Annexe: Bayesian approach for electron quantum state tomography

March 9, 2019

1 The problem

In our measurement protocol, we don't performed a direct measurement of the quantity of interest. Indeed by measuring overlaps between the search quantity and probes, we only have access to convoluted results. This convolution is express in equation ... and can be model by a linear product with a matrix H. In addition some measurement noise N is added to this and we can draw a schematic of the data generation, see figure 1

$$\mathbf{N}_n$$

$$\mathbf{\Delta W}_{S,n} \quad \boxed{\mathbf{H}_n} \quad \xrightarrow{\mathbf{H}_n \mathbf{\Delta W}_{S,n}} \quad \bigoplus \quad \xrightarrow{\mathbf{\Delta S}_n = \mathbf{H}_n \mathbf{\Delta W}_{S,n} + \mathbf{N}_n}$$
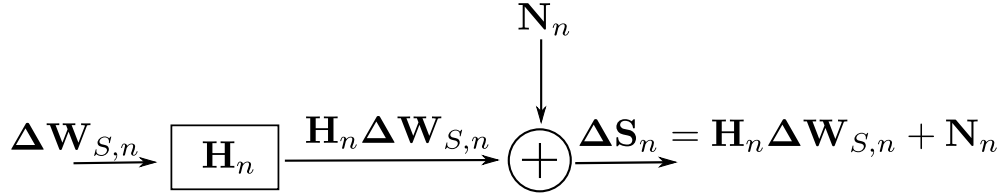
Figure 1: direct model schematic

The problem is how to find ΔW for the measurement of ΔS, so how to inverse the direct problem. A naive approach will be to discard the unknown noise and to just invert the known convolution product H, as represented in schematic 2.

$$\mathbf{\Delta S}_n \quad \boxed{\mathbf{H}_n^{-1}} \quad \xrightarrow{\mathbf{H}_n^{-1} \mathbf{\Delta S}_n = \mathbf{\Delta W}_{S,n} + \mathbf{H}_n^{-1} \mathbf{N}_n}$$

Figure 2: naive deconvolution schematic

To support the explanation in this annex we can take the example of a spectroscopy (or 0th harmonic) measurement of a 2 GHz sine wigner function. If we look at the estimated result $H^{-1}\Delta S$ in figure ... one can clearly remark that the estimator is dominated by only one oscillation. This result is clearly not reliable as it is sensible to small error perturbation as one can check if we apply to a different data set measured in the same condition and that do not seemed to show sensible difference, as tested on figure .... It also does not verify physical properties such as Pauli principle, which implies that the result plus the fremi function should be bounded between 0 and 1:

$$0 \le f + \text{fermi} \le 1 \tag{1}$$

This behaviour is explained by the fact that deconvolution is an ill-posed problem. This means that the H matrix that model the convolution has some zero or close to zero eigenvalues. This can be shown in the fourier

space where the convolution is just the product by the fourier transform of the convolution function h which posses some zeros. So inverting H is equivalent to divide some Fourier component of ΔS per 0, and even if $H\Delta W$ vanishes for these points, the noise N doesn't and the solution is completely dominated by these noise terms.

2   The field

This problem of inverting an ill-posed problem is also present in other situation such as image deblurring, reconstructiion. We benefit from these developpement to adapt and apply it to our specific case.

3   The Fourier space technique: Wiener filter

The first technique to overcome the problem presented in section 1 is to filter the Fourier components of noise divided by zero. The idea is to find the optimum filter which by averaging on realization of the noise N gives the minimum of the distance between filtered result $FH^{-1}\Delta S$ and the original signal ΔW, this distance is:

$$< ||FH^{-1}\Delta S - \Delta W||^2 > \tag{2}$$

and so by solving the equation

$$\frac{\mathrm{d}}{\mathrm{d}F}\left(< ||H^{-1}F\Delta S - \Delta W||^2 >\right) = 0 \tag{3}$$

and we find

$$F = \frac{HH^*}{HH^* + \left\langle\left|\frac{N}{\Delta W}\right|^2\right\rangle} = \frac{1}{1 + \left\langle\left|\frac{N}{H\Delta W}\right|^2\right\rangle} \tag{4}$$

- in the case of a convoluted signal higher than the noise $N \ll H\Delta W$, we get $\Delta S \sim H\Delta$ and so we don't need to filter since $H^{-1}\Delta S \sim \Delta W$ and this is the case because in this limit $F \sim 1$.

- The other case is when the convoluted signal is smaller than the noise $H\Delta W \ll N$. This implies that we measure only noise $\Delta S \sim N$ so the filter has to suppress this term. And this is the case since $F \sim \left\langle\left|\frac{H\Delta W}{N}\right|^2\right\rangle \ll 1$

To construct this filter we need two quantities first the noise spectrum $< |N|^2 >$ which can be measured if we assumed a white noise and if we use repeated measurement. The second quantity is the spectrum of the solution $|\Delta W|^2$, but as it is the one we are looking for, we don't know it. So we need to use some hypothesis to approach the optimum filter. To do so the first one is that when $\Delta S \gg N$ we can assume $\Delta W \sim H^{-1}\Delta S$ and so the filter has the form:

$$F = \frac{1}{1 + \left\langle\left|\frac{N}{\Delta S}\right|^2\right\rangle} \tag{5}$$

When $\Delta S \sim N$ we know that $H\Delta W \leq N \sim \Delta S$ so we assumed that there is a corner point where $\Delta W \leq \Delta W(\tau_c) \sim \frac{\Delta S(\tau_c)}{H(\tau_c)} \sim \frac{N}{H(\tau_c)}$ and that we use as a bound for our filter:

$$F = \frac{1}{1 + \left|\frac{H(\tau_c)}{H}\right|^2} \tag{6}$$

We end up with the following filter

$$F = \frac{1}{1 + \left\langle\left|\frac{N}{\Delta S}\right|^2\right\rangle}\,\text{when}\,H \geq H(\tau_c) \tag{7}$$

$$= \frac{1}{1 + \left|\frac{H(\tau_c)}{H}\right|^2}\,\text{when}\,H \leq H(\tau_c) \tag{8}$$

$$\tag{9}$$

The result on the example of the spectroscopy measurement and the whole wigner function of a 2 GHz sine is display in figure ...

Figure 3: Wiener deconvolution schematic

4   The Bayesian framework

The problem of the above technique is as we don't know the power spectrum of the solution, we are forced to use hypothesis that don't correspond to physical properties. To get a solution which verify known physical properties, we introduce a new problem formulation. The measurement is the mean of the probability distribution of ΔS knowing the searched quantity ΔW and the variance V_e of Gaussian noise N.

$$p\left(\Delta S | \Delta W, V_e\right) = \exp\left(-\frac{1}{2}||\Delta S - H\Delta W||^2_{V_e}\right) \tag{10}$$

But what we are looking for is ΔW so we are interested in the probability distribution of ΔW knowing the measurement ΔS and the noise variance V_e. Thanks to Bayes'formula we can link the two precedent probability distribution.

$$p\left(\Delta W | \Delta S, V_e\right) p\left(\Delta S\right) = p\left(\Delta S | \Delta W, V_e\right) p\left(\Delta W\right) \tag{11}$$

So we can expressed the probability distribution of interest $p\left(\Delta W | \Delta S, V_e\right)$, thanks to the measurement with $p\left(\Delta S | \Delta W, V_e\right)$ and to a prior knowledge on the solution with $p\left(\Delta W\right)$.
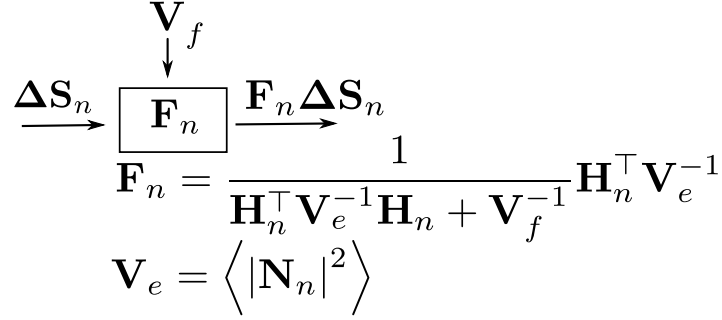
4.1 Posterior law maximization

$$\mathbf{V}_f$$

$$\Delta\mathbf{S}_n \longrightarrow \boxed{\mathbf{F}_n} \xrightarrow{\mathbf{F}_n\Delta\mathbf{S}_n}$$

$$\mathbf{F}_n = \frac{1}{\mathbf{H}_n^\top \mathbf{V}_e^{-1}\mathbf{H}_n + \mathbf{V}_f^{-1}}\mathbf{H}_n^\top\mathbf{V}_e^{-1}$$

$$\mathbf{V}_e = \left\langle |\mathbf{N}_n|^2 \right\rangle$$

Figure 4: MAP deconvolution schematic

4.2 Unsupervised technique with joint posterior law

$$\Delta\mathbf{S}_n \longrightarrow \boxed{\mathbf{F}_n} \xrightarrow{\mathbf{F}_n\Delta\mathbf{S}_n}$$

$$\alpha, \beta \longrightarrow \boxed{\mathbf{V}_f}$$
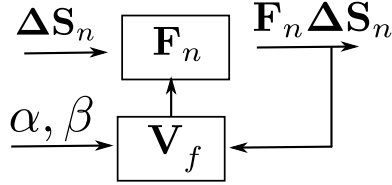
Figure 5: JMAP deconvolution schematic

4.3 Box-constraint problem with projected gradiant

$$\Delta\mathbf{S}_n \longrightarrow \boxed{\mathbf{F}_n} \xrightarrow{\mathbf{F}_n\Delta\mathbf{S}_n} \boxed{P} \quad M \quad \boxed{P\nabla} \longrightarrow$$

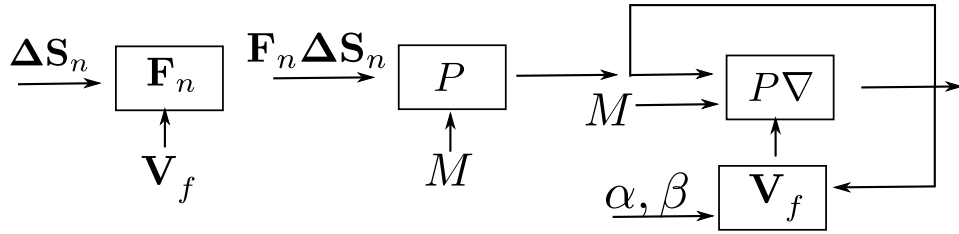$$\mathbf{V}_f \qquad M \qquad \alpha, \beta \longrightarrow \boxed{\mathbf{V}_f}$$

Figure 6: Projected Gradient deconvolution schematic

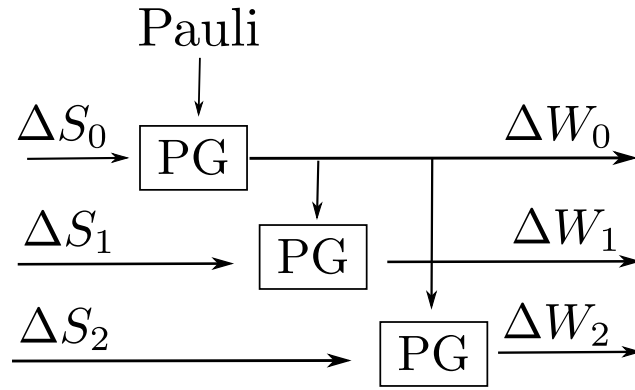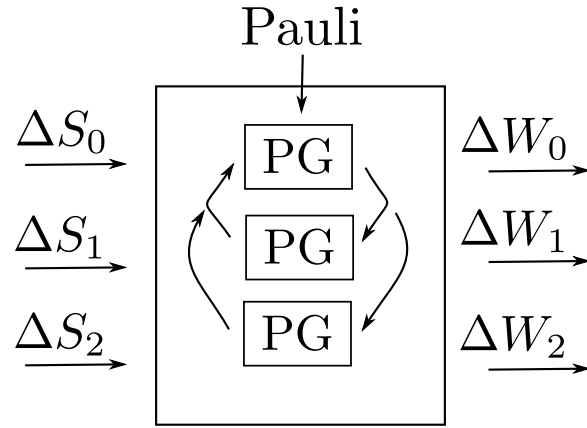## 4.4 outlook: toward a 2D treatment

Figure 7: 1D deconvolution schematic

Figure 8: 2D deconvolution schematic