ELSEVIER

Neurocomputing 49 (2002) 163–173

NEUROCOMPUTING

www.elsevier.com/locate/neucom

Extraction of a source from multichannel data using sparse decomposition ☆

Michael Zibulevsky*, Yehoshua Y. Zeevi

Department of Electrical Engineering, Technion, 32000 Haifa, Israel

Received 1 March 2001; accepted 22 October 2001

Abstract

It was discovered recently that sparse decomposition by signal dictionaries results in dramatic improvement of the qualities of blind source separation. We exploit sparse decomposition of a source in order to extract it from multidimensional sensor data, in applications where a rough template of the source is known. This leads to a convex optimization problem, which is solved by a Newton-type method. Complete and overcomplete dictionaries are considered. Simulations with synthetic evoked responses mixed into natural 122-channel MEG data show significant improvement in accuracy of signal restoration.

Keywords: ICA; Blind source separation; Sparse representation; Wavelets; EEG; MEG

1. Introduction

We consider the following problem:

$$\mathbf{x}(t) = \mathbf{a}s(t) + \boldsymbol{\xi}(t), \tag{1}$$

where $\mathbf{x}(t)$ an observed n-channel sensor signal, $s(t)$ is an unknown scalar signal of interest, $\mathbf{a}$ an unknown n-dimensional vector of weights, and $\boldsymbol{\xi}(t)$ an n-channel background signal.

We assume that a rough template $\hat{s}(t)$ of the signal $s(t)$ is known in advance. The template can be obtained using a priori knowledge about the signal, or alternatively, by averaging multiple trials. It can be for example, a rectangular pulse, which corresponds

☆ Supported in part by the Ollendorff Minerva Center and by the Israeli Ministry of Science.

* Corresponding author.

E-mail addresses: mzib@ee.technion.ac.il (M. Zibulevsky), zeevi@ee.technion.ac.il (Y.Y. Zeevi).

0925-2312/02/$ - see front matter
PII: S0925-2312(02)00515-5

to the sign of the original signal, or of its most significant part. This is a realistic assumption in many practical cases.

We also assume that a sparse representation of $s(t)$ can be obtained by means of its decomposition coefficients c_k, corresponding to the set of functions $\varphi_k(t)$:

$$s(t) = \sum_{k=1}^{K} c_k \varphi_k(t). \tag{2}$$

The functions $\varphi_k(t)$ are called *atoms* or *elements* of the dictionary of functions. These elements do not have to be linearly independent, and instead may form an overcomplete dictionary. Important examples are wavelet and wavelet-related dictionaries (wavelet packets, stationary wavelets, Gabor-type frames, etc., see for example [5,9,17] and references therein), or learned dictionaries [8,11].

Sparsity means that only a small number of coefficients c_k differ significantly from zero. It was shown in [15,4,7,16] that use of sparseness often yields much better blind source separation than other techniques. In this work, we use the same property of sparseness for extraction of one source.

There are other approaches of a single source extraction. For example, Fast ICA algorithm [6] permits the extraction sources from mixtures sequentially, using an approximation of entropy as a criterion for separation. It will not necessarily extract the source of interest first, especially when the number of data channels is large. In order to deal with this problem, it was suggested in [1] to initialize separation weights in fast ICA-type algorithm using a second-order method based on maximal correlation with a template. This approach improves the order, in which the sources are extracted, but it does not exploit to its fullest extent available information regarding the structure of a template at the stage of separation.

In our work, we combine the prior knowledge about the sparsity of a source representation with the information regarding the relevant template into one optimization criterion. The resulting optimization problem is convex (unlike problems arising in usual ICA). It leads to high-quality solution even when the number of data channels is high and total number of samples is small. In our simulations we use 512 samples of 122-channel MEG data. In this situation standard ICA techniques cannot give a meaningful separation, because the number of free parameters in the separation 122×122 matrix is much larger than the number of data samples (normally the amount of data used for blind separation of such a data by standard methods is of order 10^5 samples or more, see for example [13]).

In the sequel we use a matrix notation. Let $t = 1, 2, \ldots, T$ be a discrete time under consideration, $\mathbf{X}$ be a $T \times n$ matrix, with discrete signals $x_i(t)$ in its columns, and $\mathbf{\Phi}$ be a matrix $T \times K$ with columns $\varphi_k(t)$. Then, instead of (2), we have

$$\mathbf{s} = \mathbf{\Phi c}. \tag{3}$$

If an estimate $\tilde{s}(t)$ of the signal would be known, it could be sparsely decomposed in the dictionary $\mathbf{\Phi}$ using the following optimization [5]:

$$\min_{c} \|\tilde{\mathbf{s}} - \mathbf{\Phi c}\|^2 + \mu \sum_{k=1}^{K} h(c_k). \tag{4}$$

Here, $h(c)$ can be considered as a penalty for non-sparseness. A reasonable choice of $h(c)$ [12,11] is

$$h(c) = |c|^{1/\gamma}, \quad \gamma \geqslant 1 \tag{5}$$

or a smooth approximation thereof. Below we use a family of convex smooth approximations of the absolute value [15]

$$h_1(c) = |c| - \log(1 + |c|), \tag{6}$$

$$h_\alpha(c) = \alpha h_1(c/\alpha) \tag{7}$$

with α being a proximity parameter: $h_\alpha(c) \rightarrow |c|$ as $\alpha \rightarrow 0^+$. Other approximations can be used as well, for example

$$h_\alpha(c) = \sqrt{c^2 + \alpha}.$$

2. Second-order source extraction using correlation with a template

In this section we present a standard approach of maximum correlation with a template, which will be used as a reference point. We look for an estimate $\tilde{s}(t)$ of the signal $s(t)$ as a linear combination of the sensor signals

$$\tilde{s}(t) = \sum_i w_i x_i(t), \tag{8}$$

which in a matrix form is

$$\tilde{\mathbf{s}} = \mathbf{X}\mathbf{w}, \tag{9}$$

where $\mathbf{w}$ is a vector of weights that we would like to determine.

Suppose that we have an approximate template $\hat{\mathbf{s}}$ of the signal $\mathbf{s}$. Then one can find an estimate of the signal $\mathbf{s}$ in form (9), which has maximal correlation with the template

$$\max_{\tilde{\mathbf{s}}} \frac{\hat{\mathbf{s}}^{\mathrm{T}}\tilde{\mathbf{s}}}{\|\hat{\mathbf{s}}\| \cdot \|\tilde{\mathbf{s}}\|}.$$

It can be rewritten equivalently as

$$\begin{aligned} &\min_{\tilde{\mathbf{s}}} \quad \|\tilde{\mathbf{s}}\|^2 \\ &\text{s.t.} \quad \hat{\mathbf{s}}^{\mathrm{T}}\tilde{\mathbf{s}} = 1. \end{aligned} \tag{10}$$

Combining this with (9), we obtain

$$\begin{aligned} &\min_{\mathbf{w}} \quad \|\mathbf{X}\mathbf{w}\|^2 \\ &\text{s.t.} \quad \hat{\mathbf{s}}^{\mathrm{T}}\mathbf{X}\mathbf{w} = 1. \end{aligned} \tag{11}$$

This problem can be solved using the method of Lagrange multipliers, yielding

$$\tilde{\mathbf{w}} = \lambda \mathbf{R}_{xx}^{-1}\mathbf{X}^{\mathrm{T}}\hat{\mathbf{s}}, \tag{12}$$

where $\mathbf{R}_{xx}$ is the covariance matrix: $\mathbf{R}_{xx} = \mathbf{X}^{\mathrm{T}}\mathbf{X}$.

3. Sparse estimation with a template

Suppose now that we have the following two priors:

- sparsity of the coefficients in the representation (3); and
- an approximate template $\hat{\mathbf{s}}$ of the signal.

We look for the estimated signal $\tilde{\mathbf{s}}$ with the sparsest representation $\mathbf{c}$ according to the dictionary Φ, which has a unit covariance with the template. In the general case of overcomplete dictionary this leads to the following optimization problem:

$$\begin{aligned} \min_{c} \quad & \|\tilde{\mathbf{s}} - \mathbf{\Phi c}\|^2 + \mu \sum_{k=1}^{K} h(c_k) \\ \text{s.t.} \quad & \hat{\mathbf{s}}^{\mathrm{T}} \tilde{\mathbf{s}} = 1. \end{aligned} \tag{13}$$

In the framework of linear estimation (9) we obtain

$$\begin{aligned} \min_{\mathbf{c},\mathbf{w}} \quad & \|\mathbf{Xw} - \mathbf{\Phi c}\|^2 + \mu \sum_{k=1}^{K} h(c_k) \\ \text{s.t.} \quad & \hat{\mathbf{s}}^{\mathrm{T}} \mathbf{Xw} = 1. \end{aligned} \tag{14}$$

When the dictionary is *complete*, we obtain significant simplification of the problem: the matrix $\mathbf{\Phi}$ is invertible, and the coefficients can be estimated directly, i.e.

$$\tilde{\mathbf{c}} = \mathbf{\Phi}^{-1} \tilde{\mathbf{s}} = \mathbf{\Phi}^{-1} \mathbf{Xw}. \tag{15}$$

Combining this with (14), where the first term $\|\mathbf{Xw} - \mathbf{\Phi c}\|^2$ vanishes, and using the transformed sensor data

$$\mathbf{Y} = \mathbf{\Phi}^{-1} \mathbf{X}$$

we get

$$\begin{aligned} \min_{\mathbf{w}} \quad & \sum_{k=1}^{K} h((\mathbf{Yw})_k), \\ \text{s.t.} \quad & \hat{\mathbf{s}}^{\mathrm{T}} \mathbf{Xw} = 1. \end{aligned} \tag{16}$$

Using the method of Lagrange multipliers, we come to the following problem:

$$\min_{\mathbf{w}} \sum_{k=1}^{K} h((\mathbf{Yw})_k) - \lambda \hat{\mathbf{s}}^{\mathrm{T}} \mathbf{Xw}. \tag{17}$$

There is a potential for instability in optimization of (17): growth of the first term in any direction is asymptotically linear, therefore the minimum of the objective function may approach $-\infty$, when λ is too large, so that the second term decreases faster than the first term grows. In order to avoid this, we use a monotonic convex transformation of the second term $u(\hat{\mathbf{s}}^{\mathrm{T}} \mathbf{Xw})$, where $u(\cdot)$ is a convex monotonically decreasing function

of one variable. For example, we can use quadratic-logarithmic function [2]

$$u_\tau(t) = \begin{cases} \frac{1}{2}t^2 - t, & t \leqslant \tau, \\ -(1-\tau)^2 \log(\frac{1-2\tau+t}{1-\tau}) - \tau + \frac{1}{2}\tau^2, & t > \tau. \end{cases}$$

where $0 \leqslant \tau < 1$. The second derivative of this function *is continuous and bounded* $\forall t \in \mathbb{R}$. Thus, we are in good position for the Newton minimization. Finally, our function for optimization becomes

$$F(\mathbf{w}) = \sum_{k=1}^{K} h((\mathbf{Y}\mathbf{w})_k) + \lambda u(\hat{\mathbf{s}}^{\mathrm{T}}\mathbf{X}\mathbf{w}). \tag{18}$$

It is easy to see from the optimality conditions, that (18) yields the same solution $\bar{\mathbf{w}}$ as (17), when λ is changed by a factor of $u'(\hat{\mathbf{s}}^{\mathrm{T}}\mathbf{X}\bar{\mathbf{w}})$.

4. Computational experiments with synthetic evoked responses mixed into natural MEG recordings

In order to verify the method, we synthesized a typical evoked brain response and mixed it linearly (with random weights) into real 122-channel MEG recording taken at the rate of 256 samples/s. The synthetic evoked response (Fig. 1, top plot) is composed of a narrow positive Gaussian pulse with a standard deviation of four samples and a wide negative Gaussian pulse with a standard deviation of 10 samples. The second pulse is delayed by 20 samples with respect to the first and decreased in amplitude by a factor 0.6. Other plots in Fig. 1 show few MEG channels already mixed with our synthetic evoked response. As one can see, the response is almost invisible on the background of brain activity. As a template (Fig. 2) we used a rectangular signal, corresponding to the time interval, when the response is above 10% of its maximal positive value.

We compared two methods of recovering evoked responses: the maximum correlation method (12) and our sparse estimation method, which consists of minimization of the objective function (18). We used a wavelet basis Φ with the mother-wavelet *Symmlet-8*, which has eight vanishing moments. This basis is convenient for approximation of smooth functions, like evoked responses are (see for example [9]).

In (18) we used the parameter $\lambda = 1000$, and in (7) the parameter $\alpha = 0.01$. Our empirical observation is that the results are not that sensitive to the values of these parameters; scaling by a factor 10 up and down does not affect the results significantly. Slight improvement in the quality can be observed when λ grows and α decreases more significantly, but the problem becomes more difficult for optimization.

As a minimization procedure we used the Newton method with frozen Hessian (see for example [10]). At each iteration the Hessian matrix was computed and three consequent Newton steps were then produced by substituting current gradients into the same Cholessky decomposition of the Hessian (expressions of gradient and Hessian are presented in Appendix A). Cubic line search with bisection safeguard and early stopping by Goldstain criterion was used at every Newton step (see for example [3]).

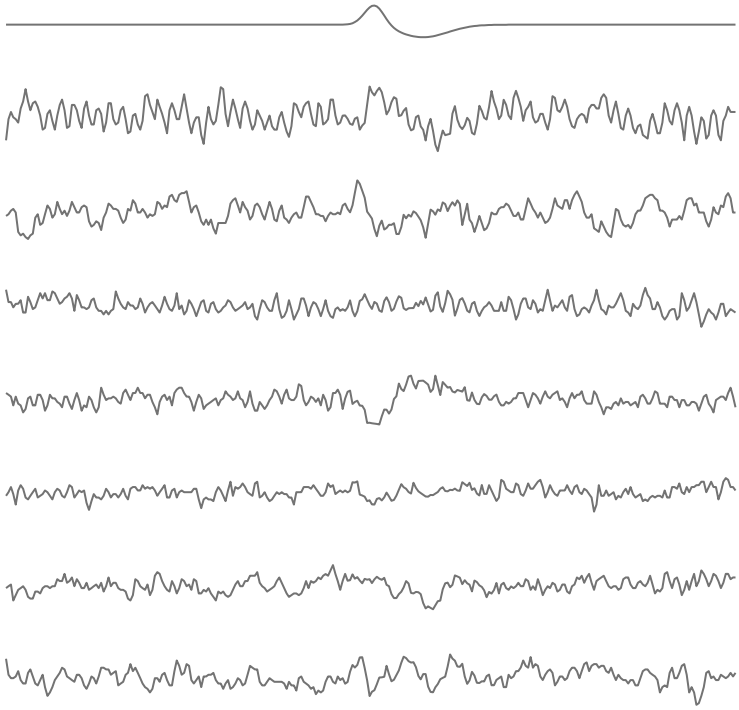

Fig. 1. Plot at the top shows synthetic evoked response; other plots show some of MEG channels already mixed with the evoked response: the response is almost invisible on the background of brain activity.

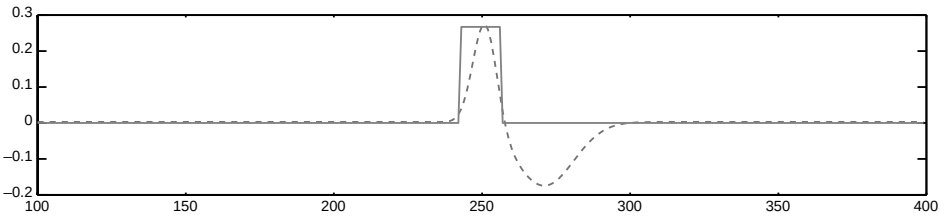

Fig. 2. A template (solid line) corresponding to the time interval, when the response (dashed line) is above 10% of its maximal value.

The results of estimation by the maximum correlation method are shown in Fig. 3 (upper frame). The signal-to-noise ratio is significantly better than that of the original sensor data (shown in Fig. 1), but the form of the pulse is corrupted, especially the negative part, which was not included in the template.

In Fig. 3, bottom, we see the recovered evoked response using our sparse estimation (18). It resembles the original pulse much more accurately, than the maximum correlation method does. Results of 50 simulated trials with random pulse position and random mixing weights are shown in Table 1. The mean-squared error is about

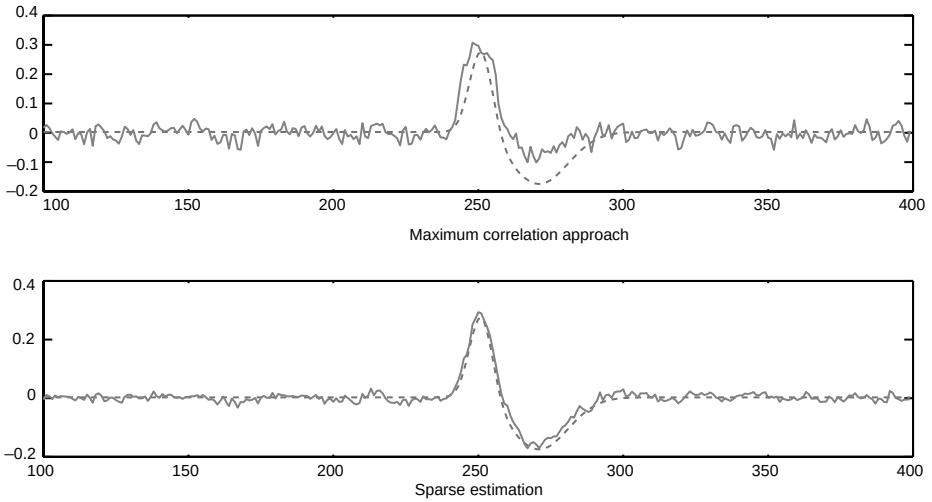

Fig. 3. (Top frame) Solid line—evoked response recovered by using the maximum correlation approach; (Bottom frame) Solid line—evoked response recovered by using sparse estimation; Dashed line in both frames—the original signal.

Table 1
Results of 50 simulated trials with random pulse position and random mixing weights

Method	Mean-squared error	Std. deviation of sq. error
Max correlation	0.38	0.0354
Sparse estimation	0.044	0.0185

Table 2
l_1 measure of sparseness of the wavelet coefficients

Original signal	3.30
Max correlation estimate	4.48
Sparse estimate	3.50

9 times smaller in the case of using our method, than with the maximum correlation approach.

We can measure sparseness of the wavelet coefficients $\tilde{c}$ of obtained estimates as a ratio $\|\tilde{c}\|_1/\|\tilde{c}\|_2$. Table 2 shows that the sparseness of the coefficients of the signals obtained by our method is better than one of the maximum correlation estimates.

Another important issue is robustness with regard to the variations in width and position of the template. Figs. 4 and 5 demonstrate superiority of sparse estimation. We should mention, that very similar results were obtained also with non-random mixing weights, corresponding to an area in visual cortex.

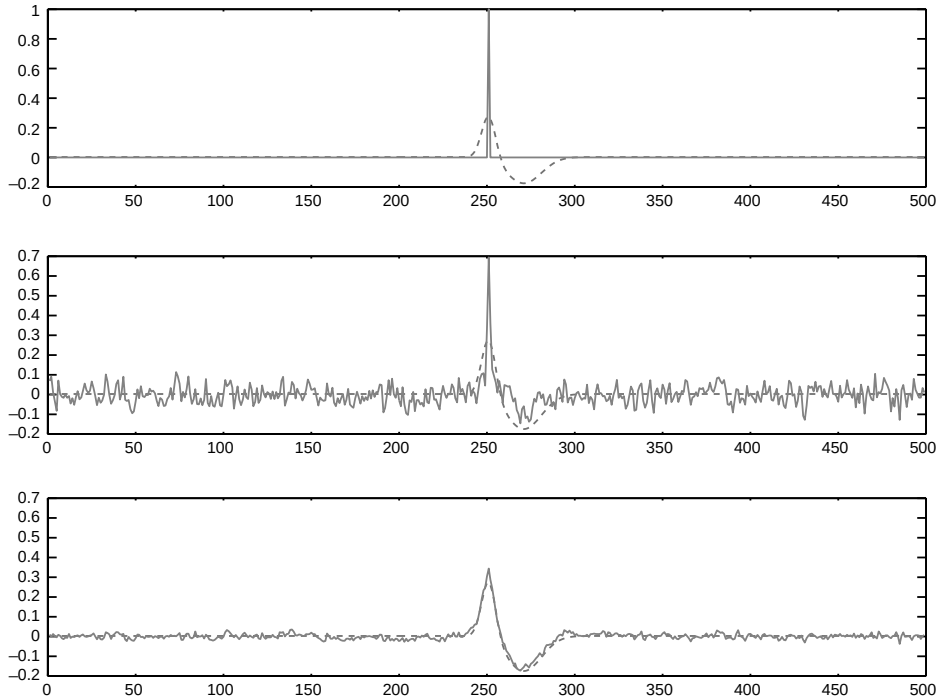

Fig. 4. A narrow template (top); max. correlation estimate (middle); sparse estimate (bottom).

We also tested some standard ICA techniques with the same data, but the results were meaningless. This can be easily understood taking into account a very small amount of data compared with the number of channels.

5. Conclusions

The proposed new approach to extraction a source from multichannel data, using a template and sparse representability of the source according to a signal dictionary is most suitable for physiological and medical, as well as wide range of other applications. Our simulations with complete dictionary demonstrate significant superiority of the method over the maximum correlation approach.

A more extensive study has yet to be conducted using overcomplete representations (14), which are more sparse, but also more expensive computationally.

The optimization problems (14) and (16) can be also reformulated as a quadratic or linear programming problems, when $h(\cdot)$ is exactly the absolute value function. This can be done in the spirit of the previous studies [5,15]. It provides a possibility of using the polynomial complexity algorithms, like *Interior Point Methods*. One can use also a special *Augmented Lagrangian* method for *sum-max* optimization problems [14], which reduces twice the number of variables as compared to the quadratic/linear

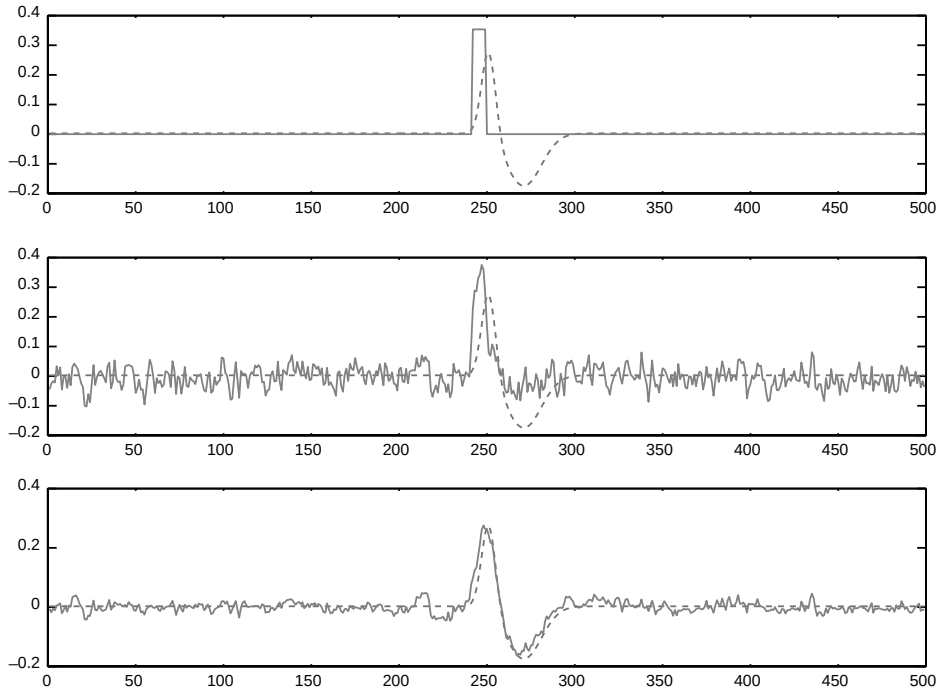

Fig. 5. Shifted template (top); max. correlation estimate (middle); sparse estimate (bottom).

programming approach, and provides better accuracy of solution. Practical comparison of all these approaches remains open for future research.

Appendix A. Gradient and Hessian of the objective function

Here we obtain derivatives of the objective function (18). Denoting $\mathbf{s} = \mathbf{Y}\mathbf{w}$, $\mathbf{z} = \mathbf{X}^{\mathrm{T}}\hat{\mathbf{s}}$, and $r = \mathbf{z}^{\mathrm{T}}\mathbf{w}$, expression (18) becomes

$$F(\mathbf{w}) = \sum_{k=1}^{K} h(s_k) + \lambda u(r). \tag{A.1}$$

A.1. Derivation of the gradient formula

Let $\mathbf{h}'(\mathbf{s})$ denotes the vector column of the first derivatives $h'(s_k)$, the differential of the objective is

$$\mathrm{d}F(\mathbf{w}) = \mathrm{d}\mathbf{s}^{\mathrm{T}}\,\mathbf{h}'(\mathbf{s}) + \lambda u'(r)\,\mathrm{d}r. \tag{A.2}$$

Recalling that $\mathrm{d}\mathbf{s} = \mathbf{Y}\,d\boldsymbol{w}$ and $\mathrm{d}r = \mathbf{z}^{\mathrm{T}}\,\mathrm{d}\mathbf{w} = \mathrm{d}\mathbf{w}^{\mathrm{T}}\,\mathbf{z}$, we get

$$\mathrm{d}F(\mathbf{w}) = \mathrm{d}\mathbf{w}^{\mathrm{T}}(\mathbf{Y}^{\mathrm{T}}\,\mathbf{h}'(\mathbf{s}) + \lambda u'(r)\mathbf{z}). \tag{A.3}$$

Let $\mathbf{g}$ denotes the gradient of F. Comparing (A.3) with

$$\mathrm{d}F(\mathbf{w}) = \mathbf{g}^{\mathrm{T}}\,\mathrm{d}\mathbf{w} = \mathrm{d}\mathbf{w}^{\mathrm{T}}\,\mathbf{g}$$

we obtain finally

$$\mathbf{g}(\mathbf{w}) = \mathbf{Y}^{\mathrm{T}}\,\mathbf{h}'(\mathbf{s}) + \lambda u'(r)\mathbf{z}. \tag{A.4}$$

A.2. Derivation of the Hessian formula

Let $\mathbf{D}$ denotes the diagonal matrix of the second derivatives $h''(s_k)$. It is easy to obtain from (A.4)

$$\mathrm{d}\mathbf{g}(\mathbf{w}) = \mathbf{Y}^{\mathrm{T}}\mathbf{D}\,d\boldsymbol{s} + \lambda u''(r)\mathbf{z}\,dr. \tag{A.5}$$

Taking into account that $\mathrm{d}\mathbf{s} = \mathbf{Y}\,d\boldsymbol{w}$ and $\mathrm{d}r = \mathbf{z}^{\mathrm{T}}\,\mathrm{d}\mathbf{w}$, we get

$$\mathrm{d}\mathbf{g}(\mathbf{w}) = \mathbf{Y}^{\mathrm{T}}\mathbf{D}\mathbf{Y}\,d\boldsymbol{w} + \lambda u''(r)\mathbf{z}\mathbf{z}^{\mathrm{T}}\,\mathrm{d}\mathbf{w}. \tag{A.6}$$

Comparing this with the known expression

$$\mathrm{d}\mathbf{g}(\mathbf{w}) = \mathbf{H}(\mathbf{w})\,\mathrm{d}\mathbf{w},$$

where $\mathbf{H}(\mathbf{w})$ is a Hessian matrix, we finally obtain

$$\mathbf{H}(\mathbf{w}) = \mathbf{Y}^{\mathrm{T}}\mathbf{D}\mathbf{Y} + \lambda u''(r)\mathbf{z}\mathbf{z}^{\mathrm{T}}.$$

References

[1] A.K. Barros, R. Vigario, V. Jousmaki, N. Ohnishi, Extraction of event-related signals from multi-channel bioelectrical measurements, IEEE Trans. Biomed. Eng. 47 (5) (2000) 583–588.
[2] A. Ben-Tal, M. Zibulevsky, Penalty/barrier multiplier methods for convex programming problems, SIAM J. Optim. 7 (2) (1997) 347–366.
[3] D.P. Bertsekas, Nonlinear Programming, 2nd Edition, Athena Scientific, Belmont, MA, 1999.
[4] P. Bofill, M. Zibulevsky, Blind separation of more sources than mixtures using the sparsity of the short-time fourier transform, International Workshop on Independent Component Analysis and Blind Signal Separation, Helsinki, Finland, June 19–20, 2000.
[5] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM J. Scientific Comput. 20(1) (1998) 33–61.
[6] A. Hyvärinen, Fast and robust fixed-point algorithms for independent component analysis, IEEE Trans. Neural Networks 10 (3) (1999) 626–634.
[7] P. Kisilev, M. Zibulevsky, Y.Y. Zeevi, B.A. Pearlmutter, Multiresolution framework for sparse blind source separation, Technical report, Department of Electrical Engineering, Technion, Haifa, Israel, 2000, http://ie.technion.ac.il/∼mcib/.
[8] M.S. Lewicki, T.J. Sejnowski, Learning overcomplete representations, Neural Comput. 12 (2) (2000) 337–365.
[9] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, New York, 1998.
[10] L. Mosheyev, M. Zibulevsky, Penalty/barrier multiplier algorithm for semidefinite programming, Optim. Methods Software 13 (4) (2000) 235–261.
[11] B.A. Olshausen, D.J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, Nature 381 (1996) 607–609.
[12] B.A. Olshausen, D.J. Field, Sparse coding with an overcomplete basis set: a strategy employed by v1? Vision Res. 37 (1997) 3311–3325.

[13] A.C. Tang, B.A. Pearlmutter, M. Zibulevsky, Blind separation of multichannel neuromagnetic responses, Neurocomputing 32–33 (2002) 1115–1120.
[14] M. Zibulevsky, Smoothing method of multipliers for sum-max problems, Technical report, Department of Electrical Engineering, Technion, Haifa, Israel, 2001. http://ie.technion.ac.il/∼mcib/.
[15] M. Zibulevsky, B.A. Pearlmutter, Blind source separation by sparse decomposition in a signal dictionary, Neural Comput. 13 (4) (2001) 863–882.
[16] M. Zibulevsky, B.A. Pearlmutter, P. Bofill, P. Kisilev, Blind source separation by sparse decomposition in a signal dictionary, in: S.J. Roberts, R.M. Everson (Eds.), Independent Components Analysis: Principles and Practice, Cambridge University Press, Cambridge, 2001.
[17] M. Zibulski, Y.Y. Zeevi, Analysis of multi-window gabor-type schemes by frame methods, Applied and Computational Harmonic Analysis 4 (1997) 188–221.

Michael Zibulevsky was born in 1959 in Ukraine. He received the M.Sc. degree in Electrical Engineering from MIIT in Moscow in 1981, and a Ph.D. degree in Operation Research (Non-linear Optimization) from the Technion—Israel Institute of Technology in Haifa in 1996. Starting from 1981 he has held research positions at the Telecommunication Lab, NIIASS, Kiev; Optimization Lab, Technion, Haifa; Brain and Computation Lab, University of New Mexico. Currently, he is with the Department of Electrical Engineering at the Technion. His areas of interests include Non-linear Optimization, Neural Networks, Tomography, Independent Component Analysis and Sparse Representations of Signals and Images.

Yehoshua Y. Zeevi is the Barbara and Norman Seiden Professor of Computer Sciences in the Department of Electrical Engineering, Technion—Israel Institute of Technology, and the Head of the Jacobs Center for Communication and Information Technologies. He also served as the Dean of the Faculty of Electrical Engineering (1994–1999). He received the B.Sc. from the Technion, the M.Sc. from the University of Rochester, NY, and the Ph.D. from the University of California, Berkeley. He was a Vinton Hayes Fellow at Harvard University and has been a regular visitor there. He was also a Visiting Professor at MIT and the CAIP Center of Rutgers university. Dr. Zeevi is the coinventor of many patents implemented in advanced medical and other technologies, including those related to the wide dynamic range adaptive sensitivity camera that mimics the eye, and the author of over 200 papers and technical reports related to vision and image sciences. He is a Fellow of the SPIE and the Rodin Academy. He is the Editor-in-Chief of the Journal of Visual Communication and Image Representation, published by Academic Press, and the co-editor of three books published by Academic Press and Springer.