

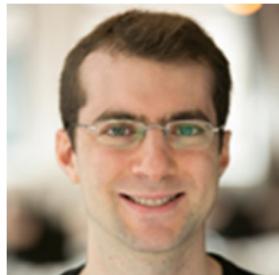
Optimal prediction in the linearly transformed spiked model

Edgar Dobriban

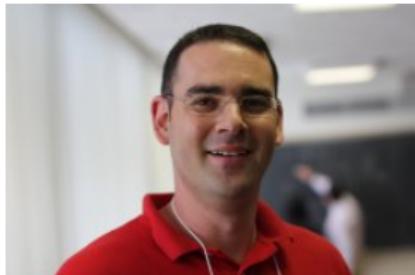
University of Pennsylvania

September 21, 2017

Collaborators



(a) William Leeb.
Princeton PACM



(b) Amit Singer.
Princeton Math+PACM

arxiv.org/abs/1709.03393

slides: github.com/dobriban (can get there from my webpage)

Overview

Introduction and examples

Our methods

Experimental results

Summary

Overview

Introduction and examples

Our methods

Experimental results

Summary

Linearly transformed spiked model

- ▶ Model $Y_i = A_i X_i + \varepsilon_i$, $i = 1, \dots, n$. “Individual regressions”
- ▶ Observe
 - ▶ $Y_i \in \mathbb{R}^{q_i}$, $A_i \in \mathbb{R}^{q_i \times p}$
 - ▶ Y_i : linear transform of unobserved signal of interest $X_i \in \mathbb{R}^p$
- ▶ Goal: recover (predict) X_i
 - ▶ X_i varies from sample to sample, but in a simple way
 - ▶ random vector lying on unknown low dimensional space
- ▶ Commutative case: $A_i^\top A_i$ diagonal

Examples

- I Standard spiked model
- II Cryo-Electron Microscopy (Cryo-EM)
- III Missing data

Example I: Standard **spiked** model

- ▶ $Y_i = X_i + \varepsilon_i$, $i = 1, \dots, n$, so $A_i = I_p$
 - ▶ $X_i \in \mathbb{R}^p$ random signal, lies on unknown low dimensional space
 - ▶ $\varepsilon_i \in \mathbb{R}^p$ noise
- ▶ PCA: X_i in direction of true PCs; how well do we estimate them by empirical PCs of Y_i ?
- ▶ Well studied in statistics and random matrix theory under high-dimensional asymptotics: e.g., Johnstone (2001); Baik et al. (2005); Paul (2007); Nadakuditi and Edelman (2008); Nadler (2008); Benaych-Georges and Nadakuditi (2012); Onatski et al. (2013); Nadakuditi (2014); Gavish and Donoho (2014); Johnstone and Onatski (2015); Hachem et al. (2015); Yao et al. (2015); Donoho et al. (2017).

Example II: Cryo-EM

- ▶ Mapping the structure of molecules without crystallizing them
- ▶ Allows imaging of heterogeneous samples with mixtures of molecules

Cryo-EM



Home News Journals Topics Careers
 Science Science Advances Science Immunology Science Robotics Science Signaling Science Translational Medicine

SHARE PERSPECTIVE BIOCHEMISTRY

The Resolution Revolution

Werner Kühlbrandt

• See all authors and affiliations

Science 30 Mar 2016;
 Vol. 352, Issue 6278, pp. 1445-1446
<http://science.org/lookup/doi/10.1126/science.1255446>

Article

Figures & Data

Info & Metrics

eLetters

PDF

[View Full Text](#)

You are currently viewing the summary.



ARTICLE TOOLS

- Email
- Print
- Alerts
- Request Permissions
- Citation tools

RELATED CONTENT

RESEARCH [arXiv](#)
 Structure of the Yeast Mitochondrial Large Ribosomal Subunit

SIMILAR ARTICLES IN:

- PubMed
- Google Scholar

CITED BY...

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

Archive | Volume 525 | Issue 7568 | News Feature | Article

NATURE | NEWS FEATURE

The revolution will not be crystallized: a new method sweeps through structural biology

Move over X-ray crystallography. Cryo-electron microscopy is kicking up a storm by revealing the hidden machinery of the cell.

Ewen Callaway

09 September 2015

[PDF](#) [Rights & Permissions](#)



Search Go Advanced search

Cell by cell



The tree of life in biology
 Scientists are striving for a deeper view of development, from embryo to adult, cell-by-cell.

Recent Read Commented

1. Massive database of 182,000 leaves is helping predict plants' family trees Nature | 07 July 2017
2. Why many scientists want better fake space dust Nature | 07 July 2017
3. Trump administration dissolves Georgia physician to lead US public-health agency Nature | 07 July 2017



The best science news from Nature and beyond, direct to your inbox every day.

Science AAAS

Home News Journals Topics Careers

Science Science Advances Science Immunology Science Robotics Science Signaling Science Translational Medicine

Authors | Members | Liberties | Advertisers

Search Go

SHARE REPORT

The 3.8 Å resolution cryo-EM structure of Zika virus

Doktor Söder*, Zheng Chen*, Lei Sun*, Thomas Kloss*, Theodore C. Pearson*, Michael G. Rossmann†, Richard J...
 • See all authors and affiliations

Science 31 Mar 2017;
 Vol. 355, issue 6320
<http://science.org/lookup/doi/10.1126/science.aai3206>

Article

Figures & Data

Info & Metrics

eLetters

PDF

You are currently viewing the abstract.

[View Full Text](#)

Abstract

The recent rapid spread of Zika virus and its unexpected linkage to birth defects and an autoimmune neurological syndrome has generated worldwide concern. Zika virus is a flavivirus like dengue, yellow fever and West Nile viruses. We present the 3.8 Å resolution

- Email
- Print
- Alerts
- Request Permissions
- Citation tools

CITED BY...

CITING ARTICLES IN:

• Web of Science (5)

• Scopus (30)

RELATED JOBS FROM SCIENCECAREERS

• Biochemistry



Deconvolution in cryo-EM

- ▶ Imaging: form 2D snapshots of molecules; add blur and noise

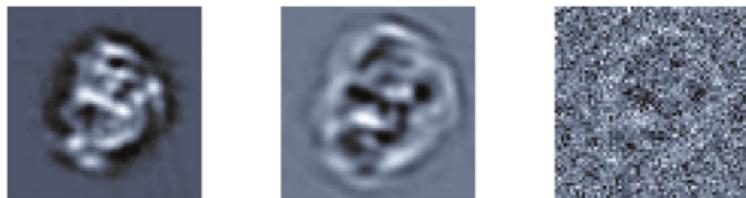


Figure : Left to right: X_i . $A_i X_i$. $A_i X_i + \varepsilon_i$

- ▶ **Goal:** Estimate unfiltered projections X_i from

$$Y_i = A_i X_i + \varepsilon_i, \quad i = 1, \dots, n.$$

- ▶ A_i blurring convolution operator. Ill-conditioned. Diagonalized by Fourier Transform.
- ▶ Deconvolution is a more general problem in signal processing

Example III: Missing data

- ▶ $Y_i = A_i X_i + \varepsilon_i$
- ▶ A_i are coordinate-selection operators.
 - ▶ k -th row of A_i selects coordinate $I_i(k)$: $A_i(k, l) = \delta_{k, I_i(k)}$
 - ▶ $A_i^\top A_i$ diagonal, 1-s for observed entries, 0-s otherwise
- ▶ High-noise matrix completion.
 - ▶ Most prior work designed for low-noise scenarios. (Candès and Recht, 2009; Candès and Tao, 2010; Keshavan et al., 2009, 2010; Koltchinskii et al., 2011; Negahban and Wainwright, 2011; Recht, 2011; Rohde et al., 2011; Jain et al., 2013)

Overview

Introduction and examples

Our methods

Experimental results

Summary

Goal

- ▶ Develop methods to predict X_i from $Y_i = A_i X_i + \varepsilon_i$
- ▶ How?

First attempt: Individual regressions

- ▶ $Y_i = A_i X_i + \varepsilon_i$ are n linear models
- ▶ Perform n individual regressions (OLS, ridge)
- ▶ Can try to “shrink” to global mean
- ▶ Problem: does not use low-rank structure. does not reduce large noise.

Our approach: idea

1. Explicitly model low-rank X_i
2. “Borrow strength” across the n samples via empirical Bayes, but only use moments
3. Estimates are low-rank

Our approach: steps

1. **Set up Bayesian model:** Best Linear Predictor depending on population parameters
2. **Moment-based empirical Bayes:** Estimate parameters using marginal
3. **Low-rank:** Equivalent to singular value shrinkage of a new or random matrix model, up to tuning.
4. **Optimal tuning:** Characterize spectrum of RM model.

1. How to get started? Bayesian model.

- ▶ “Spike” X_i : $X_i = \sum_{k=1}^r \ell_k^{1/2} z_{ik} u_k$, where $z_{ik} \sim (0, 1)$, $\ell_1 > \dots > \ell_r > 0$, and $|u_k| = 1$.
- ▶ **Best Linear Predictor (BLP)**: linear predictor $\hat{X}_i^B = LY_i$ that minimizes $\mathbb{E}|\hat{X}_i^B - X_i|^2$. Posterior mean in Gaussian-world. Linear approx in general.
 - ▶ aka “linear Bayes” (Efron)
- ▶
$$L = \text{Cov}[X_i, Y_i] \text{Cov}[Y_i, Y_i]^{-1} = \Sigma_X A_i^\top (A_i \Sigma_X A_i^\top + \Sigma_\varepsilon)^{-1}$$
- ▶ Under conditions (see paper) show $\mathbb{E}|\hat{X}_i^B - \hat{X}_i^0|^2 \rightarrow 0$, where

$$\hat{X}_i^0 = \sum_{k=1}^r \tau_k u_k u_k^\top A_i^\top Y_i$$

for some $\tau_k > 0$.

2. How to estimate unknown coefficients? Empirical BLP.

- ▶ BLP $\hat{X}_i^0 = \sum_{k=1}^r \tau_k u_k u_k^\top A_i^\top Y_i$
- ▶ u_k, τ_k parameters, must be estimated.
- ▶ **Empirical BLP**: estimate by eigenvectors \hat{u}_k of cov mx of $A_i^\top Y_i$.

$$\hat{X}_i = \sum_{k=1}^r \eta_k \hat{u}_k \hat{u}_k^\top A_i^\top Y_i$$

- ▶ “borrows strength”, could be called “second-moment empirical Bayes”
- ▶ Goal: find optimal η_k
 - ▶ if p fixed, $n \rightarrow \infty$, then $\hat{u}_k \rightarrow u_k$, so $\eta_k \rightarrow \tau_k$
 - ▶ but if $p \rightarrow \infty$, \hat{u}_k are biased for estimating u_k
 - ▶ η_k need to be adjusted

3. How to find optimal η_k ? Reduce to singular value shrinkage.

- ▶ Suppose that A_i are iid random matrices
- ▶ \hat{X}_i depend symmetrically on $(A_i, X_i, \varepsilon_i)$, thus are exchangeable
- ▶ By exchangeability,

$$\mathbb{E}|\hat{X}_i - X_i|^2 = \frac{1}{n}\mathbb{E}|\hat{X} - X|^2,$$

where $X = (X_1, \dots, X_n)^\top$, $\hat{X} = (\hat{X}_1, \dots, \hat{X}_n)^\top$.

- ▶ Since $\hat{X}_i = \sum_{k=1}^r \eta_k \hat{u}_k \hat{u}_k^\top A_i^\top Y_i$, then \hat{X} is **singular value shrinkage** of the **backprojected data** $B = (A_1^\top Y_1, \dots, A_n^\top Y_n)^\top = \sum_{k=1}^m \sigma_k \cdot \hat{u}_k \hat{v}_k^\top$, i.e.,

$$\hat{X} = \sum_{k=1}^r \eta_k \hat{u}_k \hat{u}_k^\top B = \sum_{k=1}^r \eta_k \sigma_k \cdot \hat{u}_k \hat{v}_k^\top$$

3. Reminder: Singular value shrinkage.

- ▶ When estimating a matrix, only change singular values of some “natural” estimator.
- ▶ Basic idea: singular values “spread out” due to sampling noise. So shrink them
- ▶ Dates back to Stein (1956, 1961), who derived an eigenvalue shr. covariance estimator [in addition to the more famous James-Stein mean estimator]
- ▶ Classically limited success. Complicated, unstable estimators.
- ▶ Recently more success. Low rank: Only need to work with a few singular values. In spiked models, they are asy deterministic. [concentration] “Electrostatic field” generated by spectrum of B “pushes” the spikes
- ▶ For us: Asy optimal η_k are determined by asy spectrum of B (e.g. Nadakuditi, 2014; Gavish and Donoho, 2014).

4. How to characterize spectrum?

- ▶ Back-projected data $B = (A_1^\top Y_1, \dots, A_n^\top Y_n)^\top$ is non-standard spiked random matrix model
- ▶ Typical spiked model:

$$S_i + N_i = z_i u + N_i$$

with z_i, u having iid entries, indept of N_i .

- ▶ Our model: $A_i^\top Y_i = A_i^\top A_i X_i + A_i^\top \varepsilon_i = S_i + N_i$.
 - ▶ Signal and noise are dependent due to A_i .
 - ▶ u_i are deterministic, not random.
- ▶ Many theoretical tools are inadequate.
- ▶ Fortunately, can extend approach of Benaych-Georges and Nadakuditi (2012); also adapt “deterministic equivalents” tools of Bai et al. (2007).

Overview

Introduction and examples

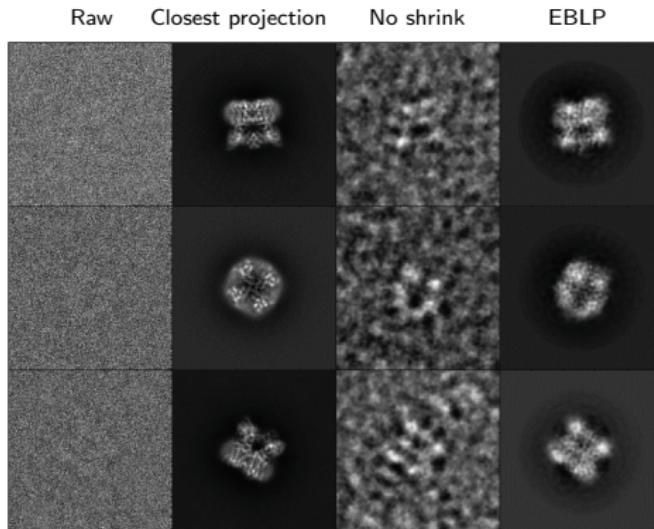
Our methods

Experimental results

Summary

Results: Cyro-EM experimental data

Bhamre, Zhang, Singer (*Journal of Structural Biology*, 2016)

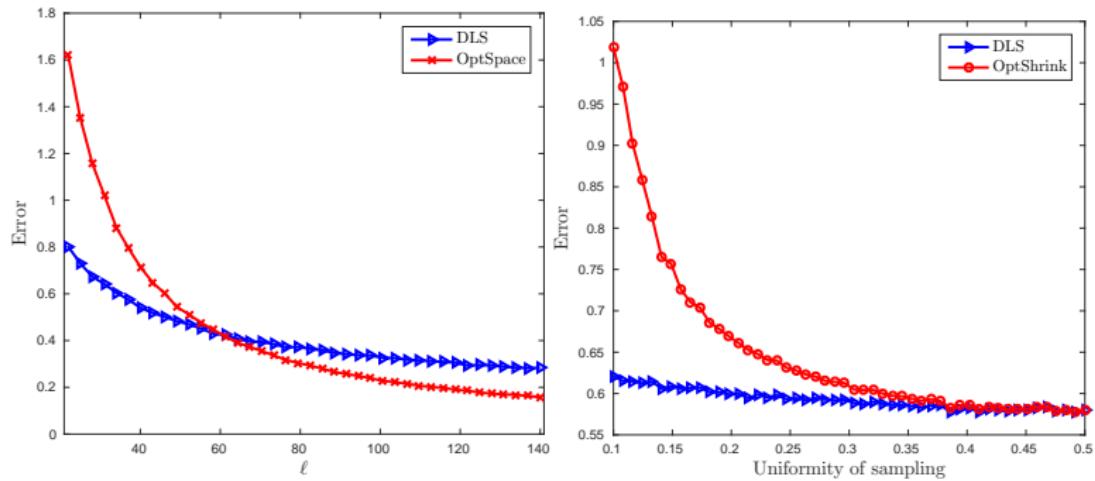


- ▶ Deconvolution in cryo-EM
- ▶ molecule: TRPV1 ion channel
- ▶ 35645 images of 256×256 pixels (Liao et al., Nature 2013)

Results: matrix completion

- We test our shrinkage method to other scalable matrix completion methods.

D, Leeb, Singer (*arXiv 2016*)



- Left: Comparison with OptSpace (Keshavan et al., 2009) for different SNRs. Our method outperforms in the high-noise regime.
- Right: Comparison with OptShrink (Nadakuditi, 2014) for different sampling uniformity levels. Our method outperforms OptShrink when the data is unevenly sampled across the rows of the matrix.

Overview

Introduction and examples

Our methods

Experimental results

Summary

Summary

- ▶ Linearly transformed spiked model: $Y_i = A_i X_i + \varepsilon_i$
 - ▶ Broad applicability: Image processing, Cryo-EM, Missing data
- ▶ Developed methods for predicting X_i
- ▶ Broad conclusion:
 - ▶ combining classical approach (EBLP) with modern random matrix theory; leads to effective prediction methods

- Bai, Z., Miao, B., and Pan, G. (2007). On asymptotics of eigenvectors of large sample covariance matrix. *The Annals of Probability*, 35(4):1532–1572.
- Baik, J., Ben Arous, G., and Péché, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability*, 33(5):1643–1697.
- Benaych-Georges, F. and Nadakuditi, R. R. (2012). The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, 111:120–135.
- Candès, E. J. and Recht, B. (2009). Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772.
- Candès, E. J. and Tao, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080.
- Donoho, D. L., Gavish, M., and Johnstone, I. M. (2017). Optimal shrinkage of eigenvalues in the spiked covariance model. *arXiv preprint arXiv:1311.0851, to appear in AoS*.
- Gavish, M. and Donoho, D. L. (2014). The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Transactions on Information Theory*, 60(8):5040–5053.
- Hachem, W., Hardy, A., and Najim, J. (2015). A survey on the eigenvalues local behavior of large complex correlated wishart matrices. *ESAIM: Proceedings and Surveys*, 51:150–174.
- Jain, P., Netrapalli, P., and Sanghavi, S. (2013). Low-rank matrix completion using alternating minimization. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 665–674. ACM.
- Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics*, 29(2):295–327.

- Johnstone, I. M. and Onatski, A. (2015). Testing in high-dimensional spiked models. *arXiv preprint arXiv:1509.07269*.
- Keshavan, R. H., Montanari, A., and Oh, S. (2010). Matrix completion from a few entries. *IEEE Transactions on Information Theory*, 56(6):2980–2998.
- Keshavan, R. H., Oh, S., and Montanari, A. (2009). Matrix completion from a few entries. In *2009 IEEE International Symposium on Information Theory*, pages 324–328. IEEE.
- Koltchinskii, V., Lounici, K., and Tsybakov, A. B. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *The Annals of Statistics*, pages 2302–2329.
- Nadakuditi, R. R. (2014). Optshrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage. *IEEE Transactions on Information Theory*, 60(5):3002–3018.
- Nadakuditi, R. R. and Edelman, A. (2008). Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples. *Signal Processing, IEEE Transactions on*, 56(7):2625–2638.
- Nadler, B. (2008). Finite sample approximation results for principal component analysis: A matrix perturbation approach. *The Annals of Statistics*, 36(6):2791–2817.
- Negahban, S. and Wainwright, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *The Annals of Statistics*, pages 1069–1097.
- Onatski, A., Moreira, M. J., and Hallin, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *The Annals of Statistics*, 41(3):1204–1231.
- Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, 17(4):1617–1642.

- Recht, B. (2011). A simpler approach to matrix completion. *Journal of Machine Learning Research*, 12:3413–3430.
- Rohde, A., Tsybakov, A. B., et al. (2011). Estimation of high-dimensional low-rank matrices. *The Annals of Statistics*, 39(2):887–930.
- Yao, J., Bai, Z., and Zheng, S. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. Cambridge University Press.