A Bayesian approach for denoising one-dimensional data: Supplementary Material

Charles R. Hogg III,[a,b]* Katharine Mullen[a] and Igor Levin[a]

[a]*Ceramics Division, National Institute of Standards and Technology (NIST), 100 Bureau Drive, M/S 8520, Gaithersburg, MD, 20899*, and [b]*Materials Science and Engineering Department, Univ. of Maryland, College Park, MD 20742-2115*.

E-mail: charles.hogg@nist.gov

(Received 0 XXXXXXX 0000; accepted 0 XXXXXXX 0000)

1. X-ray diffraction data from TiO_2 nanoparticles

To demonstrate that our technique can be applied to experimental data, we took X-ray powder diffraction curves from 20 nm TiO_2 nanoparticles. The true curve (with zero noise) is not accessible in real experiments, but we can reduce the noise arbitrarily by increasing the count time. We took 80 scans of 10 seconds each, which were all supposed to be equivalent. Our noisy dataset was constructed by summing four scans, and a low-noise comparison signal was formed from the sum of the remaining 76. The latter needed to be multiplied by a scaling factor Φ to bring it to the same level as the noisy data.

If all datasets were equivalent to one another, this factor would be the ratio of the number of scans, $\Phi = 4/76$. To test this, we ranked the scans by intensity for each of the 10769 pixels. The results for a representative subset are shown in the top half of Figure 1. The greyscale intensity in row i and column j represents the ranking of dataset i at detector pixel j: black denotes the dataset with the highest counts; white,

the dataset with the lowest counts. This simple visual test can detect both differences between datasets and correlations between neighboring pixels. Results for a simulated image with genuine independent Poisson noise are shown at bottom for comparison.

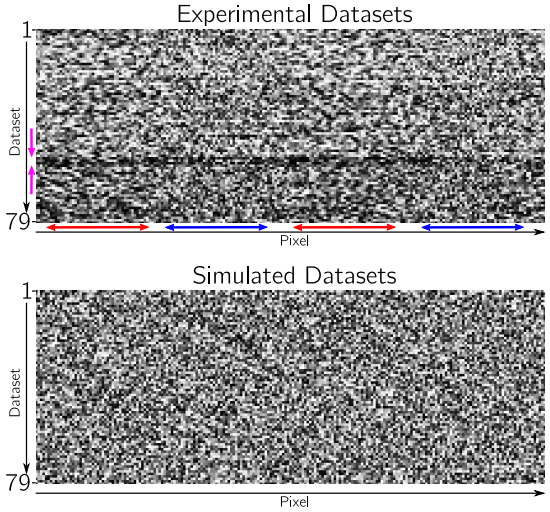

Fig. 1. Test that the datasets are statistically exchangeable. Each row represents a single noisy dataset, and each column corresponds to one value of 2θ. The greyscale intensity in a given column gives the rankings of the datasets at that value of 2θ: the dataset with the most counts is black, and that with the least counts is white. Comparing to simulated noisy data (bottom) reveals two main departures from our assumption. First, some rows are darker than others (see magenta arrows), suggesting fluctuations in the X-ray flux. Second, there is a correlation between neighboring pixels, whose lengthscale varies periodically from short (blue arrows) to long (red arrows).

First, it is clear that the datasets are *not* exactly equivalent. In particular, three rows indicated by the magenta arrows are clearly darker than the others, meaning that the corresponding datasets had disproportionately high counts. We attribute this to a fluctuation in the incident X-ray flux. Consequently, the scaling factor will be different for different datasets, and even for different regions within the same dataset, which implies the scaling factor Φ can vary with Q.

To find $\Phi(Q)$, we first took the square root of both datasets so the noise level would be the same. The ratio of the high-noise dataset (4 scans) to the low-noise dataset (76

scans) can be considered a noisy version of $\varphi(Q) \equiv \sqrt{\Phi(Q)}$; this ratio is shown as the gray bins in Figure 2. To find the underlying signal $\varphi(Q)$, we used the default settings (i.e., cubic regression splines) of the `geom_smooth` function in the `ggplot2` package (Wickham, 2009) for `R` (Figure 2, purple line). The red line is a reference showing the expected result of $\sqrt{4/76}$ if all noisy scans had been equivalent. As Figure 1 suggested, the actual $\varphi(Q)$ is generally lower than the expected value.

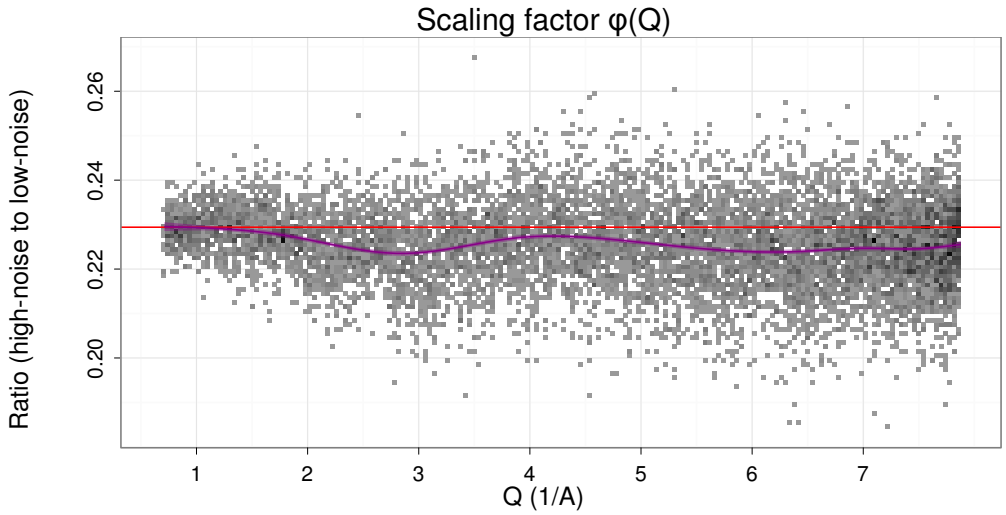

Fig. 2. The square root of the ratio of high-noise datapoints to low-noise datapoints. The gray bins show the raw ratio datapoints. The purple line shows the smoothed version $\varphi(Q)$, which we used to scale the low-noise data to get a comparison function. The red line shows the expected $\varphi(Q)$ when all scans are equivalent: $\sqrt{4/76}$.

Additionally, there is an interesting *horizontal* correlation in the experimental data. The correlation lengthscale oscillates with a period of roughly 100 pixels; two periods have been included in this figure. This oscillation does not appear to be related to any features of the signal; it seems to be due to the detector. In the short-lengthscale regions (indicated by blue arrows), the pattern appears similar to the simulated data, suggesting correlations are below the single-pixel scale; correlations in the long-lengthscale regions (red arrows) extend over multiple pixels.

If correlations between neighboring pixels blurs the measured intensity — including the noise — the practical effect would be to lower the noise variance. This is consistent with our findings, where we see the MSR with respect to noisy data is considerably smaller (≈ 0.16) than the expected value of 0.25. Thus, this X-ray data tests how well our method applies in real-world situations which violate our idealized assumptions of independent Poisson noise.

2. Novel prior for $\ell(Q)$: "hypocrisy"

$\ell(Q)$ is a function which tells functions of Q what to do: its natural interpretation is that they cannot have features near Q which are narrower than ℓ. Since $\ell(Q)$ is itself a function of Q, we can check whether it obeys its own decrees. A person who tells others to behave one way, but acts a different way, is called a hypocrite; we borrow this term for functions $\ell(Q)$ which exhibit more variation than they permit in other functions.

Further motivation for the term, "hypocrisy," comes from the way functions $f(Q)$ respond to a hypocritical $\ell(Q)$. Consider a constant $\ell_{\rm min}$ and a function $\ell(Q) > \ell_{\rm min}$ everywhere. Suppose $\ell(Q)$ has a fluctuation at Q_0 whose width $\ell_{\rm fluct}$ is much smaller than $\ell_{\rm min}$. As Gibbs (1997) pointed out, it is $\ell_{\rm fluct}$ rather than $\ell_{\rm min}$ which limits features around $f(Q_0)$. In other words: on encountering a hypocritical $\ell(Q)$, $f(Q)$ will do what it *does*, not what it *says*! This destroys the natural interpretation of the values of $\ell(Q)$, i.e., the lower limit of feature widths near Q. By using a prior which discourages hypocrisy, we hope to keep $\ell(Q)$ easily interpretable by researchers.

The goal of the remainder of this section is to derive a useful measure for the *hypocrisy* of a function. Our strategy is to build it in two stages: first, measure the hypocrisy density of a single point with respect to another point; second, integrate over all pairs of points in the function. Thus, for a suitable non-negative density

$\rho\left[\ell(Q');\ell(Q)\right]$, the total hypocrisy will be

$$H = \iint \rho\left[\ell(Q');\ell(Q)\right]\,\mathrm{d}Q'\,\mathrm{d}Q, \tag{1}$$

and the prior probability will be

$$p(\ell) = \exp\left[-H\left[\ell\right]\right]. \tag{2}$$

Conceptually, ρ is based on expectations about function values at a point Q', given the value at a point Q. As usual, we express these expectations in terms of probability. Plainly, the most probable value for $\ell(Q')$ should be the same as $\ell(Q)$, corresponding to the constant-ℓ case. The *variance* of $\ell(Q')$ should depend on the distance: it should be very small when Q' is close (in terms of $\ell(Q)$) to Q, and very large when Q' is far away. For simplicity, we choose a Gaussian distribution, which is fully specified by its mean and variance:

$$\rho\left[\ell(Q');\ell(Q)\right] = \log\left[\frac{N(\ell(Q');\ell(Q),\sigma^2)}{N(\ell(Q);\ell(Q),\sigma^2)}\right], \tag{3}$$

where $N(x;\mu,\sigma^2)$ is the Gaussian probability density function at x having mean μ and variance σ^2, and we have normalized to the most probable value to make the density zero by definition for constant functions. The problem then reduces to choosing a functional form for the variance $\sigma^2(Q';Q,\ell(Q))$.

The discussion will be clearest in terms of the *normalized* horizontal distance

$$D(Q';Q,\ell(Q)) \equiv (Q'-Q)/\ell(Q), \tag{4}$$

abbreviated to D when there is no ambiguity. The variance then depends only on D, and $\sigma^2(Q';Q,\ell_0) \equiv V(D)$ needs the following properties:

1. $\lim_{D\to\infty} V(D) = \infty$;
2. $\lim_{D\to 0} V(D) = 0$;

3. $\left[\frac{\mathrm{d}V}{\mathrm{d}D}\right]_{D=0} > 0.$

Property 1 means we are indifferent to values which are many correlation lengths away. Property 2 constrains $\ell(Q)$ to be single-valued. Property 3 allows $\ell(Q)$ to have nonzero slope.

To fulfil the third condition, we can write V in the form

$$V(D) \propto AD + g(D), \tag{5}$$

where A is the initial slope, and $g(D)$ has zero slope at $D = 0$. Now, $V(D)$ will fulfil conditions 1 and 2 whenever $g(D)$ does. We make an ansatz for $g(D)$ based on the *correlation*[1] of the two points, $C(D) = \exp(-(1/2)D^2)$:

$$g(D) = C(D)^{-n} - 1. \tag{6}$$

Thus, we have a *family* of hypocrisy measures, based on the initial slope A (higher values permit more variation), and the correlation exponent $n > 0$ (higher values constrain $\ell(Q)$ out to longer distances). Comparing the virtues of different parameter choices is beyond the scope of this paper. Here we simply choose $A = 1$ and $n = 2$ as default values, and observe that the resulting curves $\ell(Q)$ are in good agreement with the prior beliefs about smoothness which we intended to capture.

Putting it all together, the hypocrisy density is

$$\rho\left[\ell(Q');\ell(Q)\right] = \left(\frac{1}{2\tau^2}\right)\left(\frac{e}{e^{D(Q';Q,\ell(Q))^2} + D(Q';Q,\ell(Q)) - 1}\right)^2 \left(\frac{\ell(Q')}{\ell(Q)} - 1\right)^2, \tag{7}$$

where τ is the constant of proportionality in Equation 5. We call it the "tolerance for hypocrisy", because $\tau = 0$ permits only constant $\ell(Q)$, while as $\tau \to \infty$ there are no constraints on $\ell(Q)$ at all. Values for τ can be chosen by model selection techniques similar to those discussed in Section 2.3 of the paper ("Model selection and Occams razor"), though this does require Markov Chain Monte Carlo techniques. In this paper, we have used $\tau^2 = 10^{-4}$, since we observed that it gave good smoothness in $\ell(Q)$.

[1] Technically, the equivalent constant-ℓ correlation.

References

Gibbs, M. (1997). *Unpublished doctoral dissertation, University of Cambridge.*
URL: *http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.147.1130—&rep=rep1—&type=pdf*

Wickham, H. (2009). *ggplot2: elegant graphics for data analysis.* Springer New York.
URL: *http://had.co.nz/ggplot2/book*