Noise in Linear Inverse Problems

Plamen Stefanov

PURDUE UNIVERSITY

In collaboration with Samy Tindel (Purdue)

We want to solve

$$Af = g$$

with A and g given. Assume noisy measurements $g + g_{\mathrm{noise}}$. So we are solving

$$Af = g + g_{\mathrm{noise}}.$$

Say A is invertible in some sense. Since this is a linear problem, we will get back the true f plus

$$f_{\mathrm{noise}} = A^{-1} g_{\mathrm{noise}}.$$

We want to understand f_{noise}. Note that g_{noise} may not be in the range of A.

Noise could be not just additive – it could be multiplicative, modulation noise (Poisson), etc.

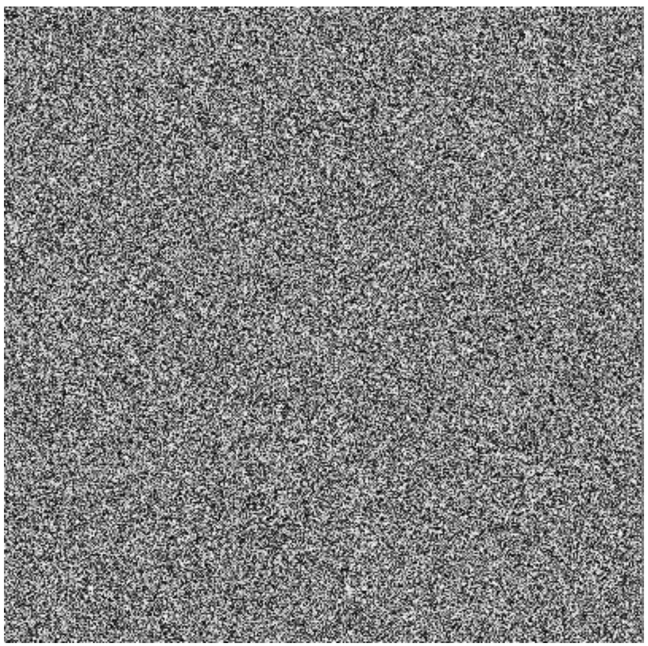
Our measurements are discrete. So are our numerical inversions. The models are “continuous”. We switch between discrete and “continuous” functions.

We assume that the noise is added at the discrete stage: either at the finitely many sensors or in (discrete) numerical simulations.

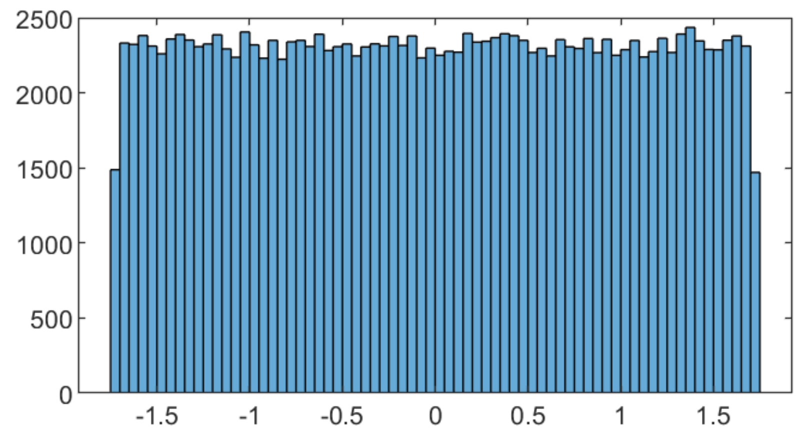
What is noise?

We want to model noisy <u>discrete</u> measurements. Assume that at each detector we have a random variable with a given distribution; and those variables are independent (independent and identically distributed random variables).

Then we get something like this.

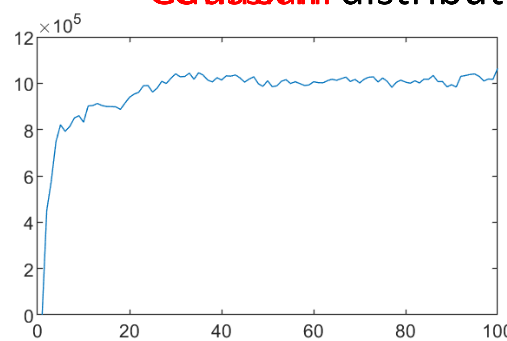


Gaussian noise
Uniform noise



histogram

Gaussian distribution
Uniform distribution

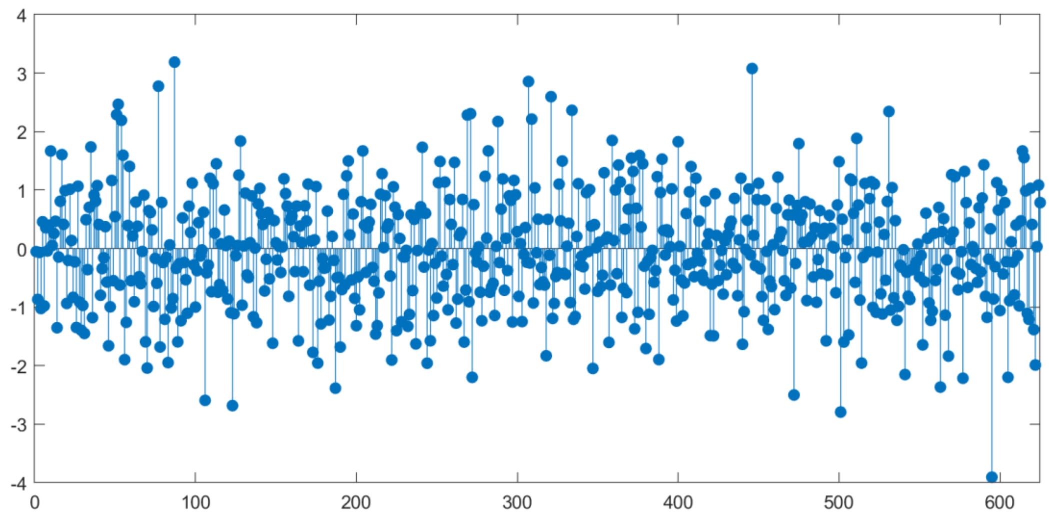


Spectrum, i.e., $\int |\hat{f}(r\omega)|d\omega$.

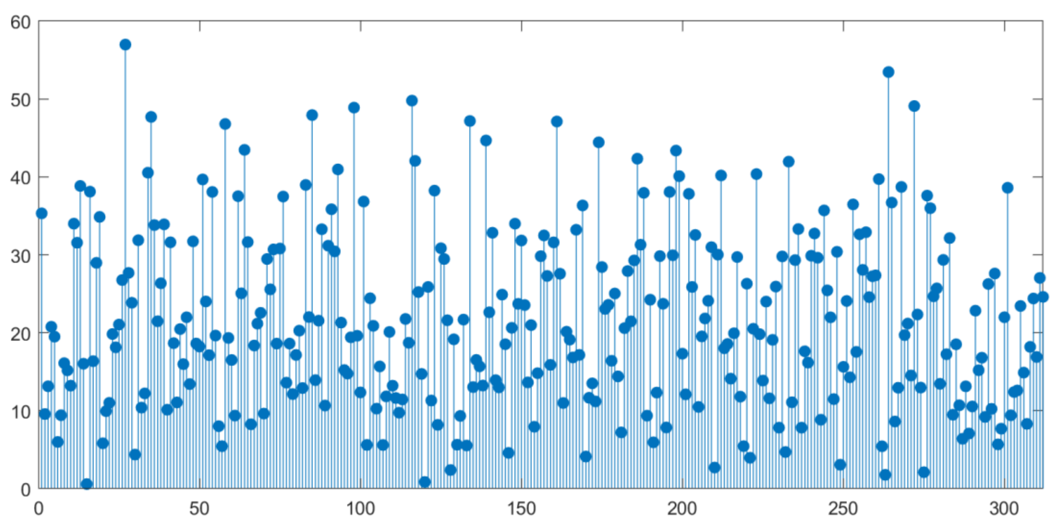
The uniform spectrum is a theorem, actually, well, kind of. It follows from the fact that the auto-correlation function is delta, so its Fourier transform, giving $|\hat{f}|^2$, must be constant. This is the reason it is called white: uniform spectrum as white light. Note that this is in the discrete setting, $\hat{f}(\xi)$ is the Discrete Fourier Transform with the number of discrete frequencies ξ being N.

Let us test it in one dimension.

How white is white noise?



White gaussian noise



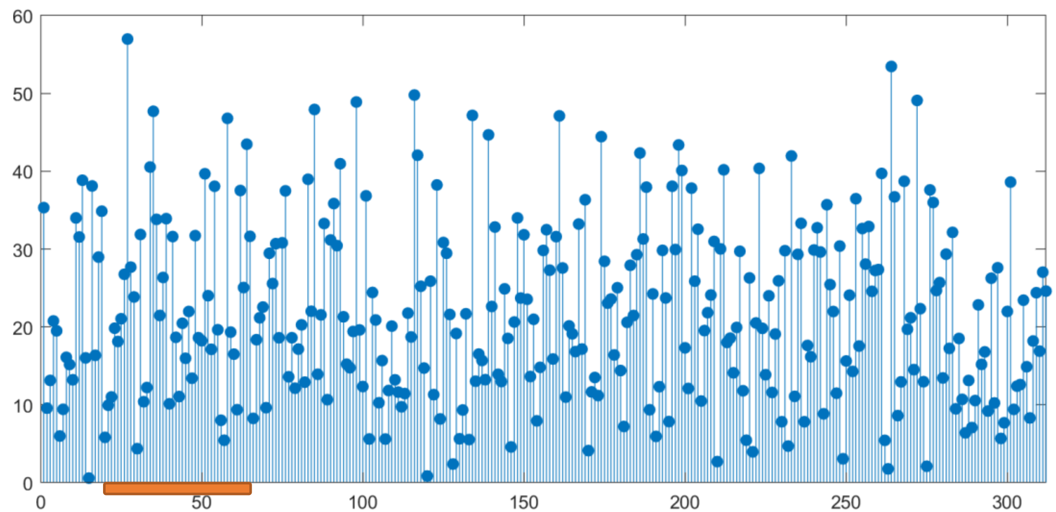
Power spectrum, $|\hat{f}|$.
Supposed to be uniform but it is … noisy.
St. deviation = mean!
"On average", it is uniform.

It does not get more uniform when $N \to \infty$!

There are two regimes:

1. Keep the number N of pixels fixed, and run the experiment many times. Then the expectation of $\left|\hat{f}(n)\right|^2$ is independent of n, and it is actually $N.\,\mathrm{VAR}(f)$. For the normalized DFT, it is just $\mathrm{VAR}(f)$.

2. Run the experiment once but keep increasing N. The spectrum stays hairy even for $N \geq 1$! On average however, it is flat.



The average over that interval tends to a constant as $N \to \infty$, and that constant is proportional to its length but it is independent of the position.

Say we have $f(x)$ on $[-1,1]^2$ discretized on an $N \times N$ grid. Set

$$h = \frac{2}{N}.$$

We think of $h > 0$ as a small parameter. The natural framework is the semi-classical one.

We want to identify functions on the discrete grid and functions of a continuous variable. We have to; any time we compute something on a discrete grid, we hope that this discretizes some function $f(x)$ and the accuracy gets better and better as $N \to \infty$, i.e., when $h \to 0$.

Shannon-Nyquist's Sampling Theory comes to the rescue. It says that if $f(x)$ has a Fourier transform $\hat{f}$ supported in the box $[-B, B]^n$ (i.e., it is band-limited), then $f(x)$ is uniquely and stably determined by its samples $f(sk), k \in \boldsymbol{Z}^n$ if the sampling rate s satisfies $0 < s \leq \pi/B$.

More precisely (Whittaker interpolation formula),

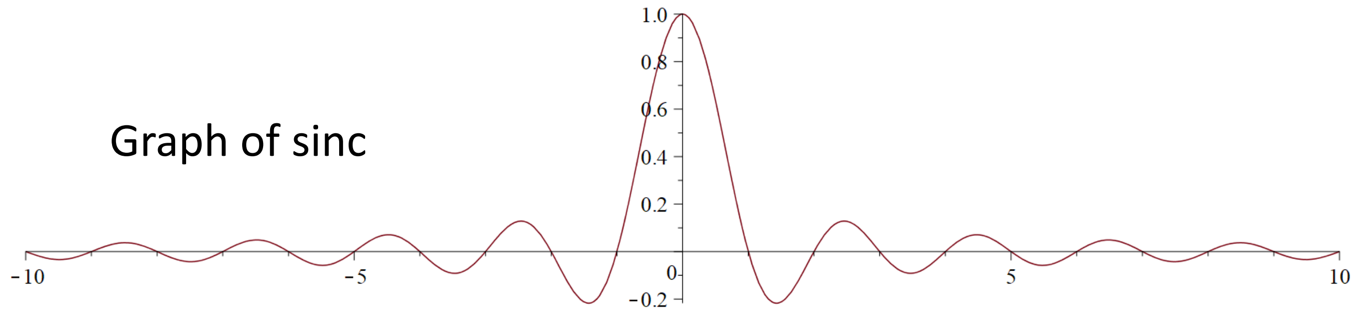
$$f(x) = \sum_{k \in \boldsymbol{Z}^n}^{\infty} f(sk) \chi\left(\frac{1}{s}(x - sk)\right), \qquad \chi(x) = \prod_j \operatorname{sinc}(x_j)$$

with $\operatorname{sinc}(x) = \sin(\pi x) / \pi x$, and (unitarity)

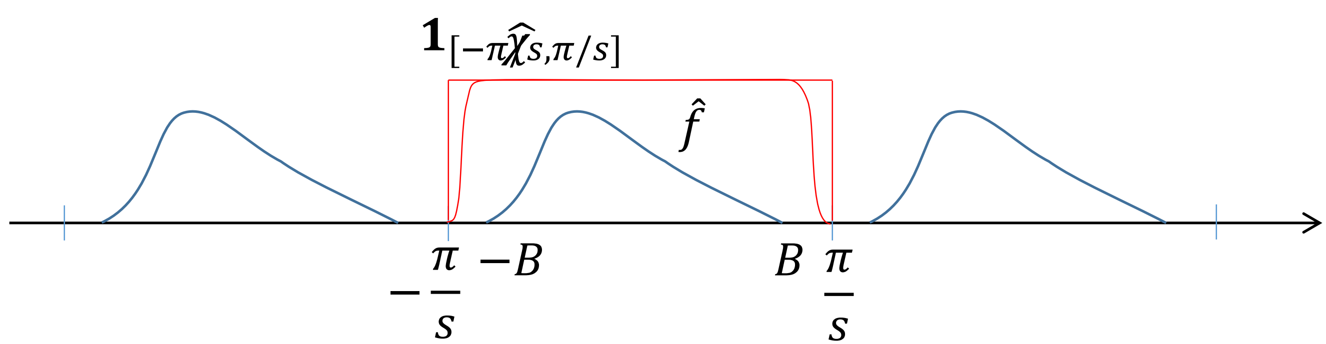
$$\|f\|^2 = s^n \sum_{k \in \boldsymbol{Z}^n}^{\infty} |f(sk)|^2 .$$

The proof is simple. Think of f as the inverse FT of $\hat{f}$. Then the samples $f(sk)$ are the Fourier coefficients of $\hat{f}$, more precisely of its periodic extension over the lattice $B\boldsymbol{Z}^n$. The interpolation kernel $\chi = \operatorname{sinc}$ comes from inverting the FT of the characteristic function of the band.

Graph of sinc



If we allow oversampling, i.e., the sampling rate s satisfies $s < \pi/B$ (strictly), we can have $\hat{\chi} = 1$ in $[-B, B]$ and $\operatorname{supp} f \subset \left[-\frac{\pi}{s}, \frac{\pi}{s}\right]$ to get χ decaying faster.



Assume it is done, and fix it. Then $\left\{\chi\left(\frac{\pi}{s}(x - sk)\right), k \in \mathbf{Z}^n\right\}$ is a partition of unity and the interpolation formula converts sequences to band-limited functions. Sampling converts them back to sequences.

Semi-classical sampling

Theorem (semi-classical sampling):

Let $f_h \in C_0^\infty(\Omega)$ with $\mathrm{WF}_h(f) \subset \Omega \times [-B, B]^n$. Then for $s \leq \pi/B$,

$$f_h(x) = \sum\nolimits_{k\in \boldsymbol{Z}^n} f(shk)\chi\left(\frac{1}{sh}(x - shk)\right) + O_S(h^\infty)\|f\|^2,$$

where χ is a product of sinc functions. Parseval's equality holds, too, up to $O(h^\infty)$.

Just a rescaled classical version, with error estimates. The condition on $\mathrm{WF}_h(f)$ is a condition on $\mathcal{F}_h f$ modulo $O(h^\infty)$.

The step size is sh with $s \leq \pi/B$.

As above, if $\mathrm{WF}_h(f) \subset \Omega \times (-B, B)^n$ (oversampling), χ can be made rapidly decreasing.

We call the projection $\Sigma_h(f)$ of $\mathrm{WF}_h(f)$ onto ξ the frequency set of f.

- Finitely many points are needed unlike the classical theorem. That number is $\sim h^{-N}$. In fact, we need at least $(2\pi h)^{-n}\mathrm{Vol}(\mathrm{WF}_h(f))$ many points even if we sample non-uniformly.

- Not exact, error $O(h^\infty)$.

- Differential operators are "bounded" on that space. In fact, $\|f\|_{H_h^s} \leq B^s\|f\| + O(h^\infty)$. Recall that D_x is naturally replaced by hD_x.

- One can have $[a,b]\times[c,d]$ (not a square), then different step-sizes w.r.t. each variable.

- One can sample over a parallelogram lattice (just apply a linear change of variables).

- The sampling rates are determined by the size of the smallest box containing $\Sigma_h(f)$, same as the essential support of $\mathcal{F}_h f$. Can be generalized.

Why work with semiclassically band limited functions $f(x,h)$ (those with a finite $\mathrm{WF}_h(f)$)? Most of the time, we have a single function $f(x)$ and we want to recover it.

Our measurements are actually averaged/blurred somehow. In TAT, we have a limited frequency and detectors not points. In optics, the wavelength of the light puts a bound of the resolution. In X-ray tomography, similarly – the rays are not infinitely thin, etc. We actually measure not Af but, say, $\phi_\epsilon * Af$, where $\phi_\epsilon(x) = \epsilon^{-n}\phi\left(\frac{x}{\epsilon}\right)$, $\phi \in C_0^\infty$.

$\phi_\epsilon *$ is an h-ΨDO (with a symbol $\phi(\xi)$). By Egorov's theorem

$$\phi_\epsilon * Af = AP(x,hD)f + \cdots$$

for some h-ΨDO $P(x,hD)$ with a compactly supported symbol. So f can be replaced by $\tilde{f}(x,h) = Q(x,hD)f$, and the new f depends on h and is s.c. band limited.

Also, let $f(x) \in C_0^\infty$ be h independent but $\hat{f}(\xi)$ is "small" for $|\xi| > B_0$. Then we can replace f with $\tilde{f}(x,h) \coloneqq \psi(x,hD)f$ with some $\psi \in C_0^\infty$ plus a "small" error.

Assume now that A is an FIO (classical). It has a canonical relation C. We have

$$\mathrm{WF}_h(Af) = C \circ \mathrm{WF}_h(f)$$

away from the zero section.

Therefore,

- Knowing the band limit of f, allows us to know the band limit of A, and then we know how to sample f.

- Given the sampling rate of Af, we know what resolution limit on f it poses.

- Aliasing exists when we undersample. It turns out that it is a simple h-FIO (a frequency shift). Aliased measurements of Af lead to non-local artifact of f. Those artifacts are given by an h-FIO as well, just compose the two.

- Locally averaged measurements can be handled.

Back to noise

So we identify discrete functions on girds like $N_1 \times N_2$ ones with $f(x,h)$ with $x^j \in [-a_j, a_j]$. Here, $N_1 \sim 1/h$, $N_2 \sim 1/h$. Then noise turns into such a s.c. band limited function $f(x,h)$. It has zero mean (as $h \to 0$).

Characterizing noise

- **Color** (spectrum)

White noise when the pixels are independent and equally distributed. Then

$$|\mathcal{F}_h f(\xi)| \approx \text{const.} \text{ (modulo lower order)}$$

over the band range, i.e., for $|\xi_j| \leq B_j$. We are not claiming that all such functions are "noise". We could have $\mathcal{F}_h f(\xi) = 1$ for $|\xi_j| \leq B_j$, so f could be the sinc function. We expect the phase to be "random" though.

We could have pink noise $|\mathcal{F}_h f(\xi)| \sim 1/|\xi|$ or blue noise $|\mathcal{F}_h f(\xi)| \sim |\xi|$.

- **Values distribution (histogram)**

This is $d_x \text{meas}(f^{-1}(-\infty, x))$. It is could be Gaussian, uniform, Poisson, etc.

Histogram and spectrum are independent characteristics.

- **Standard deviation/variance**

$$\text{VAR}_\Omega(f) = \frac{1}{|\Omega|} \int_\Omega f^2(x)\, dx, \qquad \text{STD}(f) = \sqrt{\text{VAR}(f)}.$$

A rule of a thumb is that noise with $\text{STD} = \sigma$ would “burry in the noise” signals with amplitudes $< \sigma$ and will preserve signals with amplitudes $> \sigma$. This is not rigorous, not a theorem, and not a sharp threshold.

Problem: We add noise with known characteristics (spectrum, distribution, and STD) to the data.

How will that affect the noise in the reconstruction?

Think microlocally. Noise has strength (STD) and spectrum. Why not combine them in the phase space. We want to measure its strength at each $(x, \xi) \in T^*\mathbf{R}^n \setminus 0$. Remember, the phase space is not a conic set!

We propose to use the s.c. defect measure for that.

Definition (semi-classical defect measure):

Let $a(x, \xi) \in C_0^\infty$. For every $f(x, h)$ (s.c. band limited), with $\|f(\cdot, h)\| \leq C$, there exists a measure $d\mu_f$ so that

$$\lim_{h \to 0} (a(x, hD) f_h, f_h) = \int a(x, \xi)\, d\mu_f$$

for some subsequence $h = h_j \to 0$.

Note that $\mathrm{supp}(d\mu_f) = \mathrm{WF}_h(f)$. Also, $\int d\mu_f = \left\| f_{h_j} \right\|^2 + o(1)$.

We could assume that our noise has this limit for all $h \to 0$. Also, our noise would have measures of the type $d\mu_f = \gamma_f(x, \xi) dx\, d\xi$ with $\gamma_f(x, \xi)$ regular enough. For white noise, for example, $\gamma_f(x, \xi) = \mathrm{const.}$

Now we have, for a subdomain Ω,

$$\mathrm{VAR}_\Omega(f) = \frac{1}{|\Omega|}\int_{T^*\Omega} \gamma_f(x,\xi)\, dx\, d\xi.$$

We can define variance (and STD) at a point x by

$$\mathrm{VAR}_x(f) = \int \gamma_f(x,\xi)\, d\xi.$$

Or better yet, do not touch it, and interpret $\gamma_f(x,\xi)$ as the spectral density of the variance at any point in the phase space.

Before taking the limit, the defect measure is known as the Wigner function $W_f^h(x,\xi)$ defined by

$$(a^w(x,hD)f_h, f_h) = \int a(x,\xi) W_f^h(x,\xi) dx\, d\xi\,.$$

Colin de Verdière used the expectation of the Wigner function to define the power spectrum of the noise (in geophysical applications). $W_f^h(x,\xi)$ may take negative values though!

Theorem. For white noise, $E(d\mu_f) = \mathrm{VAR}(f)$ over the Nyquist frequency range. In particular, the power spectrum is constant.

How is the defect measure transformed under an FIO?

This follows directly from Egorov's theorem. Assume from now on that F is an elliptic FIO associated with a local diffeomorphism C. Then

$$\gamma_{Ff} = \left(b\gamma_f\right) \circ C^{-1}, \qquad b = \sigma_p(F^*F).$$

Remember, we are solving

$$Af_{\text{noise}} = g_{\text{noise}}.$$

How to invert it? Say, A is a matrix, not necessarily invertible. The *Moore–Penrose inverse* is given by $A^{-1} = (A^*A)^{-1}A^*$, with $(A^*A)^{-1}$ restricted to $\mathrm{Ran}(A^*) = \mathrm{Ker}(A)^{\perp}$. Same as least-squares optimization. Assume A is an elliptic FIO associated to a local canonical diffeo. Then A^{-1} for us is just a parametrix.

$$f_{\text{noise}} = A^{-1}g_{\text{noise}}.$$

Then set $F = A^{-1}$ to get

$$\gamma_{f_{\text{noise}}} = \left(b^{-1}\gamma_{g_{\text{noise}}}\right) \circ C, \qquad b = \sigma_p(A^*A).$$

So we can compute the microlocal STD (defect measure) of the noise of the reconstruction given the microlocal STD of the data. This allows us to compute the noise even with a filter.

The Radon Transform in the plane in "parallel geometry"

Let $\mathcal{R}$ be the Radon transform in 2D

$$\mathcal{R}f(\omega, p) = \int_{x\cdot\omega=p} f(x)dl_x.$$

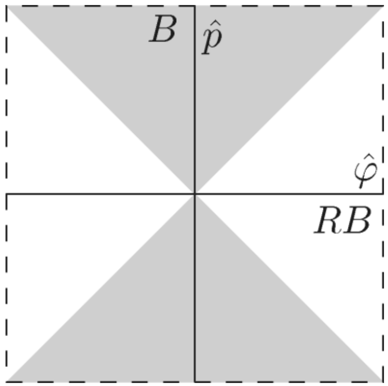
Write $\omega(\varphi) = (\cos\varphi, \sin\varphi)$. It is well known that $\mathcal{R}$ is an FIO with canonical relation $C = C_- \cup C_+$, where

$$C_\pm(x,\xi) = (\underbrace{\arg(\pm\xi)}_{\varphi}, \underbrace{\pm x\cdot\xi/|\xi|}_{p}, \underbrace{-x\cdot\xi^\perp}_{\hat{\varphi}}, \underbrace{\pm|\xi|}_{\hat{p}}).$$

Each $C_\pm$ is a diffeo, with inverse $(x,\xi) = C_\pm^{-1}(\varphi, p, \hat{\varphi}, \hat{p})$ given by

$$x = p\omega(\varphi) - (\hat{\varphi}/\hat{p})\omega^\perp(\varphi), \qquad \xi = \hat{p}\,\omega(\varphi).$$

Assume $\mathrm{WF}_h(f) \subset \{|x| < R, |\xi| < B\}$. Take C of that and project it to the $(\hat{\varphi}, \hat{p})$ variables. We get



$\Sigma_h(\mathcal{R}f)$

The smallest bounding box is

$$[-RB, RB] \times [-B, B].$$

This determines the relative sampling rates (before multiplying by h)

$$s_\varphi < \frac{\pi}{RB}, \qquad s_p < \frac{\pi}{B}.$$

For f, it is enough

$$s_{x^1} = s_{x^2} < \frac{\pi\sqrt{2}}{B}.$$

We have $\sigma_p(\mathcal{R}^*\mathcal{R}) = \frac{4\pi}{|\xi|}$. Therefore, for the solution of $\mathcal{R}f_{\text{noise}} = g_{\text{noise}}$, we get

$$\gamma_{f_{\text{noise}}} = \frac{|\xi|}{4\pi}\gamma_{g_{\text{noise}}} \circ C.$$

Assume no under-sampling, and that the recovered f will be restricted after reconstruction to its (known) frequency range. Assume white noise $\gamma_{g_{\text{noise}}} = \gamma^{\#}_{g_{\text{noise}}} = \text{const.}$ from now on. Then

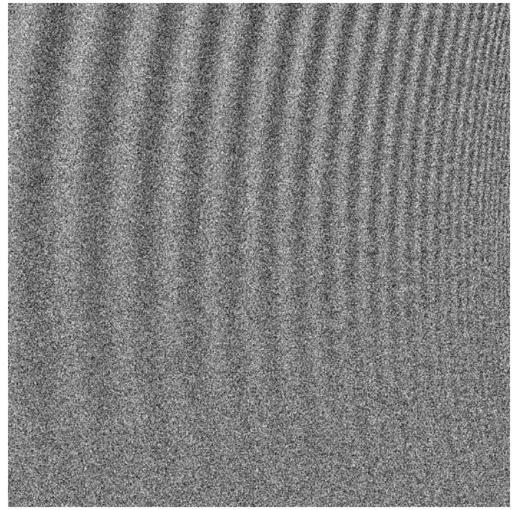
$$\gamma_{f_{\text{noise}}} = \frac{|\xi|}{4\pi}\gamma^{\#}_{g_{\text{noise}}}, \qquad |\xi| \leq B.$$

This is blue noise. Recall that the STD is a square root of that, so the STD (or the spectrum) of the reconstructed noise is like $\sim |\xi|^{\frac{1}{2}}$.

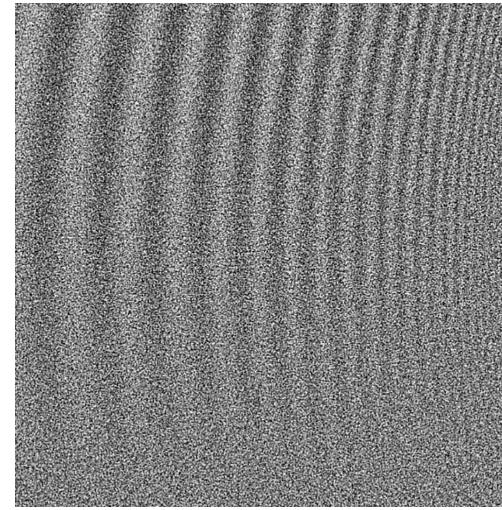
- It is x-independent
- It is isotropic in ξ (depends on $|\xi|$ only).

comparison

Original f, range in $[-1,1]$



f with white noise added, STD=1



Reconstructed f from noisy data, Hann filter, STD=1

The last two images have the same standard deviation at 800x800 only! When downsampled to a lower resolution, the third one would have less noise!

Assume a band limit for f equal to B as above, and the band limits for $\mathcal{R}f(\varphi, p)$ equal to $B_\varphi > RB$ and $B_p > B$. Those inequalities guarantee proper sampling. Then

$$\mathrm{STD}(\mathcal{R}^{-1}g) = \frac{B^{\frac{3}{2}}}{\sqrt{24 B_p B_\varphi}} \mathrm{STD}(g).$$

If $B_\varphi = RB$ and $B_p = B$ (sharp sampling), then

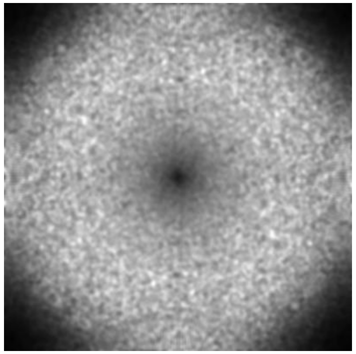
$$\mathrm{STD}(\mathcal{R}^{-1}g) = \frac{\sqrt{B}}{\sqrt{24R}} \mathrm{STD}(g).$$

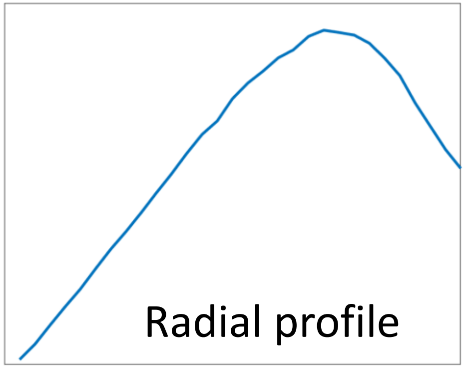
The first inequality seems strange – if we keep increasing B_p and B_φ (then we keep adding noise), the noise in the reconstruction decreases! In fact, we would have to cut the higher frequency noise in the inversion to fit it in $|\xi| < B$, which explains the paradox.

The second inequality says that the noise increases as $\sqrt{B}$.

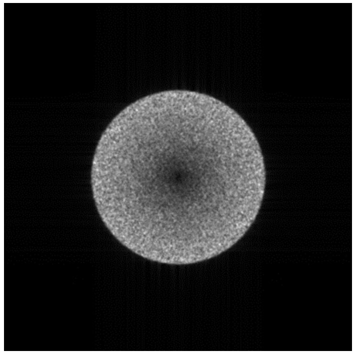
Numerical experiments

We take g to be white noise and invert it with iradon in MATLAB. Then we take $|\hat{f}|^2$ and plot it. In fact, we plot $|\hat{f}|$, not $|\hat{f}|^2$, for lower contrast.





Hm… The spectrum increases first but then it drops a bit. The reason is that iradon has some built-in smoothing. Do a higher accuracy inversion:



We get a linear increase as expected. The disk is smaller because we doubled N.

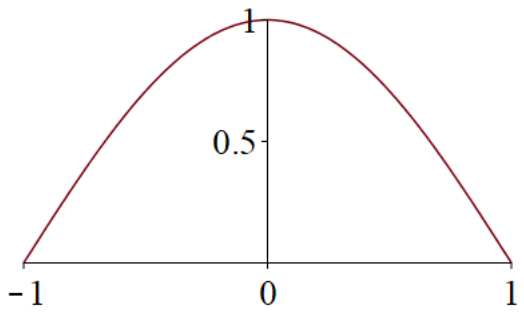
Numerical tests with $f = \mathcal{R}^{-1}g$ on an $N \times N$ grid discretizing $[-1,1]^2$, and g on a $2\pi N \times 2N$ grid equal to white noise. Here, m is the upsizing coefficient before inversion. Theoretically,

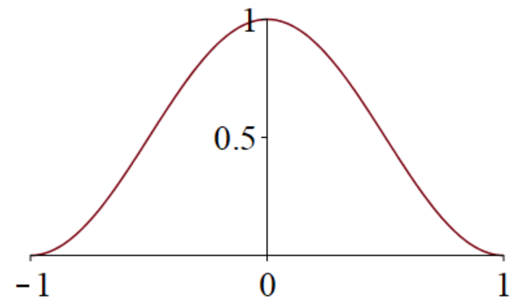
$$\text{Noise ratio} := \frac{\text{STD}(\mathcal{R}^{-1}g)}{\sqrt{N}.\,\text{STD}(g)} \approx 0.2558$$

Noise ratio with Gaussian noise. Theoretical ratio: 0.2558			
	$N = 100$	$N = 200$	$N = 300$
m=1	$0.2224 \pm 0.61\%$	$0.2223 \pm 0.32\%$	$0.2226 \pm 0.16\%$
m=2	$0.2552 \pm 0.38\%$	$0.2572 \pm 0.25\%$	$0.2578 \pm 0.17\%$
m=3	$0.2569 \pm 0.70\%$	$0.2584 \pm 0.41\%$	$0.2591 \pm 0.07\%$

Filtered Inversion

Often, the inversion is done with a low pass filter in the frequency domain.





Cosine filter $\cos\left(\frac{\xi}{B}\right)$. Here, $B = 1$.

Hann filter $\cos^2\left(\frac{\xi}{B}\right)$. Here, $B = 1$.

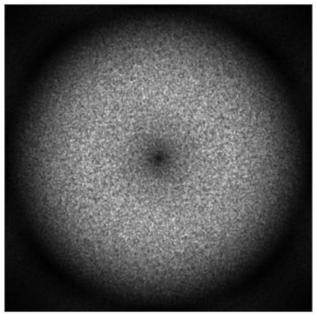
More generally, assume a filter $\chi(\xi) = \chi_0\left(\frac{\xi}{B}\right)$, so the inversion is $\mathcal{R}^{-1}\chi(|D_p|)g$.

The effect is multiplying the defect measure γ_f by $\chi_0^2\left(\frac{\xi}{B}\right)$ and multiplying the STD by the factor $\sqrt{c_\chi}$, where

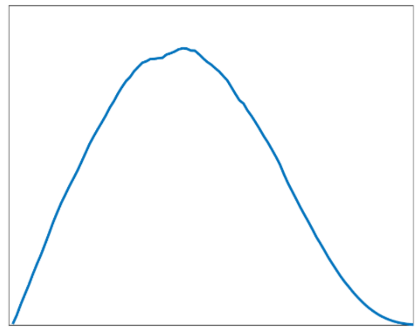
$$c_\chi = 3\int_0^1 \rho^2\chi_0^2(\rho)d\rho \leq 1.$$

For cosine, $\sqrt{c_\chi} \approx 0.4427$. For Hann, $\sqrt{c_\chi} \approx 0.3$.

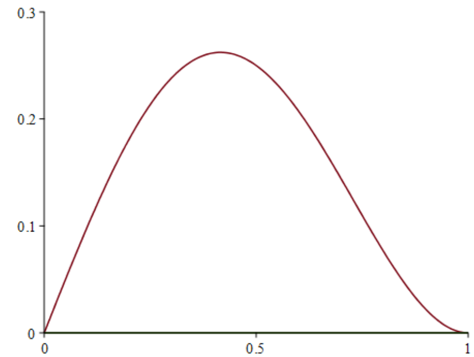
Spectrum of the noise with the <u>cosine filter</u>. Input: white noise.



Measured plot of $|\hat{f}|^2$



Measured radial profile of $|\hat{f}|^2$



Theoretical profile ($B = 1$)
$|\hat{f}(\rho\theta)|^2 = \rho . \cos^2 \frac{\pi\rho}{2}$

We should get 0.4427 of the STD compared to the non-filtered inversion.
We do. We test the Hann filter as well.

In many numerical simulations, we add a certain percentage of noise to the data and measure the percentage of noise in the reconstruction. Let us take a closer look at that.

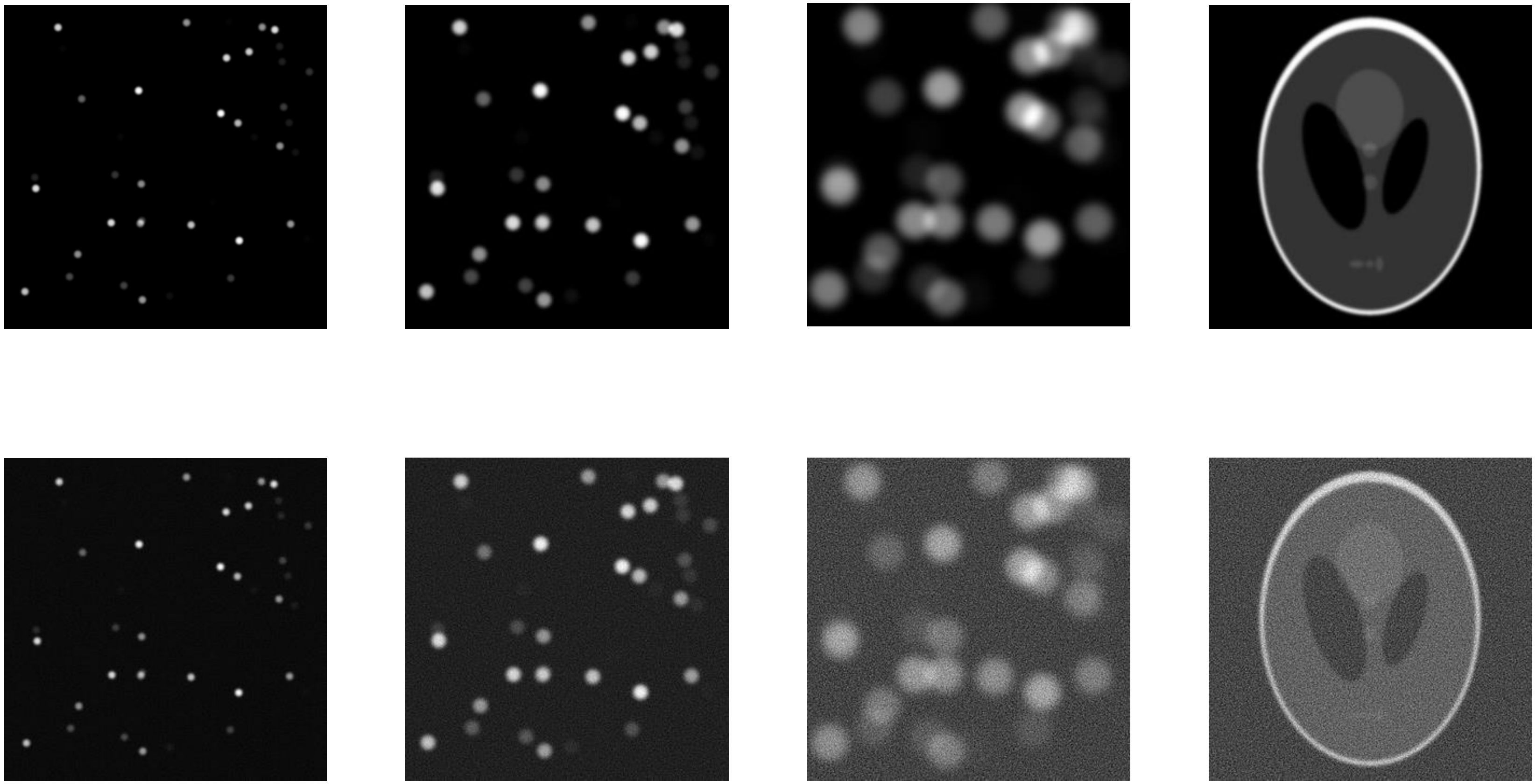
We reconstruct $f + f_{\text{noise}}$ from the data $\mathcal{R}f + g_{\text{noise}}$. Then $\|g_{\text{noise}}\|/\|\mathcal{R}f\|$ is the percentage of added noise, and $\|f_{\text{noise}}\|/\|f\|$ is the percentage of the measured one in the reconstruction. It does not matter if we replace norms by STD. We have

$$\frac{\|f_{\text{noise}}\|}{\|f\|} = K\frac{\|g_{\text{noise}}\|}{\|\mathcal{R}f\|}, \qquad K = \frac{\|f_{\text{noise}}\|}{\|g_{\text{noise}}\|}\cdot\frac{\|\mathcal{R}f\|}{\|f\|}.$$

The noise ratio we studied above. Depends on the discretization.

$= \frac{4\pi\left\||D|^{-\frac{1}{2}}f\right\|}{\|f\|}$, depends on f! Higher frequency f would yield a lower ratio. Also, functions with $\hat{f}(0) = \int f(x)dx = 0$ would gives us a very small ratio.

Examples. Top row: originals. Bottom row: reconstructed f with 20% noise added to $\mathcal{R}f$. All functions $f \geq 0$.



Added noise	25%	39.8%	74.4%	79.8%

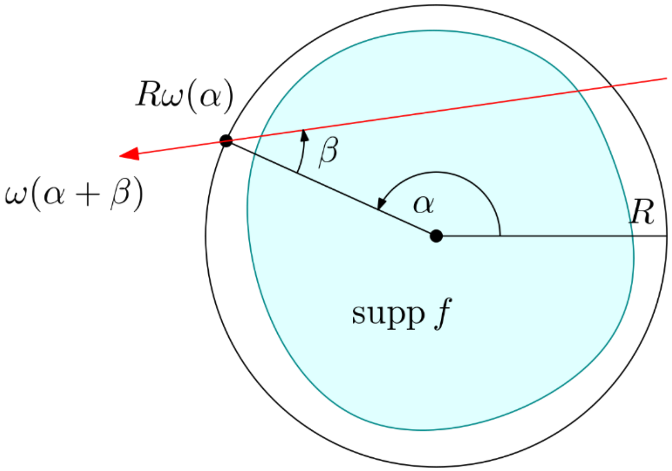
If we take f with $\int f(x)dx = 0$, then we get down to 10-12% or so.

So this is a very unreliable measure. If you want your reconstruction to look good,

- Use a coarser grid for f and much finer for $\mathcal{R}f$.
- Use high frequency images; or even better, phantoms with a zero mean.

On the other hand, for most conventional *images* like Lena, Shepp-Logan, etc., the ratio $\frac{\left\| |D|^{-\frac{1}{2}} f \right\|}{\|f\|}$ is pretty constant from image to image. This is related to some common statistical properties of the FT of such images.

The Radon Transform in the plane in fan-beam coordinates



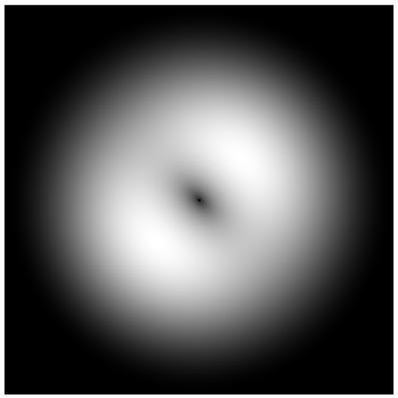
Parameterize lines by α, β as shown. Then $\mathcal{R}_{\mathrm{FB}} = \mathcal{R} \circ \Phi$, where Φ is the change of variables.

Then for the noise spectral density we get

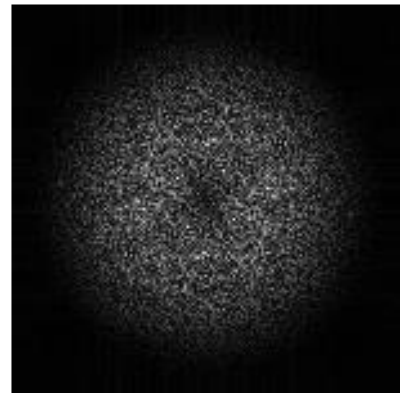
$$\gamma_{f_{\text{noise}}} = \frac{|\xi|}{4\pi}\left(1 - \frac{(x \cdot \xi)^2}{R^2|\xi|^2}\right)^{\frac{1}{2}} \gamma_{g_{\text{noise}}} \circ C.$$

The noise then depends on the position and the direction! The new factor is independent of $|\xi|$ though (in polar coordinates). Near $x = 0$, that factor is one. It gets to zero when $|x| = R$ and $x \parallel \xi$. We never get to $|x| = R$ though.

The power spectrum $\left|\hat{f}_{\mathrm{noise}}\right|^2$ of a small patch in the upper left corner centered at $(0.85R, 0.85R)$.



theoretical



measured

We compute the STD as before. Near the center same as in the previous case. Near the boundary $|x| = R$, it drops by about 20% only, based on the formula. We validate it numerically as well.

Thank you!