with an algorithm that is able to assemble data close to the information theoretic limit.

To better understand the theoretical issues, we introduce a minimal, three parameter model of crypto-tomography. A sample signal is shown in Figure 1 and consists of N one-dimensional diffraction patterns generated by N one-dimensional objects. Each of the latter comprise M independent, complex-valued resolution elements Ψ_{mn}. The diffraction signal is given by

$$w_n(\theta) = \left| \sum_{m=-M/2}^{M/2} \Psi_{mn} e^{im\theta} \right|^2 \quad n = 1, \ldots, N , \quad (1)$$

where θ is the single angle that specifies the position of the tomograph. For any θ, the numbers $w_n(\theta)$ are the time-integrated photon fluxes recorded at N detector pixels. The third parameter of the model is the mean photon count per pixel, μ. We will be interested in the limit of large M and N. In this limit, the mean photon count is related to the statistics of the ensemble of resolution elements:

$$\langle |\Psi_{mn}|^2 \rangle = \mu/M . \quad (2)$$

A single exercise in crypto-tomography consists first in selecting one $M \times N$ set of resolution elements with the above statistics; this defines the correct diffraction signal. A noisy data set is then generated by repeatedly selecting a random θ and sampling N photon counts k_1, k_2, etc. from Poisson distributions with means given by $w_1(\theta)$, $w_2(\theta)$, etc. Every data item thus consists of an N-tuple of photon counts. Given an unlimited number of such data, we are interested in determining the feasibility of reconstructing the original signal as a function of the model parameters M, N, and μ.

Huldt *et al.* [3] studied crypto-tomography from the perspective of *classifying* the recorded photon counts. For the N-tuples of data in the model above, a decision is made if a pair $\{k_1(\theta_i), \ldots\}$ and $\{k_1(\theta_j), \ldots\}$ originated from different angles, $\theta_i \neq \theta_j$, or nearly the same angle, $\theta_i \approx \theta_j$ (on the scale $2\pi/M$ since M is the highest frequency in the angular variation of the signal). This decision is based on the value of the cross-correlation

$$c(\theta_i, \theta_j) = \sum_{n=1}^{N} k_n(\theta_i) k_n(\theta_j) \quad (3)$$

whose expected value distinguishes between counts derived from a common signal ($\theta_i \approx \theta_j$) or two independent signals [3]:

$$c_{ii} = \langle c(\theta_i, \theta_i) \rangle = 2N\mu^2 \quad c_{ij} = \langle c(\theta_i, \theta_j) \rangle = N\mu^2 . \quad (4)$$

The averages are with respect to Poisson distributions of photon counts with mean values given by diffraction signals w with Wilson statistics, $p(w) = e^{-w/\mu}/\mu$. A pair
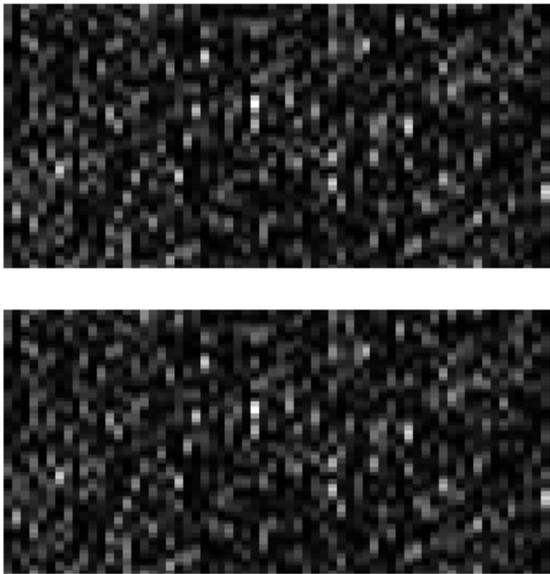
FIG. 1: Top: A diffraction signal comprising 64 continuous one dimensional signals arranged side-by-side. The tomography angle θ varies vertically. Bottom: Reconstruction of the top signal from 10^4 tomographic measurements (horizontal sections) of unknown θ. Each measurement was a 64-tuple of photon counts with mean count $\mu = 0.32$.

of N-tuples from different angles may be misidentified as originating from the same angle because the cross-correlation itself fluctuates. Evaluating the standard deviation σ_{ij} of $c(\theta_i, \theta_j)$ one finds [3]

$$\sigma_{ij}^2 = N(\mu^2 + 4\mu^3 + 3\mu^4) . \tag{5}$$

To avoid classification errors we must have $c_{ii} - c_{ij} \gg \sigma_{ij}$, which reduces to the statement

$$N\mu \gg \sqrt{N} \tag{6}$$

in the limit of small μ. This criterion, however, cannot be the fundamental limit since it makes no reference to M. As shown below, there is sufficient information for assembly with the much smaller number of detected photons given by (32). Finally, we demonstrate an assembly algorithm, not based on classification, that succeeds in this regime of high noise.

A criterion for the feasibility of crypto-tomography can be formulated in terms of the information content of an abstract function F that, given a measured set of photon counts $K = \{k_1, \ldots, k_N\}$, checks consistency with the reconstructed signal $W(\theta) = \{w_1(\theta), \ldots, w_N(\theta)\}$. For the reconstruction to be unique, F must be able to map fewer bits of input to a greater number of bits checked in the output. This is a restatement of the fact that F has access to a unique reconstruction — a positive source of information.

The number of bits in the output of F is the mutual information [2] $I(K, W)$ associated with the joint probability distribution of photon counts and signals:

$$p(K, W) = p(W) \int \frac{d\theta}{2\pi} p(K|W(\theta)) . \tag{7}$$

Here $p(W)$ is the prior distribution of signals, as specified by (1) and the statistics (2), and

$$p(K|W(\theta)) = \prod_{n=1}^{N} \frac{w_n(\theta)^{k_n}}{k_n!} e^{-w_n(\theta)} \tag{8}$$

is the Poisson distribution of photon counts at angle θ for signal W. The mutual information $I(K, W)$ gives the number of bits of information about W obtained, on average, from each measurement K at an unknown θ.

Given a model reconstruction W, the number of independent consistency checks associated with a measurement K has a size in bits given by the entropy of θ that remains, on average, after K is known:

$$I(K, \theta) = H(\theta) - H(\theta|K) . \tag{9}$$

This too is mutual information, now associated with the joint distribution $p(K, \theta) = p(K|W(\theta)) / 2\pi$. Since we are interested in the uniqueness question for average case signals, $I(K, \theta)$ should be averaged over signals with the prior distribution $p(W)$.

The crypto-tomography criterion can now be written down explicitly:

$$I(K, W) > \langle I(K, \theta) \rangle_W , \tag{10}$$

where the mutual information expressions

$$I(K, W) = \int dW p(W) \sum_K p(K|W) \log \frac{p(K|W)}{p(K)} \qquad \langle I(K, \theta) \rangle_W = \int dW p(W) \sum_K \int \frac{d\theta}{2\pi} p(K|W(\theta)) \log \frac{p(K|W(\theta))}{p(K|W)}$$

involve (8) and the marginal distributions $p(K|W) = \int p(K|W(\theta)) d\theta/2\pi$ and $p(K) = \int p(K|W)p(W)dW$. What follows is an analysis of what criterion (10) im-

plies about the parameters of our minimal model in the limit of large M and N.

The definitions above imply that he sum

$$I(K, W) + \langle I(K, \theta) \rangle_W = I(K, W(\theta)) \qquad (11)$$

corresponds to the mutual information associated with the photon counts K and the signal $W(\theta)$ at a known angle θ of measurement. The joint distribution $p(K, W(\theta)) = p(K|W(\theta)) p(W(\theta))$ involves the Wilson distribution of signal values:

$$dW p(W(\theta)) = \prod_{n=1}^{N} dw_n(\theta) e^{-w_n(\theta)/\mu}/\mu . \qquad (12)$$

Evaluating the mutual information we find $I(K, W(\theta)) = NI(\mu)$ where

$$I(\mu) = (1 + \mu) \log (1 + \mu) - \gamma\mu - \sum_{k=2}^{\infty} \left(\frac{\mu}{1 + \mu} \right)^k \log k . \qquad (13)$$

In the limit of interest, $\mu \to 0$, $I(\mu) \sim (1 - \gamma)\mu$, with the result

$$I(K, W(\theta)) \sim (1 - \gamma)N\mu , \qquad (14)$$

or about $(1-\gamma)/\log 2 \approx 0.61$ bits per photon (γ is Euler's constant).

We next evaluate $\langle I(K, \theta) \rangle_W$ in the limits of few and many detected photons $N\mu$. Since $N\mu$ is in either case large in the limit of large N, what matters is the relationship between $N\mu$ and the other parameter of the model, M. The few photon limit therefore corresponds to $N\mu$ fixed with $M \to \infty$. In this limit the complete prior distribution of signals is sampled by the process of sampling a particular, arbitrarily complex signal $W(\theta)$ at different θ. The distribution $p(K|W)$ in the expression for $\langle I(K, \theta) \rangle_W$ thus can be replaced by the distribution $p(K)$ with the result

$$\langle I(K, \theta) \rangle_W \sim I(K, W(\theta)) \qquad (M \to \infty) . \qquad (15)$$

The limit of many detected photons is an important point of reference, where the photon counts K can be assigned a unique θ up to a width defined by a Gaussian distribution. Using the symmetry of the mutual information, we can write $I(K, \theta) = I(\theta, K)$ in the form

$$I(K, \theta) = \left\langle \int d\theta \, p(\theta|K) \log 2\pi p(\theta|K) \right\rangle_K . \qquad (16)$$

In the limit of large $N\mu$, the distribution of θ is a Gaussian centered at some θ_K with standard deviation σ_K:

$$\log p(\theta|K) \sim - \log \sqrt{2\pi\sigma_K^2} - \frac{(\theta - \theta_K)^2}{2\sigma_K^2} . \qquad (17)$$

The resulting mutual information is then given by

$$I(K, \theta) \sim \left\langle \log \sqrt{\frac{2\pi}{e \, \sigma_K^2}} \right\rangle_K \sim \log \sqrt{\frac{2\pi}{e} \left\langle \frac{1}{\sigma_K^2} \right\rangle_K} , \qquad (18)$$
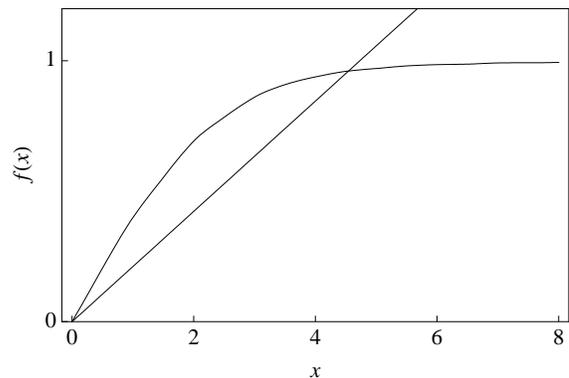

FIG. 2: Plot of the mutual information scaling function $f(x)$ and its intersection with a line of slope $(1 - \gamma)/2$.

where the average over K may be taken inside the logarithm because, as we shall see, σ_K^{-2} is the sum of N independent random terms.

Since $2\pi p(\theta|K) = p(K|W(\theta))/p(K|W)$ depends on θ only through $p(K|W(\theta))$, we have from (8) the equation

$$\log p(\theta|K) = \sum_{n=1}^{N} [k_n \log w_n(\theta) - w_n(\theta)] + \text{const.} \qquad (19)$$

Using (17), we obtain

$$\left\langle \frac{1}{\sigma_K^2} \right\rangle_K = - \left\langle \left(\frac{d}{d\theta} \right)^2 \Big|_{\theta=\theta_K} \log p(\theta|K) \right\rangle_K \qquad (20)$$

$$= \sum_{n=1}^{N} \left\langle \frac{w_n'(\theta_K)^2}{w_n(\theta_K)} \right\rangle_K \qquad (21)$$

$$= \sum_{n=1}^{N} \left\langle \frac{w_n'(\theta)^2}{w_n(\theta)} \right\rangle_\theta , \qquad (22)$$

where the last step makes use of the fact that the distribution on θ_K associated with the distribution of K is the uniform distribution.

The final step in the many photon limit of $\langle I(K, \theta) \rangle_W$ is to average (18) over signals W of the form (1). Since (22) is again a sum of N independent random terms, the average over W may be taken inside the logarithm:

$$\langle I(K, \theta) \rangle_W \sim \log \sqrt{\frac{2\pi}{e} N \left\langle \frac{w'(0)^2}{w(0)} \right\rangle_W} . \qquad (23)$$

The remaining average can be expressed in terms of random variables X and Y,

$$\frac{w'(0)^2}{w(0)} = \frac{X}{X^*}(Y^*)^2 + \frac{X^*}{X}(Y)^2 + 2|Y|^2 \qquad (24)$$

where

$$X = \sum_{m=-M/2}^{M/2} \Psi_m = \Psi_0 + \sum_{m=1}^{M/2} (\Psi_m + \Psi_{-m}) \qquad (25)$$
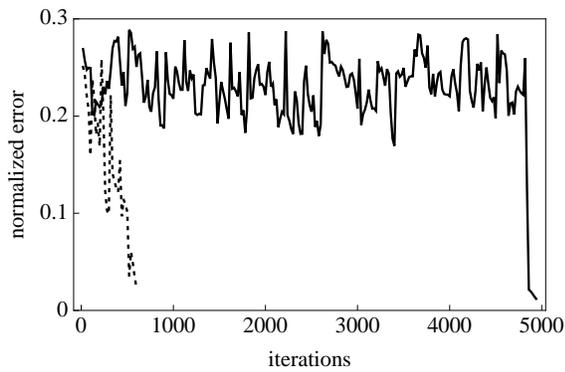
FIG. 3: Normalized difference map error versus iteration in overconstrained ($\mu = 0.32$) and underconstrained ($\mu = 0.25$, dashed curve) crypto-tomography reconstructions.

$$Y = i \sum_{m=-M/2}^{M/2} m\Psi_m = i \sum_{m=1}^{M/2} m(\Psi_m - \Psi_{-m}) . \quad (26)$$

Since Ψ_m and Ψ_{-m} are independent, so are X and Y. Associated with the arbitrariness of the angle θ is the uniformity of the phases of the Ψ_m; consequently, the phases of X and Y are uniformly distributed and $\langle (Y^*)^2 \rangle_W = \langle (Y)^2 \rangle_W = 0$. Using (2) for the third term we obtain

$$\left\langle \frac{w'(0)^2}{w(0)} \right\rangle_W = 2\langle |Y|^2 \rangle_W \sim \frac{1}{6}M^2\mu \quad (27)$$

in the limit of large M.

Combining (23) and (27), we obtain

$$\langle I(K,\theta) \rangle_W \sim I_\infty = \log\left(M\sqrt{\frac{2\pi}{6\,e}(N\mu)} \right) \quad (N\mu \to \infty) \quad (28)$$

for the many photon limit. We recognize $\log M$ as the scaling of the entropy of the tomography angle measured in speckle units of $2\pi/M$. Both of the limits (15) and (28) can be expressed in terms of a scaled photon count

$$x = N\mu/I_\infty , \quad (29)$$

and a dimensionless scaling function

$$\langle I(K,\theta) \rangle_W \sim I_\infty f(x) \quad (30)$$

behaving as $f(x) \sim (1-\gamma)x$ for small x and $f(x) \sim 1$ for large x. We have substantiated this claim by evaluating $\langle I(K,\theta) \rangle_W$ numerically for a large range of parameter values and find the results are consistent with a single function f plotted in Figure 2.

Using (11), (14) and (30), the criterion (10) takes the form

$$(1-\gamma)x > 2f(x) \quad (31)$$

which, as shown in Figure 2, requires $x > 4.5$. We therefore conclude that crypto-tomography for the three parameter model is feasible only when

$$N\mu > 4.5 \log\left(M\sqrt{\frac{2\pi}{6\,e}(N\mu)} \right) . \quad (32)$$

As an alternative to assembling a diffraction signal by first classifying its noisy tomographs, we propose using the tomographs as constraints on a *de novo* reconstruction. A general constraint satisfaction algorithm may then be able to operate right at the information theoretic limits of feasibility. We now present results obtained with the iterative difference map algorithm [4] that support this claim. A description of the algorithm and more extensive results are given elsewhere.

To test the criterion (10) we generated 10^4 sets of photon counts from random tomographs taken from an instance of the three parameter model with $M = 16$ and $N = 64$. Inequality (32) then implies that crypto-tomography is possible only for mean photon counts $\mu > 0.26$. An example of a successful reconstruction, for the case $\mu = 0.32$, is shown in Figure 1. Figure 3 shows the corresponding difference map error [4], a measure of the incompatibility of constraints at each iteration. An error that remains large during the search for the solution of a constraint satisfaction problem is an indication that the problem is overconstrained and that the solution will be unique. Correct reconstructions were obtained for μ as small as 0.29, at which point the search became difficult and required very many iterations. For $\mu < 0.26$ the behavior of the difference map error changed, decreasing to zero in few iterations (Fig. 3). This is consistent with the problem having become underconstrained, and in fact reconstructions obtained in this regime never agreed with the original diffraction signal. These observations suggest that the two forms of mutual information in the crypto-tomography criterion (10) correspond, respectively, to the numbers of constraints and free variables in a constraint satisfaction problem.

[1] R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, and J. Hajdu, Nature **406**, 752 (2000).
[2] C. E. Shannon, Bell System Tech. J. **27**, 379 (1948).
[3] G. Huldt, A. Szoke, and J. Hajdu, J. Struct. Biol. **144**, 219 (2003).
[4] V. Elser, I. Rankenburg, and P. Thibault, Proc. Natl. Acad. Sci. USA **104**, 418 (2007).