A Robust Nonlinear Kalman Smoothing Approach for Dynamic Matrix Factorization

Aleksandr Y. Aravkin, Kush R. Varshney, and Dmitry M. Malioutov
IBM Thomas J. Watson Research Center, 1101 Kitchawan Road, Yorktown Heights, New York 10598
Email: {saravkin,dmalioutov,krvarshn}@us.ibm.com

I. Introduction

The modeling of data as the product of two lower-dimensional (often matrix-valued) latent variables or factors has been widely used in applications as varied as psychometrics, finance, recommender systems, DNA microarray analysis, and foreground/background video separation. In finance, the factors often correspond to market segments and market styles; in recommender systems, the factors correspond to users and items. In many cases, the data matrices under investigation are dynamic, i.e., are functions of time.

Matrix factorization methods have a long history in statistics and signal processing, but most methods deal with the case where either both factors are not time-varying or one of the factors is not [1]–[3]. In contrast, we formulate an approach for factoring dynamic matrices into *two* time-varying low-dimensional matrices [4], [5]. The proposed method uses recent developments in robust nonlinear Kalman smoothing [6].

II. Formulation

Let a (possibly partially-observed) matrix $Z \in \mathbb{R}^{S\times T}$ be approximately factored into latent matrices $A \in \mathbb{R}^{S\times K}$ and $F \in \mathbb{R}^{T\times K}$ with K small. Moreover, let that matrix and consequently its factors be time-varying: $Z_k \approx A_k F_k^T$, $k = 1,\dots,N$. We would like to infer A_k and F_k from Z_k.

We pose this problem through a state space formulation to bring a broad set of dynamic inference tools to bear [6]. In particular, we can take both factors stacked together as the state variable X_k with some forward model g_k and some process noise W_k:

$$X_k = \begin{bmatrix} A_k; & F_k \end{bmatrix}; \quad X_k = g_k(X_{k-1}) + W_k. \tag{1}$$

The forward model could simply be the identity operation, but we include g_k for generality. The key part of the proposed state space model and a source of potential difficulty in optimization is the *nonlinear* measurement model:

$$Z_k = h_k(X_k) + V_k; \quad h_k(X_k) = \mathcal{L}(A_k F_k^T), \tag{2}$$

where a linear measurement operator $\mathcal{L}$, which can be a mask to capture partial observation, is composed with the outer product of A_k and F_k; there is also additive measurement noise V_k.

The distributions of the noises V_k and W_k are subject to design. With desiderata of tracking sudden changes unexplained by the forward model and obtaining good results in the face of artifacts in the observed data, we would often like to choose these distributions to be non-Gaussian. The Laplacian distribution and Student's t-distribution are robust alternatives.

Inference under the proposed model is a Kalman smoothing operation that can be approached by an optimization formulation [6]. The optimization problem is:

$$\min_X \rho_W\left(g(X) - \begin{bmatrix} X_0; & 0; & \cdots; & 0\end{bmatrix}\right) + \rho_V\left(h(X) - Z\right) + \rho_r\left(X\right), \tag{3}$$

where $g(X) = \begin{bmatrix} X_1; & X_2 - g_2(X_1); & \cdots; & X_N - g_N(X_{N-1})\end{bmatrix}$ and $h(X) = \begin{bmatrix} h_1(X_1); & \cdots; & h_N(X_N)\end{bmatrix}$. The functions ρ_V and ρ_W come from the choice of noise distribution, e.g. $\|\cdot\|_1$ for Laplacian and the non-convex function $\log(\nu\sigma^2 + \cdot^2)$ for Student's t. The ρ_r term may be included for extra regularization, such as to promote smoothness across time on one of the factors. Problem (3) is non-convex and thus it is difficult to have global guarantees.

We propose a matrix-free gradient-based optimization technique that takes advantage of the special structure of the problem to keep complexity per iteration moderate. Computing $A_k F_k^T$ requires $O(KST)$ operations and computing gradients of ρ functions with respect to A_k and F_k require the same. For example, the gradient of $\rho_V(Z - \mathcal{L}(A_k F_k^T))$ with respect to A_k is $\mathcal{L} * \nabla\rho_V(Z - \mathcal{L}(A_k F_k^T))(F_k \otimes I)$. The gradients may be used within the L-BFGS algorithm [7].

III. Application to Finance

A key application area for matrix factorization methods is finance. In many types of investments, including equities (e.g. stocks) and fixed income (e.g. bonds), such methods are used to reduce the dimensionality of the investment universe when modeling risk and return. For time series of stock prices, S is often in the tens of thousands, T in the hundreds, and K thirty to fifty. For fixed income yield curves, often $K = 3$ with intuitive interpretations of the three factors as level-shift, slope, and curvature changes.

Oftentimes in these applications, the matrices that are factored are static; doing so allows the analyst to produce stable factors, but such a model cannot react quickly to changes in the market. Exponential smoothing methods with fast decay can allow dynamics in the model but result in factors highly sensitive to short-lived spikes in volatility and other market anomalies. Our proposed method, since it incorporates robust Kalman smoothing, is able to both react to changes and ignore short-lived anomalies. Parametric and semiparametric methods for bond yield curves are not as flexible and rich as our proposed approach [2], [8], [9].

In practice, we have applied the proposed approach to yield curve modeling. For example, Fig. 1 shows the yield curve for eurodollar contracts from 1998 to 2010. Factors estimated using the Kalman smoothing for different example days with $K = 3$ are shown in Fig. 2. The results show that the factors do tend to have the level-shift, slope, and curvature interpretations and vary smoothly over time, thereby reacting to market changes without sensitivity to spikes.

IV. Discussion

In the big data era, modeling of high-dimensional dynamic matrices is increasingly important. We have proposed a new, efficient, and robust method for dynamic matrix factorization based on an optimization viewpoint of nonlinear Kalman smoothing that permits both factors to be time-varying. An extension of the proposed approach with orthogonality constraints on the inferred factors allows us to tackle the subspace estimation and tracking problem [10]–[12].

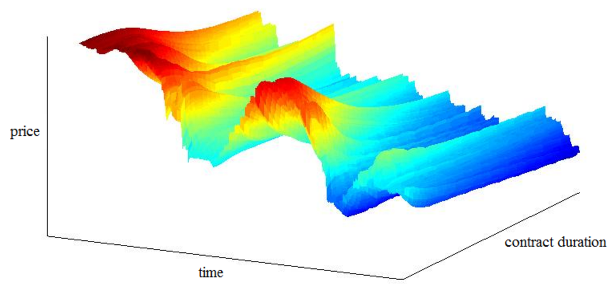

Fig. 1. Yield curve data for eurodollar contracts from 1998 to 2010. The curve for a given time represents the prices for the contracts at different available durations.

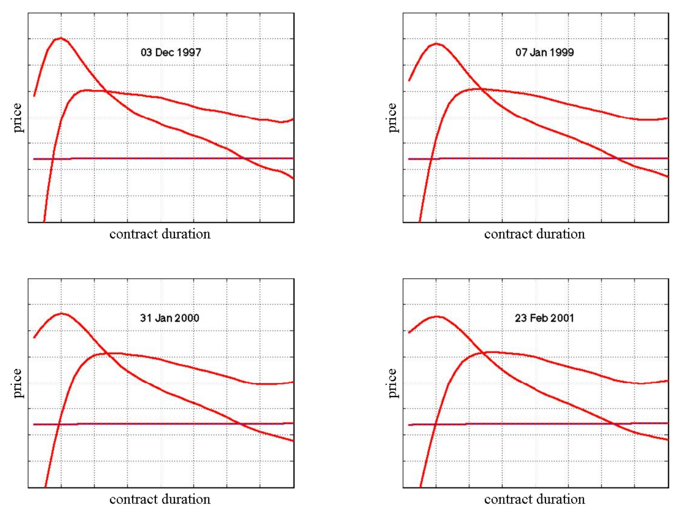

Fig. 2. Yield curve factor estimates using proposed Kalman smoothing method for four different example days. The three factors per plot have an intuitive interpretation as level-shift (flat curve), slope (mostly decreasing curve), and curvature. The factors track changes smoothly across time due to the robust Kalman smoothing formulation.

References

[1] J. Z. Sun, D. Parthasarathy, and K. R. Varshney, "Collaborative Kalman filtering for dynamic matrix factorization," *IEEE Trans. Signal Process.*, vol. 62, no. 14, pp. 3499–3509, 15 Jul. 2014.

[2] W. K. Härdle and P. Majer, "Yield curve modeling and forecasting using semiparametric factor dynamics," *Eur. J. Finance*, 2014, in press.

[3] F. C. T. Chua, R. J. Oentaryo, and E.-P. Lim, "Modeling temporal adoptions using dynamic matrix factorization," in *Proc. IEEE Int. Conf. Data Min.*, Dallas, TX, Dec. 2013, pp. 91–100.

[4] A. Y. Aravkin, K. R. Varshney, and D. M. Malioutov, "Dynamic factor modeling via robust subspace tracking," in *Industr.-Acad. Workshop Optim. Finance Risk Manage.*, Toronto, Canada, Sep. 2013.

[5] A. Y. Aravkin, D. M. Malioutov, and K. R. Varshney, "Nonparametric tracking and forecasting of multivariate data," U.S. Patent Application 14/480,704, Sep. 9, 2014.

[6] A. Y. Aravkin, J. V. Burke, and G. Pillonetto, "Optimization viewpoint on Kalman smoothing, with applications to robust and sparse estimation," in *Compressed Sensing & Sparse Filtering*, A. Y. Carmi, L. S. Mihaylova, and S. J. Godsill, Eds. Berlin, Germany: Springer, 2014, pp. 237–280.

[7] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Math. Program.*, vol. 45, no. 1–3, pp. 503–528, Aug. 1989.

[8] F. X. Diebold, G. D. Rudebusch, and S. B. Aruoba, "The macroeconomy and the yield curve: A dynamic latent factor approach," *J. Econometrics*, vol. 131, no. 1–2, pp. 309–338, Mar.–Apr. 2006.

[9] F. X. Diebold, C. Li, and V. Z. Yue, "Global yield curve dynamics and interactions: A dynamic Nelson–Siegel approach," *J. Econometrics*, vol. 146, no. 2, pp. 351–363, Oct. 2008.

[10] L. Bolzano, R. Nowak, and B. Recht, "Online identification and tracking of subspaces from highly incomplete information," http://arxiv.org/pdf/1006.4046, Jun. 2011.

[11] J. He, L. Bolzano, and J. C. S. Lui, "Online robust subspace tracking from partial information," http://arxiv.org/pdf/1109.3827, Sep. 2011.

[12] Y. Chi, Y. C. Eldar, and R. Calderbank, "PETRELS: Parallel subspace estimation and tracking by recursive least squares from partial observations," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5947–5959, 1 Dec. 2013.