Covariance Matrix Estimation for Interest-Rate Risk Modeling via Smooth and Monotone Regularization

Dmitry M. Malioutov, *Senior Member, IEEE*, Aycan A. Corum, and Mujdat Çetin, *Member, IEEE*

***Abstract*—Estimating covariance matrices in high-dimensional settings is a challenging problem central to modern finance. The sample covariance matrix is well-known to give poor estimates in high dimensions with insufficient samples, and may cause severe risk underestimates of optimized portfolios in the Markowitz framework. In order to provide useful estimates in this regime, a variety of improved covariance matrix estimates have been developed that exploit additional structure in the data. Popular approaches include low-rank (principal component and factor analysis) models, banded structure, sparse inverse covariances, and parametric models. We investigate a novel nonparametric prior for random vectors which have a spatial ordering: we assume that the covariance is monotone and smooth with respect to this ordering. This applies naturally to problems such as interest-rate risk modeling, where correlations decay for contracts that are further apart in terms of expiration dates. We propose a convex optimization (semi-definite programming) formulation for this estimation problem, and develop efficient algorithms. We apply our framework for risk measurement and forecasting with Eurodollar futures, investigate limited, missing and asynchronous data, and show that it provides valid (positive-definite) covariance estimates more accurate than existing methods.**

***Index Terms*—High-dimensional covariance estimation, Smooth-monotone regularization, Semi-definite programming.**

I. Introduction

MODELING statistical dependence among a collection of random variables is a fundamental problem in statistics, engineering, and finance [1]. Practical problems in finance have grown increasingly high-dimensional, with tens of thousands of domestic and global equities, bonds and futures contracts, and other instruments. This renders the estimation and forecasting problems needed for trading strategies and risk management very challenging. The jointly-Gaussian model, and its extensions to elliptic distributions and Gaussian copulas, use the covariance matrix to describe the strength of interaction between the random variables, and remains a dominant tool in practice. However, estimating and forecasting the covariance matrix is very difficult in high-dimensions when one is faced with limited data [2].

Manuscript received October 15, 2015; accepted January 10, 2016. Date of publication April 20, 2016; date of current version August 12, 2016. This work was presented in part at the Statistical Signal Processing Workshop, 2012. The guest editor coordinating the review of this manuscript and approving it for publication was Prof. Ali N. Akansu.

D. M. Malioutov is with the IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA (e-mail: dmalioutov@us.ibm.com).

A. A. Corum is with the Finance Department, Wharton School, University of Pennsylvania, Philadelphia, PA 19104 USA.

M. Çetin is with the Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul 34956, Turkey.

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/JSTSP.2016.2555285

We consider the problem of estimating large covariance matrices in the context of modeling risk for Markowitz portfolio selection [3]. Finance practitioners [4], [5] and researchers in random matrix theory [6] are painfully aware that using the sample covariance matrix is a disastrous choice when one is modeling large portfolios. For example, when faced with fewer historical samples than the dimension, the sample covariance matrix is rank-deficient which creates an illusion of risk-free linear combinations of financial instruments. More generally, the sample covariance matrix with scarce data produces an inconsistent estimate of the eigenvalue spectrum, and when it is used to create optimized portfolios, the solution tends to prefer those components which have underestimated risk. The end-result tends to be a vast understatement of risk of the Markowitz portfolio [6].

To alleviate this problem, one has to rely on some prior knowledge of structure in the data. One such widely used assumption stipulates that the data lie close to a low-dimensional subspace, which, for covariance estimation translates into principal component analysis (PCA) or factor analysis models [7], [8]. The covariance matrix is assumed to be composed of a low-rank term plus a diagonal noise term, thus reducing the number of parameters from $O(N^2)$ to $O(NK)$, with dimension N and assumed rank K. This approach is popular in fixed income modeling, where the three main factors have the interpretation of level-shift, slope, and curvature changes of the interest rate curve [9]. Another popular assumption on covariance structure is the sparsity of the *information matrix*, i.e., the inverse of the covariance matrix. This is known as *covariance selection* in statistics and as *Gaussian graphical model* or *Gaussian Markov Random Field (MRF)* in machine learning [10]–[12]. The pattern of nonzero elements of the information matrix captures the conditional independence structure, with their number often assumed to be bounded by a small constant K per row, again reducing the total number of parameters to $O(NK)$. *Banded covariance matrices* that allow only a few non-zero diagonals (bands) have been investigated in [13], [14]. Parametric models assume a functional form for the covariance, e.g., exponential or a power-law decay. Gaussian Processes (GP) provide a general framework for such models [15]. Shrinkage estimates [16] take a weighted combination of the sample covariance matrix and a strongly-regularized model (such as low-rank). While they do improve the expected mean-squared-error, they do not introduce any new structure beyond the one inherited from the regularized model. All of the above models have successful domains of applications, but in general they manage to reduce the required number of samples by imposing very strict assumptions on the data.

In this paper we instead allow all the elements of the covariance matrix to be treated as separate parameters, but we require the covariance matrix to be smooth[1] and monotone with respect to some ordering of the variables. An example of such ordering comes from interest-rate risk models and general term-rate models, where the ordering is given by the expiration date (or maturity) of the contract: we expect correlations between pairs of contracts with maturities further apart to be lower. The non-parametric approach, while not directly limiting the number of parameters, reduces the complexity of the space of their joint configurations: this is a regularization approach to covariance estimation. Related approaches have been studied in nonparametric statistics for applications including monotone density and function estimation, spline smoothing, etc. [19]. However, to our knowledge, our work is the first to apply this structure to covariance estimation. We presented initial ideas of smooth-monotone covariance estimation at the IEEE statistical signal processing workshop [20]. We also investigated the use of smooth-covariance estimation for elliptic distributions in [21]. The focus of this paper is on efficient smooth-monotone optimization and eigenvalue spectrum correction. We formulate the covariance smoothing problem as an instance of semidefinite programming (SDP), and describe a fast first-order approach that can be used for large-scale settings: we adapt the dual projected gradient method of [22], [23], develop a dual coordinate-descent extension for the smoothed cost, and accelerate it following the ideas in [24] and [25]. We also describe extensions to problems with missing data and asynchronous measurements, where the sample covariance matrix may be invalid (non positive definite). The well-known approach of [26] used an optimization approach to *fix* a misspecified (non-p.d.) covariance matrix by finding the closest valid one. In our approach we use smooth and monotone regularization which provides valid (p.d.) covariance matrices with significantly improved out-of-sample accuracy. We apply our approach to interest-rate curve risk modeling for the Eurodollar futures curve, and obtain promising estimation and forecasting results.

We start in Section II with an outline of the challenges of high-dimensional covariance estimation, and the distortion of the eigenvalue spectrum. In Section II-A we describe the term-rate modeling problem to be used for our experiments. Our framework is presented in Section III. In Section IV we describe the optimization formulation and propose an algorithm based on optimal first-order methods. We present experimental results with historical data for the Eurodollar futures contracts in Section VI.

II. Preliminaries: High-Dimensional Covariance Estimation

Suppose we have a collection of financial instruments $\{1,\ldots,N\}$ and let $\mathbf{x}(t)\in\mathbb{R}^N$ denote their returns, or linear changes in prices.[2] We would like to estimate the covariance matrix $P^*=E[\mathbf{x}(t)\mathbf{x}(t)^T]$. The sample covariance $\bar{P}\triangleq\frac{1}{T}\sum_i \mathbf{x}(t_i)\mathbf{x}(t_i)^T$ is an unbiased and consistent estimate in the high-sample regime, where the number of samples T far exceeds the dimension N, i.e. $T/N\to\infty$, but with scarce data it has well-documented failures [6]. In particular, the eigenvalue spectrum is biased with T/N held fixed, even as $T\to\infty$. This can be described as spectral blurring, and can be understood through the lens of random matrix theory. Consider a sample covariance matrix obtained from T i.i.d. samples from the multivariate standard normal $\mathcal{N}(0,I)$ in N-dimensions, with zero-mean and an $N\times N$ identity matrix I as the covariance matrix. Let $\rho=N/T$. The true eigenvalues are all 1. The sample eigenvalue spectrum, i.e., the distribution of the eigenvalues of $\bar{P}$, asymptotically follows the Marcenko-Pastur law [27], as illustrated in Figure 1:

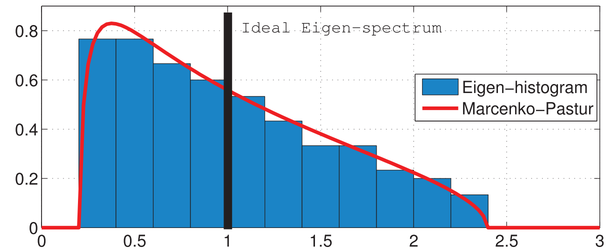

Fig. 1. Marcenko-Pastur law closely approximates the eigen-spectrum of the sample covariance matrix. Samples are from $\mathcal{N}(0,\mathcal{I})$, so eigenvalues of the true covariance matrix are all 1.

$$f_\rho(x)=\frac{1}{2\pi}\frac{\sqrt{(y_+-x)(x-y_-)}}{x},$$

where $y_\pm=(1\pm\sqrt{\rho})^2$. Hence, the smallest eigenvalue of $\bar{P}$ (corresponding to the direction which allegedly has the least risk) is a severe underestimate of its true value 1 for small or moderate T. A similar spectrum "blurring" effect happens for samples from multivariate Gaussian distributions with arbitrary covariance matrices [6]. This poses significant problems when the sample covariance matrix is used for risk-modeling in Markowitz portfolios: the optimized portfolio gets aligned with the most underestimated components of risk, but has less weight in over-estimated ones, causing severe overall risk underestimates [4]–[6]. We will see in Section VI that our smooth-monotone covariance estimate gives a much more accurate eigen-spectrum than the sample covariance, thus dramatically mitigating the problem of bad risk forecasts in optimized portfolios.

A. Term-Structure Risk Modeling

We now briefly describe the interest-rate risk modeling problem that we will use to illustrate our smooth-monotone covariance estimation framework. An interest rate curve describes the available interest rate as a function of the duration for which the investment is locked in. There are a variety of interest-rate

[1] Prior work in [17] used smoothness of covariance functions via local-cosine basis expansions. [18] used smoothness of the covariance function to efficiently approximate variances in large-scale Gaussian MRFs.

[2] This choice depends on whether we use log-normal or normal model for prices (geometric Brownian motion or the Ornstein-Uhlenbeck model), which in turn depends on the instrument class, and holding time (trading frequency).

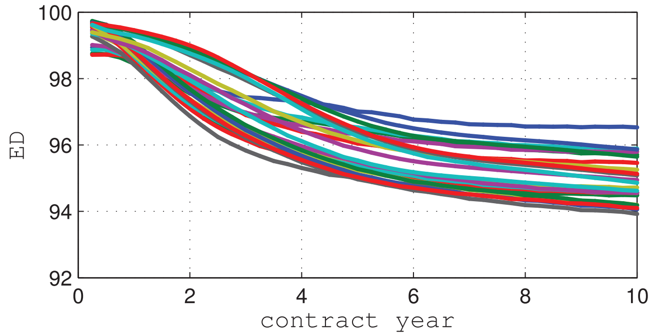

Fig. 2. Sample ED curve settle-prices for several days, linearly interpolated.

curves: US treasury curve, curves for bonds issued by other sovereign nations, municipal bonds, corporate bonds at different credit-ratings, swap-curves, inter-bank lending curves, etc.[3] [28]. For simplicity we consider a generic interest-rate curve and will use the example of Eurodollar futures throughout the paper. The curve changes with time and takes on a variety of different shapes. We expect the correlations between variables in the curve to be monotonic with respect to the difference in the expiration dates of the contracts, and also expect not to have persistent discontinuities in the correlation structure – thus fitting well with monotone and smooth assumptions for the framework in this paper.

In Figure 2 we illustrate interest-rate curves based on the Eurodollar (ED) futures contracts, with 40 quarterly expirations, i.e., up to 10 years ahead. For historical reasons ED contracts are priced as $100 - x$ where x is the interest rate. We plot the curves for a few different dates (the curve of available rates will change from day to day) with linear interpolation in between the contracts[4]. A popular model for term-rate curves is based on principal component analysis (PCA), and approximates the covariance by three main principal components, having informal interpretation of level, slope, and curvature [9]. However, for some applications such as statistical arbitrage and portfolios with high leverage, a more accurate covariance matrix estimate may be desired. Simply increasing the number of principal components typically does not produce good results, as higher-order principal components tend to be much less stable. Next, we formally describe the smooth-monotone covariance estimation framework, and study its performance for modeling risk of the Eurodollar curve movements in Section VI.

III. Smooth Isotonic Regression

We now introduce our setting for covariance matrix estimation. Our starting assumption is that the random variables of interest have an explicit ordering – in our example for interest rate risk the ordering will be by contract expiration[5]. We aim to estimate the spatial (cross-sectional over contracts) covariance matrix for these random variables from a scarce number of samples. We consider a non-parametric approach which stipulates that the desired correlation structure "respects" the variable ordering – namely – the covariance matrix is well-behaved with the distance between the variables – it is monotonic and smooth. Both of these are natural assumptions when dealing with spatial data. For our interest-rate example: if the expiration date of the i-th contract is closer to that of the j-th contract than to the k-th one, then we expect the correlation $P_{i,j}$ to be higher than $P_{i,k}$ (e.g., we expect the January and February contracts to be more correlated than January and October of the same year). Also, there is rarely any economic reason to expect discontinuities in the correlation structure of contracts having expiration dates many months or years in the future, thus justifying our second assumption of smoothness.

We now formulate the regularization problem for smooth isotonic covariances with a linear (one-dimensional) ordering of random variables. Suppose we have a zero-mean random vector $\mathbf{x}(t)$, where $\mathbf{x}(t) = (x_1(t), \ldots, x_N(t))^T$. We are interested in the spatial covariance matrix of $\mathbf{x}$, $P^* = E[\mathbf{x}\mathbf{x}^T]$, and also in the matrix of the correlation coefficients, $C^*_{ij} \propto \frac{P^*_{ij}}{\sqrt{P^*_{ii} P^*_{jj}}}$. Suppose that only a small number of samples $\mathbf{x}(t_1), \ldots, \mathbf{x}(t_T)$ is available with T comparable or even smaller than N. We do not model temporal dynamics in this paper, so we assume that the samples are i.i.d. We aim to leverage the assumptions of monotonicity and smoothness to get a better estimate of P^* than the ordinary sample covariance matrix $\bar{P}$. Let $\mathcal{M}$ be the class of monotone positive-definite (p.s.d.) covariance matrices:

$$\mathcal{M} = \{P | P \succeq 0, P_{ij} \geq P_{ik} \text{ for } i < j < k\}. \tag{1}$$

Then, we can obtain an improved estimate of the covariance by finding the monotone covariance matrix in class $\mathcal{M}$ that is closest to the sample covariance matrix $\bar{P}$:

$$\min_P D(P, \bar{P}) \text{ such that } P \in \mathcal{M} \tag{2}$$

where $D(P, \bar{P})$ is a convex error metric of our choice: we will use the Frobenius norm, but KL-divergence and the operator norm are also applicable. Note that the constraint set $\mathcal{M}$ is a convex set, with linear and positive definite constraints, and for natural choices of the metric D the objective will also be convex. When D is either the operator norm or the Frobenius norm, our regularizer can be found as a solution to a semi-definite programming problem (SDP) [29].

Remark: If the true covariance P^* indeed belongs to $\mathcal{M}$, then projecting the sample covariance onto $\mathcal{M}$ is guaranteed to decrease the error[6] due to the contraction property of projections onto convex sets: $\|\Pi_{\mathcal{M}} \bar{P} - \Pi_{\mathcal{M}} P^*\| \leq \|\bar{P} - P^*\|$. Here, we use $\Pi_{\mathcal{M}} \bar{P}$ to denote the projection of $\bar{P}$ onto the set $\mathcal{M}$, i.e., the solution $\arg\min_{\mathcal{M}} D(P, \bar{P})$ of (2).

We preview a computational example that we describe in detail in Section VI. In Figure 3 we plot (a) a ground-truth smooth-monotone covariance, and (b) the sample estimate (based on samples drawn from the ground-truth) exhibiting finite-sample noise. In plot (c) we solve (2) and see that

[3]Also it is common to look at several views of the interest-rate curve: yield-curve vs. spot curve vs. the forward curve. Similar modeling techniques also apply to futures curves.

[4]Data used with permission from the Wall Street Journal online.

[5]We note that the approach does not directly apply to the equity space, as there is no natural way to order equities.

[6]If the projection is done with respect to the same metric with which we evaluate errors, and we use either the Frobenius or the spectral norm.

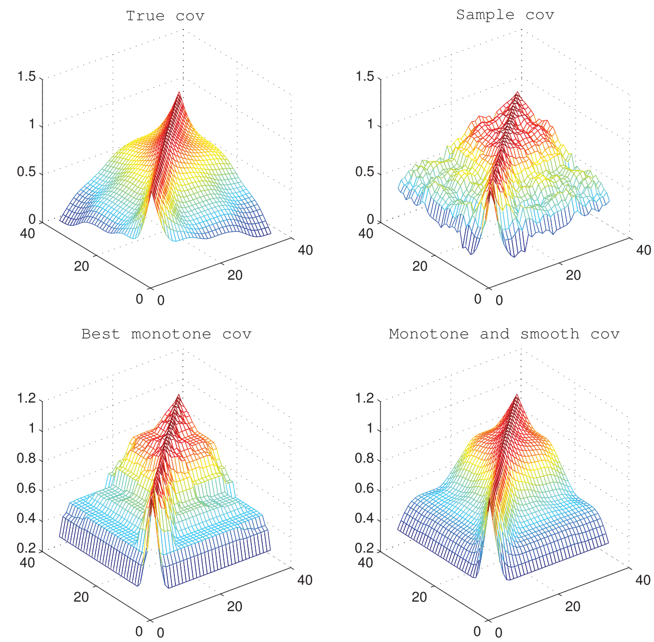

Fig. 3. Term-rate covariances: (a) true and (b) sample estimate. Covariance regularization: (c) monotone (d) monotone and smooth.

simply restricting the covariance matrices to be monotone produces a significantly improved covariance estimate. However it appears to suffer from a "staircase"-like piecewise-continuous effect. We do not expect natural phenomena to exhibit such discontinuous behavior, and so we further require that the covariance matrices have some degree of smoothness. To that end we penalize the curvature over the surface of the covariance function, namely:

$$S(P) = \int\int_U (\nabla^2 P(x_1, x_2))^2 dx_1 dx_2 \tag{3}$$

where $U = \{(x_1, x_2)|x_2 > x_1\}$ is the upper-triangular part of the covariance function. This means that we encourage smoothness in the upper-triangular part (and by symmetry in the lower-triangular part) but we do not impose the smoothness constraint over the diagonal entries, which may include idiosyncratic effects such as fluctuations in liquidity and bid-ask spreads, etc. To implement this numerically, over a discrete grid that corresponds to available expiration days, we use the discrete version of the Laplacian operator on the grid at the point of interest v, square and sum over all v:

$$S(P) = \sum_v \left(\nabla_v^2 P\right)^2, \text{ where } \nabla_v^2 P = \sum_{u \in N(v)} (P(u) - P(v)) \tag{4}$$

Here, $N(v)$ is the set of neighbors of point v: for the vertex $v = (i, j)$ that corresponds to the entry P_{ij} of the covariance matrix, the neighbors can be taken as the 4 neighboring entries, i.e., $(i \pm 1, j)$, and $(i, j \pm 1)$. The optimization problem is now:

$$\min_P D(P, \bar{P}) + \lambda \sum_v (\nabla_v^2 P)^2, \quad \text{such that } P \in \mathcal{M} \tag{5}$$

where the parameter λ trades off smoothness with data-fidelity, and should ideally be chosen automatically, e.g., via cross-validation. The problem is still convex: the objective is convex quadratic, and the constraint set is semi-definite, keeping the problem an SDP. To see the benefit of enforcing smoothness we contrast covariance estimates in Figure 3(c) and (d) and we see that (5) produces much smoother, and, as we will see in Section VI, more accurate estimates.

IV. Numerical Solution

The optimization problem in (5) is not only convex, it also can be represented as a semidefinite optimization problem [29]:

$$\min \|P - \bar{P}\|_f^2 + \lambda \|D_2 Vec(P)\|_2^2, \quad \text{such that } P \succeq 0, D\, Vec(P) \geq 0 \tag{6}$$

where the operation $Vec(P)$ denotes stacking the columns of P into a vector, $\|P\|_f^2 = \sum_{i,j} P_{ij}^2$ denotes the squared Frobenius norm, and $\|\mathbf{x}\|_2^2 = \sum_i x_i^2$ is the squared ℓ_2-norm. The matrices D_2 and D compute differences of relevant entries of P encoding smoothness and monotonicity constraints, respectively. The resulting problem for small N can be readily solved via an interior point method using one of the standard SDP optimization packages, e.g., SDPT3 [30]. Note that it is straightforward to add additional constraints, e.g., positivity of correlations, or $P_{ii} = 1$ to deal with correlation coefficient matrices.

Solving SDP via an interior point method can become unduly computationally expensive for large covariance matrices, as it involves computing the Hessian matrix. Alternatively, the problem can be solved via optimal first-order methods, an exciting recent development in optimization, generalizing classical gradient projection by a clever use of smoothing and acceleration techniques [24], [31], [32]. An important requirement to use such methods is that the projection onto the constraint set can be done efficiently. This can be achieved by considering the dual of the problem in (6) where these projections correspond to singular value thresholding. In Section V we first describe a dual first-order method for our monotone problem based on gradient projection [22], and then develop a faster optimal first-order version based on acceleration ideas of [24]. The monotone and smooth version of the problem follows the same lines, and a dual projected gradient descent solution has been described in [33].

V. Efficient Solution for Large-Scale Covariances

We now describe fast first-order optimization algorithms for the smooth and monotone covariance regularization that do not require computing the Hessian matrix.

A. *Projected Gradient Solution*

We adapt the approach of [22] and outline a dual projected gradient solution for the monotone version of our problem in (2). The primal version of the monotone problem can be written as:

$$\min \frac{1}{2}\|P - \bar{P}\|_f^2 \text{ such that } P \succeq 0, \quad (7)$$
$$\text{and } \mathrm{Tr}(D_j P) \geq 0, \quad j = 1, .., M.$$

Here $\mathrm{Tr}(D_j P)$ represents the j-th row of the constraint $D\ Vec(P)$ in (6), and M is the number of these linear constraints. Introducing dual variables μ for the inequality constraints, and Z for the p.s.d. constraint, the dual problem is:

$$\max -\frac{1}{2}\| - D(\mu) + Z + \bar{P}\|_f^2, \quad (8)$$
$$\text{such that } Z \succeq 0, \quad \mu \geq 0,$$

where $D(\mu) = -\sum \mu_j D_j$. Strict duality holds for this problem as a strictly feasible solution exists with $P \succ 0$. Given μ^*, the primal solution can then be found as $P = (\bar{P} - D(\mu^*))_+$. A closed form solution for optimization over Z simplifies the dual further:

$$\max -\frac{1}{2}\|(-D(\mu) + \bar{P})_+\|_f^2, \quad \mu \geq 0, \quad (9)$$

where $(W)_+$ is the positive definite part of the matrix W. As described in [22], the gradients of this objective can be evaluated:

$$\frac{1}{2}\nabla\|(W)_+\|_f^2 = W_+. \quad (10)$$

A dual projected gradient algorithm for the problem alternates steps along the gradient direction with projections onto the constraint set (initially we set μ to 0, and choose a step-size α):

i. Compute $D(\mu) = -\sum \mu_j D_j$, and set $P = (\bar{P} - D(\mu))_+$.
ii. Compute gradients: $\frac{\partial f}{\partial \mu_i} = \mathrm{Tr}(D_j P)$, and set $\mu_j \rightarrow (\mu_j + \alpha(\mathrm{Tr}(D_j P)))_+$.

The smooth and monotone problem can be solved in a similar fashion, except that the step corresponding to (8) does not allow a closed form solution over Z. Instead we use a projected coordinate descent algorithm, alternating descent over Z and descent over μ.

B. Optimal First Order Methods

The projected gradient algorithm avoids computing the Hessian, but it is plagued by slow convergence, with error decreasing as $O(1/k)$, where k is the iteration number. Nesterov [31] has shown that it is possible to obtain $O(1/k^2)$ convergence for a multi-step first-order method by a careful combination of the current and previous gradients. An extension of Nesterov's method to projected gradients was developed in [24], called FISTA. Convergence of the FISTA algorithm is guaranteed with step size $\alpha = \frac{1}{L}$, where L is the Lipschitz constant for ∇f (see [22]). Applying FISTA to the dual of our objective, we obtain the following algorithm:

Init: Set $\mu = 0$, $P = \bar{P}$. Compute the Lipschitz constant L. Iterate:

i. Let $\eta_j \rightarrow (\mu_j + \frac{1}{L}(\mathrm{Tr}(D_j P)))_+$, compute $D(\eta)$ and set $P \rightarrow (\bar{P} - D(\eta))_+$.
ii. Let $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$ and let $\mu_{k+1} = \eta_k + \frac{t_k - 1}{t_{k+1}}(\eta_k - \eta_{k-1})$.

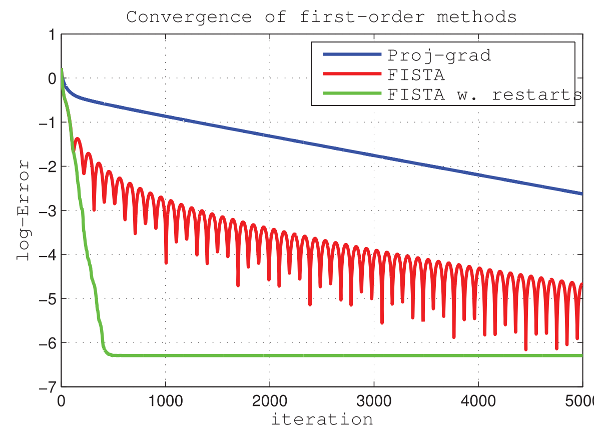

Fig. 4. Convergence speed (in terms of log-error from the interior-point solution) for projected gradient, FISTA, and FISTA with momentum restarts. Momentum restarts dramatically improve convergence of FISTA and rectify the problematic oscillations.

The main complexity per iteration is evaluating the singular value decomposition (SVD) in $(\bar{P} - D(\eta))_+$. In comparison with interior point methods, our dual FISTA implementation is on average 30 times faster than SDPT3 [30] with $1e^{-3}$ tolerance, $N = 40$. We obtain similar speedups for the accelerated dual coordinate descent method for the smooth and monotone problem, which was mentioned in Section V-A. However, FISTA has to be modified to reach higher accuracy solutions. As many researchers noticed, plain FISTA suffers from an oscillation phenomenon in practice, and a heuristic using momentum restarts can provide a dramatic improvement as we describe next.

C. Adaptive Restarts for Optimal First Order Methods

The development of FISTA was met with a lot of excitement, but in practice the algorithm does not achieve the promised fast convergence rate due to oscillations where the momentum term grows too large and the algorithm overshoots. An analysis of these oscillations [25] had shown them to be due to the difficulty of estimating the local Lipschitz constants and the local strong-convexity parameters of the smooth part of the objective function. The authors also proposed a heuristic approach to restart the momentum term in FISTA (to erase the memory of past gradients) when an early onset of the overshooting phenomenon is detected. We use the following simple rule from [25] to restart the momentum term in the algorithm in Section V-B. We use the update $\mu_{k+1} = \eta_k + \frac{t_k - 1}{t_{k+1}}(\eta_k - \eta_{k-1})$ provided that $(\mu_{k-1} - \eta_{k-1})'(\eta_k - \eta_{k-1}) < 0$, and otherwise we set $\mu_{k+1} = \eta_k$ to restart the momentum. In Figure 4 we show that adding these momentum restarts provides a dramatic convergence speed improvement compared to plain FISTA, and removes its oscillations due to momentum overshooting. We reach the accuracy of the interior point methods after a few hundred iterations as opposed to plain projected gradients and FISTA which do not reach this accuracy even after 5000 iterations.

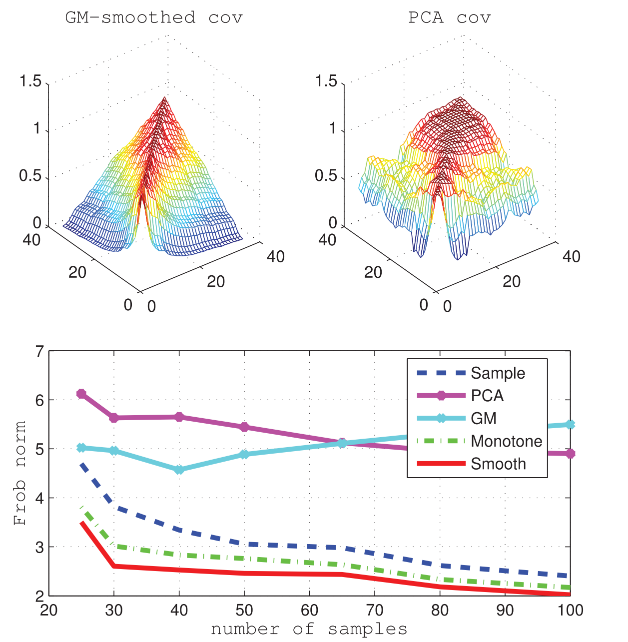

Fig. 5. (a) MRF estimate (b) PCA estimate. (c) Errors of sample covariance, monotone, smooth, GM and PCA estimates.

VI. Experimental Results

We now apply our smooth and monotone regularization to interest rate risk modeling. We focus on the Eurodollar (ED) curve, and specifically look at ED spreads: if $y_i(t)$ is the price of the i-th ED contract at time t, then the i-th spread is $x_i(t) = y_i(t) - y_{i+1}(t)$. Focusing on spreads is akin to immunizing the portfolio to parallel shifts of the ED-curve, and it makes covariance estimation more challenging by removing the dominant first principal component. The risk for a portfolio of ED spreads is defined via the covariances for daily changes in prices of the spreads.

In our first experiment we generate a number of samples from a ground-truth known[7] smooth monotone covariance P^*, and use them to generate a covariance estimate $\hat{P}$. In Figure 3 we show (a) the true covariance, and (b) the sample estimate, for the case of $N = 40$, $T = 40$. We apply (c) our monotone and (d) smooth-monotone regularization[8] from (2) and (5), respectively. We can see that while the monotone version suffers from the staircase effect, the smooth version looks qualitatively close to the true covariance.

For comparison we compute the PCA estimate with $K = 3$ principal components, and a Markov random field (MRF) model estimate where the information matrix is restricted to have $K = 5$ non-zero diagonals[9]. Figure 5(a), (b) shows that the estimated covariance matrices exhibit only a rough similarity to the original. We compute the average Frobenius error over 25 trials and present the results in Figure 5(c) for all the methods, as a function of the number of available samples. The Gaussian MRF and the PCA methods are not consistent for fixed K: the estimates do not improve with more samples. However, the monotone and the smooth-monotone estimates provide a significant improvement in accuracy over the sample covariance. We present additional evidence on out-of-sample correlation structure forecasting experiments in Section VI-C.

[7]In practice one never has the 'true' covariance – here we took a sample covariance matrix from ED data, and applied smoothing to it, as a proxy for the true one.

[8]For simplicity λ was set by trial and error and fixed for all experiments.

[9]The MRF is learned by maximum likelihood via iterative proportional fitting (IPF) optimization [34].

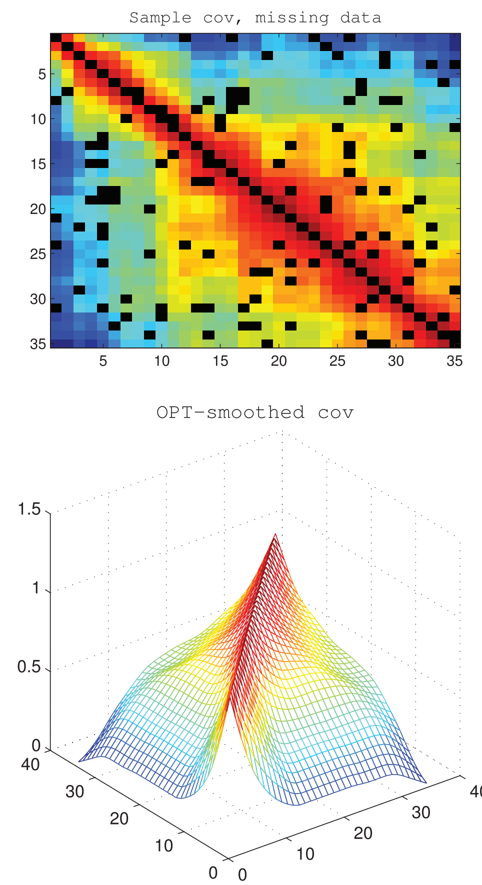

Fig. 6. (a) Sample covariance with missing data. (b) Recovered smooth-monotone covariance.

A. Missing Data

Missing data plagues all of applied science and also has many incarnations in finance, such as illiquid instruments, mismatches in expirations and holiday schedules for foreign markets. We consider an example where some entries in the sample covariance matrix are missing (unknown). Suppose that we have $\bar{P}$ for only some subset $\mathcal{I}$ of entries: $(i, j) \in \mathcal{I} \subset \{1, .., N\}^2$, and no observations for the rest. Our smooth-monotone regularization formulation can be immediately extended to this setting:

$$\min_P D_{\mathcal{I}}(P, \bar{P}) + \lambda \sum_v (\nabla_v^2(P))^2, \quad \text{such that } P \in \mathcal{M}, \tag{11}$$

where, e.g., in the Frobenius case, we define $D_{\mathcal{I}}(P, \bar{P}) = \sum_{(i,j)\in\mathcal{I}} (P_{i,j} - \bar{P}_{i,j})^2$. This does not affect the convexity of the problem, and can be solved using the same optimization

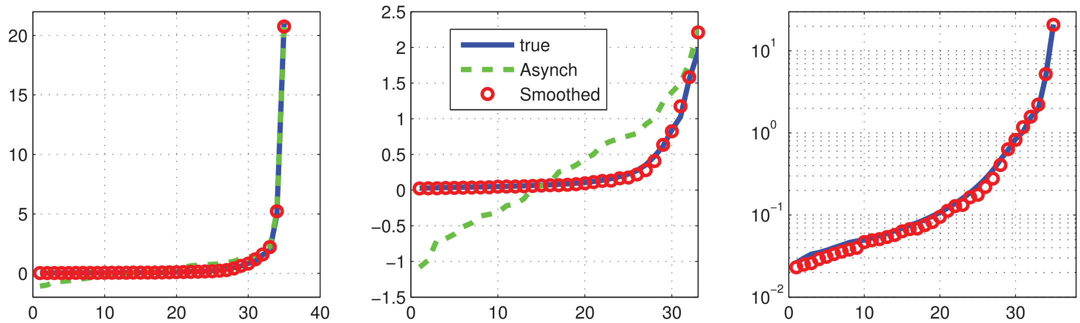

Fig. 7. (a) True, sample, and smooth-monotone eigen-spectra. (b) Detail (c) log-scale of true and smooth-monotone eigen spectra.

methods. We continue our interest-rate curve example, with $N = 40$, and $T = 50$ samples, and 10-percent of the entries missing in Figure 6. The approach is very robust against moderate missing data and recovers a practically indistinguishable estimate compared to the fully observed case.

B. Spectral Correction

As we mentioned, one of the symptoms of the catastrophic breakdown of the sample covariance estimate in the high-dimensional setting with limited data is the inconsistency of the eigenvalue spectrum. A particularly important and challenging case for correcting the eigenvalue spectrum is the *asynchronous setting* where different entries of the covariance matrix are obtained from distinct samples, for example in intraday high-frequency trading. Estimating volatility and correlations from high-frequency data is an active field of research [35], [36]. When several time series occur at different temporal resolutions, it is easiest to consider each pair of time series, align them[10], and compute the pairwise covariance. However, once this is done for each pair, the covariance matrix is no longer guaranteed to be positive definite. An existing solution to fix this defect projects the covariance matrix onto the space of p.d. matrices [26]:

$$\min D(P, \bar{P}) \text{ such that } P \succ 0 \tag{12}$$

A closed form solution based on the eigen-decomposition applies with the Frobenius norm error $D(P, \bar{P}) = \|P - \hat{P}\|_f^2$. Compute the eigen-value decomposition $\bar{P} = U\Lambda U^T$, where U is orthogonal, and Λ is diagonal. Then the solution to (12) simply sets the negative eigenvalues to zero, $\hat{P} = U \max(\Lambda, 0)U^T$, i.e., $\bar{P}$ is projected onto the boundary of the p.s.d. cone. In contrast, our approach uses the side-information of smoothness and monotonicity, and guides the solution into the interior of the p.s.d. cone and closer to the correct solution.

We consider a numerical example with $N = 36$, and $T = 40$ asynchronous samples: each $\bar{P}_{ij}$ is estimated from a pairwise sample $\{x_i(t), x_j(t)\}_{t \in 1,..,T}$, drawn independently of other pairs. We normalize $\bar{P}$ to be unit-diagonal, and use our approach in (5) with the asynchronous covariance estimate $\bar{P}$. In Figure 7 we plot eigenvalues of the (i) true covariance matrix, (ii) asynchronous sample covariance (iii) smooth-monotone fit to the sample covariance. The left plot shows the spectra, with the detail shown in the middle plot. The sample-covariance spectrum breaks down completely, with about half of the eigenvalues negative. Projection onto the p.s.d. cone would simply set the negative eigenvalues to zero, leaving the positive mis-estimated eigenvalues intact. However, the smooth-monotone eigenvalue spectrum follows the true one closely. A log-plot of the true and smooth-monotone spectra appears in the right-most plot, and we see that indeed our proposed approach matches the spectrum very closely! This experiment suggests that smooth-monotone regularization can be very effective in spectral correction for covariance estimation, and is especially valuable for asynchronous settings.

[10]For example one can re-sample the time series to include time points from both. This would be very costly for much more than two series together.

C. Out-of-Sample Covariance Prediction

We now present a study of forecasting future correlation coefficient matrices over several years of historical data of ED prices. The accuracy of this prediction is crucial for portfolio selection methods, such as Markowitz portfolios, to optimally allocate assets. We estimate the correlation coefficient matrix using our proposed method as well as alternative methods over a training window of T_{TR} business days. For the purposes of this paper we use a very simple forecast assuming that the correlation structure is slowly changing – by extending the correlation estimate on the training window to the test window and not modeling dynamics. We compute the realized matrix of correlation coefficients over a test window of T_{TEST} business days immediately following the training window, and compare it to our forecasts. We use running windows with shifts of 5 business days over the course of five years ending in December 2010.

In Figure 8(a) we show the forecast error (in Frobenius norm) for the sample correlation coefficient matrix and our smooth-monotone estimate over the course of five years. In plot (b) we show the forecast error as a function of the running window size T_{TR} with T_{TEST} set to 50. We observe that P_{SM} produces significantly smaller errors than both the sample correlation coefficient matrix estimate and the PCA-based estimate. Smooth and monotone regularization appears especially valuable for small T_{TR}, demonstrating robustness in forecasting risk in scenarios with severely limited data. We aim to extend this simple forecast to incorporate dynamics for both variances, through multivariate variations of GARCH [37], [38], and for correlations [2].

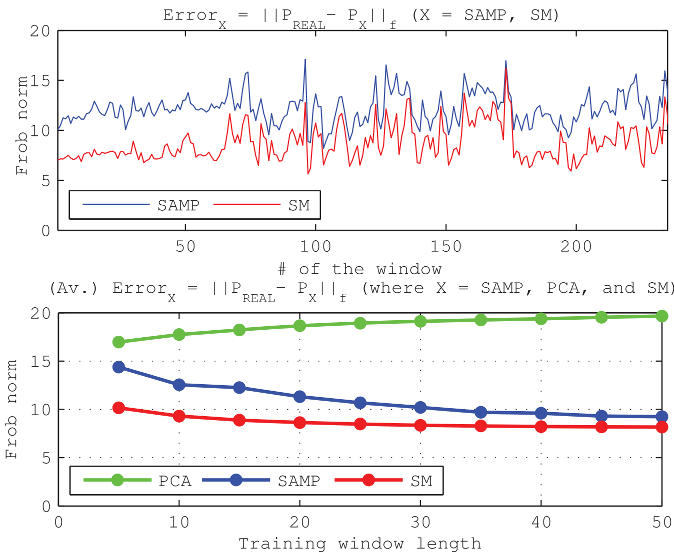

Fig. 8. (a) Absolute Frobenius error over running windows. (b) Average Frobenius error vs. the training window length.

VII. Conclusion

We have described a simple and effective framework to estimate covariance matrices satisfying monotonicity and smoothness, with applications to interest-rate-curve risk modeling in econometrics. We formulated the problem as an SDP, and described a dual projected gradient and an optimal first-order method for its solution. We applied our approach to examples with limited, missing, and even asynchronous data, and showed a significant performance improvement over existing methods. This is valuable in constructing and managing risk with optimized portfolios of interest-rate products. The framework can be directly extended from simple linear orderings to two-dimensional grids, for example to model volatility surfaces for equity options. In future work we aim to analyze convergence rates, and study temporally-correlated and time-varying settings, along the lines of ideas in [2], [37], [38].

Acknowledgment

We thank Peter Christofferson and Rahul Mazumder for valuable discussions.

References

[1] P. Embrechts, R. Frey, and A. McNeil, *Quantitative Risk Management*, Princeton Univ. Press, Princeton, NJ, USA, 2005.

[2] T. G. Andersen, T. Bollerslev, P. F. Christoffersen, and F. X. Diebold, "Volatility and correlation forecasting," in *Handbook of Economic Forecasting*, 2006, vol. 1, pp. 777–878.

[3] H. Markowitz, "Portfolio selection," *J. Finance*, vol. 7, no. 1, pp. 77–791, 1952.

[4] V. Chopra and W. T. Ziemba, "The effects of errors in means, variances, and covariances on optimal portfolio choice," *J. Portfolio Manage.*, vol. 19, no. 2, pp. 6–11, 1993.

[5] S. Ceria and R. A. Stubbs, "Incorporating estimation errors into portfolio selection: Robust portfolio construction," *J. Asset Manage.*, vol. 7, no. 2, pp. 109–127, 2006.

[6] N. El Karoui, "Spectrum estimation for large dimensional covariance matrices using random matrix theory," *Ann. Stat.*, vol. 36, no. 6, pp. 2757–2790, 2008.

[7] I. T. Jolliffe, *Principal Component Analysis*. New York, NY, USA: Springer, 2002.

[8] J. Fan, Y. Fan, and J. Lv, "High dimensional covariance matrix estimation using a factor model," *J. Econometrics*, vol. 147, pp. 186–197, 2008.

[9] F. X. Diebold and C. Li, "Forecasting the term structure of government bond yields," *J. Econometrics*, vol. 130, no. 2, pp. 337–364, 2006.

[10] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton, FL, USA: CRC, 2005.

[11] A. P. Dempster, "Covariance selection," *Biometrics*, vol. 28, pp. 157–175, Mar. 1972.

[12] J. Fan, J. Lv, and L. Qi, "Sparse high dimensional models in economics," *Ann. Rev. Econ.*, vol. 3, pp. 291–317, 2011.

[13] P. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *Ann. Stat.*, vol. 36, no. 1, pp. 199–227, 2008.

[14] T. T. Cai and H. H. Zhou, "Minimax estimation of large covariance matrices under l1-norm," *Stat. Sin.*, vol. 22, pp. 1319–1378, 2012.

[15] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.

[16] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivariate Anal.*, vol. 88, no. 2, pp. 365–411, Feb. 2004.

[17] S. Mallat, G. Papanicolaou, and Z. Zhang, "Adaptive covariance estimation of locally stationary processes," *Ann. Stat.*, vol. 26, no. 1, 1998.

[18] D. M. Malioutov, J. Johnson, M. J. Choi, and A. S. Willsky, "Low-rank variance approximation in GMRF models: Single and multi-scale approaches," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4621–4634, Oct. 2008.

[19] T. Robertson, F. T. Wright, and R. L. Dykstra, *Order Restricted Statistical Inference*. Hoboken, NJ, USA: Wiley, 1998.

[20] D. M. Malioutov, "Smooth isotonic covariances," in *Proc. IEEE Stat. Signal Process. Workshop*, 2011, pp. 29–32.

[21] X. Zhou, D. Malioutov, F. Fabozzi, and S. Rachev, "Smooth monotone covariance for elliptical distributions and applications in finance," *Quant. Finance*, vol. 14, no. 9, pp. 1555–1571, 2014.

[22] S. Boyd and L. Xiao, "Least-squares covariance matrix adjustment," *SIAM J. Matrix Anal. Appl.*, vol. 27, no. 2, pp. 532–546, 2005.

[23] J. Malick, "A dual approach to semidefinite least-squares problems," *SIAM J. Matrix Anal. Appl.*, vol. 26, pp. 272–284, 2005.

[24] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[25] B. O'Donoghue and E. Candes, "Adaptive restart for accelerated gradient schemes," *Found. Comput. Math.*, vol. 15, no. 3, pp. 715–732, 2013.

[26] N. Higham, "Computing the nearest correlation matrix—A problem from finance," *IMA J. Numer. Anal.*, vol. 22, no. 3, pp. 329–343, 2002.

[27] V. A. Marčenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Sbornik Math.*, vol. 1, no. 4, pp. 457–483, 1967.

[28] D. Brigo and F. Mercurio, *Interest Rate Models-Theory and Practice: With Smile, Inflation and Credit*. New York, NY, USA: Springer, 2007.

[29] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. Philadelphia, PA, USA: SIAM, 2001.

[30] R. Tutuncu, K. Toh, and M. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3," *Math. Program. B*, vol. 95, pp. 189–217, 2003.

[31] Y. E. Nesterov, "A method of solving a convex programming problem with convergence rate of $O\left(1/k^2\right)$," *Sov. Math. Dokl.*, vol. 27, pp. 372–376, 1983.

[32] V. Cevher, S. Becker, and M. Schmidt, "Convex optimization for big data: Scalable, randomized, and parallel algorithms for big data analytics," *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 32–43, Sep. 2014.

[33] A. A. Corum, "Fast algorithms for smooth and monotone covariance matrix estimation," M.S. thesis, Elect. Eng., Sabanci Univ., Istanbul, Turkey, 2012.

[34] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA, USA: MIT Press, 2009.

[35] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys, "Modeling and forecasting realized volatility," *Econometrica*, vol. 71, no. 2, pp. 579–625, 2003.

[36] Y. Ait-Sahalia, J. Fan, and D. Xiu, "High-frequency covariance estimates with noisy and asynchronous financial data," *J. Amer. Stat. Assoc.*, vol. 105, no. 492, pp. 1504–1517, 2010.

[37] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *J. Econometrics*, vol. 31, no. 3, pp. 307–327, 1986.

[38] R. Engle, "Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models," *J. Bus. Econ. Statist.*, vol. 20, no. 3, pp. 339–350, 2002.

Dmitry M. Malioutov (S'02–M'09–SM'15) received the S.M. and Ph.D. degrees in electrical engineering and computer science from Massachusetts Institute of Technology, Cambridge, MA, USA (where he was part of the Laboratory for Information and Decision Systems). He is a Research Staff Member of the Cognitive Algorithms Department, IBM Research. Prior to joining IBM, he spent several years as an Applied Researcher in high-frequency trading in DRW Trading, Chicago, and as a Postdoctoral Researcher with Microsoft Research, Cambridge, U.K. His research interests include interpretable machine learning, sparse signal representation, inference and learning in graphical models, message passing algorithms, Statistical risk modeling, robust covariance estimation, and portfolio optimization. He serves on the IEEE-SPS machine learning for signal processing technical committee, and is an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and a Guest Editor of the IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING. He was the recipient of the 2010 IEEE Signal Processing Society Best 5-Year Paper Award, the 2006 IEEE ICASSP Student Paper Award, and the MIT Presidential Fellowship.

Aycan A. Corum received the bachelor's and master of science degrees in electronics engineering and from Sabanci University, Istanbul, Turkey in 2012 and 2010, respectively. He is currently pursuing the Ph.D. degree at the Finance Department, Wharton School of Business, University of Pennsylvania, Philadelphia, PA, USA. He was the recipient of University of Pennsylvania Dean's Fellowship for Distinguished Merit.

Müjdat Çetin (S'98–M'02) received the Ph.D. degree from Boston University, Boston, MA, USA, in 2001. He has held visiting faculty positions at Massachusetts Institute of Technology (MIT), Cambridge. MA, USA, Northeastern University, and Boston University. He has been a Faculty Member of Sabancı University, Istanbul, Turkey, since 2005. From 2001 to 2005, he was with the Laboratory for Information and Decision Systems, MIT. His research interests include statistical signal processing, particular focus on computational imaging, sparse signal and image representation, brain-computer interfaces, biomedical image analysis, and image segmentation. He is currently an Associate Editor for the IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING. He has been an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE SIGNAL PROCESSING LETTERS, and IEEE TRANSACTIONS ON CYBERNETICS, a Guest Editor for *Pattern Recognition Letters*, and an Area Editor for the *Journal of Advances in Information Fusion*. He served as the Technical Program Co-Chair for the IEEE Image, Video, and Multidimensional Signal Processing (IVMSP) Workshop in 2016, the International Conference on Information Fusion (FUSION) in 2013 and 2016, the International Conference on Pattern Recognition (ICPR) in 2010, and the IEEE Turkish Conference on Signal Processing, Communications, and their Applications in 2006. He was one of the keynote speakers for the 2015 International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar, and Remote Sensing. He was the recipient of the 2010 IEEE Signal Processing Society Best Paper Award, the 2007 EURASIP/Elsevier Signal Processing Best Paper Award, the 2013 IET Radar, Sonar and Navigation Premium Best Paper Award, and the 2008 Turkish Academy of Sciences Distinguished Young Scientist Award.