EM for SR-MRA

Tamir Bendory

December 2018

We assume to collect N independent observations from the model

$$y = PR_\ell x + \varepsilon, \tag{1}$$

where $x \in \mathbb{R}^L$, R_ℓ is the circular shift (drawn from random distribution) and $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$. The operator P denotes down-sampling by a factor of K. It can be thought of as a diagonal matrix, whose diagonal composed of 1's every K entry, and zero otherwise. We also assume that $x \sim \mathcal{N}(0, \Sigma)$ (this is the prior).

In the **E-step**, given the current signal estimate x_{est}, we compute the probability that each measurement came from each shift. In our case, we compute the weights

$$w_{i,\ell} = C_i e^{-\frac{1}{2\sigma^2}\|y_i - PR_\ell x_{est}\|_2^2}, \quad i = 1, \ldots, N, \quad \ell = 0, \ldots, L-1, \tag{2}$$

where C_i denotes a normalization constant (per measurement).

This step can be computed efficiently using fft. It is useful to think of x_{est} as being composed of K sub-signals and then the norm $\|PR_\ell x_{est}\|_2^2$ can be computed separately for each of them. To compute the cross-terms, we again split the signal into sub-signals. Each one of them can be computed using fft (since it is a correlation). After computing the weights to each one of the K sub-signals, we interleave them to get the full weight matrix (see the code).

In the **M-step**, we need to solve the optimization problem

$$\max_x Q(x|x_{est}), \tag{3}$$

where

$$Q(x|x_{est}) = -\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{\ell=0}^{L-1} w_{i,\ell}\|y_i - PR_\ell x\|_2^2 - \frac{1}{2}x^T\Sigma^{-1}x, \tag{4}$$

where the second term is the prior. Computing the gradient, the maximum is obtained by

$$-\frac{1}{\sigma^2}\sum_{i=1}^{N}\sum_{\ell=0}^{L-1} w_{i,\ell}(R_\ell^T PR_\ell x - R_\ell^T P y_i) - \Sigma^{-1}x = 0, \tag{5}$$

and rearranging the equations we get

$$Ax = b, \tag{6}$$

where

$$A = \Sigma^{-1} + \frac{1}{\sigma^2} \sum_{i=1}^{N} \sum_{\ell=0}^{L-1} w_{i,\ell} R_\ell^T P R_\ell, \tag{7}$$

$$b = \frac{1}{\sigma^2} \sum_{i=1}^{N} \sum_{\ell=0}^{L-1} w_{i,\ell} R_\ell^T P y_i. \tag{8}$$

Now, because of the uniform distribution of shifts

$$A \approx \Sigma^{-1} + \frac{NK}{\sigma^2} I. \tag{9}$$

This approximation is being used to accelerate the code. The vector b can be computed efficiently using fft by splitting the computation into different sub-signals (see code).