Small MRI is boring

Tamir Bendory

November 2, 2017

We consider the problem of estimating a signal $x \in \mathbb{R}^L$ from one measurement $y \in \mathbb{R}^N$, $N \gg L$, that contains many repetitions of this signal, at unknown locations, swallowed in noise.

Formally, our model read

$$y[n] = \sum_{i=1}^{K} x_{\mathrm{zp}}[n - n_i] + \varepsilon[n], \tag{1}$$

where:

- $y \in \mathbb{R}^N$ is the noisy acquired data (measurement),

- $x_{\mathrm{zp}} \in \mathbb{R}^N$ is a zero-padded version of $x \in \mathbb{R}^L$, $L \ll N$, defined as $x_{\mathrm{zp}} = [x, \underbrace{0, 0, \ldots, 0}_{N-L \text{ zeros}}]$,

- $n_i, \quad i = 1, \ldots, K$, are the translations of x_{zp},

- ε is i.i.d. normal noise with mean zero and variance σ^2 .

Note that both x and the translations n_i are unknown. Therefore, essentially, this is a blind deconvolution problem. Particularly, let $s \in \{0, 1\}^N$ be the "support" signal, which takes the value 1 at the locations n_i are zero otherwise. Then,

$$y = x_{\mathrm{zp}} * s + \varepsilon, \tag{2}$$

and the goal is then to estimate x from y although s is unknown as well.

For small values of σ, the problem is easy. In this case, one can identify x in y by simple detection algorithms (e.g., thresholding). Once several copies of signal were detected and cropped from y, then one can improve the SNR by averaging. We are interest in the more challenging regime of high noise level, in which the signal is completely swamped in the noise, and one cannot detect

1

the signal's appearances. In this case, one need to use the prior information that the signal appears many times across the measurement.

To perform this blind deconvolution, we use the invariant approach proposed for multireference alignment in a previous paper. Let us first introduce notation. We crop the measurement $y \in \mathbb{R}^N$ into a series of analysis windows:

$$y_i = y[(i-1)W + 1 : iW] \in \mathbb{R}^W, \quad i = 1, \ldots, N/W, \qquad (3)$$

To ease notation, we assume that W divides N so that there are exactly N/W windows. We also assume that $|n_i - n_j| > W$ so that each window contains no more than one signal (in general, we can introduce overlapping windows. The code enables it, but it does not seem to improve performance.). For each one of the windows, we compute its first three translation-invariant features, namely, mean, power spectrum and bispectrum. We then hope that averaging over these features will produce a reliable estimation of the invariant features of the signal itself. We recall that given accurate estimators of those features, we can estimate the signal itself reliably. Therefore, the question boils down to the accuracy of these feature estimations.

Before moving ahead to the feature estimators, we need to estimate the norm of the signal. This is necessary since many analysis windows do not contain information on the signal, but pure noise. We will use the norm to scale the the signal's estimator. This can be done as follows. For short, we denote $x_K[n] = \sum_{i=1}^{K} x_{\text{zp}}[n - n_i]$ and note that $\|x_K\|_2^2 = K\|x\|_2^2$. Then,

$$\|y\|_2^2 = \|x_K\|_2^2 + \|\varepsilon\|_2^2 + 2\varepsilon^T x_K. \qquad (4)$$

As $N \to \infty$, we can estimate

$$\begin{aligned} \mathbb{E}\left\{\|y\|_2^2\right\} &= \|x_K\|_2^2 + \mathbb{E}\left\{\|\varepsilon\|_2^2\right\} + 2\mathbb{E}\left\{\varepsilon^T x_K\right\} \\ &= K\|x\|_2^2 + N\sigma^2. \end{aligned} \qquad (5)$$

And therefore,

$$\|x\|_2^2 \cong \frac{\|y\|_2^2 - N\sigma^2}{K}. \qquad (6)$$

Next, we analyze the feature estimation in the asymptotic regime of $N, \sigma, K \to \infty$ and fixed L, W. We introduce the "sparsity factor" defined as $S = \frac{N}{WK} < \infty$, WK is a measure for the "active" windows. We also define another zero padded signal $x_W = [x, \underbrace{0, 0, \ldots, 0}_{N-W \text{ zeros}}] \in \mathbb{R}^W$.

We want to estimate the features of this signal, and focus on the bispectrum for the sake of simplicity (the estimation of the power spectrum and mean obeys the same methodology). In order to estimate the bispectrum we

average over the bispectra of the windows, namely, (for the power spectrum, we need to do unbiasing. In addition, we can replace the mean by more robust estimator, like the median.)

$$\hat{B}_{xW} = \frac{W}{N} \sum_{i=1}^{N/W} B_{y_i}. \tag{7}$$

We split the sum into three groups of the windows:

$$\hat{B}_{xW} = B_{\text{signal}} + B_{\text{clutter}} + B_{\text{noise}}, \tag{8}$$

where

- B_{noise} - sum of the bispectra of the segments that contain pure noise (no signal),

- B_{signal} - sum of the bispectra of the segments that contain a full signal (one appearance of x) and noise,

- B_{clutter} - sum of the bispectra of the segments that contain only part of x and noise.

We first observe that $B_{\text{noise}} \to 0$ (note that this is the zero matrix). More precisely, the variance of the noise goes to zero at rate $\mathcal{O}\left(\frac{W}{N}\left(1 - \frac{1}{S}\right)\sigma^6\right)$. In addition, under the assumption that n_i (if appear) are distributed uniformly in the window (we need to come up with precise statistical model; but it would be rather easy), there are (asymptotically) $K(1 - L/W)$ windows with complete signals. Therefore, $B_{\text{signal}} \to \frac{KW(1-L/W)}{N}B_x$ at rate $\mathcal{O}\left(\frac{\sigma^6}{K(1-L/W)}\right)$. (Here we need to be a bit careful with the mean of the signal, or to assume it is zero.) Note that the scaling would be corrected at the last stage of the algorithm, by rescaling according to the estimated norm.

The more interesting term is the "clutter" — the cropped signals. Since each such signal appears in two windows, we have $2KL/W$ clutter segments. Of course, the noise in this segments goes to zero at rate $\mathcal{O}\left(\frac{W\sigma^6}{2KL}\right)$. However, we have an additional error term arises from the cropped signals. To analyze the effect of this term, let us assume that N/σ^6 is great enough so that $B_{\text{noise}} \to 0$. Then,

$$\left\|\hat{B}_{xW} - B_{\text{signal}}\right\|_{\text{F}} \cong \|B_{\text{clutter}}\|_{\text{F}} = \left\|\frac{W}{N} \sum_{i \in \text{clutter}} B_{y_i}\right\|_{\text{F}}$$

$$\leq \frac{W}{N}\frac{2KL}{W}\|B_x\|_{\text{F}} = \frac{2L}{SW}\|B_x\|_{\text{F}}. \tag{9}$$

Therefore, for large sparsity term S or large W/L, we get an accurate estimation of a scaled version of the signal's bispectrum.

Few insights from the analysis: (need to check numerically)

- Larger S reduces the clutter error, however, increases the variance of the noise estimation.

- Larger W/L reduces the clutter error, however, increases the variance of the noise of B_{noise}.

- We must keep W/L large enough to see enough full signals.