Improved M-FOCUSS Algorithm With Overlapping Blocks for Locally Smooth Sparse Signals

Rafal Zdunek, *Member, IEEE*, and Andrzej Cichocki, *Senior Member, IEEE*

***Abstract*—The FOCal Underdetermined System Solver (FOCUSS) algorithm has already found many applications in signal processing and data analysis, whereas the regularized M-FOCUSS algorithm has been recently proposed by Cotter *et al.* for finding sparse solutions to an underdetermined system of linear equations with multiple measurement vectors. In this paper, we propose three modifications to the M-FOCUSS algorithm to make it more efficient for sparse and locally smooth solutions. First, motivated by the simultaneously autoregressive (SAR) model, we incorporate an additional weighting (smoothing) matrix into the Tikhonov regularization term. Next, the entire set of measurement vectors is divided into blocks, and the solution is updated sequentially, based on the overlapping of data blocks. The last modification is based on an alternating minimization technique to provide data-driven (simultaneous) estimation of the regularization parameter with the generalized cross-validation (GCV) approach. Finally, the simulation results demonstrating the benefits of the proposed modifications support the analysis.**

***Index Terms*—FOCal Underdetermined System Solver (FOCUSS), generalized cross-validation (GCV), smooth signals, sparse solutions, underdetermined systems.**

I. Introduction

THE need for finding sparse solutions to systems of linear equations has been motivated by many applications such as those in signal and image processing [1]–[7], blind source separation [8]–[16], brain activity imaging [17]–[25], sparse component analysis [26], subband decomposition [27], [28], dictionary learning [29]–[32], sparse Bayesian learning [33], [34], sparse coding [35], sparse audio signal representation [28], [36], sparse regression [37]–[40], inverse synthetic aperture radar (ISAR) imaging [41], low-bit-rate compression [42], decision feedback equalization [43], or echo cancellation [44].

In all cases, the underlying inverse problem can be expressed in terms of finding the l_0-based sparsest solution, i.e., to find a minimal number of basis vectors that represent the solution (signal) of interest. Despite the problem being generally regarded as NP-hard [43], a number of computational strategies have been developed to find solutions with low computational complexity. Examples include the basis pursuit [45], greedy algorithms [43], [46]–[53], iterative-thresholding algorithms [8], [10], [11], [54], [55], FOCal Underdetermined System Solver (FOCUSS) algorithm [17], [18], and its extensions such as M-FOCUSS [56], [57], FOCUSS-CNDL [58], LORETA-FOCUSS [22], together with other improvements [59]–[61]. Obviously, the solutions provided by these algorithms are not always equivalent to the l_0 sparsest solution, and this has been extensively analyzed in the literature [62]–[64].

The iterative scheme used by the FOCUSS algorithm is based on the iteratively reweighted least square (IRLS) algorithm that was analyzed by Karlovitz [65]. The FOCUSS algorithm was proposed to minimize the l_p diversity measure, which enforces a certain degree of sparsity (with parameter p), subject to the equality constraints that are given by the system of linear equations whose sparse solution is sought. The M-FOCUSS algorithm proposed by Cotter *et al.* [57] extends the standard FOCUSS algorithm to the case of simultaneous processing of multiple measurement vectors (MMV) with a common sparsity profile. Examples of applications include magnetoencephalography/electroencephalography (MEG/EEG) [20], [21], [23], [24], sparse Bayesian learning [34], nonparametric spectrum analysis, and blind signal separation [66].

Unfortunately, in many real applications, the sparsity profile is time-varying, and noisy disturbances may lead to the nonzero entries being spuriously recovered. To increase robustness of the M-FOCUSS with respect to spiky disturbances, we can exploit the locally smooth nature of signals, e.g., in MEG/EEG, where the locality is restricted to only few samples. By incorporating such information to the prior that models the sparsity constraints in the M-FOCUSS algorithm, the effects of high-frequency disturbances, such as noise and other interferences, can be considerably reduced. Smoothness constraints can be imposed to the solution in various ways, for instance, in image restoration or tomographic image reconstruction, smoothness is enforced with Markov random field (MRF) [67]–[70], simultaneously autoregressive (SAR) or conditionally autoregressive (CAR) models [71]–[74]. All these models exploit spatial pairwise interactions among the pixels that belong to a given neighborhood.

We propose to model local smoothness within signals (time series) by exploiting pairwise interactions between the samples in a short window, and our approach is motivated by the SAR model that can be easily combined with the Gaussian prior in the M-FOCUSS algorithm.

MRF models with nonquadratic (or even concave) potential functions are generally more robust in image processing but intrinsically involve much more computational effort, mostly due to the need of estimating the associated hyperparameters. In this respect, the SAR or CAR models are computationally

Manuscript received May 29, 2007; revised May 17, 2008. First published July 9, 2008; current version published September 17, 2008. The associate editor coordinating the review of this paper and approving it for publication was Dr. Thierry Blu.

R. Zdunek is with the Laboratory for Advanced Brain Signal Processing, Brain Science Institute, RIKEN, Saitama 351-0198, Japan and also with the Institute of Telecommunications, Teleinformatics, and Acoustics, Wroclaw University of Technology, 50-370 Wroclaw, Poland (e-mail: rafal.zdunek@pwr.wroc.pl).

A. Cichocki is with the Laboratory for Advanced Brain Signal Processing, Brain Science Institute, RIKEN, Saitama 351-0198, Japan, with the IBS, Polish Academy of Sciences (PAN), 01-447 Warsaw, Poland and also with the Warsaw University of Technology, 00-661 Warsaw, Poland (e-mail: cia@brain.riken.jp).

Digital Object Identifier 10.1109/TSP.2008.928160

more tractable, but their efficiency critically depends on the related hyperparameter estimation scheme. In [74] and [75], the maximum *a posteriori* (MAP) problem associated with the SAR and CAR models is solved with a hierarchical Bayesian paradigm. The hyperparameters are associated with the Gamma hyperpriors, which simplifies the marginalization procedure that is performed simultaneously with iterative updates for the solution.

Neglecting the parameter p (which only affects the degree of sparsity) in the l_p diversity measure, we need to estimate only one parameter that is the regularization parameter. Limiting the number of hyperparameters to be estimated considerably simplifies the related hyperparameter estimation scheme. Rao *et al.* [76] proposed to combine the discrepancy principle with the L-curve technique to estimate the regularization parameter in the FOCUSS. Because our modified M-FOCUSS algorithm involves the singular value decomposition (SVD) in each iterative step, the generalized cross-validation (GCV) technique [77], [78], which provides an estimate for the regularization parameter when the variance of disturbances is unknown, seems to be more suitable for our task. Moreover, the GCV is superior to the L-curve method, at least in the sense of overcoming the problems of finding the L-curve corner.

We assume the sparsity profile may be different in each vector from the set of MMV, but due to the smoothness constraints, the changes in a sparsity profile between the successive vectors are quasi-constant. Thus, we divide the whole set of MMV into blocks of a few samples, for which the sparsity profile is assumed to have nearly the same structure. Moreover, our solution is updated sequentially using overlapping data blocks, which is motivated by frame overlapping [79], [80].

The organization of this paper is as follows. In Section II, we formulate the problem, and then we describe the basic M-FOCUSS algorithm. Section III contains our new algorithmic results: starting from the overlapping technique in the M-FOCUSS algorithm, and ending at the SOB-M-FOCUSS algorithm, which includes smoothness constraints. The numerical tests are presented in Section IV. Finally, some conclusions are drawn in Section V.

II. M-FOCUSS ALGORITHM

A. Problem Formulation

The problem is to find a possible good approximation of the true solution to an undetermined system of linear equations subject to sparsity and smoothness constraints. The system in the matrix form is given by

$$\boldsymbol{AX} + \boldsymbol{N} = \boldsymbol{B} \tag{1}$$

where $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ is a system matrix with $M \leq N$, $\operatorname{rank}\{\boldsymbol{A}\} = M$, $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T] \in \mathbb{R}^{N\times T}$ (possibly $T > N$) is the true solution, $\boldsymbol{N} \in \mathbb{R}^{M\times T}$ represents additive noise or errors, and $\boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_T] \in \mathbb{R}^{M\times T}$ is an observation matrix. Each column $\boldsymbol{b}_t$ or $\boldsymbol{x}_t$ $(1 \leq t \leq T)$ contains the tth sample from the observations (the tth vector from MMV) or the solution, respectively, where T is a total number of samples. The rows of $\boldsymbol{X}$ and $\boldsymbol{B}$ are considered as corresponding signal representations.

If $M < N$, the nullspace of $\boldsymbol{A}$ is nontrivial, and the inverse problem has many solutions. Therefore, additional constraints are needed to select the true solution. We assume that the sparsity profile in the solution varies slowly with the vectors $\boldsymbol{x}_t$, $t = 1, \ldots, T$. Moreover, we also assume that the signal representations expressed by the rows in $\boldsymbol{X}$ are locally smooth.

Remark 1: It was proved in [57] that the sparse solution to the consistent system (1) with $\boldsymbol{N} = 0$ is unique, if $\operatorname{rank}\{\boldsymbol{B}\} = T$ with $T \leq M$, any M columns of $\boldsymbol{A}$ are linearly independent (unique representation property (URP) condition [18]), and $\forall\, t$: $\boldsymbol{x}_t$ has at most $\lceil (M+T)/2 \rceil - 1$ nonzero entries, where $\lceil \cdot \rceil$ is a ceil function.

B. M-FOCUSS

The M-FOCUSS algorithm iteratively solves the following equality constrained problem:

$$\min_X J^{(p)}(\boldsymbol{X}), \qquad \text{s.t.} \quad \boldsymbol{AX} = \boldsymbol{B} \tag{2}$$

where

$$J^{(p)}(\boldsymbol{X}) = \sum_{j=1}^{N} \|\tilde{\boldsymbol{x}}_j\|_2^p = \sum_{j=1}^{N} \left(\sum_{t=1}^{T} x_{jt}^2 \right)^{(p/2)}, \qquad p \in [0, 2] \tag{3}$$

is the l_p diversity measure that is related to the joint sparsity [48], [52], [81]. The parameter p indicates the degree of sparsity, and $\tilde{\boldsymbol{x}}_j$ denotes the jth row of $\boldsymbol{X}$. For the inconsistent case (noisy case), Cotter *et al.* have developed the regularized M-FOCUSS algorithm, which in a single iterative step solves the Tikhonov regularized least squares problem

$$\boldsymbol{X}^{(k)} = \arg\min_X \Psi(\boldsymbol{X}|\boldsymbol{X}^{(k-1)}) \tag{4}$$

where

$$\Psi(\boldsymbol{X}|\boldsymbol{X}^{(k-1)}) = \|\boldsymbol{B} - \boldsymbol{AX}\|_F^2 + \lambda \|\boldsymbol{W}^{-1}\boldsymbol{X}\|_F^2 \tag{5}$$

$\boldsymbol{W} = \operatorname{diag}\left\{ w_j^{1-(p/2)} \right\}$ with

$$w_j = \left(\sum_{t=1}^{T} \left(x_{jt}^{(k-1)} \right)^2 \right)^{(1/2)} = \left\| \tilde{\boldsymbol{x}}_j^{(k-1)} \right\|_2$$

the current iterative step is denoted by k, λ is a regularization parameter, and $\boldsymbol{I}_M \in \mathbb{R}^{M\times M}$ is an identity matrix.

III. METHODOLOGY

A. M-FOCUSS With Overlapped Blocks

According to Remark 1, one of the assumptions to get a unique solution with the M-FOCUSS algorithm is to have $T \leq M$. Because we allow for a very large number of MMV, i.e., $T \gg N$, it is efficient to split the measurements into blocks of MMV. Let $\boldsymbol{B} = [\boldsymbol{B}^{(1)}, \ldots, \boldsymbol{B}^{(R)}]$, where $\boldsymbol{B}^{(r)}$: $r = 1, \ldots, R$ is the rth block of the MMV (submatrix of $\boldsymbol{B}$), and R is a number of the blocks.

A typical strategy is to split $\boldsymbol{B}$ into disjoint blocks of length L, i.e., the number of samples in one block $\boldsymbol{B}^{(r)}$ for $r < R$. According to Remark 1, we assume $L \leq M$. The total number of blocks is equal to $R = \lceil T/L \rceil$. Thus, $\boldsymbol{B}^{(r)} = [\boldsymbol{b}_{\Pi^{(r)}}]$, where

$\Pi^{(r)}$ is a set of indices of MMV that belong to $\boldsymbol{B}^{(r)}$, and it is given by

$$\Pi^{(r)} = \begin{cases} \{\phi_r + 1, \ldots, \phi_r + L\} \in \mathbb{N}^L, & \text{for } r < R \\ \{\phi_R + 1, \ldots, T\} \in \mathbb{N}^{T-\phi_R}, & \text{for } r = R \end{cases} \quad (6)$$

$$\phi_r = \lfloor (r-1)L \rfloor \quad (7)$$

where $\lfloor \cdot \rfloor$ denotes the floor function.

Our strategy of overlapping blocks, which is motivated with the overlapping frame technique in [79] and [80], assumes the blocks are overlapped with the fixed rate of overlapping $\theta \in [0, 99]$ (percentages). This gives the number of overlapping samples $\eta = \lfloor \theta L/100 \rfloor$, and consequently, $R = \lceil (T-\eta)/(L-\eta) \rceil$ and

$$\phi_r = \lfloor (r-1)(L-\eta) \rfloor. \quad (8)$$

The set of the vectors in $\boldsymbol{B}^{(r)}$ is given by (6).

Note that each data block $\boldsymbol{B}^{(r)}$ updates only the corresponding solution block $\boldsymbol{X}^{(r)} = [x_{*,\Pi^{(r)}}]$ with $\Pi^{(r)}$ being defined by (6)–(7). The column vectors from $\boldsymbol{B}$ and $\boldsymbol{X}$ that belong to the respective blocks $\boldsymbol{B}^{(r)}$ and $\boldsymbol{X}^{(r)}$ are uniquely determined by the set of indices (6). The block updating is sequential, so when the blocks are overlapped, a certain number of the column vectors from $\boldsymbol{X}^{(r)}$ are treated as the initial samples for updating the successive block $\boldsymbol{X}^{(r+1)}$.

The overlapping blocks are justified intuitively by the fact that the source signals are expected to be locally smooth during the observation period. If the blocks are disjoint, smoothness on the block's borders might be seriously perturbed. Also, a sparsity profile changes more smoothly for the overlapping blocks.

B. Smoothness Constraints

The solution to the problem $\min_X \Psi(\boldsymbol{X}|\boldsymbol{X}^{(k-1)})$, with $\Psi(\boldsymbol{X}|\boldsymbol{X}^{(k-1)})$ defined by (5), can be regarded as the MAP estimate, where the likelihood function $p(\boldsymbol{B}|\boldsymbol{X})$ is expressed in terms of the joint Gaussian distribution

$$p(\boldsymbol{B}|\boldsymbol{X}) = \frac{1}{\left((2\pi)^M \det\{\boldsymbol{\Sigma}_N\}\right)^{T/2}} \times \exp\left\{-\frac{1}{2}\operatorname{tr}\left\{(\boldsymbol{B}-\boldsymbol{A}\boldsymbol{X})^T \boldsymbol{\Sigma}_N^{-1}(\boldsymbol{B}-\boldsymbol{A}\boldsymbol{X})\right\}\right\}. \quad (9)$$

The columns $\boldsymbol{n}_1, \ldots, \boldsymbol{n}_T$ from $\boldsymbol{N}$ in (1) are assumed to be independent and identically distributed (i.i.d.) vectors of random variables from a normal distribution with a constant variance, i.e., $\boldsymbol{\Sigma}_N = \sigma_N^2 \boldsymbol{I}_M \in \mathbb{R}^{M \times M}$. The conditional prior $p(\boldsymbol{X}|\boldsymbol{X}^{(k-1)})$ is also modeled with the joint Gaussian distribution

$$p(\boldsymbol{X}|\boldsymbol{X}^{(k-1)}) = \frac{1}{\left((2\pi)^N \det\{\boldsymbol{\Sigma}_X\}\right)^{T/2}} \times \exp\left\{-\frac{1}{2}\operatorname{tr}\left\{\boldsymbol{X}^T \boldsymbol{\Sigma}_X^{-1} \boldsymbol{X}\right\}\right\} \quad (10)$$

where $\boldsymbol{\Sigma}_X = \lambda^{-1}\boldsymbol{W}^2 \in \mathbb{R}^{N \times N}$ is the covariance matrix with the scaling factor λ introduced to control the degree of sparsity simultaneously for all the row vectors in $\boldsymbol{X}$. Note that $\boldsymbol{W}$ is a diagonal matrix that only depends on $\boldsymbol{X}^{(k-1)}$. Let $\tilde{\boldsymbol{x}}_j^{(k)}$ be the jth row of $\boldsymbol{X}^{(k)}$, thus w_j is proportional to a standard deviation assigned to the variables in $\tilde{\boldsymbol{x}}_j^{(k)}$, and it could be interpreted as a sparsity measure of $\tilde{\boldsymbol{x}}_j^{(k-1)}$. To model local smoothness in each row vector $\tilde{\boldsymbol{x}}_j \in \mathbb{R}^{1 \times L}$ of the rth block $\boldsymbol{X}^{(r)} \in \mathbb{R}^{N \times L}$, we use the SAR model [71]–[73] that belongs to a class of MRF models. It is widely used in many scientific fields [73], [74] to represent interaction among spatial data with a Gaussian noise. The SAR models the distributions of the random variables in $\tilde{\boldsymbol{x}}_j = \left[\tilde{x}_1^{(j)}, \ldots, \tilde{x}_L^{(j)}\right] \in \mathbb{R}^{1 \times L}$ with the stochastic equations

$$\tilde{\boldsymbol{x}}_j^T = \boldsymbol{S}\tilde{\boldsymbol{x}}_j^T + \boldsymbol{\epsilon}_j \quad (11)$$

where $\boldsymbol{S} = [s_{ml}] \in \mathbb{R}^{L \times L}$ is a symmetric matrix of spatial dependencies between the random variables, $\boldsymbol{\epsilon}_j \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_X^2 \boldsymbol{I}_L\right)$ is a vector of a Gaussian noise, and $\boldsymbol{I}_L \in \mathbb{R}^{L \times L}$ is an identity matrix.

There are many possibilities for defining the spatial dependence matrix $\boldsymbol{S}$ in (11) (see, e.g., [72]). Following [74] and [75], we assumed $\boldsymbol{S} = \gamma \boldsymbol{Z}$, where γ is a constant to ensure the matrix $\boldsymbol{C} = \boldsymbol{I}_L - \boldsymbol{S}$ being positive definite, and $\boldsymbol{Z} = [z_{mn}]$ is a binary symmetric band matrix that indicates the neighboring column vectors in $\boldsymbol{X}^{(r)}$. Assuming first-order interactions, we have $z_{m,m-1} = 1$ and $z_{m,m+1} = 1$ for $m \in \{2, \ldots, L\}$, and $z_{mn} = 0$, otherwise. For P-order interactions, each sample $\boldsymbol{x}_l$ has the following set of neighbors: $\{\boldsymbol{x}_{l-\nu}\}$ and $\{\boldsymbol{x}_{l+\nu}\}$ with $\nu = 1, \ldots, P$, and consequently, $\boldsymbol{Z}$ is a symmetric band matrix with P subdiagonals and P superdiagonals whose entries are equal to ones, and zeros otherwise. To keep the matrix $\boldsymbol{C}$ positive definite, it should be $\gamma < (2P)^{-1}$ for P-order interactions [74], [75]. We selected $\gamma = (2P)^{-1} - \tilde{\epsilon}$, where $\tilde{\epsilon}$ is a small constant, e.g., $\tilde{\epsilon} = 10^{-16}$.

In many applications of the SAR to image reconstruction and restoration [67], [69], [70], [74], [75], the order of interactions is fixed and usually $P = 2$. Obviously, higher order interactions attenuate high-frequency components stronger, but on the other hand, a computational complexity increases. In our case, $1 \leq P \leq \lfloor (L-1)/2 \rfloor$. Typically, $P = 2$ is a good tradeoff between a degree of smoothness and preservation of the desired high-frequency components.

From (11), the joint probability density function (pdf) for the variables in $\tilde{\boldsymbol{x}}_j$ is proportional to

$$p(\tilde{\boldsymbol{x}}_j|\beta_j) \propto \exp\left\{-\frac{1}{2}\beta_j \tilde{\boldsymbol{x}}_j \boldsymbol{C}^T \boldsymbol{C} \tilde{\boldsymbol{x}}_j^T\right\} \quad (12)$$

where β_j is a shape hyperparameter. According to (10), each row vector $\tilde{\boldsymbol{x}}_j$ is scaled as $\lambda^{(1/2)} w_j^{-1} \tilde{\boldsymbol{x}}_j$, and in consequence, $\beta_j = \lambda w_j^{-2}$. Assuming a common row smoothness profile, the joint conditional prior for $\boldsymbol{X}^{(r)}$ with the smoothness constraints can be expressed by

$$p\left(\boldsymbol{X}^{(r)}|\boldsymbol{X}_{(r)}^{(k-1)}, \lambda\right) = \prod_{j=1}^{N} p(\tilde{\boldsymbol{x}}_j|\beta_j) \propto \exp\left\{-\frac{1}{2}\lambda \sum_{j=1}^{N} w_j^{-2} \tilde{\boldsymbol{x}}_j \boldsymbol{C}^T \boldsymbol{C} \tilde{\boldsymbol{x}}_j^T\right\} = \exp\left\{-\frac{1}{2}\lambda \|\boldsymbol{W}^{-1}\boldsymbol{X}^{(r)}\boldsymbol{C}\|_F^2\right\}. \quad (13)$$

Applying the Bayesian framework to the prior (13) and the likelihood function (9) for each block $\boldsymbol{B}^{(r)}$, the posterior distribution has the form

$$p(\boldsymbol{X}|\boldsymbol{B},\boldsymbol{X}^{(k-1)},\lambda)=\frac{p(\boldsymbol{B}|\boldsymbol{X})p(\boldsymbol{X}|\boldsymbol{X}^{(k-1)},\lambda)}{p(\boldsymbol{B}|\boldsymbol{X}^{(k-1)},\lambda)}. \tag{14}$$

The superscript r is deliberately dropped for notational simplicity. The MAP estimate in the kth iterative update can be obtained with solving the following regularized least squares minimization problem:

$$\boldsymbol{X}^{(k)}=\arg\min_{X}\Psi(\boldsymbol{X}|\boldsymbol{B},\boldsymbol{X}^{(k-1)},\lambda) \tag{15}$$

where

$$\begin{aligned}\Psi(\boldsymbol{X}|\boldsymbol{B},\boldsymbol{X}^{(k-1)},\lambda)&=-2\ln p(\boldsymbol{X}|\boldsymbol{B},\boldsymbol{X}^{(k-1)},\lambda)\\&=\|\boldsymbol{B}-\boldsymbol{A}\boldsymbol{X}\|_F^2+\lambda\|\boldsymbol{W}^{-1}\boldsymbol{X}\boldsymbol{C}\|_F^2+c\end{aligned} \tag{16}$$

where c is a constant.

The stationary point of (16) expressed in the vectorized form is as follows:

$$\begin{aligned}\boldsymbol{x}_*&=\mathrm{vec}\{\boldsymbol{X}_*\}\\&=(\boldsymbol{C}^{-1}\otimes\boldsymbol{W})\boldsymbol{G}^T(\boldsymbol{G}\boldsymbol{G}^T+\lambda\boldsymbol{I}_{ML})^{-1}\mathrm{vec}\{\boldsymbol{B}\}\end{aligned} \tag{17}$$

where $\boldsymbol{G}=\boldsymbol{C}^{-1}\otimes\boldsymbol{A}\boldsymbol{W}\in\mathbb{R}^{ML\times NL}$, $\boldsymbol{I}_{ML}$ is an identity matrix, and $\otimes$ denotes the Kronecker product. The detailed calculations of the stationary point can be found in Appendix I. The smoothing matrix $\boldsymbol{C}$ has a band structure, so its inversion can be performed efficiently by a recursive procedure [82].

If $\lambda>0$ and $\mathrm{rank}(\boldsymbol{C})=L$, the point $\boldsymbol{X}_*=\mathrm{Matrix}\{\boldsymbol{x}_*\}\in\mathbb{R}^{N\times L}$, where $\mathrm{Matrix}\{\cdot\}$ means the matricized form of a vector and $\boldsymbol{x}_*\in\mathbb{R}^{NL}$ is given by (17), is the solution to the problem (15). Applying the affine transform $\boldsymbol{Q}=\boldsymbol{W}^{-1}\boldsymbol{X}$ as in Appendix I, the first term in (26) is convex by convexity of the Frobenius norm, and the second term is strictly convex because $\boldsymbol{C}$ is a positive-definite matrix. Unfortunately, the inverse transform $\boldsymbol{X}=\boldsymbol{W}\boldsymbol{Q}$ is not always guaranteed to be injective because $\boldsymbol{W}$ might be a singular matrix by row sparsity, and hence, the stationary point $\boldsymbol{X}_*$ may not be the global minimum of the functional (16). However, when the iterative approximations are still far from the exact (sparse) solution, we may assume small perturbations of the diagonal entries in $\boldsymbol{W}$, i.e., $\forall\, j: w_j\neq 0$. Under such a practical assumption, the affine transform $\boldsymbol{Q}=\boldsymbol{W}^{-1}\boldsymbol{X}$ is bijective, the functional (26) is strictly convex, and the stationary point $\boldsymbol{X}_*$ is the global minimum.

Thus, we have $\Psi(\boldsymbol{X}^{(k+1)}|\boldsymbol{X}^{(k)},\lambda)<\Psi(\boldsymbol{X}^{(k)}|\boldsymbol{X}^{(k-1)},\lambda)$ if $\boldsymbol{X}^{(k+1)}\neq\boldsymbol{X}^{(k)}$, which justifies usage of the (17) to perform the kth iterative step in our algorithm.

To simplify the inversion in (17), we use the SVD of $\boldsymbol{G}$, which is also motivated by the algorithm used for estimation of the regularization parameter.

Let $\boldsymbol{G}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$, where $\boldsymbol{\Sigma}=\mathrm{diag}\{\sigma_i\}$ contains the singular values, and $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal matrices. Thus, from (17), we have

$$\begin{aligned}&\mathrm{vec}\{\boldsymbol{X}^{(k)}\}\\&\quad=(\boldsymbol{C}^{-1}\otimes\boldsymbol{W})\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{U}^T(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\boldsymbol{U}^T+\lambda\boldsymbol{I}_{ML})^{-1}\\&\qquad\times\mathrm{vec}\{\boldsymbol{B}\}\\&\quad=(\boldsymbol{C}^{-1}\otimes\boldsymbol{W})\boldsymbol{V}_{ML}\,\mathrm{diag}\left\{\left[\frac{\sigma_i}{\sigma_i^2+\lambda}\right]\right\}\\&\qquad\times\boldsymbol{U}^T\mathrm{vec}\{\boldsymbol{B}\}\end{aligned} \tag{18}$$

where $\boldsymbol{V}_{ML}=[\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{ML}]\in\mathbb{R}^{ML\times ML}$.

C. Estimation of the Regularization Parameter

We assume that the noisy disturbances described by the matrix $\boldsymbol{N}$ in (1) are Gaussian with zero-mean and covariance matrix $\sigma^2\boldsymbol{I}_M$, where the variance σ^2 is not known. For such a case, the regularization parameter in (18) can be typically estimated with the discrepancy principle, L-curve, or GCV methods [77].

Rao *et al.* [76] proposed to estimate the regularization parameter in the FOCUSS algorithm with the modified L-curve method that combines the discrepancy principle and the linear scale L-curve method. The proposed modification assumes that the boundary values of the regularization parameter are known or they can be estimated given a prior value of SNR.

In our approach, we propose to use the GCV because this technique seems to be more flexible for integration with the main algorithm, which updates the solution $\boldsymbol{X}$. Also, it is easier to automatically search for the minimum of the GCV function than the corner on the L-curve. Rao *et al.* [76] use an analytical expression to calculate the curvature of the L-curve, however, its maximization involves computation of the solution for several samples of the parameter resulting in an increase in computational cost. Using the GCV combined with a suitable minimization algorithm, we can numerically estimate the parameter with better computational efficiency, without resorting to solving the problem (15) for each sample of the parameter in a given range. Additionally, the GCV usually gives a more accurate estimate of the regularization parameter (the distribution of samples is sharper) than the L-curve method [83].

The GCV assumes that an optimum value of λ should be chosen to minimize the GCV function, which for our model in (1) takes the following form:

$$V(\lambda)=\frac{\|\mathrm{vec}\{\boldsymbol{B}\}-(\boldsymbol{I}_L\otimes\boldsymbol{A})\mathrm{vec}\{\boldsymbol{X}^{(k)}\}\|_2^2}{\frac{1}{ML}\left(\mathrm{tr}\{\boldsymbol{I}_{ML}-\boldsymbol{P}_\lambda\}\right)^2} \tag{19}$$

where $\boldsymbol{P}_\lambda$ is the influence matrix that can be derived from

$$(\boldsymbol{I}_L\otimes\boldsymbol{A})\mathrm{vec}\{\boldsymbol{X}^{(k)}\}=\boldsymbol{P}_\lambda\mathrm{vec}\{\boldsymbol{B}\} \tag{20}$$

where $\mathrm{vec}\{\boldsymbol{X}^{(k)}\}$ is the estimated solution given by (18) in the kth iterative step.

Note that λ should be estimated for each iterative step and each block $\boldsymbol{X}^{(r)}$. Applying the SVD to the matrix $\boldsymbol{G}$, and after straightforward calculations, shown in Appendix II, we have

$$V(\lambda) = \frac{\sum_{i=1}^{ML} \xi_i^2 \left(\frac{\lambda}{\sigma_i^2+\lambda}\right)^2}{\frac{1}{ML}\left(\sum_{i=1}^{ML} \frac{\lambda}{\sigma_i^2+\lambda}\right)^2} \tag{21}$$

with $\boldsymbol{\xi} = [\xi_i] = \boldsymbol{U}^T \mathrm{vec}\{\boldsymbol{B}\} \in \mathbb{R}^{ML}$.

Remark 2: For $\boldsymbol{C} = \boldsymbol{I}_L$, we arrive at the standard M-FOCUSS algorithm: $\boldsymbol{x} = \boldsymbol{W}\boldsymbol{A}^T\left(\boldsymbol{A}\boldsymbol{W}^2\boldsymbol{A}^T + \lambda \boldsymbol{I}_M\right)^{-1}\boldsymbol{B}$, for which the GCV function becomes

$$V(\lambda) = \frac{\sum_{i=1}^{M} \xi_{ii} \left(\frac{\lambda}{\sigma_i^2+\lambda}\right)^2}{\frac{1}{M}\left(\sum_{i=1}^{M} \frac{\lambda}{\sigma_i^2+\lambda}\right)^2} \tag{22}$$

where $\boldsymbol{\Xi} = [\xi_{ij}] = \boldsymbol{U}^T\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{U} \in \mathbb{R}^{M\times M}$ and $\boldsymbol{A}\boldsymbol{W} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ with $\boldsymbol{\Sigma} = \mathrm{diag}\{\sigma_i\}$.

To minimize $V(\lambda)$, we solve the constrained problem

$$\hat{\lambda} = \arg \min_{\lambda_{\min} \le \lambda \le \lambda_{\max}} V(\lambda) \tag{23}$$

with the algorithm based on the golden section (GS) search and parabolic interpolation [84], which is closely related to the Fibonacci search. We used Matlab's implementation of the GS search, which is available in the Optimization toolbox (function f_{minbnd}). Despite the fact that the GS search finds only a local extremum, the function $V(\lambda)$ is unimodal, and the GS search always finds a desired solution, provided that the boundary values $\lambda_{\min}$ and $\lambda_{\max}$ determine a sufficiently wide range of search. Selection of the boundaries closer to the minimum accelerates the GS searching. Rough estimation of the boundaries can be achieved following the method by Rao *et al.* in [76], which uses prior knowledge about the SNR. In all our simulations, we used $\lambda_{\max} = 10^3$ and $\lambda_{\min} = 10^{-10}$.

The following alternating minimization technique was used for the update:

$$\hat{\lambda} = \arg \min_{\lambda_{\min} \le \lambda \le \lambda_{\max}} V(\lambda | \boldsymbol{X}^{(k-1)}) \tag{24}$$

$$\boldsymbol{X}^{(k)} = \arg \min_{\mathbf{X}} \Psi(\boldsymbol{X} | \boldsymbol{X}^{(k-1)}, \hat{\lambda}) \tag{25}$$

which is motivated by the hierarchical Bayesian paradigm [75].

Finally, Algorithm 1, which is based on (24) and (25), combines all the proposed approaches: the overlapping blocks, SAR-based smoothing, and simultaneous GCV-based hyperparameter estimation. The algorithm is initialized with $\boldsymbol{X}^{(0)} = \boldsymbol{A}^+\boldsymbol{B}$, where $\boldsymbol{A}^+$ is a pseudoinverse to $\boldsymbol{A}$. The SOB-M-FOCUSS denotes the M-FOCUSS algorithm with smooth overlapped blocks (SOB).

Algorithm 1: (**SOB-M-FOCUSS**)

Set $p \in [0, 2]$, $\lambda_{\min}$, $\lambda_{\max}$, $L \in [1, T]$ (window length), $\theta \in [0, 99]$ (overlapping),

$\boldsymbol{C}^{-1}$ (inverse of the smoothing matrix),

$\boldsymbol{X}^{(0)} = \boldsymbol{A}^+\boldsymbol{B}$,

$\eta = \lfloor \theta L/100 \rfloor$, % No. of overlapping samples

$R = \lceil (T-\eta)/(L-\eta) \rceil$, % No. of blocks

For $k = 1, 2, \ldots,$ until convergence **do**

For $r = 1, 2, \ldots, R$ % Block sequences

Use ϕ_r with (8) and $\Pi^{(r)}$ with (6),

$w_j^{(k)} = \left(\sum_{t\in\Pi^{(r)}} \left(x_{jt}^{(k-1)}\right)^2\right)^{(1/2)}$

$\boldsymbol{W}^{(k)} = \mathrm{diag}\left\{\left(w_j^{(k)}\right)^{1-(p/2)}\right\}$

$\boldsymbol{B}^{(r)} = [b_{*,\Pi^{(r)}}] \in \mathbb{R}^{M\times L}$

$\boldsymbol{G} = \boldsymbol{C}^{-1} \otimes \boldsymbol{A}\boldsymbol{W} \in \mathbb{R}^{ML\times NL}$

$[\boldsymbol{U}, \boldsymbol{\Sigma}, \boldsymbol{V}] = \mathtt{svd}(\boldsymbol{G})$

Compute GCV function $V(\lambda)$ with (21),

$\hat{\lambda} \leftarrow \mathtt{f_{minbnd}}\left(V(\lambda), \lambda_{\min}, \lambda_{\max}\right)$,

Compute $\boldsymbol{x}$ with (18),

$\left[x_{*,\Pi^{(r)}}^{(k)}\right] = \mathrm{Matrix}\{\boldsymbol{x}\} \in \mathbb{R}^{N\times L}$,

End % Block sequences

End % Main M-FOCUSS loop

IV. Simulation Examples

Quantitative and qualitative performance of the proposed algorithms was evaluated on the experiments of several types of signals.

The small-scale benchmark contains some multichannel electrocardiography (ECG) signals that were generated with the algorithm proposed by Sameni *et al.* [85]. The signals, which are shown in Fig. 1(a), have the following parameters: $N = 6$, $T = 712$, $d = 2$, where d is a maximal number of nonzero entries in each column vector $\boldsymbol{x}_t$ of $\boldsymbol{X}$. We assume four linear noisy mixtures, where the mixing matrix $\boldsymbol{A} \in \boldsymbol{R}^{4\times 6}$, with $\mathrm{cond}(\boldsymbol{A}) = 6.3$ and linearly independent columns, is generated randomly from a uniform statistics.

According to Remark 1, the unique sparse solution can be obtained for $d = 2$ and $M = 4$, if $2 \le L \le 4$. Because besides the sparsity constraints we also introduce the smoothness constraints, the assumptions on width of the processing window can be relaxed. For a common smoothness profile, L should be equal at least to a number of the samples that belong to the neighborhood. For second-order interactions, we should set $L = 5$. The experiments demonstrated that for our data the performance of

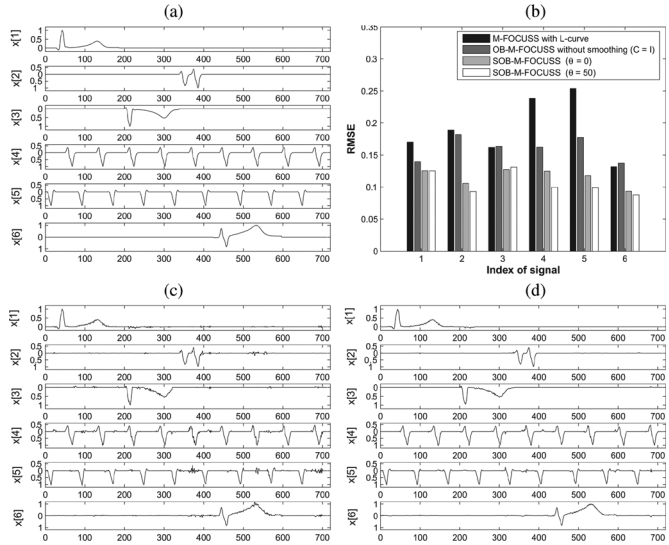

Fig. 1. Small-scale benchmark ($N = 6, T = 712, d = 2$, four noisy mixtures ($M = 4$) with SNR = 20 dB): (a) six true signals; (b) RMSE of each signal representation (x-axis variable) estimated with the following algorithms ($p = 0$, $L = 10$): M-FOCUSS with L-curve, OB-M-FOCUSS (without smoothing: $\boldsymbol{C} = \boldsymbol{I}_L$ and $\theta = 0$), SOB-M-FOCUSS with $\theta = 0$, SOB-M-FOCUSS with $\theta = 50$, respectively; (c) signal representations estimated with the OB-M-FOCUSS: $p = 0$, $L = 10$, $\theta = 0$; (d) signal representations estimated with the SOB-M-FOCUSS: $p = 0$, $L = 10$, $\theta = 50$.

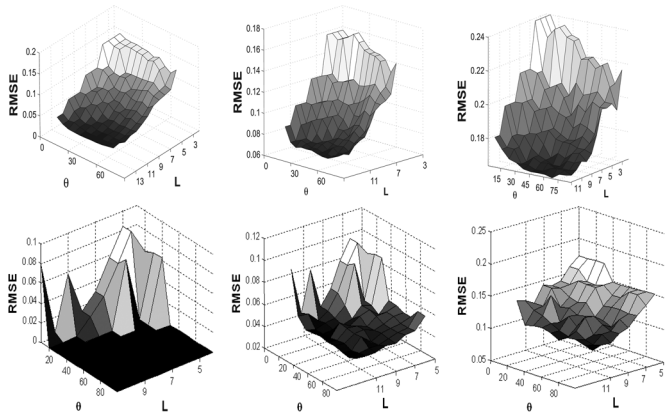

Fig. 2. RMSE plots versus θ and L for the cases: (left) noise-free; (middle) SNR = 30 dB; (right) SNR = 20 dB; (top) M-FOCUSS with the L-curve; (bottom) SOB-M-FOCUSS.

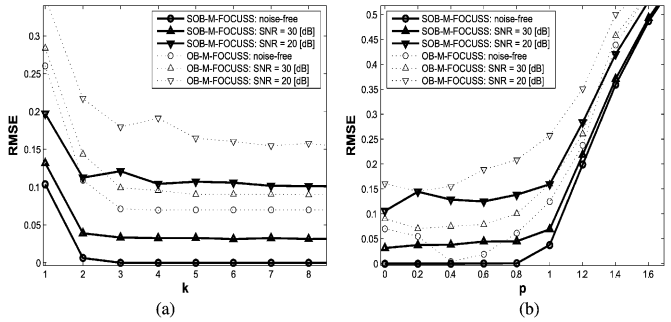

Fig. 3. RMSE plots versus: (a) k (iterations) at $p = 0$, $L = 10$; (b) p (sparsification) at $k = 6$, $L = 10$. SOB-M-FOCUSS with $\theta = 50$, OB-M-FOCUSS ($\boldsymbol{C} = \boldsymbol{I}_L$) with $\theta = 0$.

our algorithm increases with L until about $L = 20$. Hence, for a majority of the experiments, we set $L = 10$.

The noisy data was generated by adding to the exact mixtures a zero-mean homoscedastic Gaussian noise with σ^2 adjusted to have a desired value of SNR, where SNR $= 20\log_{10}(\|\boldsymbol{B}_{\text{exact}}\|_F/\|\boldsymbol{N}\|_F)$ [dB].

We compare the following algorithms: SOB-M-FOCUSS, OB-M-FOCUSS, and M-FOCUSS with the L-curve. The OB-M-FOCUSS algorithm is obtained from the SOB-M-FOCUSS by setting $\boldsymbol{C} = \boldsymbol{I}_L$ in Algorithm 1. This case removes the smoothness constraints whereas the regularization parameter is still adaptively updated with the GCV technique. In the M-FOCUSS with the L-curve, the regularization parameter is estimated adaptively (in each iterative step and processing window) using the L-curve-based estimation technique that was discussed in [76]. For the case of MMV, we search the corner of the L-shaped curve of $\log\|\boldsymbol{W}^{-1}\boldsymbol{X}^{(r)}\|_F^2$ versus $\log\|\boldsymbol{B}^{(r)} - \boldsymbol{A}\boldsymbol{X}^{(r)}\|_F^2$. For our data, we have found that the L-curve in a log–log scale usually has a distinct corner, and for this case, we could use the function `l_corner.m` from the Hansen's Matlab toolbox [86].

For quantitative performance evaluation of the tested algorithms, the root mean squared error (RMSE), which is calculated as RMSE $= \|\boldsymbol{X}_{\text{estim}} - \boldsymbol{X}_{\text{exact}}\|_F/\|\boldsymbol{X}_{\text{exact}}\|_F$, is used. Fig. 1(b)–(d) presents the results of recovering the true signals illustrated in Fig. 1(a), from the noisy data with SNR = 20 dB. The RMSEs of the rows of $\boldsymbol{X}_{\text{estim}}$ estimated with the tested algorithms are plotted in Fig. 1(b). The signal representations obtained with the OB-M-FOCUSS (without smoothing) and SOB-M-FOCUSS algorithms (with smoothing) are shown in Fig. 1(c) and (d), respectively. Note that the SOB-M-FOCUSS algorithm tends to smooth signals, which considerably reduces the effect of noise and other spiky disturbances. Moreover, the estimations are not oversmoothed, and lower values of RMSE can be obtained than with the OB-M-FOCUSS. Oversmoothing usually occurs with the L-curve-based technique [83]. The GCV gives rather sharper distributions of parameter's samples, but with longer tails.

The plots in Figs. 2 and 3 show the dependence of RMSE on the parameters: θ, L, k, p, and SNR. The overlapping block technique discussed in Section III-A can be also applied to the standard M-FOCUSS algorithm. Hence, we also compare the SOB-M-FOCUSS with the M-FOCUSS (Fig. 2) in which the regularization parameter is estimated with the L-curve technique.

It is easy to notice that the SOB-M-FOCUSS algorithm gives the estimates that have lower RMSE values than the M-FOCUSS algorithm. Tremendous difference occurs for the noise-free data, when the overlapping block technique is used for $\theta > 30$. The results demonstrate (Fig. 2) that usage of the overlapping always leads to the better performance, even if it is applied only to the standard M-FOCUSS. However, the improvement is extremely high only if the data has a very low level of disturbances. As shown in Fig. 2 (middle), for SNR = 30 dB, θ should be higher than 50 in the SOB-M-FOCUSS. For the OB-M-FOCUSS, an increase in θ results in gradual improvement of RMSE for each noisy case. Both algorithms give the estimations with lower values of RMSE when L is higher.

Fig. 3(a) and (b) gives the information on the behavior of RMSE versus the number of iterations and the value of the

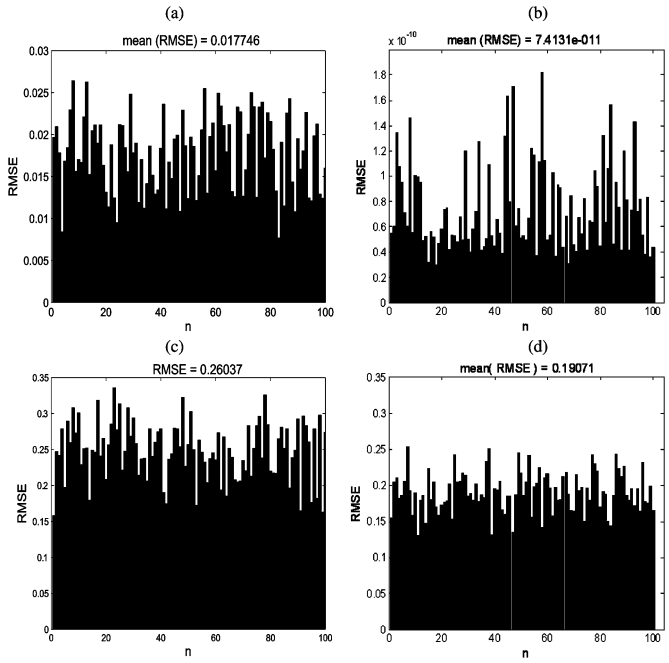

Fig. 4. Middle-scale benchmark ($N = 100, T = 10\,000, d = 10, M = 50$). RMSE of each signal ($n$) for the cases: (a) OB-M-FOCUSS ($\boldsymbol{C} = \boldsymbol{I}_L$) with $\theta = 0$ applied to noise-free data; (b) SOB-M-FOCUSS with $\theta = 50$ applied to noise-free data; (c) OB-M-FOCUSS ($\boldsymbol{C} = \boldsymbol{I}_L$) with $\theta = 0$ applied to noisy data with SNR = 20 dB; (d) SOB-M-FOCUSS with $\theta = 50$ applied to noisy data with SNR = 20 dB.

sparsity parameter p, respectively. The RMSE is lower with the SOB-M-FOCUSS than with the OB-M-FOCUSS, for each noisy case. Usually, the stagnation of RMSE occurs after a few iterations, however, in the experiments shown in Figs. 1, 2, 3(b), and 4, the iterative process is terminated when the following stopping criterion is satisfied: $\|\boldsymbol{X}^{(k)} - \boldsymbol{X}^{(k-1)}\|_F \leq \epsilon$. The threshold ϵ is set up experimentally—usually the truncation occurs after six iterative steps. Fig. 3(b) shows that for our data $p = 0$ gives the best results with respect to RMSE, which means that $J^{(p)}(\boldsymbol{X})$ in (3) measures the number of nonzero entries in $\boldsymbol{X}$. For smaller values of p, the solution with the SOB-M-FOCUSS usually better approximates the true solution but the risk of getting stuck in local minima is higher. Fig. 3(b) demonstrates that the SOB-M-FOCUSS algorithm gives satisfactory results even for higher values of p (less than 0.5).

In the middle-scale benchmark, we used 100 sparse trapezoidal signals, normalized to the unit magnitude, for which $T = 10\,000$ and $d = 10$. The signals were mixed with a random uniformly distributed matrix $\boldsymbol{A} \in \mathbb{R}^{50\times 100}$ to give 50 observed mixed signals. The noisy mixtures are generated similarly as in the previous example. The RMSEs of the estimated signal representations with the OB-M-FOCUSS and SOB-M-FOCUSS algorithms are plotted in Fig. 4.

V. Conclusion

This paper presents a modified M-FOCUSS algorithm that outperforms the M-FOCUSS for signals with a smooth temporal structure. The algorithm combines several strategies. If the observations are slightly corrupted with noise (SNR $>$ 30 dB), the important strategy is to use the overlapping block technique that considerably improves the performance of the algorithm. For this case, the estimation of the regularization parameter is not essential, and to speed up the computations, the GCV-based estimation can be switched off. Another strategy involves the smoothing technique with the SAR model that reduces spiky disturbances of the signals of interest. Our experiments demonstrated that if the true signals have a smooth temporal structure, such smoothing improves the performance independently on strength of noisy disturbances. The usage of the prior knowledge such as smoothing additionally stabilizes the solution. The strategy of simultaneous GCV-based estimation of the regularization parameter, even though it involves high computational complexity, is very important in case of real data. The experiments demonstrated that quite exact estimation of the parameter is needed, especially if the data are noisy (real).

The proposed approach is aimed to make the point that incorporating additional prior information (such as smoothing) to the M-FOCUSS algorithm may considerably improve the performance. Obviously, we have presented only one of the ways how to purse the aim. In general, many other more powerful smoothness constraints (e.g., MRF smoothing with more robust potential functions) can be considered. There are also many other techniques for incorporating the smoothness constraints to the objective function, e.g., as an additive term. To speed up the computations substantially, the parameter estimation can be performed, e.g., on a small subset of samples that are good representatives of the whole set. These approaches will be considered in our future research.

Appendix I
Derivation of the Stationary Point to (16)

Let $\boldsymbol{Q} = \boldsymbol{W}^{-1}\boldsymbol{X}$, thus (16) can be expressed as

$$\Psi(\boldsymbol{Q}) = \|\boldsymbol{AWQ} - \boldsymbol{B}\|_F^2 + \lambda\|\boldsymbol{QC}\|_F^2. \tag{26}$$

The stationary point $\boldsymbol{Q}_*$ to (26) satisfies the condition $\nabla_Q\Psi(\boldsymbol{Q}_*) = 0$, from which we get the Lyapunov equation

$$\boldsymbol{WA}^T\boldsymbol{AWQ}_* + \lambda\boldsymbol{Q}_*\boldsymbol{CC}^T = \boldsymbol{WA}^T\boldsymbol{B}. \tag{27}$$

In the vectorized form, from (27), we have (28), shown at the top of the next page, where $\boldsymbol{G} = \boldsymbol{C}^{-1} \otimes \boldsymbol{AW} \in \mathbb{R}^{ML\times NL}$, and $\boldsymbol{C}$ is assumed to be a symmetric matrix. Because $\mathrm{vec}\{\boldsymbol{Q}_*\} = \mathrm{vec}\{\boldsymbol{W}^{-1}\boldsymbol{X}_*\} = (\boldsymbol{I}_L \otimes \boldsymbol{W}^{-1})\mathrm{vec}\{\boldsymbol{X}_*\}$, and $\boldsymbol{W}$ does not depend on $\boldsymbol{X}_*$ (in the current iteration), we have (17), which is the stationary point to (16).

Appendix II
Derivation of the GCV Function in (21)

The influence matrix $\boldsymbol{P}_\lambda$ in (20) has the following form[1]:

$$\begin{aligned}\boldsymbol{P}_\lambda &= (\boldsymbol{I}_L \otimes \boldsymbol{A})(\boldsymbol{C}^{-1} \otimes \boldsymbol{W})\boldsymbol{G}^T(\boldsymbol{GG}^T + \lambda\boldsymbol{I}_{ML})^{-1} \\ &= \boldsymbol{GG}^T(\boldsymbol{GG}^T + \lambda\boldsymbol{I}_{ML})^{-1}.\end{aligned} \tag{29}$$

Let $\tilde{\boldsymbol{G}} = \boldsymbol{GG}^T$, then $\tilde{\boldsymbol{G}} = \boldsymbol{U}\tilde{\boldsymbol{\Sigma}}\boldsymbol{U}^T$, where $\tilde{\boldsymbol{\Sigma}} = \mathrm{diag}\ \{\sigma_i^2\} \in \mathbb{R}^{ML\times ML}$, and $\boldsymbol{P}_\lambda$ in (29) takes the form

$$\boldsymbol{P}_\lambda = \boldsymbol{U}\ \mathrm{diag}\ \left\{\left[\frac{\sigma_i^2}{\sigma_i^2 + \lambda}\right]\right\}\boldsymbol{U}^T. \tag{30}$$

[1]In the literature [82], the influence matrix usually has the form $\boldsymbol{P}_\lambda = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A} + \lambda\boldsymbol{I})\boldsymbol{A}^T$, and it refers to the overdetermined systems of linear equations: $\boldsymbol{Ax} = \boldsymbol{b}$, where $M > N$.

$$\begin{aligned}
\mathrm{vec}\{\boldsymbol{Q}_*\} &= \left[\boldsymbol{I}_L \otimes \boldsymbol{W}\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{W} + \lambda(\boldsymbol{C}\boldsymbol{C}^T \otimes \boldsymbol{I}_N)\right]^{-1} \mathrm{vec}\{\boldsymbol{W}\boldsymbol{A}^T\boldsymbol{B}\} \\
&= \left[(\boldsymbol{I}_L \otimes \boldsymbol{W}\boldsymbol{A}^T)(\boldsymbol{I}_L \otimes \boldsymbol{A}\boldsymbol{W}) + \lambda(\boldsymbol{C} \otimes \boldsymbol{I}_N)(\boldsymbol{C}^T \otimes \boldsymbol{I}_N)\right]^{-1} (\boldsymbol{I}_L \otimes \boldsymbol{W}\boldsymbol{A}^T)\mathrm{vec}\{\boldsymbol{B}\} \\
&= (\boldsymbol{C}^{-T} \otimes \boldsymbol{I}_N)\left[(\boldsymbol{C}^{-1} \otimes \boldsymbol{W}\boldsymbol{A}^T)(\boldsymbol{C}^{-T} \otimes \boldsymbol{A}\boldsymbol{W}) + \lambda \boldsymbol{I}_{NL}\right]^{-1} (\boldsymbol{C}^{-1} \otimes \boldsymbol{W}\boldsymbol{A}^T)\mathrm{vec}\{\boldsymbol{B}\} \\
&= (\boldsymbol{C}^{-T} \otimes \boldsymbol{I}_N)(\boldsymbol{C}^{-1} \otimes \boldsymbol{W}\boldsymbol{A}^T)\left[(\boldsymbol{C}^{-T} \otimes \boldsymbol{A}\boldsymbol{W})(\boldsymbol{C}^{-1} \otimes \boldsymbol{W}\boldsymbol{A}^T) + \lambda \boldsymbol{I}_{ML}\right]^{-1} \mathrm{vec}\{\boldsymbol{B}\} \\
&\underset{=}{\overset{\mathbf{C}=\mathbf{C}^T}{=}} (\boldsymbol{C}^{-1} \otimes \boldsymbol{I}_N)(\boldsymbol{C}^{-1} \otimes \boldsymbol{A}\boldsymbol{W})^T\left[(\boldsymbol{C}^{-1} \otimes \boldsymbol{A}\boldsymbol{W})(\boldsymbol{C}^{-1} \otimes \boldsymbol{A}\boldsymbol{W})^T + \lambda \boldsymbol{I}_{ML}\right]^{-1} \mathrm{vec}\{\boldsymbol{B}\} \\
&= (\boldsymbol{C}^{-1} \otimes \boldsymbol{I}_N)\boldsymbol{G}^T(\boldsymbol{G}\boldsymbol{G}^T + \lambda \boldsymbol{I}_{ML})^{-1}\mathrm{vec}\{\boldsymbol{B}\} \qquad (28)
\end{aligned}$$

Thus, the trace operation in the denominator in (19) can be simplified to the form

$$\mathrm{tr}\{\boldsymbol{I}_{ML} - \boldsymbol{P}_\lambda\} = \sum_{i=1}^{ML} \frac{\lambda}{\sigma_i^2 + \lambda}. \qquad (31)$$

Let ζ denote the nominator in (19), thus we have

$$\begin{aligned}
\zeta &= \left\| \mathrm{vec}\{\boldsymbol{B}\} - (\boldsymbol{I}_L \otimes \boldsymbol{A})\mathrm{vec}\{\boldsymbol{X}^{(k)}\} \right\|_2^2 \\
&= (\mathrm{vec}\{\boldsymbol{B}\})^T \boldsymbol{U}\ \mathrm{diag}\left\{ \left[\frac{\lambda}{\sigma_i^2 + \lambda} \right]^2 \right\} \boldsymbol{U}^T \mathrm{vec}\{\boldsymbol{B}\}. \qquad (32)
\end{aligned}$$

Assuming $\boldsymbol{\xi} = \boldsymbol{U}^T\mathrm{vec}\{\boldsymbol{B}\} = [\xi_1, \ldots, \xi_{ML}]^T \in \mathbb{R}^{ML}$, then

$$\zeta = \sum_{i=1}^{ML} \xi_i^2 \left(\frac{\lambda}{\sigma_i^2 + \lambda} \right)^2 \qquad (33)$$

and finally, we have the GCV function expressed by (21).

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

REFERENCES

[1] D. L. Donoho and P. B. Starck, "Uncertainty principles and signal recovery," *SIAM J. Appl. Math.*, vol. 49, no. 3, pp. 906–931, 1989.

[2] D. L. Donoho and B. F. Logan, "Signal recovery and the large sieve," *SIAM J. Appl. Math.*, vol. 52, no. 2, pp. 571–599, 1992.

[3] D. L. Donoho, "Superresolution via sparsity constraints," *SIAM J. Math. Anal.*, vol. 23, pp. 1309–1331, 1992.

[4] S. Alliney and S. A. Ruzinsky, "An algorithm for the minimization of mixed l_1 and l_2 norms with application to Bayesian estimation," *IEEE Trans. Signal Process.*, vol. 42, no. 3, pp. 618–627, Mar. 1994.

[5] S. Alliney, "Digital filters as absolute norm regularizers," *IEEE Trans. Signal Process.*, vol. 40, no. 6, pp. 1548–1562, Jun. 1992.

[6] B. D. Jeffs, "Sparse inverse solution methods for signal and image processing applications," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Seattle, WA, 1998, vol. III, pp. 1885–1888.

[7] B. D. Jeffs and M. Gunsay, "Restoration of blurred star field images by maximally sparse optimization," *IEEE Trans. Image Process.*, vol. 2, no. 2, pp. 202–211, Apr. 1993.

[8] J. Bobin, J. Fadili, Y. Moudden, and J.-L. Starck, "Morphological diversity and sparsity: New insights into multivariate data analysis," *Proc. SPIE*, vol. 6701, 2007, doi: 10.1117/12.731589, D. Van De Ville, V. K. Goyal, M. Papadakis, eds.

[9] J. Bobin, Y. Moudden, J. Fadili, and J.-L. Starck, "Morphological diversity and sparsity in blind sources separation," in *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 2007, vol. 4666, pp. 349–356.

[10] J. Bobin, J.-L. Starck, J. Fadili, and Y. Moudden, "Sparsity and morphological diversity in blind source separation," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2662–2674, Nov. 2007.

[11] J. Bobin, Y. Moudden, J.-L. Starck, and M. Elad, "Morphological diversity and source separation," *IEEE Signal Process. Lett.*, vol. 13, no. 7, pp. 409–412, Jul. 2006.

[12] M. Zibulevsky and B. A. Pearlmutter, "Blind separation of sources with sparse representations in a given signal dictionary," *Neural Comput.*, vol. 13, no. 4, pp. 863–882, 2001.

[13] J. Murray and K. Kreutz-Delgado, "An improved FOCUSS-based learning algorithm for solving blind sparse linear inverse problems," in *Conf. Record 35th Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, Nov. 4–7, 2001, vol. 1, pp. 347–351.

[14] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. West Sussex, U.K.: Wiley, 2003.

[15] R. Gribonval, "Sparse decomposition of stereo signals with matching pursuit and application to blind separation of more than two sources from a stereo mixture," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Orlando, FL, 2002, pp. 3057–3060.

[16] Y. Li, S.-I. Amari, A. Cichocki, D. W. C. Ho, and S. Xie, "Underdetermined blind source separation based on sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 423–437, Feb. 2006.

[17] I. F. Gorodnitsky and B. D. Rao, "A recursive weighted minimum-norm algorithm: Analysis and applications," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Minneapolis, MN, Apr. 1993, vol. III, pp. 1331–1334.

[18] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstructions from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.

[19] J. W. Phillips, R. M. Leahy, and J. C. Mosher, "Meg-based imaging of focal neuronal current sources," *IEEE Trans. Med. Imag.*, vol. 16, no. 3, pp. 338–348, Jun. 1997.

[20] J. Han and K. S. Park, "Regularized FOCUSS algorithm for EEG/MEG source imaging," in *Proc. 26th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, San Francisco, CA, 1993, vol. 1, pp. 122–124.

[21] P. Xu, Y. Tian, H. Chen, and D. Yao, "LP norm iterative sparse solution for EEG source localization," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 3, pp. 400–409, Mar. 2007.

[22] H. Liu, P. H. Schimpf, G. Dong, X. Gao, F. Yang, and S. Gao, "Standardized shrinking LORETA-FOCUSS (SSLOFO): A new algorithm for spatio-temporal EEG source reconstruction," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 10, pp. 1681–1691, Oct. 2005.

[23] C.-H. Im, H.-K. Jung, K.-Y. Jung, and S. Y. Lee, "Reconstruction of continuous and focalized brain functional source images from electroencephalography," *IEEE Trans. Magn.*, vol. 43, no. 4, pp. 1709–1712, Apr. 2007.

[24] J. E. Moran, S. M. Bowyer, and N. Tepley, "Multi-resolution FOCUSS: A source imaging technique applied to MEG data," *Brain Topography*, vol. 18, no. 1, pp. 1–17, 2005.

[25] M. J. Fadili and E. T. Bullmore, "Penalized partially linear models using sparse representations with an application to fMRI time series," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3436–3448, Sep. 2005.

[26] Y. Li, A. Cichocki, and S. Amari, "Sparse component analysis for blind source separation with less sensors than sources," in *Proc. 4th Int. Symp. Independent Component Anal. Blind Signal Separat.*, Kyoto, Japan, Apr. 2003, pp. 89–94.

[27] M. E. Davies and L. Daudet, "Sparsifying subband decompositions," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2003, pp. 107–110.

[28] M. E. Davies and L. Daudet, "Sparse audio representations using the MCLT," *Signal Process.*, vol. 86, no. 3, pp. 457–470, 2006.

[29] K. Kreutz-Delgado and B. D. Rao, "FOCUSS-based dictionary learning algorithms," in *Proc. Wavelet Appl. Signal Image Process.*, Bellingham, WA, Jul.-Aug. 2000, vol. 4119, pp. 459–473.

[30] K. Kreutz-Delgado, J. Murray, B. Rao, K. Engan, T. Lee, and T. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neural Comput.*, vol. 15, pp. 349–396, 2003.

[31] M. Aharon, M. Elad, and A. Bruckstein, "The K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[32] K. Engan, K. Skretting, and J. H. Husoy, "Family of iterative LS-based dictionary learning algorithms, ILS-DLA, for sparse signal representation," *Digital Signal Process.*, vol. 17, no. 1, pp. 32–49, 2007.

[33] D. Wipf and B. Rao, "l_0-norm minimization for basis selection," in *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 2005.

[34] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.

[35] S. Lesage, R. Gribonval, F. Bimbot, and L. Benaroya, "Learning unions of orthonormal bases with thresholded singular value decomposition," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Philadelphia, PA, Mar. 18–23, 2005, vol. 5, pp. v/293–v/296.

[36] G.-M. Zhang, D. M. Harvey, and D. R. Braden, "Effect of sparse basis selection on ultrasonic signal representation," *Ultrasonics*, vol. 45, pp. 82–91, 2006.

[37] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. B.*, vol. 58, pp. 267–288, 1996.

[38] M. Osborne, B. Presnell, and B. Turlach, "A new approach to variable selection in least squares problems," *IMA J. Numer. Anal.*, vol. 20, pp. 389–404, 2000.

[39] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, 2004.

[40] T. Simila, "Majorize-minimize algorithm for multiresponse sparse regression," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process*, Honolulu, HI, 2007, vol. II, pp. II-553–II-556.

[41] P. Cheng, Y. Jiang, and R. Xu, "ISAR imaging based on sparse signal representation with multiple measurement vectors," in *Proc. Int. Conf. Radar*, Shanghai, China, 2006, pp. 1–4.

[42] T. Ryen, S. O. Aase, and J. H. Husoy, "Finding sparse representation of quantized frame coefficients using 0–1 integer programming," in *Proc. 2nd Int. Symp. Image Signal Process. Anal.*, Pula, Croatia, 2001, pp. 541–544.

[43] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, Apr. 1995.

[44] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 508–518, Sep. 2000.

[45] S. S. Chen, D. L. Donoho, and M. A. Saunder, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[46] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[47] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive Approx.*, vol. 13, no. 1, pp. 57–98, 1997.

[48] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation, Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, 2006.

[49] R. A. DeVore and V. N. Temlyakov, "Some remarks on greedy algorithms," *Adv. Comput. Math.*, vol. 12, pp. 213–227, 1996.

[50] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Annu. Asilomar Conf. Signals Syst. Comput.*, Nov. 1–3, 1993, vol. 1, pp. 40–44.

[51] P. S. Huggins and S. W. Zucker, "Greedy basis pursuit," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 55, no. 7, pp. 3760–3772, Jul. 2007.

[52] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.

[53] I. Takigawa, M. Kudo, and J. Toyama, "Performance analysis of minimum l_1-norm solutions for underdetermined source separation," *IEEE Trans. Signal Process.*, vol. 52, no. 3, pp. 582–591, Mar. 2004.

[54] K. K. Herrity, A. C. Gilbert, and J. A. Tropp, "Sparse approximation via iterative thresholding," in *Proc. IEEE Int.Conf. Acoust. Speech Signal Process.*, Toulouse, France, May 14–19, 2006, vol. III, pp. 624–627.

[55] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

[56] B. Rao, K. Engan, and S. Cotter, "Diversity measure minimization based method for computing sparse solutions to linear inverse problems with multiple measurement vectors," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Montreal, QC, Canada, May 17–21, 2004, vol. 2, pp. II–369.

[57] S. F. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.

[58] J. Murray and K. Kreutz-Delgado, "Learning sparse overcomplete codes for images," *J. VLSI Signal Process.*, vol. 45, no. 1–2, pp. 97–110, Nov. 2006.

[59] B. D. Rao and K. Kreutz-Delgado, "Deriving algorithms for computing sparse solutions to linear inverse problems," in *Proc. 31st Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, Nov. 2–5, 1997, vol. 1, pp. 955–959.

[60] J. Murray and K. Kreutz-Delgado, "An improved FOCUSS-based learning algorithm for solving sparse linear inverse problems," in *Conf. Record 35th Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, Nov. 4–7, 2001, vol. 1, pp. 347–351.

[61] B. D. Rao, "Analysis and extensions of the FOCUSS algorithm," in *Proc. 30th Asilomar Conf. Signals Syst. Comput.*, 1996, vol. 2, pp. 1218–1223.

[62] D. L. Donoho and M. Elad, "Maximal sparsity representation via l_1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, pp. 2197–2202, 2003.

[63] D. L. Donoho, "For most large underdetermined systems of equations, the minimal l_1-norm near-solution approximates the sparsest near-solution," *Commun. Pure Appl. Math.*, vol. 59, no. 7, pp. 907–934, 2006.

[64] Y. Q. Li, A. Cichocki, and S. Amari, "Analysis of sparse representation and blind source separation," *Neural Comput.*, vol. 16, pp. 1193–1234, 2004.

[65] L. A. Karlovitz, "Construction of nearest points in the $l^p p$ even and l^∞ norms," *J. Approx.Theory*, vol. 3, pp. 123–127, 1970.

[66] A. Cichocki, "Blind source separation: New tools for extraction of source signals and denoising," in *Proc. SPIE*, Bellingham, WA, 2005, vol. 5818, pp. 11–25.

[67] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Jun. 1984.

[68] J. Besag, "Toward Bayesian image analysis," *J. Appl. Statist.*, vol. 16, pp. 395–407, 1989.

[69] P. J. Green, "Bayesian reconstruction from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imag.*, vol. 9, no. 1, pp. 84–93, Mar. 1990.

[70] A. Lopez, R. Molina, A. Katsaggelos, and J. Mateos, "Spect image reconstruction using compound models," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 16, no. 3, pp. 317–330, 2002.

[71] P. Whittle, "On stationary processes in the plane," *Biometrika*, vol. 41, no. 3, pp. 434–449, 1954.

[72] J. Besag, "Spatial interactions and the statistical analysis of lattice systems," *J. Roy. Statist. Soc. B*, vol. 36, pp. 192–236, 1974.

[73] B. D. Ripley, *Spatial Statistics*. New York: Wiley, 1981.

[74] R. Molina, A. Katsaggelos, and J. Mateos, "Bayesian and regularization methods for hyperparameter estimation in image restoration," *IEEE Trans. Image Process.*, vol. 8, no. 2, pp. 231–246, Feb. 1999.

[75] N. Galatsanos, V. Mesarovic, R. Molina, and A. Katsaggelos, "Hierarchical Bayesian image restoration for partially-known blurs," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1784–1797, Oct. 2000.

[76] B. D. Rao, K. Engan, S. F. Cotter, J. Palmer, and K. Kreutz-Delgado, "Subset selection in noise based on diversity measure minimization," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 760–770, Mar. 2003.

[77] G. Wahba, "Practical approximate solution to linear operator equations when the data are noisy," *SIAM J. Numer. Anal.*, vol. 14, pp. 651–667, 1977.

[78] G. H. Golub, M. T. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, pp. 215–223, 1979.

[79] K. Skretting, "Sparse signal representation using overlapping frames" Ph.D. dissertation, Faculty Inf. Technol., Math Electr. Eng., Norwegian Univ. Sci. Technol., Hogskolen i Stavanger, Norway, 2002.

[80] K. Skretting, J. H. Husoy, and S. Aase, "General design algorithm for sparse frame expansions," *Signal Process.*, vol. 86, no. 1, pp. 117–126, 2006.

[81] J. A. Tropp, "Algorithms for simultaneous sparse approximation: Part II: Convex relaxation," *Signal Process.*, vol. 86, no. 3, pp. 589–602, 2006.

[82] A. Björck, *Numerical Methods for Least-Squares Problems*. Philadelphia, PA: SIAM, 1996.

[83] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*. Philadelphia, PA: SIAM, 1998.
[84] G. E. Forsythe, M. A. Malcolm, and C. Moler, *Computer Methods for Mathematical Computations*. Englewood Cliffs, NJ: Prentice-Hall, 1976.
[85] R. Sameni, G. D. Clifford, C. Jutten, and M. B. Shamsollahi, "Multichannel ECG and noise modeling: Application to maternal and fetal ECG signals," *EURASIP J. Adv. Signal Process.*, 2007, to be published.
[86] P. C. Hansen, "Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems," *Numer. Algorithms*, vol. 6, pp. 1–35, 1994.

Rafal Zdunek (M'07) received the M.Sc. and Ph.D. degrees in telecommunications from Wroclaw University of Technology, Wroclaw, Poland, in 1997 and 2002, respectively.

Since 2002, he has been a Lecturer at the Institute of Telecommunications, Teleinformatics and Acoustics, Wroclaw University of Technology. In 2004, he was a Visiting Associate Professor at the Institute of Statistical Mathematics, Tokyo, Japan. From 2005 to 2007, he worked as a Research Scientist in RIKEN Brain Science Institute, Saitama, Japan. He has published over 50 journal and conference papers. His area of interests includes numerical methods and inverse problems in application to nonnegative matrix factorization, digital signal processing, tomographic image reconstruction, and wideband code-division multiple-access (WCDMA) network optimization.

Andrzej Cichocki (M'96–SM'07) received the M.Sc. (with honors), Ph.D. and Dr.Sc. (Habilitation) degrees, all in electrical engineering from Warsaw University of Technology, Warsaw, Poland, in 1972, 1976, and 1982, respectively

Since 1972, he has been with the Institute of Theory of Electrical Engineering, Measurement and Information Systems, Faculty of Electrical Engineering at the Warsaw University of Technology, where he obtained a title of a Full Professor in 1995. He spent several years at University Erlangen-Nuerenberg, Germany, at the Chair of Applied and Theoretical Electrical Engineering directed by Prof. R. Unbehauen, as an Alexander-von-Humboldt Research Fellow and Guest Professor. In 1995–1997, he was a team leader of the Laboratory for Artificial Brain Systems, Frontier Research Program RIKEN, Japan, in the Brain Information Processing Group. He is currently the Head of the Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Japan, in the Brain-Style Computing Group directed by Prof. Shun-ichi Amari. He is coauthor of more than 250 scientific papers and four monographs (two of them translated to Chinese): *Nonnegative Matrix and Tensor Factorization and Beyond* (New York: Wiley, 2009), *Adaptive Blind Signal and Image Processing* (New York: Wiley, 2003, rev. ed.), *CMOS Switched-Capacitor and Continuous-Time Integrated Circuits and Systems* (New York: Springer-Verlag, 1989), and *Neural Networks for Optimizations and Signal Processing* (New York: Teubner-Wiley, 1994).

Dr. Cichocki is the Editor-in-Chief of *Computational Intelligence and Neuroscience*.