Observing the selection of the cost function for gradient-based decomposition of surface electromyograms

A. Holobar¹, D. Zazula²

¹Laboratory of Engineering of Neuromuscular System and Motor Rehabilitation, Politecnico di Torino, Italy
²System Software Laboratory, Faculty of Electrical Engineering and Computer Science, University of Maribor, Slovenia

Abstract

Recently, gradient Convolution Kernel Compensation method was introduced for blind assessment of sparse pulse sequences (PSs) out of their convolutive mixtures [2]. The method employs multichannel recordings, is fully automatic and is minimally biased by assumptions underling mixing process. In the first step, the unknown mixing channels (convolution kernels) are compensated, whereas in the second step the gradient method is utilized to blindly optimize the estimated PSs. This contribution discusses the selection of the cost functions for aforementioned gradient-based optimization and provides analytical framework for their mutual comparison.

Data Model

Suppose M convolutive measurements are simultaneously observed and denote their sampled vector by \( y(t) = (y_1(t), y_2(t), \ldots, y_M(t)) \), where \( y(t) \) stands for the t-th sample of the i-th measurement. In the case of linear time-invariant (LTI) multiple-input-multiple-output (MIMO) system, \( y(t) \) can be written as:

\[
y(t) = H_{(i,j)} + n(t)
\]

where \( n(t) = (n_1(t), n_2(t), \ldots, n_M(t)) \) is a zero-mean spatially and temporally white additive noise vector, \( H_{(i,j)} = \{h_{i1}, h_{i2}, \ldots, h_{iM}\} \) is the extended version of vector of input signals from N sources \( s(t) = \{s_1(t), s_2(t), \ldots, s_N(t)\} \) is the mixing matrix \( H \) comprises all the channel responses \( h_{ij} = [h_{i0}, h_{i1}, \ldots, h_{L-1}] \) of length of L samples:

\[
H = \left[ \begin{array}{cccc}
    h_{00} & h_{01} & \cdots & h_{0L-1} \\
    h_{10} & h_{11} & \cdots & h_{1L-1} \\
    \vdots  & \vdots & \ddots & \vdots \\
    h_{L-10} & h_{L-11} & \cdots & h_{L-1L-1}
\end{array} \right]
\]

In surface electromyography, the channel response \( h_{ij} \) corresponds to the J-th motor unit (MU) action potential (MUAP), as detected by the i-th measurement, whereas each input PS \( s_{ij}(t) \), modelled as a sum of Dirac delta functions, \( s_{ij}(t) = \sum_{i=0}^{L-1} d_{ij}(t-i) \) determines the MUAP triggering times, where \( f(t) \) is the time instant in which the k-th MUAP of the i-th MU appears.

Selection of the Cost Function

Denote \( f_i(t) \) and \( F_i(t) \), then the second factor in the update rule (4) simplifies to:

\[
\sum_{k=1}^M e_{ik} e_{ik} = \sum_{k=1}^M \| H_{(i,j)} \|_2^2 \| \mathbf{w}^*_k \|_2^2 = \sum_{k=1}^M \| H_{(i,j)} \|_2^2 \| \mathbf{w}^*_k \|_2^2
\]

Now, assume (6) has converged to \( e = e_x \), where \( e_x = 0, \ldots, 0, e_y, 0, \ldots, 0 \) is a vector with the J-th position equal to 1 and K is vector of errors. Then, using the Taylor expansion of \( f_i(t) \), the updates of the i-th and k-th rows \( (j) \) in (6) yield:

\[
t_{f_{i,j}} = \frac{(f_{i,j} - f_{i,j} + c_{ij} \cdot \Delta \| F_{i,j} \|_2^2)}{\sum_{k=1}^M (f_{i,k} - f_{i,k} + c_{ij} \cdot \Delta \| F_{i,k} \|_2^2)}
\]

According to (7), convergence and stability of (6) depend on the values of the first few derivatives of \( f_i(t) \) at points \( t = 0 \) and \( t = 1 \). The values of the first few derivatives of cost functions \( f(t) \), \( f(t) \), and \( f(t) \) are depicted in Fig. 1. Functions \( f(t) \), \( f(t) \), and \( f(t) \) were proposed by Hyvärinen [3] and are implemented in the popular fastICA algorithm. Function \( f(t) \) was selected empirically and exhibits high robustness to outliers.

Experimental results

Surface EMG signals were recorded by a 2D matrix of 5×13 surface electrodes (inter-electrode distance 3.5 mm), from 2 subjects. Simulated PSs were generated with mean inter-pulse interval IPI = 100 samples and normally distributed zero-mean values \( f_i(k) \), \( k = 1, 2, \ldots, 100 \), with SD of 5 samples. The length of generated PSs was 10,000 samples. Zero-mean mixing matrix \( H \) was generated, with \( L = 11 \) samples long random channel responses. Average channel number of M was 100 (mean±SD). Gaussian zero-mean noise was added to simulated signals. Each mixture was decomposed three times - by gradient CKC with cost function set to \( f(t) \), \( f(t) \), and \( f(t) \), respectively. The results are depicted in Figs. 3 and 4.

Conclusions

We derived and validated an analytical expression that captures the convergence properties of different cost functions in gradient CKC. We demonstrated that two frequently used cost functions are not optimal for selection of cost function for gradient-based optimization of cost function \( C(t) \). General iteration step for gradient optimization of \( C(t) \) is then defined as:

\[
c(t) = [\mathbf{F}(t_m) + \eta \Delta \mathbf{F}(t_m)]
\]

and \( f(t) \) is shown in Fig. 1. Functions \( f(t) \), \( f(t) \), and \( f(t) \) were proposed by Hyvärinen [3] and are implemented in the popular fastICA algorithm. Function \( f(t) \) was selected empirically and exhibits high robustness to outliers. 

References


Grants

This research was supported by a Marie Curie Intra-European Fellowship within the 6th European Community Framework Programme (Contract No. MEIF-CT-2006-023537).