Order-Preserving Factor Analysis—Application to Longitudinal Gene Expression

Arnau Tibau Puig, Ami Wiesel, Aimee K. Zaas, Chris W. Woods, Geoffrey S. Ginsburg, Gilles Fleury, and Alfred O. Hero, III, *Fellow, IEEE*

***Abstract*—We present a novel factor analysis method that can be applied to the discovery of common factors shared among trajectories in multivariate time series data. These factors satisfy a precedence-ordering property: certain factors are recruited only after some other factors are activated. Precedence-ordering arise in applications where variables are activated in a specific order, which is unknown. The proposed method is based on a linear model that accounts for each factor's inherent delays and relative order. We present an algorithm to fit the model in an unsupervised manner using techniques from convex and nonconvex optimization that enforce sparsity of the factor scores and consistent precedence-order of the factor loadings. We illustrate the order-preserving factor analysis (OPFA) method for the problem of extracting precedence-ordered factors from a longitudinal (time course) study of gene expression data.**

***Index Terms*—Dictionary learning, genomic signal processing, misaligned data processing, structured factor analysis.**

I. Introduction

With the advent of high-throughput data collection techniques, low-dimensional matrix factorizations have become an essential tool for preprocessing, interpreting or compressing high-dimensional data. They are widely used in a variety of signal processing domains including electrocardiogram [1], image [2], or sound [3] processing. These methods can take advantage of a large range of *a priori* knowledge on the form of the factors, enforcing it through constraints on sparsity or patterns in the factors. However, these methods do not work well when there are unknown misalignments between subjects in the population, e.g., unknown subject-specific time shifts. In such cases, one cannot apply standard patterning constraints without first aligning the data; a difficult task. An alternative approach, explored in this paper, is to impose a factorization constraint that is invariant to factor misalignments but preserves the relative ordering of the factors over the population. This order-preserving factor analysis is accomplished using a penalized least squares formulation using shift-invariant yet order-preserving model selection (group lasso) penalties on the factorization. As a byproduct the factorization produces estimates of the factor ordering and the order-preserving time shifts.

In traditional matrix factorization, the data is modeled as a linear combination of a number of factors. Thus, given an $n \times p$ data matrix $\boldsymbol{X}$, the Linear Factor model is defined as

$$\boldsymbol{X} = \boldsymbol{M}\boldsymbol{A} + \boldsymbol{\epsilon} \tag{1}$$

where $\boldsymbol{M}$ is a $n \times f$ matrix of factor loadings or dictionary elements, $\boldsymbol{A}$ is a $f \times p$ matrix of scores (also called coordinates) and ϵ is a small residual. For example, in a gene expression time course analysis, n is the number of time points, and p is the number of genes in the study, the columns of $\boldsymbol{M}$ contain the features summarizing the genes' temporal trajectories and the columns of $\boldsymbol{A}$ represent the coordinates of each gene on the space spanned by $\boldsymbol{M}$. Given this model, the problem is to find a parsimonious factorization that fits the data well according to selected criteria, e.g., minimizing the reconstruction error or maximizing the explained variance. There are two main approaches to such a parsimonious factorization. One, called factor analysis, assumes that the number of factors is small and yields a low-rank matrix factorization [4], [5]. The other, called dictionary learning [6], [7] or sparse coding [8], assumes that the loading matrix $\boldsymbol{M}$ comes from an overcomplete dictionary of functions and results in a sparse score matrix $\boldsymbol{A}$. There are also hybrid approaches such as sparse factor analysis [1], [2], [9] that try to enforce low rank and sparsity simultaneously.

In many situations, we observe not one but several matrices $\boldsymbol{X}_s$, $s = 1, \ldots, S$ and there are physical grounds for believing that the $\boldsymbol{X}_s$'s share an underlying model. This happens, for instance, when the observations consist of different time-blocks of sound from the same music piece [3], [10], when they consist of time samples of gene expression microarray data from different individuals inoculated with the same virus [11], or when they arise from the reception of digital data with code, spatial and temporal diversity [12]. In these situations, the fixed factor model (1) is overly simplistic.

Manuscript received September 21, 2010; revised January 12, 2011, and April 25, 2011; accepted April 26, 2011. Date of publication May 19, 2011; date of current version August 10, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Z. Jane Wang. This work was supported in part by DARPA under the PHD program. The work of G. Fleury and A. Tibau Puig was partially supported by the Digiteo DANSE project.

A. T. Puig and A. O. Hero, III, are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: atibaup@umich.edu; hero@umich.edu).

A. Wiesel is with the School of Computer Science and Engineering, The Hebrew University of Jerusalem, Israel (e-mail: amiw@cs.huji.ac.il).

A. K. Zaas and G. S. Ginsburg are with the Institute for Genome Sciences and Policy, Duke University, Durham, NC 27706 USA (e-mail: aimee.zaas@duke.edu; Geoffrey.Ginsburg@duke.edu).

A. K. Zaas and C. W. Woods are with the Division of Infectious Diseases and International Health, Department of Medicine, Duke University School of Medicine, Durham, NC 27706 USA (e-mail: woods004@mc.duke.edu).

G. Fleury and A. T. Puig are with the E3S—Supélec Systems Sciences/Signal Processing and Electronic Systems Department, Supélec, France (e-mail: gilles.fleury@supelec.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2011.2157146

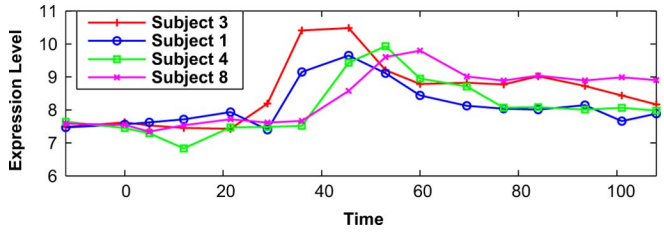


Fig. 1. Example of temporal misalignment across subjects of upregulated gene *CCRL2* . Subject 6 and subject 10 show the earliest and the latest upregulation responses, respectively.

An example, which is the main motivation for this work is shown in Fig. 1, which shows the effect of temporal misalignment across subjects in a viral challenge study reported in [11]. Fig. 1 shows the expression trajectory for a particular gene that undergoes an increase (upregulation) after viral inoculation at time 0, where the moment when upregulation occurs differs over the population. Training the model (1) on this data will produce poor fit due to misalignment of gene expression onset times.

A more sensible approach for the data in Fig. 1 would be to separately fit each subject with a translated version of a common upregulation factor. This motivates the following extension of model (1), where the factor matrices $\boldsymbol{M}_s, \boldsymbol{A}_s$ are allowed to vary across observations. Given a number S of $n \times p$ data matrices $\boldsymbol{X}_s$, we let:

$$\boldsymbol{X}_s = \boldsymbol{M}_s \boldsymbol{A}_s + \epsilon_s \; s = 1, \ldots, S. \tag{2}$$

Following the gene expression example, here n is the number of time points, p is the number of genes in the study and S is the number of subjects participating in the study. Hence, the $n \times f$ matrices $\boldsymbol{M}_s$ contain the translated temporal features corresponding to the sth subject and the $f \times p$ matrices $\boldsymbol{A}_s$ accommodate the possibility of subjects having different mixing weights. For different constraints on $\boldsymbol{M}_s$, $\boldsymbol{A}_s$, this model specializes to several well-known paradigms such as principal components analysis (PCA) [4], sparse PCA [1], k-SVD [6], structured PCA [2], nonnegative matrix factorization (NNMF) [13], maximum-margin matrix factorization (MMMF) [14], sparse shift-invariant models [3], parallel factor analysis (PARAFAC) [5], [15], or higher-order SVD (HOSVD) [16]. In this paper, we will restrict the columns of $\boldsymbol{M}_s$ to be translated versions of a common set of factors, where these factors have onsets that occur in some relative order that is consistent across all subjects. Our model differs from previous shift-invariant models considered in [3], [10], [17] in that it restricts the possible shifts to those which preserve the relative order of the factors among different subjects. We call the problem of finding a decomposition (2) under this assumption the order preserving factor analysis (OPFA) problem.

The contributions of this paper are the following. First, we propose a nonnegatively constrained linear model that accounts for temporally misaligned factors and order restrictions. Second, we give a computational algorithm that allows us to fit this model in reasonable time. Finally, we demonstrate that our methodology is able to successfully extract the principal features in a simulated dataset and in a real gene expression dataset. In addition, we show that the application of OPFA produces factors that can be used to significantly reduce the variability in clustering of gene expression responses.

This paper is organized as follows. In Section II we present the biological problem that motivates OPFA and introduce our mathematical model. In Section III, we formulate the nonconvex optimization problem associated with the fitting of our model and give a simple local optimization algorithm. In Section IV we apply our methodology to both synthetic data and real gene expression data. Finally we conclude in Section V. For lack of space many technical details are left out of our presentation but are available in the accompanying technical report [18].

II. Motivation: Gene Expression Time-Course Data

In this section we motivate the OPFA mathematical model in the context of gene expression time-course analysis. Temporal profiles of gene expression often exhibit motifs that correspond to cascades of upregulation/downregulation patterns. For example, in a study of a person's host immune response after inoculation with a certain pathogen, one would expect genes related to immune response to exhibit consistent patterns of activation across pathogens, persons and environmental conditions.

A simple approach to characterize the response patterns is to encode them as sequences of a few basic motifs such as (see, for instance, [19]):

- *Upregulation*: Gene expression changes from low to high.
- *Down-regulation*: Gene expression changes from a high to a low level.
- *Steady*: Gene expression does not vary.

If gene expression is coherent over the population of several individuals, e.g., in response to a common viral insult, the response patterns can be expected to show some degree of consistency across subjects. Human immune system response is a highly evolved system in which several biological pathways are recruited and organized over time. Some of these pathways will be composed of genes whose expressions obey a precedence-ordering, e.g., virally induced ribosomal protein production may precede toll-like receptor activation and antigen presentation [20]. This consistency exists despite temporal misalignment: even though the order is preserved, the specific timing of these events can vary across the individuals. For instance, two different persons can have different inflammatory response times, perhaps due to a slower immune system in one of the subjects. This precedence-ordering of motifs in the sequence of immune system response events is invariant to time shifts that preserve the ordering. Thus if a motif in one gene precedes another motif in another gene for a few subjects, we might expect the same precedence relationship to hold for all other subjects. Fig. 2 shows two genes from [11] whose motif precedence-order is conserved across 3 different subjects. This conservation of order allows one to impose ordering constraints on (2) without actually knowing the particular order or the particular factors that obey the order-preserving property.

Often genes are coregulated or coexpressed and have highly correlated expression profiles. This can happen, for example, when the genes belong to the same signaling pathway. Fig. 3 shows a set of different genes that exhibit a similar expression pattern (upregulation motif). The existence of high correlation

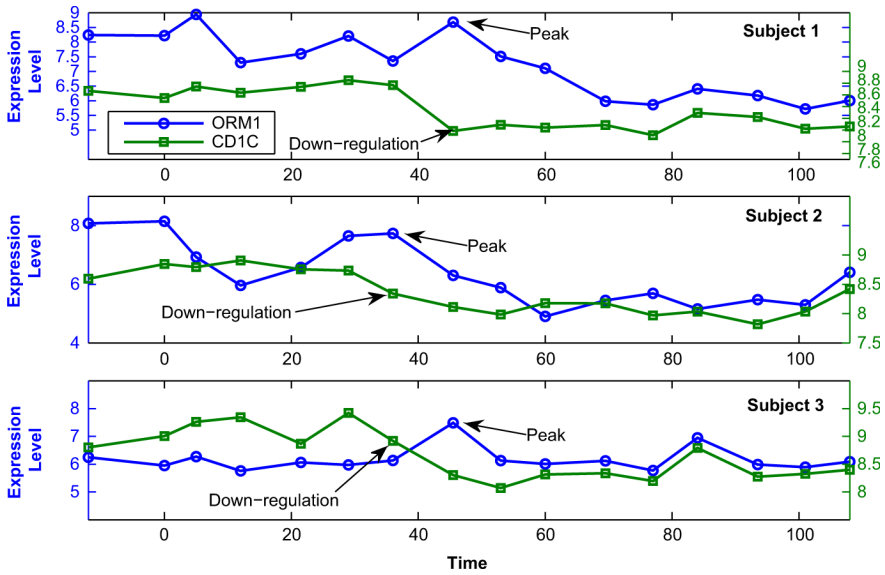


Fig. 2. Example of gene patterns with a consistent precedence-order across 3 subjects. The downregulation motif of gene *CD1C* precedes the peak motif of gene *ORM1* across these three subjects.

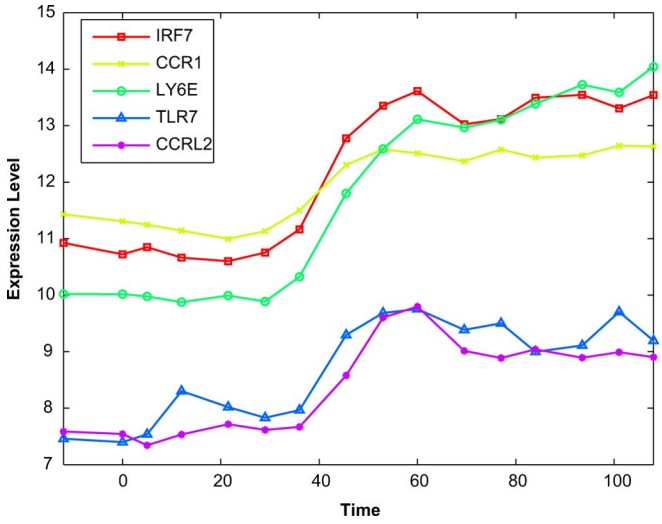


Fig. 3. Example of gene patterns exhibiting coexpression for a particular subject in the viral challenge study in [11].

between large groups of genes allows one to impose a low rank property on the factorization in (2).

In summary, our OPFA model is based on the following assumptions:

- *A1: Motif consistency across subjects*: Gene expression patterns have consistent (though not-necessarily time aligned) motifs across subjects undergoing a similar treatment.
- *A2: Motif sequence consistency across subjects*: If motif X precedes motif Y for subject s, the same precedence must hold for subject $t \neq s$.
- *A3: Motif consistency across groups of genes*: There are (not necessarily known) groups of genes that exhibit the same temporal expression patterns for a given subject.
- *A4: Gene Expression data is nonnegative*: Gene expression on a microarray is measured as an abundance and standard normalization procedures, such as RMA [21], preserve the nonnegativity of this measurement.

A few microarray normalization software packages produce gene expression scores that do not satisfy the nonnegativity assumption A4. In such cases, the nonnegativity constraint in the algorithm implementing (6) can be disabled. Note that in general, only a subset of genes may satisfy assumptions *A1-A3*.

III. OPFA Mathematical Model

In the OPFA model, each of the S observations is represented by a linear combination of *temporally aligned* factors. Each observation is of dimension $n \times p$, where n is the number of time points and p is the number of genes under consideration. Let $\boldsymbol{F}$ be an $n \times f$ matrix whose columns are the f common *alignable* factors and let $\boldsymbol{M}(\boldsymbol{F}, \boldsymbol{d})$ be a matrix valued function that applies a circular shift to each column of $\boldsymbol{F}$ according to the vector of shift parameters $\boldsymbol{d}$, as depicted in Fig. 4. Then, we can refine model (2) by restricting $\mathbf{M}_s$ to have the form

$$\boldsymbol{M}_s = \boldsymbol{M}(\boldsymbol{F}, \boldsymbol{d}^s) \tag{3}$$

where $\boldsymbol{d}^s \in \{0, \ldots, d_{\max}\}^f$ and $d_{\max} \leq n$ is the maximum shift allowed in our model. This model is a generalization of a simpler one that restricts all factors to be aligned but with a common delay

$$\boldsymbol{M}_s = \boldsymbol{U}_s \boldsymbol{F} \tag{4}$$

where $\boldsymbol{U}_s$ is a circular shift operator. Specifically, the fundamental characteristic of our model (3) is that each column can have a different delay, whereas (4) is a restriction of (3) with $d_i^s = d_j^s$ for all s and all i, j.

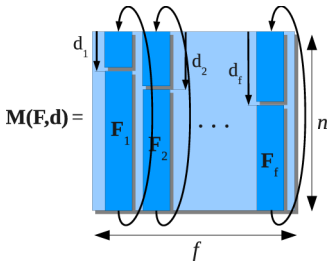


Fig. 4. Each subject's factor matrix $\boldsymbol{M}_s$ is obtained by applying a circular shift to a common set of factors $\boldsymbol{F}$ parameterized by a vector $\boldsymbol{d}$.

The circular shift is not restrictive. By embedding the observation into a larger time window it can accommodate transient gene expression profiles in addition to periodic ones, e.g., circadian rhythms [18]. There are several ways to do this embedding. One way is to simply extrapolate the windowed, transient data to a larger number of time points $n_F = n + d_{\max}$. This is the strategy we follow in the numerical experiments of Section IV-B.

This alignable factor model parameterizes each observation's intrinsic temporal dynamics through the f-dimensional vector $\boldsymbol{d}^s$. The precedence-ordering constraint *A2* is enforced by imposing the condition

$$d^{s_1}_{j_1} \leq d^{s_1}_{j_2} \Leftrightarrow d^{s_2}_{j_1} \leq d^{s_2}_{j_2} \ \forall s_2 \neq s_1 \tag{5}$$

that is, if factor j_1 precedes factor j_2 in subject s_1, then the same ordering will hold in all other subjects. Since the indexing of the factors is arbitrary, we can assume without loss of generality that $d^s_i \leq d^s_{i+1}$ for all i and all s. This characterization constrains each observation's delays $\boldsymbol{d}^s$ independently, allowing for a computationally efficient algorithm for fitting model (3).

A. OPFA as an Optimization Problem

OPFA tries to fit the model (2)–(5) to the data $\{\boldsymbol{X}_s\}_{s=1}^S$. For this purpose, we define the following penalized and constrained least squares problem:

$$\begin{aligned} \min \quad & \sum_{s=1}^{S} \left|\left|\boldsymbol{X}_s - \boldsymbol{M}\left(\boldsymbol{F}, \boldsymbol{d}^s\right)\boldsymbol{A}_s\right|\right|_F^2 \\ & + \lambda P_1\left(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_S\right) + \beta P_2\left(\boldsymbol{F}\right) \\ \text{s.t.} \quad & \boldsymbol{d}^s \in \mathcal{K}, \boldsymbol{A}_s \in \mathcal{A}_s, \ s = 1, \cdots, S, \boldsymbol{F} \in \mathcal{F} \end{aligned} \tag{6}$$

where $||\cdot||_F$ is the Frobenius norm, λ and β are regularization parameters and the set $\mathcal{K}$ constrains the delays $\boldsymbol{d}^s$ to be order-preserving

$$\mathcal{K} = \left\{\boldsymbol{d} \in \{0, \ldots, d_{\max}\}^f : d_{i+1} \geq d_i, \forall i\right\}. \tag{7}$$

where $d_{\max} \leq n$. The other soft and hard constraints are briefly described as follows.

For the gene expression application we wish to extract factors $\boldsymbol{F}$ that are smooth over time and nonnegative. Smoothness will be captured by the constraint that $P_2(\boldsymbol{F})$ is small where $P_2(\boldsymbol{F})$ is the squared total variation operator

$$P_2\left(\boldsymbol{F}\right) = \sum_{i=1}^{f} ||\boldsymbol{W}\boldsymbol{F}_{\cdot,i}||_2^2 \tag{8}$$

TABLE I
MODELS CONSIDERED IN SECTION IV-A

Model	$\boldsymbol{M}_s$	$\boldsymbol{A}_s$
OPFA	$\boldsymbol{M}_s = \boldsymbol{M}\left(\boldsymbol{F}, \boldsymbol{d}^s\right)$ $\boldsymbol{d}^s \in \mathcal{K}$, $\boldsymbol{F}$ smooth and non-negative	Non-negative sparse $\boldsymbol{A}_s$
OPFA-C	$\boldsymbol{M}_s = \boldsymbol{M}\left(\boldsymbol{F}, \boldsymbol{d}^s\right)$ $\boldsymbol{d}^s \in \mathcal{K}$, $\boldsymbol{F}$ smooth and non-negative	Non-negative sparse $\boldsymbol{A}_1 = \cdots = \boldsymbol{A}_S$
SFA	$\boldsymbol{M}_s = \boldsymbol{M}\left(\boldsymbol{F}, \boldsymbol{d}^s\right)$, $\boldsymbol{d}^s = \boldsymbol{0}$,$\boldsymbol{F}$ smooth and non-negative	Non-negative sparse $\boldsymbol{A}_s$

where $\boldsymbol{W}$ is an appropriate weighting matrix and $\boldsymbol{F}_{\cdot,i}$ denotes the ith column of matrix $\boldsymbol{F}$. From *A4* , the data is nonnegative and hence nonnegativity is enforced on $\boldsymbol{F}$ and the loadings $\boldsymbol{A}_s$ to avoid masking of positive and negative valued factors whose overall contribution sums to zero. To avoid numerical instability associated with the scale invariance $\boldsymbol{M}\boldsymbol{A} = \frac{1}{\alpha}\boldsymbol{M}\alpha\boldsymbol{A}$ for any $\alpha > 0$, we constrain the Frobenius norm of $\boldsymbol{F}$. This leads to the following constraint sets:

$$\begin{aligned} \mathcal{F} &= \left\{\boldsymbol{F} \in \mathbb{R}_+^{n \times f} : ||\boldsymbol{F}||_F \leq \delta\right\} \\ \mathcal{A}_s &= \mathbb{R}_+^{f \times p}, s = 1, \ldots, S. \end{aligned} \tag{9}$$

The parameter δ above will be fixed to a positive value as its purpose is purely computational and has little practical impact. Since the factors $\boldsymbol{F}$ are common to all subjects, assumption *A3* requires that the number of columns of $\boldsymbol{F}$ (and therefore, its rank) is small compared to the number of genes p. In order to enforce *A1* we consider two different models. In the first model, which we shall name OPFA, we constrain the columns of $\boldsymbol{A}_s$ to be sparse and the sparsity pattern to be consistent across different subjects. Notice that *A1* does not imply that the mixing weights $\boldsymbol{A}_s$ are the same for all subjects as this would not accommodate magnitude variability across subjects. We also consider a more restrictive model where we constrain $\boldsymbol{A}_1 = \cdots = \boldsymbol{A}_S = \boldsymbol{A}$ with sparse $\boldsymbol{A}$ and we call this model OPFA-C, the *C* standing for the additional constraint that the subjects share the same sequence $\mathbf{A}$ of mixing weights. The OPFA-C model has a smaller number of parameters than OPFA, possibly at the expense of introducing bias with respect to the unconstrained model. A similar constraint has been successfully adopted in [22] in a factor model for multiview learning.

Similarly to the approach taken in [23] in the context of simultaneous sparse coding, the common sparsity pattern for OPFA is enforced by constraining $P_1\left(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_S\right)$ to be small, where P_1 is a mixed-norm group-Lasso type penalty function [24]. For each of the $p \times f$ score variables, we create a group containing its S different values across subjects:

$$P_1\left(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_S\right) = \sum_{i=1}^{p}\sum_{j=1}^{f} ||\left[\boldsymbol{A}_1\right]_{j,i} \ldots \left[\boldsymbol{A}_S\right]_{j,i}||_2. \tag{10}$$

Table I summarizes the constraints of each of the models considered in this paper.

Following common practice in factor analysis, the nonconvex problem (6) is addressed using Block Coordinate Descent, displayed in the figure labeled Algorithm 1, which iteratively minimizes (6) with respect to the shift parameters $\{\boldsymbol{d}^s\}_{s=1}^S$, the

scores $\{\boldsymbol{A}_s\}_{s=1}^S$ and the factors $\boldsymbol{F}$ while keeping the other variables fixed. This algorithm is guaranteed to monotonically decrease the objective function at each iteration. Since both the Frobenius norm and $P_1(\cdot)$, $P_2(\cdot)$ are nonnegative functions, this ensures that the algorithm converges to a (possibly local) minima or a saddle point of (6).

Algorithm 1: BCD Algorithm for Finding a Local Minima of **(6)**. OPFAObjective $(\boldsymbol{F}, \{\boldsymbol{A}_s\}_{s=1}^S, \{\boldsymbol{d}^s\}_{s=1}^S)$ Denotes the Objective Function in **(6)**

Input: Initial estimate of $\boldsymbol{F}$ and $\{\boldsymbol{A}_s\}_{s=1}^S$, ϵ, λ, β.

Output: $\boldsymbol{F}$, $\{\boldsymbol{A}_s\}_{s=1}^S$, $\{\boldsymbol{d}^s\}_{s=1}^S$

$c^0 = \infty$

$c^1 \leftarrow$ OPFAObjective $\left(\boldsymbol{F}, \{\boldsymbol{A}_s\}_{s=1}^S, \{\boldsymbol{d}^s\}_{s=1}^S\right)$

$t = 1$

while $c^{t-1} - c^t \geq \epsilon$ **do**

$\{\boldsymbol{d}^s\}_{s=1}^S \leftarrow$ EstimateDelays $\left(\boldsymbol{F}, \{\boldsymbol{A}_s\}_{s=1}^S\right)$

$\{\boldsymbol{A}_s\}_{s=1}^S \leftarrow$ EstimateScores $\left(\boldsymbol{F}, \{\boldsymbol{d}^s\}_{s=1}^S\right)$

$\boldsymbol{F} \leftarrow$ EstimateFactors $\left(\{\boldsymbol{A}_s\}_{s=1}^S, \{\boldsymbol{d}^s\}_{s=1}^S\right)$

$c^t \leftarrow$ OPFAObjective $\left(\boldsymbol{F}, \{\boldsymbol{A}_s\}_{s=1}^S, \{\boldsymbol{d}^s\}_{s=1}^S\right)$

$t \leftarrow t+1$

The subroutines EstimateFactors and EstimateScores solve the following penalized regression problems:

$$\min_{\boldsymbol{F}} \quad \sum_{s=1}^{S} ||\boldsymbol{X}_s - \boldsymbol{M}(\boldsymbol{F}, \boldsymbol{d}^s)\boldsymbol{A}_s||_F^2 + \beta \sum_{i=1}^{f} ||\boldsymbol{W}\boldsymbol{F}_{\ldots,i}||_2^2$$

$$\text{s.t.} \quad \begin{cases} ||\boldsymbol{F}||_{\mathcal{F}}^2 \leq \delta \\ \boldsymbol{F}_{i,j} \geq 0 \qquad i = 1, \ldots, n, \cdot \\ \qquad\qquad\quad j = 1, \ldots, f \end{cases}$$

and

$$\min_{\{\boldsymbol{A}_s\}_{s=1}^S} \quad \sum_{s=1}^{S} ||\boldsymbol{X}_s - \boldsymbol{M}(\boldsymbol{F}, \boldsymbol{d}^s)\boldsymbol{A}_s||_F^2 + \lambda \sum_{i=1}^{p} \sum_{j=1}^{f} ||\,[\boldsymbol{A}_1]_{j,i} \ldots [\boldsymbol{A}_S]_{j,i}\,||_2$$

$$\text{s.t.} \quad \begin{cases} [\boldsymbol{A}_s]_{j,i} \geq 0 \quad i = 1, \ldots, p, \\ \qquad\qquad\quad j = 1, \ldots, f, \\ \qquad\qquad\quad s = 1, \ldots, S. \end{cases}$$

Notice that in OPFA-C, we also incorporate the constraint $\boldsymbol{A}_1 = \cdots = \boldsymbol{A}_S$ in the optimization problem above. The first is a convex quadratic problem with a quadratic and a linear constraint over a domain of dimension fn. In the applications considered here, both n and f are small and hence this problem can be solved using any standard convex optimization solver. EstimateScores is trickier because it involves a nondifferentiable convex penalty and the dimension of its domain is equal to[1] Sfp, where p can be very large. In our implementation, we use an efficient first-order method [25] designed for convex problems involving a quadratic term and a nonsmooth penalty. These procedures are standard and therefore we focus on the EstimateDelays subroutine. EstimateDelays is a discrete optimization that is solved using a branch-and-bound (BB) approach [26]. In this approach a binary tree is created by recursively dividing the feasible set into subsets ("branch"). On each of the nodes of the tree lower and upper bounds ("bound") are computed. When a candidate subset is found whose upper bound is less than the smallest lower bound of previously considered subsets these latter subsets can be eliminated ("prune") as candidate minimizers. Whenever a leaf (singleton subset) is obtained, the objective is evaluated at the corresponding point. If its value exceeds the current optimal value, the leaf is rejected as a candidate minimizer, otherwise the optimal value is updated and the leaf included in the list of candidate minimizers. Details on the application of BB to OPFA are given here.

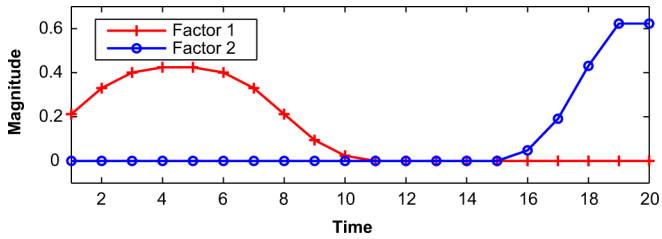


Fig. 5. Dictionary used to generated the 2-factor synthetic data of Section IV.

The subroutine EstimateDelays solves S uncoupled problems of the form

$$\min_{\boldsymbol{d} \in \mathcal{K}} ||\boldsymbol{X}_s - \boldsymbol{M}(\boldsymbol{F}, \boldsymbol{d})\boldsymbol{A}_s||_F^2 \tag{11}$$

where the set $\mathcal{K}$ is defined in (7). The "branch" part of the optimization is accomplished by recursive splitting of the set $\mathcal{K}$ to form a binary tree. The recursion is initialized by setting $\mathcal{S}_o = \{0, \ldots, d_{\max}\}^f$, $\mathcal{I}_o = \{\boldsymbol{d} \in \mathcal{K} \cap \mathcal{S}_o\}$. The splitting of the set $\mathcal{I}_o$ into two subsets is done as follows:

$$\begin{aligned} \mathcal{I}_1 &= \{\boldsymbol{d} \in \mathcal{K} \cap \mathcal{S}_o : d_{\omega_1} \leq \gamma_1\} \\ \mathcal{I}_2 &= \{\boldsymbol{d} \in \mathcal{K} \cap \mathcal{S}_o : d_{\omega_1} > \gamma_1\} \end{aligned} \tag{12}$$

and we update $\mathcal{S}_1 = \{\boldsymbol{d} \in \mathcal{S}_o : d_{\omega_1} \leq \gamma_1\}$, $\mathcal{S}_2 = \{\boldsymbol{d} \in \mathcal{S}_o : d_{\omega_1} > \gamma_1\}$. Here γ_1 is an integer $0 \leq \gamma_1 \leq d_{\max}$ and $\omega_1 \in \{1, \ldots, f\}$. $\mathcal{I}_1$ contains the elements $\boldsymbol{d} \in \mathcal{K}$ whose ω_1th component is strictly larger than γ_1 and $\mathcal{I}_2$ contains the elements whose ω_1th component is smaller than γ_1. The same kind of splitting procedure is then subsequently applied to $\mathcal{I}_1$, $\mathcal{I}_2$ and its resulting subsets. After $k-1$ successive applications of this decomposition there will be 2^{k-1} subsets and the kth split will be

$$\begin{aligned} \mathcal{I}_t &:= \{\boldsymbol{d} \in \mathcal{K} \cap \mathcal{S}_t\} \\ \mathcal{I}_{t+1} &:= \{\boldsymbol{d} \in \mathcal{K} \cap \mathcal{S}_{t+1}\} \end{aligned} \tag{13}$$

[1]This refers to the OPFA model. In the OPFA-C model, the additional constraint $\boldsymbol{A}_1 = \cdots = \boldsymbol{A}_S = \boldsymbol{A}$ reduces the dimension to fp.

where

$$\mathcal{S}_t = \{\boldsymbol{d} \in \mathcal{S}_{\pi_k} : d_{\omega_k} \leq \gamma_k\}$$
$$\mathcal{S}_{t+1} = \{\boldsymbol{d} \in \mathcal{S}_{\pi_k} : d_{\omega_k} > \gamma_k\}$$

and $\pi_k \in \{1, \ldots, 2^{k-1}\}$ denotes the parent set of the two new sets t and $t+1$, i.e., $\mathrm{pa}(t) = \pi_k$ and $\mathrm{pa}(t+1) = \pi_k$. In our implementation the splitting coordinate ω_k is the one corresponding to the coordinate in the set $\mathcal{I}_{\pi_k}$ with largest interval. The decision point γ_k is taken to be the middle point of this interval. The "bound" part of the optimization is as follows. Denote $g(\boldsymbol{d})$ the objective function in (11) and define its minimum over the set $\mathcal{I}_t \subset \mathcal{K}$

$$g_{\min}(\mathcal{I}_t) = \min_{\boldsymbol{d} \in \mathcal{I}_t} g(\boldsymbol{d}). \tag{14}$$

A lower bound for this value can be obtained by relaxing the constraint $\boldsymbol{d} \in \mathcal{K}$ in (13)

$$\min_{\boldsymbol{d} \in \mathcal{S}_t} g(\boldsymbol{d}) \leq g_{\min}(\mathcal{I}_t). \tag{15}$$

Letting $\boldsymbol{X}_s = \boldsymbol{X}_s^{\perp} + \boldsymbol{X}_s^{\|}$ where $\boldsymbol{X}_s^{\|} = \boldsymbol{X}_s \boldsymbol{A}_s^{\dagger} \boldsymbol{A}_s$ and $\boldsymbol{X}_s^{\perp} = \boldsymbol{X}_s \left(\boldsymbol{I} - \boldsymbol{A}_s^{\dagger} \boldsymbol{A}_s\right)$, we have

$$\begin{aligned} &\|\boldsymbol{X}_s - \boldsymbol{M}(\boldsymbol{F}, \boldsymbol{d}) \boldsymbol{A}_s\|_F^2 \\ &\quad = \left\|\left(\boldsymbol{X}_s \boldsymbol{A}_s^{\dagger} - \boldsymbol{M}(\boldsymbol{F}, \boldsymbol{d})\right) \boldsymbol{A}_s\right\|_F^2 + \left\|\boldsymbol{X}_s{}^{\perp}\right\|_F^2 \end{aligned}$$

where $\boldsymbol{A}_s^{\dagger}$ denotes the pseudoinverse of $\boldsymbol{A}_s$. This leads to

$$\underline{\lambda}\left(\boldsymbol{A}_s \boldsymbol{A}_s^T\right) \left\|\boldsymbol{X}_s \boldsymbol{A}_s^{\dagger} - \boldsymbol{M}(\boldsymbol{F}, \boldsymbol{d})\right\|_F^2 + \left\|\boldsymbol{X}_s{}^{\perp}\right\|_F^2 \leq g(\boldsymbol{d}) \tag{16}$$

where $\underline{\lambda}(\boldsymbol{A}_s \boldsymbol{A}_s^T)$ denotes the smallest eigenvalue of the symmetric matrix $\boldsymbol{A}_s \boldsymbol{A}_s^T$. Combining the relaxation in (15) with (16), we obtain a lower bound on $g_{\min}(\mathcal{I}_t)$

$$\begin{aligned} \Phi_{lb}(\mathcal{I}_t) &= \min_{\boldsymbol{d} \in \mathcal{S}_t} \underline{\lambda}\left(\boldsymbol{A}_s \boldsymbol{A}_s^T\right) \left\|\boldsymbol{X}_s \boldsymbol{A}_s^{\dagger} - \boldsymbol{M}(\boldsymbol{F}, \boldsymbol{d})\right\|_F^2 \\ &\quad + \left\|\boldsymbol{X}_s{}^{\perp}\right\|_F^2 \\ &\leq g_{\min}(\mathcal{I}_t) \end{aligned} \tag{17}$$

which can be evaluated by performing f *decoupled* discrete grid searches. At the kth step, the splitting node π_k will be chosen as the one with smallest $\Phi_{lb}(\mathcal{I}_t)$. Finally, this lower bound is complemented by the upper bound

$$g_{\min}(\mathcal{I}_t) \leq \Phi_{ub}(\mathcal{I}_t) = g(\boldsymbol{d}) \ \text{ for } \forall \boldsymbol{d} \in \mathcal{I}_t. \tag{18}$$

These bounds enable the branch-and-bound optimization of (11).

B. Selection of the Tuning Parameters f, λ, and β

From (6), it is clear that the OPFA factorization depends on the choice of f, λ, and β. This is a paramount problem in unsupervised learning and several heuristic approaches have been devised for simpler factorization models [9], [27], [28]. These approaches are based on training the factorization model on a subset of the elements of the data matrix (training set) to subsequently validate it on the excluded elements (test set).

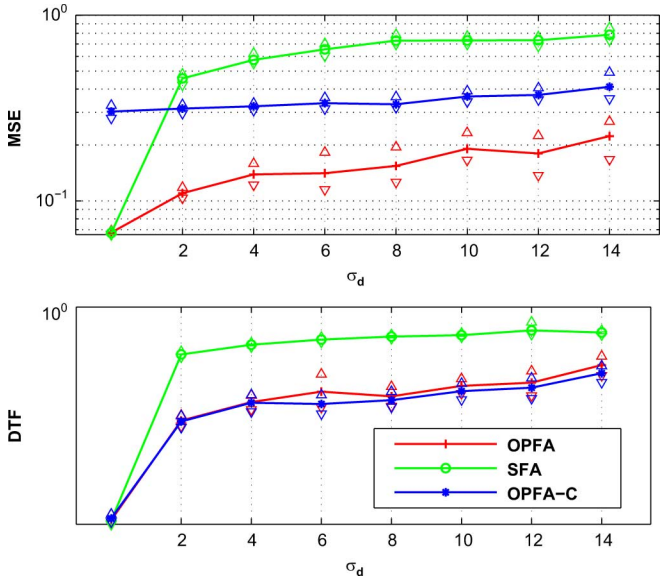


Fig. 6. MSE (top) and DTF (bottom) as a function of delay variance σ_d^2 for OPFA and SFA. These curves are plotted with 95% confidence intervals. For $\sigma_d^2 > 0$, OPFA outperforms SFA both in MSE and DTF, maintaining its advantage as σ_d increases. For large σ_d, OPFA-C outperforms the other two.

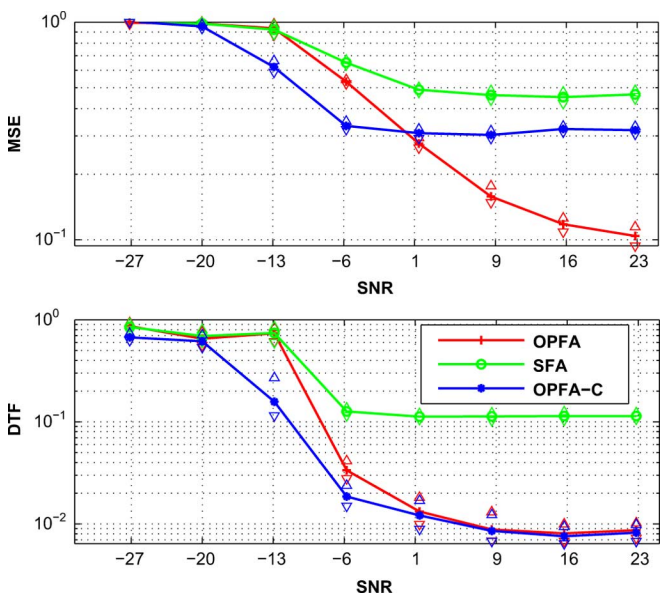


Fig. 7. Same as Fig. 6 except that the performance curves are plotted with respect to SNR for fixed $\sigma_d^2 = 5$.

The variational characterization of the OPFA decomposition allows for the presence of missing variables, i.e., missing elements in the observed matrices $\{\boldsymbol{X}_s\}_{s=1}^S$. In such case, the Least Squares fitting term in (6) is only applied to the observed set of indices[2]. We will hence follow the approach in [9] and train the

[2]While the EstimateFactors and EstimateScores remain largely the same under the presence of missing values, a little care is required to adapt EstimateDelays. Particularly, the lower bound in (16) is no longer valid in this setting, although an equally simple bound can be obtained with a bit more work. See [18] for details.

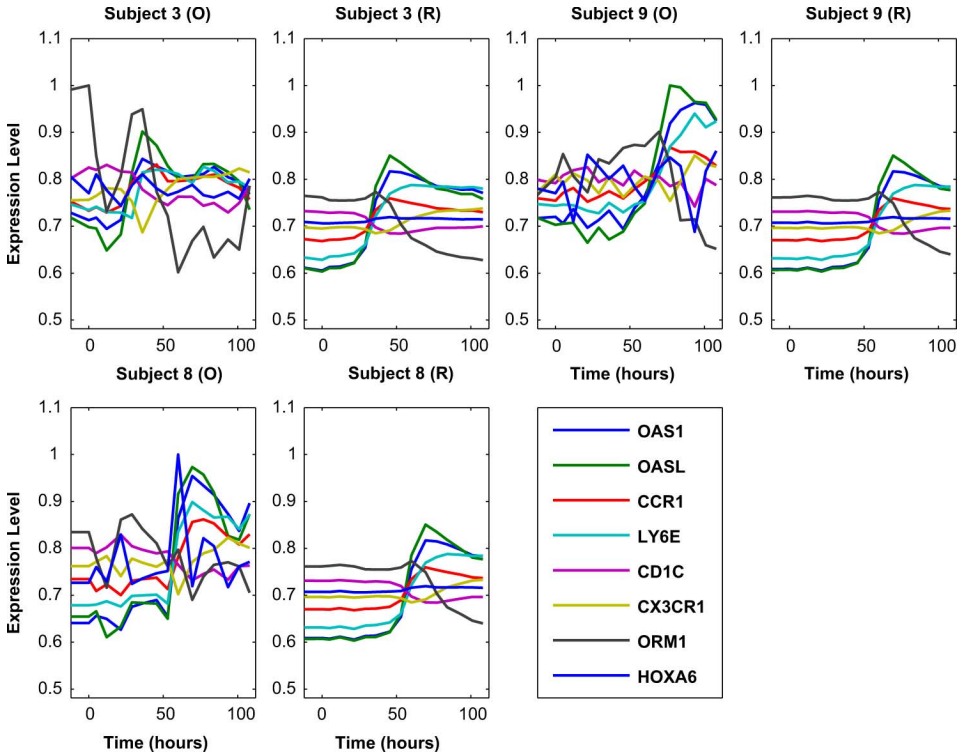


Fig. 8. Comparison of observed (O) and fitted responses (R) for three of the subjects and a subset of genes in the PHD data set. Gene expression profiles for all subjects were reconstructed with a relative residual error below 10%. The trajectories are smoothed while respecting each subject's intrinsic delay.

OPFA model over a fraction $1-\delta$ of the entries in the observations $\boldsymbol{X}_s$. Let Ω_s denote the set of $\delta\,(n \times p)$ excluded entries for the sth observation. These entries will constitute our test set and thus our Cross-Validation error measure is

$$\mathrm{CV}\,(f,\lambda,\beta) = \frac{1}{S}\sum_{s=1}^{S}\left|\left|\left[\boldsymbol{X}_s - \boldsymbol{M}\left(\hat{\boldsymbol{F}},\hat{\boldsymbol{d}}^s\right)\hat{\boldsymbol{A}}_s\right]_{\Omega_s}\right|\right|_F^2$$

where $\hat{\boldsymbol{F}}, \left\{\hat{\boldsymbol{d}}^s\right\}_{s=1}^{S}, \left\{\hat{\boldsymbol{A}}_s\right\}_{s=1}^{S}$ are the OPFA estimates obtained on the training set excluding the entries in $\{\Omega_s\}_{s=1}^{S}$, for a given choice of f, λ and β.

IV. Numerical Results

A. Synthetic Data: Periodic Model

First we evaluate the performance of the OPFA algorithm for a periodic model observed in additive Gaussian white noise

$$\boldsymbol{X}_s = \boldsymbol{M}\,(\boldsymbol{F},\boldsymbol{d}^s)\,\boldsymbol{A}_s + \epsilon_s \;\; s = 1,\ldots,S. \tag{19}$$

Here $\epsilon_s \sim \mathcal{N}_{n\times p}\left(\mathbf{o},\sigma_\epsilon^2\boldsymbol{I}\right)$, $\boldsymbol{d}^s = \mathrm{sort}\,(\boldsymbol{t}^s)$ where σ_ϵ^2 is the variance of ϵ_s and $\boldsymbol{t}^s \sim \mathcal{U}\left(0,\sqrt{12\sigma_d^2+1}\right)$ are i.i.d. The $f=2$ columns of $\boldsymbol{F}$ are nonrandom smooth signals from the predefined dictionary shown in Fig. 5. The scores $\boldsymbol{A}_s$ are generated according to a consistent sparsity pattern across all subjects and its non zero elements are i.i.d. normal truncated to the nonnegative orthant.

Here the number of subjects is $S = 10$, the number of variables is $p = 100$ and the number of time points is $n = 20$. In these experiments, we choose to initialize the factors $\boldsymbol{F}$ with temporal profiles obtained by hierarchical clustering of the data. Hierarchical clustering [29] is a standard unsupervised learning technique that groups the p variables into increasingly finer partitions according to the normalized euclidean distance of their temporal profiles. The average expression patterns of the clusters found are used as initial estimates for $\boldsymbol{F}$. The loadings $\{\boldsymbol{A}_s\}_{s=1}^{S}$ are initialized by regressing the obtained factors onto the data.

We compare OPFA and OPFA-C to a standard sparse factor analysis (SFA) solution, obtained by imposing $d_{\max}=0$ in the original OPFA model. Table I summarizes the characteristics of the three models considered in the simulations. We fix $f=2$ and choose the tuning parameters (λ,β) using the Cross-Validation procedure of Section III-B with a 5×3 grid and $\delta = .1$.

In these experiments, we consider two measures of performance, the mean square error (MSE) with respect to the generated data

$$\mathrm{MSE}:= \frac{1}{S}\sum_{s=1}^{S} E\left|\left|\boldsymbol{D}_s - \hat{\boldsymbol{D}}_s\right|\right|_F^2$$

where E is the expectation operator, $\boldsymbol{D}_s = \boldsymbol{M}\,(\boldsymbol{F},\boldsymbol{d}^s)\,\boldsymbol{A}_s$ is the generated noiseless data, and $\hat{\boldsymbol{D}}_s = \boldsymbol{M}\left(\hat{\boldsymbol{F}},\hat{\boldsymbol{d}}^s\right)\hat{\boldsymbol{A}}_s$ is the estimated data and the distance to the true factors (DTF), defined as

$$\mathrm{DTF} := 1 - \frac{1}{f}\sum_{i=1}^{f} E\frac{\boldsymbol{F}_{\cdot,i}^T\hat{\boldsymbol{F}}_{\cdot,i}}{||\boldsymbol{F}_{\cdot,i}||_2\left|\left|\hat{\boldsymbol{F}}_{\cdot,i}\right|\right|_2}$$

TABLE II
SENSITIVITY OF THE OPFA ESTIMATES TO THE INITIALIZATION CHOICE WITH RESPECT TO THE RELATIVE NORM OF THE PERTURBATION (ρ)

	DTF [mean (standard deviation)] $\times 10^{-3}$		
SNR	$\rho = 0.002$	$\rho = 1.08$	$\rho = 53.94$
22.8	0.0 (0.0)	3.4 (9.4)	1.9 (3.2)
-2	1.3 (0.5)	1 (9.4)	1.25 (1.5)
-27.1	46 (20)	58 (17)	63 (8)
	MSE [mean (standard deviation $\times 10^{-3}$)]		
SNR	$\rho = 0.002$	$\rho = 1.08$	$\rho = 53.94$
22.8	0.02 (1.5)	0.05 (69)	0.11 (99)
-2	0.35 (7.9)	0.36(22)	0.38 (32)
-27.1	0.96 (19)	0.99 (18)	1.00 (24)

TABLE III
CROSS VALIDATION RESULTS FOR SECTION IV-B

	$f=1$	$f=2$	$\mathbf{f=3}$	$f=4$	$f=5$
$\min \mathrm{CV}(f,\lambda,\beta)$	20.25	13.66	**12.66**	12.75	12.72
Relative residual error ($\times 10^{-3}$)	7.2	4.8	**4.5**	4.5	4.4
$\hat{\lambda}(\times 10^{-8})$	5.99	1	**1**	1	35.9
$\hat{\beta}$ ($\times 10^{-6}$)	3.16	3.16	**0.01**	0.01	100

where $\boldsymbol{F}$, $\hat{\boldsymbol{F}}$ are the generated and the estimated factor matrices, respectively.

Fig. 6 shows the estimated MSE and DTF performance curves as a function of the delay variance σ_d^2 for fixed $\mathrm{SNR} = 15$ dB (which is defined as $\mathrm{SNR} = 10\log\left(\frac{1}{S}\sum_s \frac{E\left(\|\boldsymbol{M}(\boldsymbol{F},\boldsymbol{d}^s)\boldsymbol{A}_s\|_F^2\right)}{np\sigma_\epsilon^2}\right)$). OPFA and OPFA-C perform at least as well as SFA for zero delay ($\sigma_d = 0$) and significantly better for $\sigma_d > 0$ in terms of DTF. OPFA-C outperforms OPFA for high delay variances σ_d^2 at the price of a larger MSE due to the bias introduced by the constraint $\boldsymbol{A}_1 = \cdots = \boldsymbol{A}_S$. In Fig. 7 the performance curves are plotted as a function of SNR, for fixed $\sigma_d^2 = 5$. Note that OPFA and OPFA-C outperform SFA in terms of DTF and that OPFA is better than the others in terms of MSE for $\mathrm{SNR} > 0$ db. Again, OPFA-C shows increased robustness to noise in terms of DTF.

We also performed simulations to demonstrate the value of imposing the order-preserving constraint in (11). This was accomplished by comparing OPFA to a version of OPFA for which the constraints in (11) are not enforced. Data were generated according to the model (19) with $S = 4$, $n = 20$, $f = 2$, and $\sigma_d^2 = 5$. The results of our simulations (not shown) were that, while the order-preserving constraints never degrade OPFA performance, the constraints improve performance when the SNR is small (below 3 dB for this example).

Finally, we conclude here by studying the sensitivity of the final OPFA estimates with respect to the initialization choice. To this end, we initialize the OPFA algorithm with the correct model perturbed with a random gaussian vector of increasing variance. We analyze the performance of the estimates in terms of MSE and DTF as a function of the norm of the model perturbation relative to the norm of the noiseless data, which we denote by ρ. Notice that larger ρ corresponds to increasingly random initialization. The results in Table II show that the MSE and DTF of the OPFA estimates are very similar for a large range of values of ρ and therefore are robust to the initialization.

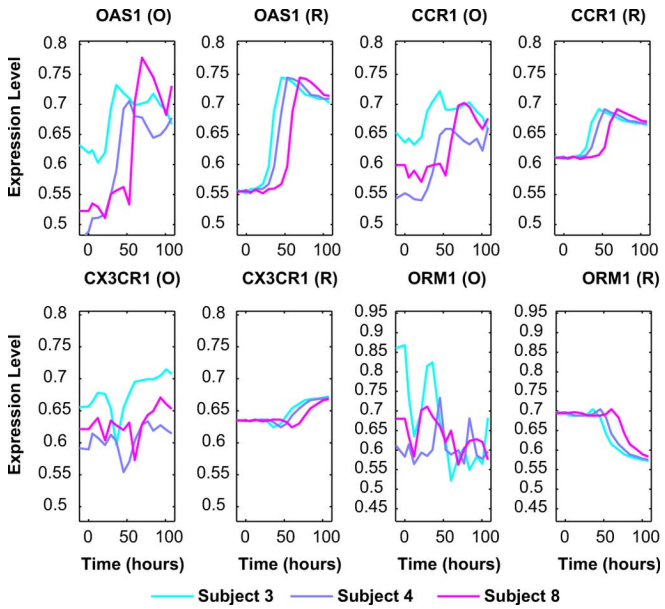


Fig. 9. Comparison of observed (O) and fitted responses (R) for four genes (*OAS1*, *CCR1*, *CX3CR1*, *ORM1*) showing upregulation and downregulation motifs and three subjects in the PHD dataset. The gene trajectories have been smoothed while conserving their temporal pattern and their precedence-order. The OPFA-C model revealed that *OAS1* upregulation occurs consistently after *ORM1* downregulation among all subjects.

B. Experimental Data: Predictive Health and Disease (PHD)

The PHD data set was collected as part of a viral challenge study that is described in [11]. In this study 20 human subjects were inoculated with live H3N2 virus and Genechip mRNA gene expression in peripheral blood of each subject was measured over 16 time points. The raw Genechip array data was preprocessed using robust multi-array analysis [21] with quantile normalization [30]. In this section we show results for the constrained OPFA model (OPFA-C). While not shown here, we have observed that OPFA-C gives very similar results to unconstrained OPFA but with reduced computation time.

Specifically, we use OPFA-C to perform the following tasks:

1) *Subject Alignment*: Determine the alignment of the factors to fit each subject's response, therefore revealing each subject's intrinsic response delays.
2) *Gene Clustering*: Discover groups of genes with common expression signature by clustering in the low-dimensional space spanned by the aligned factors. Since we are using the OPFA-C model, the projection of each subject's data on this lower dimensional space is given by the scores $\boldsymbol{A} := \boldsymbol{A}_1 = \cdots = \boldsymbol{A}_S$. Genes with similar scores will have similar expression signatures.
3) *Symptomatic Gene Signature discovery*: Using the gene clusters obtained in step 2 we construct temporal signatures common to subjects who became sick.

The data was normalized by dividing each element of each data matrix by the sum of the elements in its column. Since the data is nonnegative valued, this will ensure that the mixing weights of different subjects are within the same order of magnitude, which is necessary to respect the assumption that $\boldsymbol{A}_1 = \cdots = \boldsymbol{A}_S$ in OPFA-C. In order to select a subset of strongly varying genes, we applied one-way Analysis-Of-Variance [31]

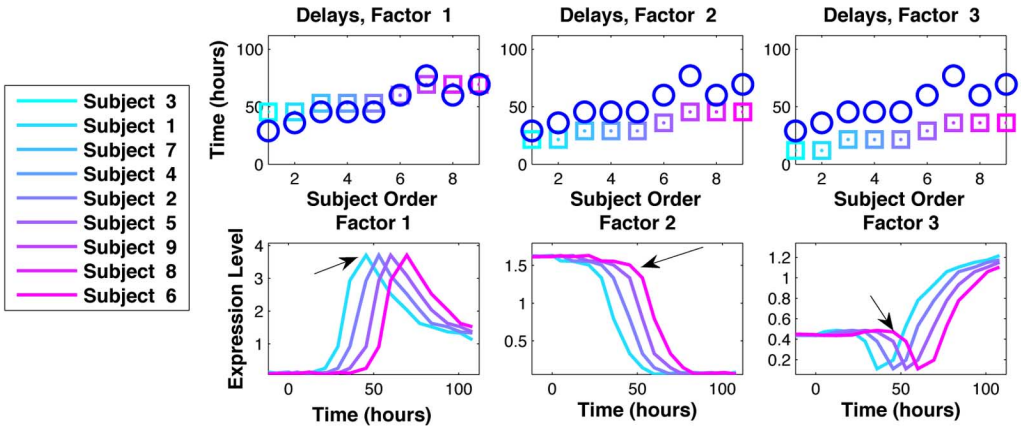


Fig. 10. Top plots: Motif onset time for each factor (□) and peak symptom time reported by expert clinicians (O). Bottom plots: Aligned factors for each subject. Factor 1 and 3 can be interpreted as upregulation motifs and factor 2 is a strong downregulation pattern. The arrows show each factor's motif onset time.

to test for the equality of the mean of each gene at 4 different groups of time points and selected the first $p = 300$ genes ranked according to the resulting F-statistic. To these gene trajectories we applied OPFA-C to the $S = 9$ symptomatic subjects in the study. In this context, the columns in $\boldsymbol{F}$ are the set of signals emitted by the common immune system response and the vector $\boldsymbol{d}^s$ parameterizes each subject's characteristic onset times for the factors contained in the columns of $\boldsymbol{F}$. To avoid wrap-around effects, we worked with a factor model of dimension $n = 24$ in the temporal axis.

The OPFA-C algorithm was run with 4 random initializations and retained the solution yielding the minimum of the objective function (6). For each $f = 1, \ldots, 5$ (number of factors), we estimated the tuning parameters (λ, β) following the Cross-Validation approach described in III-B over a 10×3 grid. The resulting results, shown in Table III resulted in selecting $\beta = 1 \times 10^{-8}$, $\lambda = 1 \times 10^{-8}$ and $f = 3$. The choice of three factors is also consistent with an expectation that the principal gene trajectories over the period of time studied are a linear combination of increasing, decreasing or constant expression patterns [11].

To illustrate the goodness-of-fit of our model, we plot in Fig. 8 the observed gene expression patterns of 13 strongly varying genes and compare them to the OPFA-C fitted response for three of the subjects, together with the relative residual error. The average relative residual error is below 10% and the plots demonstrate the agreement between the observed and the fitted patterns. Fig. 9 shows the trajectories for each subject for four genes having different regulation motifs: upregulation and downregulation. It is clear that the gene trajectories have been smoothed while conserving their temporal pattern and their precedence-order, e.g., the upregulation of gene *OAS1* consistently follows the downregulation of gene *ORM1*.

In Fig. 10 we show the 3 factors along with the factor delays and factor loading discovered by OPFA-C. The three factors, shown in the three bottom panels of the figure, exhibit features of three different motifs: factor 1 and factor 3 correspond to upregulation motifs occurring at different times; and factor 2 is a strong downregulation motif. The three top panels show the onset times of each motif as compared to the clinically determined peak symptom onset time. Note, for example, that the strong upregulation pattern of the first factor coincides closely with the onset peak time. Genes strongly associated to this factor have been closely associated to acute antiviral and inflammatory host response [11]. Interestingly, the downregulation motifs associated with factor 2 consistently precedes this upregulation motif.

Finally, we consider the application of OPFA as a preprocessing step preceding a clustering analysis. Here the goal is to find groups of genes that share similar expression patterns and determine their characteristic expression patterns. In order to obtain gene clusters, we perform hierarchical clustering on the raw data $(\{\boldsymbol{X}_s\}_{s=1}^S)$ and on the lower dimensional space of the estimated factor scores $(\{\boldsymbol{A}_s\}_{s=1}^S)$, obtaining two different sets of 4 well-differentiated clusters. We then compute the average expression signatures of the genes in each cluster by averaging over the observed data $(\{\boldsymbol{X}_s\}_{s=1}^S)$ and averaging over the data after OPFA correction for the temporal misalignments. Fig. 11 illustrates the results. Clustering using the OPFA-C factor scores produces a very significant improvement in cluster concentration as compared to clustering using the raw data $\{\boldsymbol{X}_s\}_{s=1}^S$. The first two columns in Figure compare the variation of the gene profiles over each cluster for the temporally realigned data (labeled "A") as compared to the profile variation of these same genes for the misaligned observed data (labeled "M"). For comparison, the last column shows the results of applying hierarchical clustering directly to the original misaligned dataset $\{\boldsymbol{X}_s\}_{s=1}^S$. It is clear that clustering on the low-dimensional space of the OPFA-C scores unveils interesting motifs from the original noisy temporal expression trajectories.

V. Conclusion

We have proposed a general method of order-preserving factor analysis that accounts for possible temporal misalignments in a population of subjects undergoing a common treatment. We have described a simple model based on circular-shift translations of prototype motifs and have shown how to embed transient gene expression time courses into this periodic model. The OPFA model can significantly improve interpretability of complex misaligned data. The method is

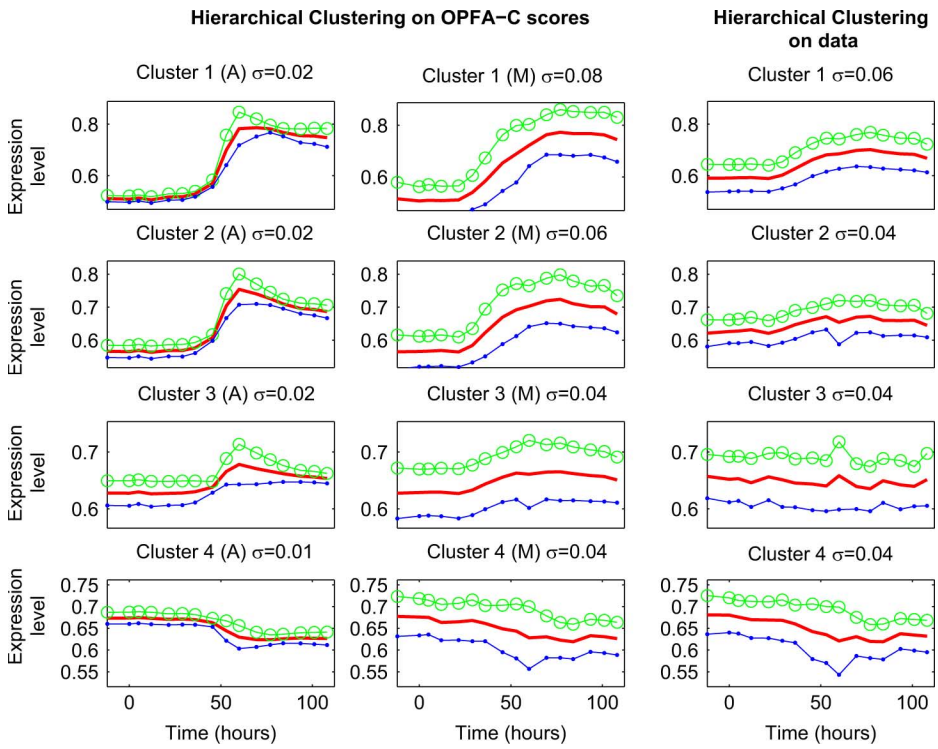


Fig. 11. The first two columns show the average expression signatures and their estimated upper/lower confidence intervals for each cluster of genes obtained by: Averaging the *estimated Aligned* expression patterns over the $S = 9$ subjects (A) and directly averaging the misaligned observed data for each of the gene clusters obtained from the OPFA-C scores (M). The confidence intervals are computed according to $\pm$ the estimated standard deviation at each time point. The cluster average standard deviation (σ) is computed as the average of the standard deviations at each time point. The last column shows the results of applying hierarchical clustering directly to the original misaligned dataset $\{\boldsymbol{X}_s\}_{s=1}^{S}$. In the first column, each gene expression pattern is obtained by mixing the estimated aligned factors $\boldsymbol{F}$ according to the estimated scores $\boldsymbol{A}$. The alignment effect is clear and interesting motifs become more evident.

applicable to other signal processing areas beyond gene expression time course analysis.

A Matlab package implementing OPFA and OPFA-C will be available at the Hero Group Reproducible Research page (http://tbayes.eecs.umich.edu).

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their insightful suggestions.

REFERENCES

[1] I. Johnstone and A. Lu, "On consistency and sparsity for principal components analysis in high dimensions," *J. Amer. Statist. Assoc.*, vol. 104, no. 486, pp. 682–693, 2009.

[2] R. Jenatton, G. Obozinski, and F. Bach, "Structured sparse principal component analysis," in *Proc. 13th Int. Conf. Artif. Intell. Statist. (AISTATS)*, 2010, vol. 9.

[3] T. Blumensath and M. Davies, "Sparse and shift-invariant representations of music," *IEEE Trans. Speech Audio Process.*, vol. 14, no. 1, pp. 50–50, 2006.

[4] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *Philosoph. Mag. Ser. 6*, vol. 2, no. 11, pp. 559–572, 1901.

[5] J. Carroll and J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.

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

[7] K. Kreutz-Delgado, J. Murray, B. Rao, K. Engan, T. Lee, and T. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neural Comput.*, vol. 15, no. 2, pp. 349–396, 2003.

[8] B. Olshausen and D. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?," *Vision Res.*, vol. 37, no. 23, pp. 3311–3325, 1997.

[9] D. Witten, R. Tibshirani, and T. Hastie, "A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis," *Biostatistics*, vol. 10, no. 3, pp. 515–515, 2009.

[10] B. Mailhé, S. Lesage, R. Gribonval, F. Bimbot, and P. Vandergheynst, "Shift-invariant dictionary learning for sparse representations: Extending K-SVD," in *Proc. 16th Eur. Signal Process. Conf.*, 2008, vol. 4.

[11] A. Zaas *et al.*, "Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans," *Cell Host Microbe*, vol. 6, no. 3, pp. 207–217, 2009.

[12] N. Sidiropoulos, G. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 810–823, 2000.

[13] D. Lee and H. Seung, "Learning the parts of objects by nonnegative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[14] N. Srebro, J. Rennie, and T. Jaakkola, "Maximum-margin matrix factorization," *Adv. Neural Inf. Process. Syst.*, vol. 17, pp. 1329–1336, 2005.

[15] T. Kolda and B. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[16] G. Bergqvist and E. Larsson, "The higher-order singular value decomposition: Theory and an application (lecture notes)," *IEEE Signal Process. Mag.*, vol. 27, pp. 151–154, 2010.

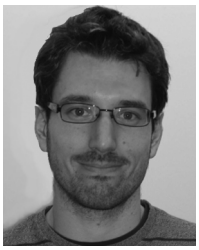
[17] M. Lewicki and T. Sejnowski, "Coding time-varying signals using sparse, shift-invariant representations," *Adv. Neural Inf. Process. Syst.*, pp. 730–736, 1999.

[18] A. T. Puig, A. Wiesel, and A. Hero, Order-Preserving Factor Analysis Univ. Michigan, Ann Arbor, 2010.

[19] L. Sacchi, C. Larizza, P. Magni, and R. Bellazzi, "Precedence temporal networks to represent temporal relationships in gene expression data," *J. Biomed. Informat.*, vol. 40, no. 6, pp. 761–774, 2007.

[20] A. Aderem and R. Ulevitch, "Toll-like receptors in the induction of the innate immune response," *Nature*, vol. 406, no. 6797, pp. 782–787, 2000.

[21] R. Irizarry, B. Hobbs, F. Collin, Y. Beazer-Barclay, K. Antonellis, U. Scherf, and T. Speed, "Exploration, normalization and summaries of high density oligonucleotide array probe level data," *Biostatistics*, vol. 4, no. 2, pp. 249–249, 2003.
[22] Y. Jia, M. Salzmann, and T. Darrell, Factorized Latent Spaces with Structured Sparsity EECS Dep., Univ. Calif., Berkeley, 2010, Tech. Rep. UCB/EECS-2010-99.
[23] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. IEEE 12th Int. Conf. Comput. Vision* , 2010, pp. 2272–2279.
[24] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Royal Statist. Soc. Ser. B Statist. Methodol.*, vol. 68, no. 1, pp. 49–49, 2006.
[25] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci*, vol. 2, pp. 183–202, 2009.
[26] E. Lawler and D. Wood, "Branch-and-bound methods: A survey," *Operat. Res.*, vol. 14, no. 4, pp. 699–719, 1966.
[27] A. Owen and P. Perry, "Bi-cross-validation of the SVD and the non-negative matrix factorization," *Ann. Appl. Statist.*, vol. 3, no. 2, pp. 564–594, 2009.
[28] S. Wold, "Cross-validatory estimation of the number of components in factor and principal components models," *Technometrics*, vol. 20, no. 4, pp. 397–405, 1978.
[29] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. New York: Springer, 2005.
[30] B. Bolstad, R. Irizarry, M. Astrand, and T. Speed, "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias," *Bioinformatics*, vol. 19, no. 2, p. 185, 2003.
[31] J. Neter *et al., Applied Linear Statistical Models*. Chicago, IL: Irwin Burr Ridge, 1996.



Arnau Tibau Puig received the Electrical Engineering and Diplôme d'Ingénieur degrees from the Technical University of Catalonia (UPC) and the École Supérieure d'Électricité (Supélec) in 2007.

He is currently pursuing the Ph.D. degree at the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, in partnership with the Signal Processing and Electronic Systems Department, Supélec, Gif-sur-Yvette. In 2007, he performed research in blind estimation for passive radar with the SONDRA Laboratory, National University of Singapore. His current research interests are in efficient methods for high-dimensional time series analysis.

Ami Wiesel received the B.Sc. and M.Sc. degrees in electrical engineering from Tel-Aviv University, Tel-Aviv, Israel, in 2000 and 2002, respectively, and the Ph.D. degree in electrical engineering from the Technion-Israel Institute of Technology, Haifa, Israel, in 2007.

He was a Postdoctoral Fellow with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, during 2007-2009. Since January 2010, he has been a Senior Lecturer with the School of Computer Science and Engineering, Hebrew University, Jerusalem, Israel.

Dr. Wiesel received the Young Author Best Paper Award for a 2006 paper in the IEEE TRANSACTIONS IN SIGNAL PROCESSING and a Student Paper Award for the 2005 Workshop on Signal Processing Advances in Wireless Communications (SPAWC) paper. He was awarded the Weinstein Study Prize in 2002, the Intel Award in 2005, the Viterbi Fellowship in 2005 and 2007, and the Marie Curie Fellowship in 2008.

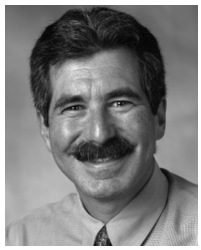
Aimee K. Zaas received the M.D. degree from the Feinberg School of Medicine, Northwestern University, Evanston, IL. She received her Postdoctoral training in infectious diseases, as well as a Master's of Health Sciences from Duke University, Durham, NC.

She completed a residency and chief residency in internal medicine at The Johns Hopkins Hospital, Baltimore, MD. She is an Associate Professor of Medicine, Division of Infectious Diseases and International Health, with the Department of Medicine, Duke University School of Medicine. Her research interests include use of host gene expression to diagnose and classify infectious diseases, as well as risk factors for infections in immunocompromised patients.

Chris W. Woods received the M.D. degree from Duke University School of Medicine in 1994.

He is the Co-Director of the Hubert-Yeargan Center for Global Health, a member of the Duke University Global Health Institute, Director of Graduate Studies for the M.Sc. in Global Health, and lead of the Institute's Signature Research Initiative on Emerging Infections. He is an Associate Professor with the Departments of Medicine and Pathology, Duke University, Durham, NC; an adjunct Assistant Professor in Epidemiology with the University of North Carolina at Chapel Hill School of Public Health; and Chief of Infectious Diseases and clinical microbiology and hospital epidemiologist for the Durham VA Medical Center. He has published more than 60 peer-reviewed articles and has a particular interest in the development of medical microbiology capacity in the developing world and the epidemiology of emerging and reemerging infectious diseases. He is a partner in the Southeastern Center for Emerging Biological Threats and coordinates the preparedness core for the Southeast Research Center for Excellence in Emerging Infections and Biodefense (SERCEB).

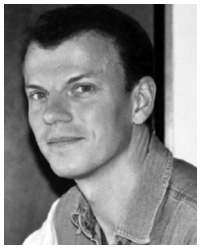
Dr. Woods is board-certified in infectious diseases and medical microbiology. He is also part of several DOD and National Institutes of Health (NIH)-funded research projects and training grants and has served as an advisor and consultant for WHO, CDC , and other Emerging Infections programs.



Geoffrey S. Ginsburg received the M.D. and Ph.D. degrees in biophysics from Boston University in 1984.

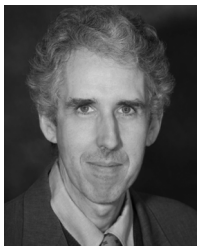
He is the Director of the Center for Genomic Medicine, Duke Institute for Genome Sciences and Policy, the Executive Director of the Center for Personalized Medicine, Duke Medicine, and a Professor of Medicine and Pathology at the Duke University Medical Center, Durham, NC. His research focuses on the development of translational pathways for genomics to human health. Prior to coming to Duke University in September 2004, he was on the faculty of Harvard Medical School, Cambridge, MA, and he was Vice President of Molecular and Personalized Medicine at Millennium Pharmaceuticals, Inc. He is a Coeditor of *Genomic and Personalized Medicine* (Elsevier: 2009, 1st ed.).

Dr. Ginsburg is a founding member of the Personalized Medicine Coalition, Editorial Advisory Board Member for Science Translational Medicine and MedScape Genomic Medicine, Senior Consulting Editor for *The Journal of the American College of Cardiology*, Editor-in-Chief of *The HUGO Journal*. He is currently a member of the Board of External Experts for the National Heart, Lung and Blood Institute, the Institute of Medicine's Roundtable on Genome-Based Research to Human Health, the National Advisory Council for Human Genome Research, and a member of the External Scientific Panel for the Pharmacogenomics Research Network.



Gilles Fleury was born in Bordeaux, France, on January 8, 1968. He received the M.S. degree in electrical engineering from Supélec in 1990, the Ph.D, and the HDR ("Habilitation a Diriger des Recherches") degrees, both in sciences physiques, from the Université de Paris-Sud, in 1994 and 2003, respectively.

He has been a Full Professor with the Université de Paris-Sud, since 2003. Since 2007, he has been the Head of the Department of Signal Processing and Electronics Systems. In 2009, he was appointed to the Head of the "E3S" team (Supelec Systems Sciences) grouping five Research Departments of Supélec. He has worked in the areas of inverse problems for many industrial applications and optimal design. His current research interests include bioinformatics, optimal nonlinear modelling and nonuniformly sampled signals processing. He is specifically interested in on-line system identification with irregularly sampled data, optimization of expensive-to-evaluate functions and uncertainty propagation.



Alfred O. Hero, III (F'98) received the B.S. degree (*summa cum laude*) from Boston University, Boston, MA, in 1980, and the Ph.D. degree from Princeton University, Princeton, NJ, in 1984, both in electrical engineering.

Since 1984, he has been with the University of Michigan, Ann Arbor, where he is the R. Jamison and Betty Williams Professor of Engineering, where his primary appointment is with the Department of Electrical Engineering and Computer Science. He has also has appointments, by courtesy, with the Department of Biomedical Engineering and the Department of Statistics. In 2008, he was awarded the Digiteo Chaire d'Excellence, sponsored by Digiteo Research Park, Paris, France, located at the École Supérieure d'Électricité, Gif-sur-Yvette. He has held other visiting positions with LIDS Massachusetts Institute of Technology (MIT) during 2006, Boston University during 2006, I3S University of Nice, Sophia-Antipolis, France, during 2001, the École Normale Supérieure de Lyon, in 1999, the École Nationale Supérieure des Télécommunications, Paris, in 1999, with Lucent Bell Laboratories in 1999, Scientific Research Labs of the Ford Motor Company, Dearborn, MI, in 1993, the École Nationale Superieure des Techniques Avancees (ENSTA), the École Supérieure d'Électricité, Paris, in 1990, and the M.I.T. Lincoln Laboratory, during 1987–1989. His recent research interests have been in detection, classification, pattern analysis, and adaptive sampling for spatio-temporal data. Of particular interest are applications to network security, multimodal sensing and tracking, biomedical imaging, and genomic signal processing.

Dr. Hero has been plenary and keynote speaker at major workshops and conferences. He has received several Best Paper awards including: a IEEE Signal Processing Society Best Paper Award (1998), the Best Original Paper Award from the *Journal of Flow Cytometry* (2008), and the Best Magazine Paper Award from the IEEE Signal Processing Society (2010). He received the IEEE Signal Processing Society Meritorious Service Award (1998), the IEEE Third Millenium Medal (2000), and the IEEE Signal Processing Society Distinguished Lecturership (2002). He was President of the IEEE Signal Processing Society (2006–2007). He is on the Board of Directors of the IEEE (2009–2011) where he is Director Division IX (Signals and Applications).