Data Analysis - Project

Lee Drori, Ran Finkelstein and Ohad Michel

September 27, 2017

1   Statistical Model

Our H_0 in real space is given by:

$$H_0 : N = F_n T * P_n + \alpha F_n \delta(q) * P_n + \epsilon_n \tag{1}$$

Where hat stands for the real space functions and $*$ for convolution. On the other hand, H_1 is given by:

$$H_1 : N = F_n T * P_n + \alpha F_n \delta(q + \Delta q) * P_n + \epsilon_n \tag{2}$$

In both cases the Reference image is given by:

$$H_0, H_1 : R = F_r T * P_r + \alpha F_r \delta(q) * P_r + \epsilon_r \tag{3}$$

We will work mainly in Fourier space where the Hypothesis become:

$$\begin{cases} H_0 : & \hat{N} = F_n \hat{T} \cdot \hat{P}_n + \alpha F_n \exp(-ikq) \cdot \hat{P}_n + \hat{\epsilon}_n \\ H_1 : & \hat{N} = F_n \hat{T} \cdot \hat{P}_n + \alpha F_n \exp(-ik(q + \Delta q)) \cdot \hat{P}_n + \hat{\epsilon}_n \\ H_0, H_1 : & \hat{R} = F_r \hat{T} \cdot \hat{P}_r + \alpha F_r \exp(-ikq) \cdot \hat{P}_r + \hat{\epsilon}_r \end{cases} \tag{4}$$

From the last equation in (4) we can find $\hat{T}$ as a function of $\hat{R}$:

$$\hat{T} = \frac{\hat{R} - \hat{\epsilon}_r}{F_r \cdot \hat{P}_r} - \alpha \exp(-ikq) \tag{5}$$

We substitute $\hat{T}$ in to $\hat{N}$ and find:

$$H_0, \hat{R} : \hat{N} = \frac{F_n \hat{R} \cdot \hat{P}_n - F_n \hat{\epsilon}_r \cdot \hat{P}_n}{F_r \cdot \hat{P}_r} + \hat{\epsilon}_n \tag{6}$$

And H_1 is given by:

$$\begin{aligned} H_1, R : \hat{N} &= \frac{F_n \hat{R} \cdot \hat{P}_n - F_n \hat{\epsilon}_r \cdot \hat{P}_n}{F_r \cdot \hat{P}_r} + \alpha F_n [\exp(-ik(q + \Delta q)) - \exp(-ikq)] \cdot \hat{P}_n + \hat{\epsilon}_n \\ &= \frac{F_n \hat{R} \cdot \hat{P}_n - F_n \hat{\epsilon}_r \cdot \hat{P}_n}{F_r \cdot \hat{P}_r} + \alpha F_n \exp(-ikq) [\exp(-ik\Delta q) - 1] \cdot \hat{P}_n + \hat{\epsilon}_n \end{aligned} \tag{7}$$

The most powerful statistic for deciding between two simple hypotheses is the likelihood ratio test:

$$\mathcal{L}(q, \Delta q, \alpha) = \frac{\mathcal{P}(N, R|H_1)}{\mathcal{P}(N, R|H_0)} \tag{8}$$

Using base theorem:

$$\mathcal{L}(q, \Delta q, \alpha) = \frac{\mathcal{P}(N|R, H_1) \cdot \mathcal{P}(R|H_1)}{\mathcal{P}(N|R, H_0) \cdot \mathcal{P}(R|H_0)} = \frac{\mathcal{P}(N|R, H_1)}{\mathcal{P}(N|R, H_0)} \tag{9}$$

Where we used the fact that $\mathcal{P}\left(R|H_1\right) = \mathcal{P}\left(R|H_0\right)$. First let's compute the log of the numerator:

$$
\begin{aligned}
\log\left(\mathcal{P}\left(N|R,H_0\right)\right) &= \sum_k \frac{-\left|\hat{N} - \frac{F_n\hat{R}\cdot\hat{P}_n}{F_r\cdot\hat{P}_r}\right|^2}{2V\left(\hat{\epsilon}_n + \frac{F_n\hat{\epsilon}_r\cdot\hat{P}_n}{F_r\cdot\hat{P}_r}\right)} \\
&= \frac{1}{2}\sum_k \frac{-\left|F_r\hat{N}\cdot\hat{P}_r - F_n\hat{R}\cdot\hat{P}_n\right|^2}{\sigma_n^2 F_r^2\left|\hat{P}_r\right|^2 + \sigma_r^2 F_n^2\left|\hat{P}_n\right|^2}
\end{aligned}
\tag{10}
$$

For the denominator we find:

$$
\begin{aligned}
\log\left(\mathcal{P}\left(N|R,H_1\right)\right) &= \sum_k \frac{-\left|\hat{N} - \frac{F_n\hat{R}\cdot\hat{P}_n}{F_r\cdot\hat{P}_r} - \alpha F_n e^{-ikq}\left[\exp\left(-ik\Delta q\right) - 1\right]\cdot\hat{P}_n\right|^2}{2V\left(\hat{\epsilon}_n + \frac{F_n\hat{\epsilon}_r\cdot\hat{P}_n}{F_r\cdot\hat{P}_r}\right)} \\
&= \frac{1}{2}\sum_k \frac{-\left|F_r\hat{N}\cdot\hat{P}_r - F_n\hat{R}\cdot\hat{P}_n - \alpha F_n F_r e^{-ikq}\left[\exp\left(-ik\Delta q\right) - 1\right]\cdot\hat{P}_n\cdot\hat{P}_r\right|^2}{\sigma_n^2 F_r^2\left|\hat{P}_r\right|^2 + \sigma_r^2 F_n^2\left|\hat{P}_n\right|^2}
\end{aligned}
\tag{11}
$$

We are now ready to compute $\log\mathcal{L}\left(q,\Delta q,\alpha\right)$:

$$
\begin{aligned}
\log\mathcal{L}\left(q,\Delta q,\alpha\right) &= \log\left(\mathcal{P}\left(N|R,H_1\right)\right) - \log\left(\mathcal{P}\left(N|R,H_0\right)\right) \\
&= \frac{1}{2}\sum_k\left[-\frac{\left|F_r\hat{N}\cdot\hat{P}_r - F_n\hat{R}\cdot\hat{P}_n - \alpha F_n F_r e^{-ikq}\left[\exp\left(-ik\Delta q\right) - 1\right]\cdot\hat{P}_n\cdot\hat{P}_r\right|^2}{\sigma_n^2 F_r^2\left|\hat{P}_r\right|^2 + \sigma_r^2 F_n^2\left|\hat{P}_n\right|^2} + \frac{\left|F_r\hat{N}\cdot\hat{P}_r - F_n\hat{R}\cdot\hat{P}_n\right|^2}{\sigma_n^2 F_r^2\left|\hat{P}_r\right|^2 + \sigma_r^2 F_n^2\left|\hat{P}_n\right|^2}\right] \\
&= \sum_k \frac{\mathcal{R}\left[\left(F_r\hat{N}\cdot\hat{P}_r - F_n\hat{R}\cdot\hat{P}_n\right)\left(\alpha F_n F_r e^{-ikq}\left[\exp\left(-ik\Delta q\right) - 1\right]\cdot\hat{P}_n\cdot\hat{P}_r\right)^*\right]}{\sigma_n^2 F_r^2\left|\hat{P}_r\right|^2 + \sigma_r^2 F_n^2\left|\hat{P}_n\right|^2}
\end{aligned}
\tag{12}
$$

Where in the last step we used $|a+b|^2 = |a|^2 + |b|^2 + 2\mathcal{R}\left[ab^*\right]$ for the first term in the square bracket. The $|a|^2$ term is canceled by the second term in the square brackets and we disregarded $|b|^2$ (why??). Since α is a scalar multiplier we can define the sufficient statistic $S\left(q,\Delta q\right)$:

$$
\begin{aligned}
S\left(q,\Delta q\right) \equiv \frac{\log\mathcal{L}\left(q,\Delta q,\alpha\right)}{\alpha} &= \mathcal{R}\left[\sum_k e^{ikq}\frac{\left(F_r\hat{N}\cdot\hat{P}_r - F_n\hat{R}\cdot\hat{P}_n\right)\left(F_n F_r\left[\exp\left(-ik\Delta q\right) - 1\right]\cdot\hat{P}_n\cdot\hat{P}_r\right)^*}{\sigma_n^2 F_r^2\left|\hat{P}_r\right|^2 + \sigma_r^2 F_n^2\left|\hat{P}_n\right|^2}\right] \\
&= \mathcal{R}\left[\mathcal{F}^{-1}\left[\frac{\left(F_r\hat{N}\cdot\hat{P}_r - F_n\hat{R}\cdot\hat{P}_n\right)\left(F_n F_r\left[\exp\left(-ik\Delta q\right) - 1\right]\cdot\hat{P}_n\cdot\hat{P}_r\right)^*}{\sigma_n^2 F_r^2\left|\hat{P}_r\right|^2 + \sigma_r^2 F_n^2\left|\hat{P}_n\right|^2}\right]\left(q\right)\right]
\end{aligned}
\tag{13}
$$

We define:

$$
\begin{aligned}
\hat{D} &= \frac{F_r\hat{N}\cdot\hat{P}_r - F_n\hat{R}\cdot\hat{P}_n}{\sqrt{\sigma_n^2 F_r^2\left|\hat{P}_r\right|^2 + \sigma_r^2 F_n^2\left|\hat{P}_n\right|^2}} \\
\hat{P}_D &= \frac{F_n F_r\hat{P}_n\cdot\hat{P}_r}{F_D\sqrt{\sigma_n^2 F_r^2\left|\hat{P}_r\right|^2 + \sigma_r^2 F_n^2\left|\hat{P}_n\right|^2}} \\
F_D &= \frac{F_n F_r}{\sqrt{F_n^2\sigma_r^2 + F_r^2\sigma_n^2}} \\
\hat{M}\left(\Delta q\right) &= \exp\left(-ik\Delta q\right) - 1
\end{aligned}
\tag{14}
$$

Where D is the proper subtraction image, P_D is the PSF of this image and F_D is its normalization factor. The matrix M is the movement matrix which represent the difference in our transient position in Fourier representation. Using these notations we can write:

$$
\begin{aligned}
S(q, \Delta q) &= \mathcal{R}\left[\mathcal{F}^{-1}\left[F_D \hat{D} \cdot \hat{P}_D^{\ *} \cdot \hat{M}^*\left(\Delta q\right)\right](q)\right] \\
&= \mathcal{R}\left[\mathcal{F}^{-1}\left[F_D \hat{D} \cdot \hat{P}_D^{\ *}\right](q + \Delta q) - \mathcal{F}^{-1}\left[F_D \hat{D} \cdot \hat{P}_D^{\ *}\right](q)\right]
\end{aligned}
\tag{15}
$$

2 Expected performance of the statistics

Find proper threshold for detection

We want to know the distribution of noise within model H_0. Each pixel is drawn from a Poisson distribution with variance λ, for a large number of pixels this can be approximated as Gaussian noise. The optimal statistic is some function of the difference image $D(k)$. Since the latter is some linear combination of the reference image R and new image N we can find its variance from the variance of the two images. Here we use the fact that the variance in Fourier space is the number of pixels times the variance in real space.

$$
Var\left[\hat{D}\right] = Var\left[\frac{F_r \hat{N} \cdot \hat{P}_r - F_n \hat{R} \cdot \hat{P}_n}{\sqrt{\sigma_n^2 F_r^2 \left|\hat{P}_r\right|^2 + \sigma_r^2 F_n^2 \left|\hat{P}_n\right|^2}}\right] = \frac{n^2 \sigma_n^2 F_r^2 \left|\hat{P}_r\right|^2 + n^2 \sigma_r^2 F_n^2 \left|\hat{P}_n\right|^2}{\sigma_n^2 F_r^2 \left|\hat{P}_r\right|^2 + \sigma_r^2 F_n^2 \left|\hat{P}_n\right|^2} = n^2
\tag{16}
$$

So the proper subtraction image under H_0 has zero mean and unity STD. We now turn to the variance of the optimal statistic $S(q, \Delta q)$ under H_0. For a general displacement, $\hat{S}$ is again a subtraction of two Gaussian distributed variables which itself is a Gaussian distributed variable with variance:

$$
Var\left[\hat{S}\right] = 2F_D^2 n^2 \left|\hat{P}_D\right|^2 = \frac{2n^2 F_n^2 F_r^2 \left|\hat{P}_n\right|^2 \left|\hat{P}_r\right|^2}{\sigma_n^2 F_r^2 \left|\hat{P}_r\right|^2 + \sigma_r^2 F_n^2 \left|\hat{P}_n\right|^2}
\tag{17}
$$

in the limit of $\sigma_r \to 0$:

$$
Var\left[\hat{S}\right] = \frac{2n^2 F_n^2 \left|\hat{P}_n\right|^2}{\sigma_n^2}
\tag{18}
$$

$$
\begin{aligned}
Var\left[S\right] &= \frac{1}{n^4}\sum_k Var\left[\hat{S}\right] = \frac{2F_n^2}{\sigma_n^2}\frac{1}{n^2}\sum_k |P_n|^2 = \frac{2F_n^2}{\sigma_n^2}\mathcal{F}^{-1}\left[\left|\hat{P}_n\right|^2\right](0) \\
&= \frac{2F_n^2}{\sigma_n^2}\mathcal{F}^{-1}\left[\hat{P}_n\right](0) * \mathcal{F}^{-1}\left[\hat{P}_n^{\ *}\right](0) \\
&= \frac{2F_n^2}{\sigma_n^2}|P_n(0)|^2 = \frac{2F_n^2}{\sigma_n^2}\cdot\frac{1}{2\pi\sigma_n^2} = \boxed{\frac{F_n^2}{\pi\sigma_n^4}}
\end{aligned}
\tag{19}
$$