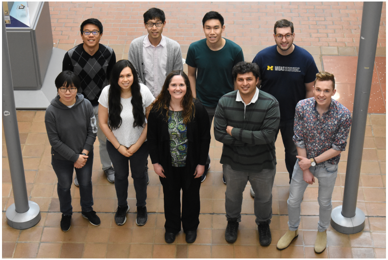
Optimization for Low Dimensional Modeling

Laura Balzano

University of Michigan

IMSI Reduced Order and Surrogate Modeling Workshop
2025 November 12

SPADA lab



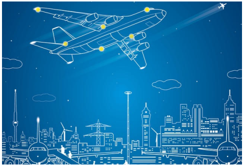
Structure in Big Data

Prompt

```
11 + 26 -> 37
33 + 13 -> quarante-six
62 + 54 -> one hundred and sixteen
74 + 59 -> 133
69 + 10 -> setenta y nueve
40 + 43 -> ochenta y tres
24 + 40 -> soixante-quatre
   ⋮
30 + 25 -> fifty-five
91 + 83 ->
```

Output Probabilities (Llama-3)

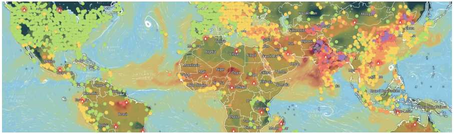
174:	52.17%
one hundred and seventy-four	13.16%
cent soixante-quatorze:	11.10%
ciento setenta y cuatro:	21.69%



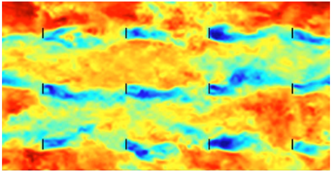
CT Scanner: https://medicine.umich.edu/sites/default/files/RAD_NewsletterFall2011.pdf
LLMs demonstrate In-Context learning: https://arxiv.org/abs/2410.05603
Power Grid: https://news.engin.umich.edu/2019/02/how-air-conditioners-could-advance-a-renewable-power-grid/
Traffic images and markings from the “Hopkins 155” Dataset, R. Vidal lab, Johns Hopkins University.
Aviation Sensing: http://interactive.aviationtoday.com/smart-sensors-expand-in-variety-scope/
Air quality sensing: http://www.livescience.com/27992-portable-pollution-sensors-improve-data-nsf-ria.html

Low-dimensional Structure in Matrix Data

CT scans



Air quality around the world

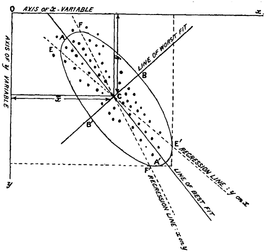


Simulated air flow around wind turbines

- In all these applications, we believe there is some structure in the data.
- That structure helps us:
  - predict and learn,
  - interpret and understand,
  - impute,
  - detect anomalies,
  - compress for memory and computational efficiency
  - etc.

Subspace Model for Matrix Data

For 124 years, Principal Component Analysis (PCA) has been used to investigate linear structure in a dataset
[Pearson, 1901, Spearman, 1904, Hotelling, 1933]:



[Pearson, 1901]

Village School, 24 Oldest Children.

A. Original Data.

Sex.	Age.		Discriminative Threshold.			Intellectual Rank.		
	Years	Months	Pitch 1/3 v. d.	Light 1:200	Weight 1:200	Common Sense out of School. (A)	Common Sense out of School. (B)	Cleverness in School.
f	11	6	8	4	4	6	5	2
m	12	11	15	3	4	11	7	22
f	12	8	14	6	4	16	10	7
f	13	8	13	4	9	1	1	1
m	11	4	5	14	7	3	2	3
f	11	11	25	7	4	10	14	9
f	11	3	10	19	8	8	19	12
f	13	1	10	12	10	2	4	6
m	12	5	18	11	9	5	6	11
m	12	7	14	30	7	21	22	19
f	12	8	60	3	10	12	9	4
f	13	10	20	12	10	13	12	18

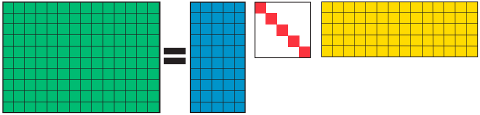
[Spearman, 1904]

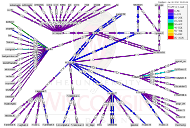
Summary of data for Corsican pine, East Anglia
(Number of props = 180)

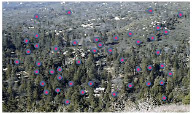
Variable	*Minimum*	*Mean*	*Maximum*	*Standard deviation*
Topdiam	2·38	4·21	6·00	0·98
Length	28·75	46·88	60·90	11·29
Moist	14·7	114·6	213·1	57·1
Testsg	0·365	0·877	1·217	0·231
Ovensg	0·289	0·415	0·993	0·070
Ringtop	7	13·3	23	3·24
Ringbut	8	16·3	26	3·95
Bowmax	0·13	0·65	2·50	0·43
Bowdist	7	23·4	52	9·06
Whorls	1	2·49	7	1·17
Clear	1	10·7	31	6·27
Knots	0	5·45	9	1·65
Diaknot	0	0·82	1·65	0·325

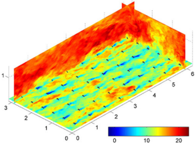
The analysis therefore suggests that there are probably six major components of the physical variables, accounting for about 87% of the variability, and ... **focuses the attention of the research worker on the basic dimensions of which his variables are only first approximations.** [Jeffers, 1967]

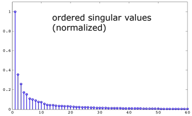
Subspace Model for Matrix Data



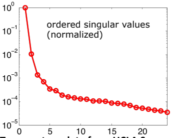




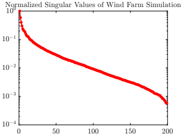


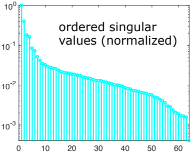


Byte Count data from UW network



Temperature data from UCLA Sensornet





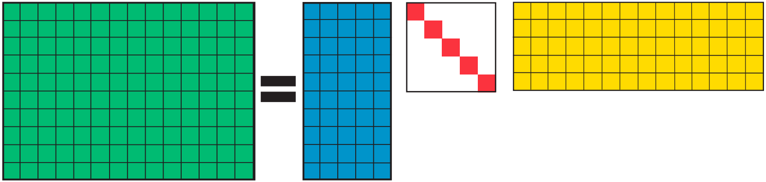
Why are big data matrices approximately low rank? [Udell and Townsend, 2019]

Subspace Model for Matrix Data

Principal components of a data matrix Y are computed with the Singular Value Decomposition (SVD):

$$Y = U\Sigma V^T$$

where U, V have orthonormal columns and Σ is diagonal.



Linear Low Rank Subspaces via SVD

If your data matrix Y is not exactly low-rank but you wish to find the best low-dimensional linear structure that models the data, you can use the SVD; that makes it a useful exploratory data analysis tool. The SVD gives the solution to the following nonconvex optimization problem for any k[1]:

$$\begin{aligned} \underset{U\in\mathbb{R}^{d\times k}, W\in\mathbb{R}^{n\times k}}{\text{minimize}} \quad & \|UW^T - Y\|_F^2 \\ \text{subject to} \quad & U \in \mathcal{G}(k,d) \end{aligned} \tag{1}$$

where $\mathcal{G}(k,d)$ is the Grassmannian, the space of all k-dimensional subspaces of $\mathbb{R}^d$.

[1] This result was discovered independently by first Schmidt in 1907 [Schmidt, 1907, Stewart, 1993, Stewart, 2011] and then Eckart and Young [Eckart and Young, 1936].

Modern Generalizations

- An observation function (*e.g.*, missing or sketched data, or calibration) represented as $g(\cdot)$
- Deeper factorizations $UW_L \cdots W_1^T$
- Generative probabilistic coefficient and noise models (non-iid-Gaussian)
- Different loss functions $\ell(\cdot)$
- Any regularization or constraint (*e.g.*, to encourage structure in the factors) represented as $h(\cdot)$
- Other matrix manifolds of k dimensions, represented as $\mathcal{M}_k$

$$\begin{aligned} \underset{U, W_L, \ldots, W_1}{\text{minimize}} \quad & \ell(g(UW_L \cdots W_1^T), g(Y))) \\ \text{subject to} \quad & h(U, W_L, \ldots, W_1) \leq \tau \\ & U \in \mathcal{M}_k \end{aligned}$$

Modern Generalizations

If we change the model slightly or add anything to the cost function, how do we adjust the SVD computation?

$$\begin{aligned} \underset{U, W_L, \ldots, W_1}{\text{minimize}} \quad & \ell(g(UW_L \cdots W_1^T), g(Y))) \\ \text{subject to} \quad & h(U, W_L, \ldots, W_1) \leq \tau \\ & U \in \mathcal{M}_k \end{aligned} \tag{2}$$

With the lack of an obvious extension to SVD computations, we turn to nonconvex optimization!

Our work probing this framework

- An observation function represented as $g(\cdot)$

  Matrix completion [Balzano et al., 2010, Kennedy et al., 2016, Balzano and Wright, 2014, Zhang and Balzano, 2016, Liu et al., 2024], Sketching [Alarcon et al., 2016, Zhang and Balzano, 2018], Variety completion [Ongie et al., 2017, Ongie et al., 2021], Ordinal embedding [Bower et al., 2018]

- Deeper factorizations $UW_L \cdots W_1^T$

  Deep matrix completion [Kwon et al., 2024] and Deep low-rank adaptation [Yaras et al., 2024]

- Generative probabilistic coefficient and noise models (non-Gaussian)

  Heteroscedastic PCA [Hong et al., 2018, Hong et al., 2021, Hong et al., 2023, Cavazos et al., 2023, Gilman et al., 2025a, Gilman et al., 2025b], Markov Online Dictionary Learning [Lyu et al., 2020]

- Different loss functions $\ell(\cdot, \cdot)$

  Robust PCA [He et al., 2012, He et al., 2014, Gilman and Balzano, 2019], Subspace clustering [Lipor and Balzano, 2017, Gitlin et al., 2018, Lipor et al., 2021, Wang et al., 2022], Cross entropy loss [Yaras et al., 2022, Yaras et al., 2023], Reduced order modeling [Newton et al., 2023]

- Any regularization or constraint represented as $h(\cdot)$ or other matrix manifolds of k dimensions

  Supervised PCA [Ritchie et al., 2020], Sparse PCA [Xiao and Balzano, 2016], Stiefel Manifold Optimization [Ritchie et al., 2020, Hong et al., 2019, Blocker et al., 2023]

Outline

- Introduction
- Convergence guarantees
- Matrix completion and streaming SVD
- Heteroscedastic PCA
- Reduced order modeling

Convergence

Can we guarantee convergence for nonconvex problems? (assuming differentiable)

- Global convergence to a stationary point of the objective function (from any initialization)
- Local convergence (within a region of the stationary point, usually a local minima)
- Identification of a guaranteed good initialization within the basin of attraction of a global minima
- Convergence to global minima

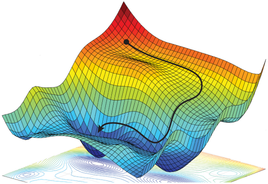


Figure courtesy Science Magazine

Nonconvex Provable Global Results (or “hidden” convexity)

- Low-rank matrix completion/sensing [Candès and Recht, 2009, Chi et al., 2019]
- Low-rank tensor decomposition [Kileel et al., 2021]
- Phase Retrieval [Chen et al., 2019]
- Dictionary Learning [Sun et al., 2017a, Sun et al., 2017b]
- Deep Linear Networks with MSE loss [Yaras et al., 2024]
- Nonnegative/Sparse/Robust PCA [Wright et al., 2009]
- Mixed Linear Regression (two component) [Chen et al., 2017]
- Blind Deconvolution/Calibration [Ling and Strohmer, 2019, Bilen et al., 2014]
- Superresolution [Candès and Fernandez-Granda, 2013, Yang et al., 2016]
- System identification for (hybrid) linear dynamical systems

  [Feng et al., 2010, Hardt et al., 2018]

Also see `https://sunju.org/research/nonconvex/` (last updated 2021)

Convergence

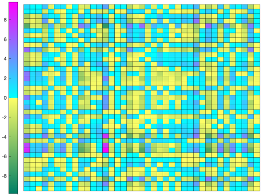
Block Variable (Alternating) MM and/or Riemannian methods [Li et al., 2023, Li et al., 2024]

Methods	Manifold	Objective	Constraints	Blocks	Complexity	Inexact comp.
Euclidean BMM (Hong et al., 2017) Euclidean Block PGD (Beck and Tetruashvili, 2013)	Euclidean	convex	convex	many	$\widetilde{O}(\varepsilon^{-1})$	✗
Euclidean BMM-DR (Lyu and Li, 2025)	Euclidean	non-convex	convex	many	$\widetilde{O}(\varepsilon^{-2})$	✓
Riemannian prox. (Li et al., 2009)	Hadamard	g-convex	g-convex	1	-	✗
Riemannian prox. (Bento et al., 2017)	Hadamard	g-convex	g-convex	1	$\widetilde{O}(\varepsilon^{-1})$	✗
Riemannian Prox-linear (line search)(Chen et al., 2020)	Riemannian & Compact	non-convex & smooth†	N/A	1	$\widetilde{O}(\varepsilon^{-2})$	✗
Block Riemannian GD (Exp) (Gutman and Ho-Nguyen, 2023)	Riemannian	non-convex	N/A	many	$\widetilde{O}(\varepsilon^{-2})$	✗
BMM on manifolds (Peng and Vidal, 2023)	Riemannian & compact	non-convex	N/A	many	$\widetilde{O}(\varepsilon^{-2})$	✗
RBMM (**Ours**) with surr.:						
g-smooth (Thm. 10)	Riemannian	non-convex & non-smooth	g-convex	many	$\widetilde{O}(\varepsilon^{-2})$	✓
Riemannian proximal (Thm. 7)	Riemannian	non-convex & non-smooth	g-convex	many	$\widetilde{O}(\varepsilon^{-2})$	✓
Euclidean proximal (Thm. 7)	Riemannian $\subseteq$ Euclidean	non-convex & non-smooth	g-convex & compact	many	$\widetilde{O}(\varepsilon^{-2})$	✓
Smooth (Cor. 11)	Euclidean/ Stiefel	non-convex & non-smooth	convex/ g-convex	many	$\widetilde{O}(\varepsilon^{-2})$	✓

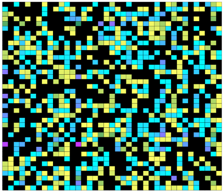
Outline

- Introduction
- Convergence guarantees
- Matrix completion and streaming SVD
- Heteroscedastic PCA
- Reduced order modeling

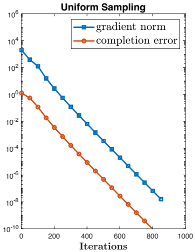
Low Rank Matrix Completion



original rank 2 matrix

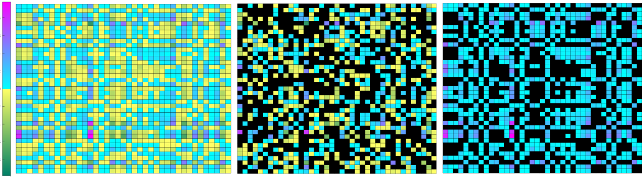


uniform sampling

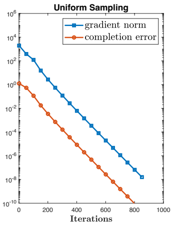


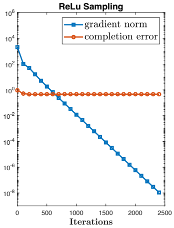
- The optimization landscape is benign for nice enough sampling (and you need very few samples for a low-rank reconstruction!)
- The objective function is strongly convex around a planted low-rank matrix
- You almost always get global convergence to the global min using all kinds of solution methods (though not always easy to prove)

Low Rank Matrix Completion



original rank 2 matrix — uniform sampling — ReLU sampling





- What happens with less nice sampling? Consider observing largest 50% entries.
- The objective function is still strongly convex around a planted low-rank matrix
- With a nice initialization you will converge to the global minimum

Use case: missing data and online setting

- Often in wireless sensing, due to sensor failure or sensors going offline for calibration, we have missing observations.
- Additionally, processing in batch could be prohibitive, due to computation and/or memory storage, **or** the data streams in real-time.

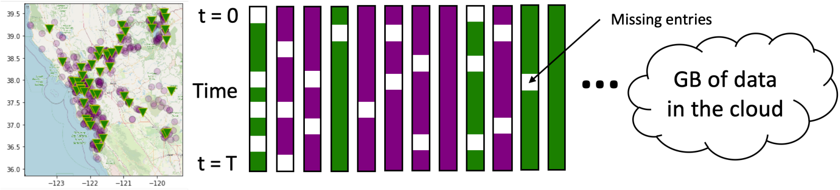


Figure: AQI monitoring with streaming, incomplete samples.

The Incremental SVD for Euclidean Subspace Estimation

Given matrix $Y = U\Sigma V^T$, form the SVD of[2] $\begin{bmatrix} Y & v_t \end{bmatrix}$.

Compute the weights: $w = \arg\min_a \|Ua - v_t\|_2^2$

Compute the residual: $r_t = v_t - Uw$.

Update the SVD:

$$\begin{bmatrix} Y & v_t \end{bmatrix} = \begin{bmatrix} U & \frac{r_t}{\|r_t\|} \end{bmatrix} \begin{bmatrix} \Sigma & w \\ 0 & \|r_t\| \end{bmatrix} \begin{bmatrix} V & 0 \\ 0 & 1 \end{bmatrix}^T$$

and diagonalize the center matrix [Bunch and Nielsen, 1978].

[2] You can also add a row or remove a row or column.

The Incremental SVD with Missing Data

Update the SVD:

$$\begin{bmatrix} U & \frac{r_t}{\|r_t\|} \end{bmatrix} \begin{bmatrix} \Sigma & w \\ 0 & \|r_t\| \end{bmatrix} \begin{bmatrix} V & 0 \\ 0 & 1 \end{bmatrix}^T$$

and diagonalize the center matrix.

- The inner matrix grows if the residual is always nonzero.
- Truncating the singular values as we stream is a heuristic.
- We could sketch [Ghashami et al., 2016].
- What about matrix completion?

The Incremental SVD with Missing Data

Given matrix $Y = U\Sigma V^T$, form the SVD of $\begin{bmatrix} Y & v_t \end{bmatrix}$.

Estimate the weights: $w = \arg\min_a \|P_{\Omega_t}(Ua - v_t)\|_2^2$.

Compute the residual: $r_t = v_t - Uw$ on Ω_t ; zero otherwise.

Update the SVD:

$$\begin{bmatrix} U & \frac{r_t}{\|r_t\|} \end{bmatrix} \begin{bmatrix} \Sigma & w \\ 0 & \|r_t\| \end{bmatrix} \begin{bmatrix} V & 0 \\ 0 & 1 \end{bmatrix}^T$$

and diagonalize the center matrix.

Incremental SVD with Missing Data: SAGE GROUSE

Given matrix $Y = U\Sigma V^T$, form the SVD of $\begin{bmatrix} Y & v_t \end{bmatrix}$.

Estimate the weights: $w = \arg\min_a \|P_{\Omega_t}(Ua - v_t)\|_2^2$.

Compute the residual: $r_t = v_t - Uw$ on Ω_t ; zero otherwise.

Update the SVD:

$$\begin{bmatrix} U & \frac{r_t}{\|r_t\|} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_k & w \\ 0 & \|r_t\| \end{bmatrix} \begin{bmatrix} V & 0 \\ 0 & 1 \end{bmatrix}^T$$

and take the SVD of the center matrix. This is equivalent to the incremental gradient method on the Grassmannian (GROUSE or Oja's method) for a particular step size [Balzano and Wright, 2013, Balzano, 2022].

Incremental SVD with Missing Data Options

projection weights $w = \arg\min_a \|P_{\Omega_t}(U_t a - v_t)\|_2^2$;
residual: $r_t = v_t - Uw$ on Ω_t ; zero otherwise.

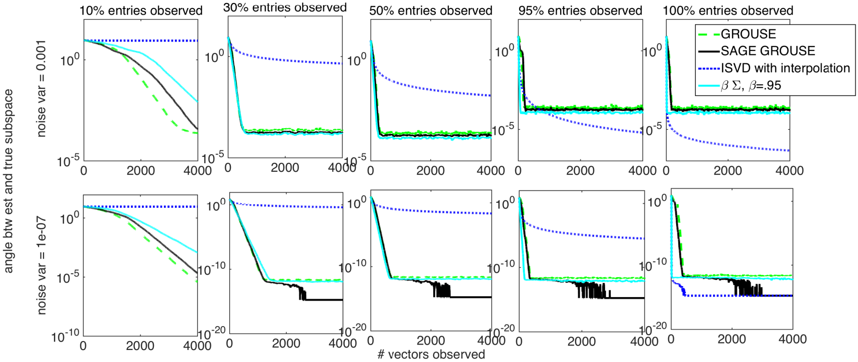
ISVD with interpolation: $\begin{bmatrix} U & \frac{r_t}{\|r_t\|} \end{bmatrix} \begin{bmatrix} \Sigma & w \\ 0 & \|r_t\| \end{bmatrix} \begin{bmatrix} V & 0 \\ 0 & 1 \end{bmatrix}^T$

SAGE GROUSE: $\begin{bmatrix} U & \frac{r_t}{\|r_t\|} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_k & w \\ 0 & \|r_t\| \end{bmatrix} \begin{bmatrix} V & 0 \\ 0 & 1 \end{bmatrix}^T$

[Brand, 2002]$(\beta \leq 1)$: $\begin{bmatrix} U & \frac{r_t}{\|r_t\|} \end{bmatrix} \begin{bmatrix} \beta\Sigma & w \\ 0 & \|r_t\| \end{bmatrix} \begin{bmatrix} V & 0 \\ 0 & 1 \end{bmatrix}^T$

There are also variants that handle more ill-conditioned data [Kennedy et al., 2014]; perform updates for a robust (ℓ_1) loss function [He et al., 2012]; generalize to RLS approach [Chi et al., 2012] with a more complex update.

Incremental SVD with Missing Data Performance



Incremental SVD has been used in reduced order modeling!

Outline

- Introduction
- Convergence guarantees
- Matrix completion and streaming SVD
- Heteroscedastic PCA
- Reduced order modeling

Modern data are heteroscedastic

Modern data are often corrupted by **heteroscedastic** noise.

varying radiation levels

varying atmosphere

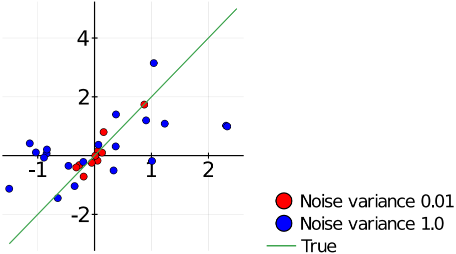
varying sensor quality

We want to be able to use the predictive power of these data combined, instead of doing analysis on each dataset separately.

---

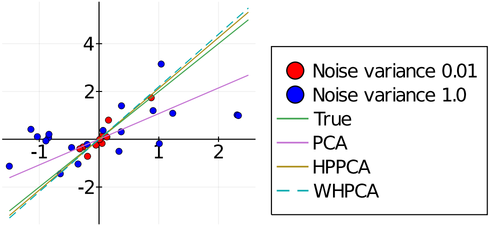
http://www.medicalnewstoday.com/articles/153201.php
https://www.nasa.gov/multimedia/imagegallery/iotd.html
http://www.livescience.com/27992-portable-pollution-sensors-improve-data-nsf-ria.html

What happens when we apply PCA?



Suppose we seek the first principal component of these data, blue and red combined. How will it look?

What happens when we apply PCA?



PCA is not robust, but our methods can handle the high variance data.

Two Approaches

Weighted PCA

$$\min_{U,z} \sum_{j=1}^{n} \alpha_j^2 \|y_j - U z_j\|_2^2$$

$$\alpha_j \in \{w_1, \ldots, w_L\}$$

- Use vanilla SVD algorithm
- Amenable to analysis:
  - asymptotic recovery results
  - optimal weights to maximize asymptotic recovery
- But is it the "right" thing to do?
- Requires knowledge or estimate of unknown parameters

Heterosced Probabilistic PCA

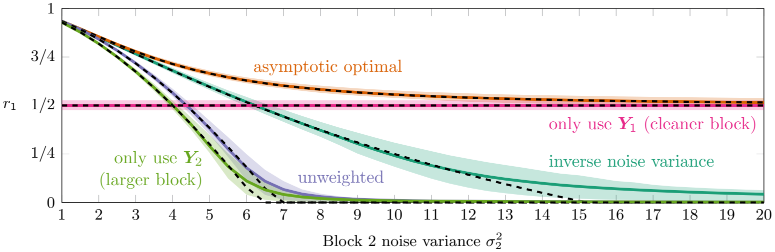
$$y_j = U\Theta z_j + \eta_j \epsilon_j \ ,$$
$$z_j \sim \mathcal{N}(0, I_k), \epsilon_j \sim \mathcal{N}(0, I_d)$$

- Maximum Likelihood can incorporate estimates of all parameters
  - Including noise variances!
- Non-concave but probably "nice"
- Algorithms don't have guarantees (yet)
- Makes distributional assumptions

Numerical Simulation - Weighted SVD

Optimal weights are not inverse noise variance [Hong et al., 2023]

- Asymptotic analysis with problem size
- Intuition: PCA is not robust to heavy outliers; must downweight more

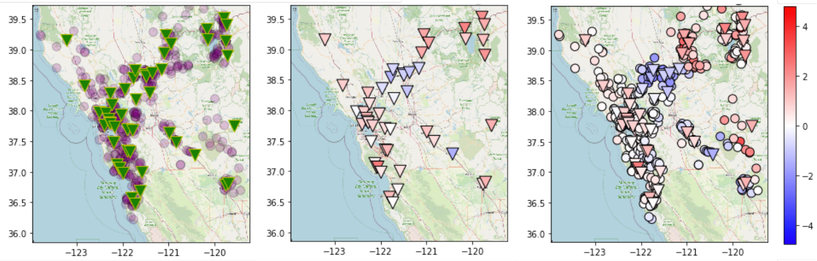


Two blocks $Y_1 \in \mathbb{R}^{d \times n_1}$ and $Y_2 \in \mathbb{R}^{d \times n_2}$ with $d = 10^3$, $n_1 = 10^3$, $n_2 = 10^4$, and signal component variance $\theta_1^2 = 2$. The first block noise variance is $\sigma_1^2 = 1$ and the second block noise variance is on the x-axis. For each weighting scheme, the dashed black curve is the predicted asymptotic performance, the solid colored curve is the average from 400 trials, and the ribbon indicates the corresponding interquartile interval.

Solving PPCA

We have several alternatives:

- Alternating MM [Hong et al., 2021]
- Nuclear norm optimization and a difference of convex approximation [Cavazos et al., 2023]
- Streaming Stochastic MM [Gilman et al., 2025b]



AirNow & PurpleAir sensor locations

PC loadings: AirNow

PC loadings: AirNow + PurpleAir

Air quality sensing

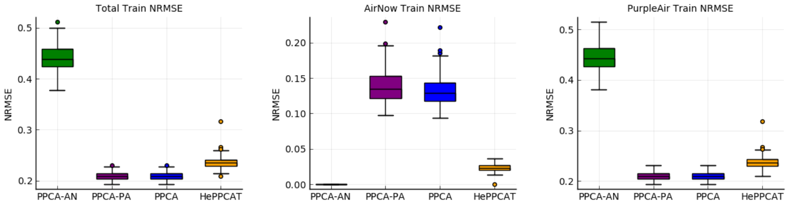


Figure: Box plots showing Normalized RMSE for the 30-dimensional subspace learned over each of 200 train/test splits of the data.

Air quality sensing

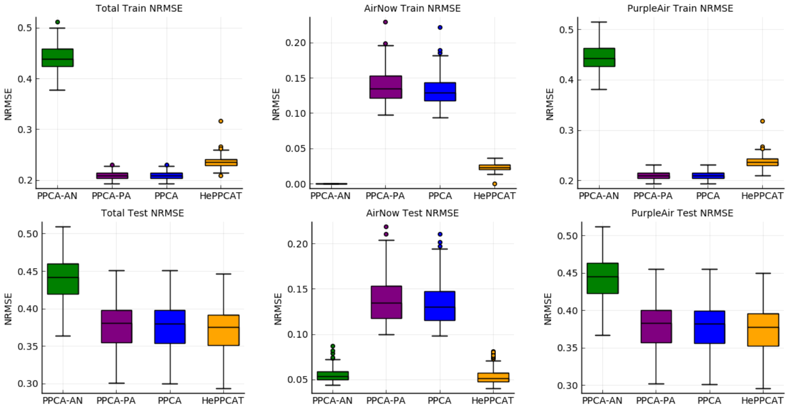


Figure: Box plots showing Normalized RMSE for the 30-dimensional subspace learned over each of 200 train/test splits of the data.

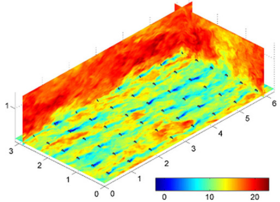
Outline

Model-based data-driven ROM

Consider a (potentially non-linear) system of difference equations

$$x_{k+1} = f(x_k, u_k)$$
$$y_k = h(x_k, u_k)$$

with states $x_k \in \mathbb{R}^{n_x}$, input $u_k \in \mathbb{R}^{n_u}$, and output $y_k \in \mathbb{R}^{n_y}$.



Wind farm simulation with 6.3 million states [Meyers and Meneveau, 2012].

Collect data through simulation or experimentation as *snapshots* of the system

$$X_0 = \begin{bmatrix} x_0 & x_1 & \dots & x_{n_s-1} \end{bmatrix} \in \mathbb{R}^{n_x \times n_s}$$
$$X_1 = \begin{bmatrix} x_1 & x_2 & \dots & x_{n_s} \end{bmatrix} \in \mathbb{R}^{n_x \times n_s}$$
$$U_0 = \begin{bmatrix} u_0 & u_1 & \dots & u_{n_s-1} \end{bmatrix} \in \mathbb{R}^{n_u \times n_s}$$
$$Y_0 = \begin{bmatrix} y_0 & y_1 & \dots & y_{n_s-1} \end{bmatrix} \in \mathbb{R}^{n_y \times n_s}$$

Linearizing the System

Using this data, the non-linear system can be approximated as the linear system

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k \\ y_k &= Cx_k + Du_k \end{cases} \tag{3}$$

by solving the least-squares problem[3]

$$\begin{bmatrix} \hat{A} & \hat{B} \\ \hat{C} & \hat{D} \end{bmatrix} = \arg \min_{A,B,C,D} \left\| \begin{bmatrix} X_1 \\ Y_0 \end{bmatrix} - \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} X_0 \\ U_0 \end{bmatrix} \right\|_F^2 \tag{4}$$

[3]Note that this is the best-fit linear system in an ℓ_2-sense, which may or may not be what we really need for downstream applications.

Reduced-Order Model (ROM)

Going further, we can project our state onto r-dimensional subspace spanned by semi-orthogonal matrix $Q \in \mathbb{R}^{n_x \times r}$

$$z := Q^\top x \in \mathbb{R}^r$$

to get the low-order system

$$\begin{aligned} z_{k+1} &\approx F z_k + G u_k \\ y_k &\approx H z_k + D u_k \end{aligned}$$

where $F := Q^\top A Q, G := Q^\top B, H := CQ$.

$$\begin{aligned} \underset{Q,F,G,H,D}{\text{minimize}} \quad & \left\| \begin{bmatrix} X_1 \\ Y_0 \end{bmatrix} - \begin{bmatrix} Q & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} F & G \\ H & D \end{bmatrix} \begin{bmatrix} Q^\top & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} X_0 \\ U_0 \end{bmatrix} \right\|_F^2 \\ \text{subject to} \quad & Q \in \mathcal{G}(n, r) \end{aligned} \tag{5}$$

Reduced-Order Model (ROM)

Going further, we can project our state onto r-dimensional subspace spanned by semi-orthogonal matrix $Q \in \mathbb{R}^{n_x \times r}$

$$z := Q^\top x \in \mathbb{R}^r$$

For a given projection matrix $\hat{Q}$, the ROM that best fits this data is given by the solution to the least-squares problem

$$\begin{bmatrix} \hat{F} & \hat{G} \\ \hat{H} & \hat{D} \end{bmatrix} := \arg \min_{F,G,H,D} \left\| \begin{bmatrix} \hat{Q}^\top X_1 \\ Y_0 \end{bmatrix} - \begin{bmatrix} F & G \\ H & D \end{bmatrix} \begin{bmatrix} \hat{Q}^\top X_0 \\ U_0 \end{bmatrix} \right\|_F^2$$

which leads to...

Reduced-Order Model (ROM)

Going further, we can project our state onto r-dimensional subspace spanned by semi-orthogonal matrix $Q \in \mathbb{R}^{n_x \times r}$

$$z := Q^\top x \in \mathbb{R}^r$$

$$\underset{Q}{\text{minimize}} \quad \left\| \begin{bmatrix} X_1 \\ Y_0 \end{bmatrix} - \begin{bmatrix} Q & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} Q^\top X_1 \\ Y_0 \end{bmatrix} \begin{bmatrix} Q^\top X_0 \\ U_0 \end{bmatrix}^\dagger \begin{bmatrix} Q^\top & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} X_0 \\ U_0 \end{bmatrix} \right\|_F^2 \tag{5}$$

subject to $\quad Q \in \mathcal{G}(n, r)$

Illustrative Example

Example Problem

$$\begin{cases} x_{k+1} = \begin{bmatrix} -0.7 & -0.7 \\ 0.5 & 0.5 \end{bmatrix} x_k + \begin{bmatrix} 0.5 \\ 0.7 \end{bmatrix} u_k \end{cases}$$

Parameterize $Q \in \mathbb{R}^{2\times 1}$ as

$$Q = \begin{bmatrix} \sin(\theta) \\ \cos(\theta) \end{bmatrix}$$

and visualize the landscape as a function of θ.

Illustrative Example

Example Problem

$$\begin{cases} x_{k+1} = \begin{bmatrix} -0.7 & -0.7 \\ 0.5 & 0.5 \end{bmatrix} x_k + \begin{bmatrix} 0.5 \\ 0.7 \end{bmatrix} u_k \end{cases}$$

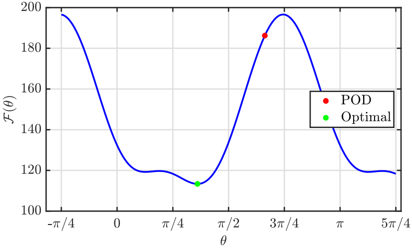
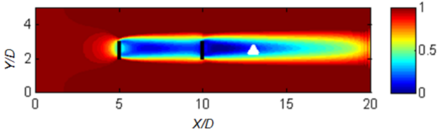


Figure: Cost function (5) as a function of θ, the angle of the orthogonal projection matrix Q.

Wind Farm Experiment

Problem Setup borrowed from [Annoni and Seiler, 2017].



- Input is front turbine axial factor.
- Output is wind speed at white triangle.

Table: Wind Farm Model Parameters

Parameter	Value
n_{turb}	2 turbines
x-by-y grid	201-by-101
n_x	40602 states
n_u	1 input
n_y	1 output
n_s	200 samples

Wind Farm Experiment

Problem Setup borrowed from [Annoni and Seiler, 2017].

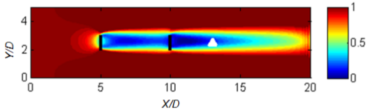


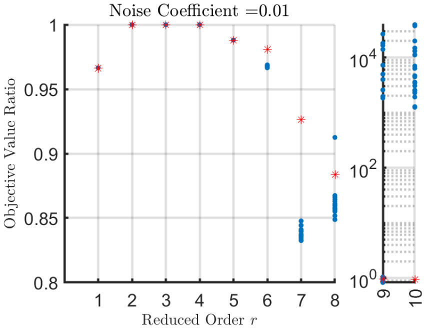
Table: Comparison of Metrics for Wind Farm Model

Method	Objective Value
POD	4380.41
GGD	4188.06

But ... POD performs well!

Optimization rarely has much room for improvement over POD in random examples (in terms of *best linear* approximations).

Not only that, but our true end goal is not ℓ_2 best fit but usually some downstream task, such as control.



Data-Driven ROM LQR Pipeline

1. Choose subspace $\hat{Q}$ and compute $\hat{F}, \hat{G}$ as the least-squares solution:

$$\begin{bmatrix} \hat{F} & \hat{G} \end{bmatrix} = \begin{bmatrix} \hat{Q}^\top X_1 \end{bmatrix} \begin{bmatrix} \hat{Q}^\top X_0 \\ U_0 \end{bmatrix}^\dagger$$

2. Design LQR controller for ROM (with LQR cost matrices P, R):

$$\hat{K}_r = \mathrm{LQR}(\hat{F}, \hat{G}, \hat{Q}^\top P \hat{Q}, R)$$

3. Lift controller to full state-space:

$$\hat{K}_f = \hat{K}_r \hat{Q}^\top$$

Data-Driven ROM LQR Pipeline

1. Choose subspace $\hat{Q}$ using POD and compute $\hat{F}, \hat{G}$ as the least-squares solution:

$$\begin{bmatrix} \hat{F} & \hat{G} \end{bmatrix} = \begin{bmatrix} \hat{Q}^\top X_1 \end{bmatrix} \begin{bmatrix} \hat{Q}^\top X_0 \\ U_0 \end{bmatrix}^\dagger$$

2. Design LQR controller for ROM (with LQR cost matrices P, R):

$$\hat{K}_r = \mathrm{LQR}(\hat{F}, \hat{G}, \hat{Q}^\top P \hat{Q}, R)$$

3. Lift controller to full state-space:

$$\hat{K}_f = \hat{K}_r \hat{Q}^\top$$

How well does POD actually work for this method?

Lifted Low-Rank Structures

1. LQR regularity: (A, B) is stabilizable and (A, P) is detectable.
2. Low-rank: there exists matrices $F \in \mathbb{R}^{r \times r}$, $G \in \mathbb{R}^{r \times n_u}$, $P_r \in \mathbb{R}^{r \times r}$ and $Q \in \mathbb{R}^{n_x \times r}$ with orthonormal columns for $r \leq n_x$ such that

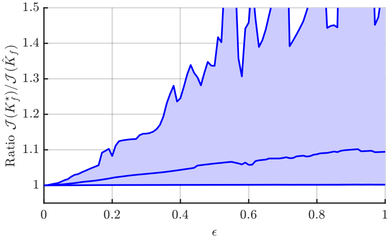
$$A = QFQ^\top \ , \ B = QG \ , \ P = QP_rQ^\top . \tag{6}$$

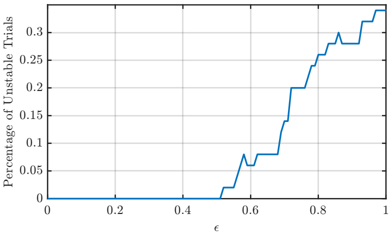
3. Identifiability: the pair (F, G) in (2) is controllable.
4. Initial-state subspace consistency: $x_0 \in \mathrm{span}(Q)$.

Lifted Low-Rank Systems

Theorem 1 (Informal Theorem Details**)**

For a lifted low-rank LQR problem with sufficiently exciting input, the POD controller found with the data-driven LQR pipeline is equivalent to the optimal LQR controller for the full-order system.





Parameters: $n_x = 100$, $n_u = 5$, $r = 20$, $n_s = 125$, $x_0 = 0$

Future Work

- More work incorporating missing data methods and streaming data SVD methods into ROM
- ROM for heterogeneous fidelity data
- Optimization for ROM
  1. Theory for other controllers besides LQR, and other downstream tasks
  2. Perturbation theory for approximately low-rank systems
  3. Techniques for identifying:
     - when POD is sufficient
     - when POD is not sufficient but an optimization-improved subspace is sufficient
     - when nonlinear model structure is needed (see also [Geelen et al., 2024])

References I

Alarcon, D. P., Balzano, L., and Nowak, R. (2016).
Necessary and sufficient conditions for sketched subspace clustering.
In *proceedings of the Allerton Conference on Communications, Control, and Computing*.

Anderson, B. D. and Moore, J. B. (2007).
Optimal control: linear quadratic methods.
Dover.

Annoni, J. and Seiler, P. (2017).
A method to construct reduced-order parameter-varying models.
International Journal of Robust and Nonlinear Control, 27(4):582–597.

Balzano, L. (2022).
On the equivalence of oja's algorithm and grouse.
In *International Conference on Artificial Intelligence and Statistics*, pages 7014–7030. PMLR.

Balzano, L., Nowak, R., and Recht, B. (2010).
Online identification and tracking of subspaces from highly incomplete information.
In *Proceedings of the Allerton conference on Communication, Control, and Computing*.

Balzano, L. and Wright, S. J. (2013).
On GROUSE and incremental SVD.
In *IEEE Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*.

References II

Balzano, L. and Wright, S. J. (2014).
Local convergence of an algorithm for subspace identification from partial data.
Foundations of Computational Mathematics, pages 1–36.

Bilen, Ç., Puy, G., Gribonval, R., and Daudet, L. (2014).
Convex optimization approaches for blind sensor calibration using sparsity.
IEEE Transactions on Signal Processing, 62(18):4847–4856.

Blocker, C. J., Raja, H., Fessler, J. A., and Balzano, L. (2023).
Dynamic subspace estimation with grassmannian geodesics.
arXiv preprint arXiv:2303.14851.

Bower, A., Jain, L., and Balzano, L. (2018).
The landscape of non-convex quadratic feasibility.
In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3974–3978. IEEE.

Brand, M. (2002).
Incremental singular value decomposition of uncertain data with missing values.
In *European Conference on Computer Vision*, pages 707–720. Springer.

Bunch, J. and Nielsen, C. (1978).
Updating the singular value decomposition.
Numerische Mathematik, 31:111–129.
10.1007/BF01397471.

References III

Candès, E. J. and Fernandez-Granda, C. (2013).
Super-resolution from noisy data.
Journal of Fourier Analysis and Applications, 19(6):1229–1254.

Candès, E. J. and Recht, B. (2009).
Exact matrix completion via convex optimization.
Foundations of Computational mathematics, 9(6):717–772.

Cavazos, J. S., Fessler, J. A., and Balzano, L. (2023).
Alpcah: Sample-wise heteroscedastic pca with tail singular value regularization.
In *2023 International Conference on Sampling Theory and Applications (SampTA)*, pages 1–6. IEEE.

Chen, Y., Chi, Y., Fan, J., and Ma, C. (2019).
Gradient descent with random initialization: Fast global convergence for nonconvex phase retrieval.
Mathematical Programming, 176(1):5–37.

Chen, Y., Yi, X., and Caramanis, C. (2017).
Convex and nonconvex formulations for mixed regression with two components: Minimax optimal rates.
IEEE Transactions on Information Theory, 64(3):1738–1766.

Chi, Y., Eldar, Y. C., and Calderbank, R. (2012).
Petrels: Parallel estimation and tracking of subspace by recursive least squares from partial observations.
submitted to IEEE Trans. Sig. Proc., arXived.

References IV

Chi, Y., Lu, Y. M., and Chen, Y. (2019).
Nonconvex optimization meets low-rank matrix factorization: An overview.
IEEE Transactions on Signal Processing, 67(20):5239–5269.

Eckart, C. and Young, G. (1936).
The approximation of one matrix by another of lower rank.
Psychometrika, 1(3):211–218.

Feng, C., Lagoa, C. M., Ozay, N., and Sznaier, M. (2010).
Hybrid system identification: An SDP approach.
In *Decision and Control (CDC), 2010 49th IEEE Conference on*, pages 1546–1552. IEEE.

Geelen, R., Balzano, L., Wright, S., and Willcox, K. (2024).
Learning physics-based reduced-order models from data using nonlinear manifolds.
Chaos: An Interdisciplinary Journal of Nonlinear Science, 34(3).

Ghashami, M., Liberty, E., Phillips, J. M., and Woodruff, D. P. (2016).
Frequent directions: Simple and deterministic matrix sketching.
SIAM Journal on Computing, 45(5):1762–1792.

Gilman, K. and Balzano, L. (2019).
Panoramic video separation with online grassmannian robust subspace estimation.
In *Proceedings of the IEEE International Conference on Computer Vision Workshops*.

References V

Gilman, K., Burer, S., and Balzano, L. (2025a).
A semidefinite relaxation for sums of heterogeneous quadratic forms on the stiefel manifold.
SIAM journal on matrix analysis and applications, 46(2):1091–1116.

Gilman, K., Hong, D., Fessler, J. A., and Balzano, L. (2025b).
Streaming heteroscedastic probabilistic pca with missing data.
Transactions on machine learning research.

Gitlin, A., Tao, B., Balzano, L., and Lipor, J. (2018).
Improving k-subspaces via coherence pursuit.
Accepted to the Journal of Selected Topics in Signal Processing.

Hardt, M., Ma, T., and Recht, B. (2018).
Gradient descent learns linear dynamical systems.
Journal of Machine Learning Research, 19(29):1–44.

He, J., Balzano, L., and Szlam, A. (2012).
Incremental gradient on the grassmannian for online foreground and background separation in subsampled video.
In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

He, J., Zhang, D., Balzano, L., and Tao, T. (2014).
Iterative grassmannian optimization for robust image alignment.
Image and Vision Computing, 32(10):800–813.

References VI

Hong, D., Balzano, L., and Fessler, J. A. (2018).
Asymptotic performance of pca for high-dimensional heteroscedastic data.
Journal of Multivariate Analysis, 167:435 – 452.

Hong, D., Balzano, L., and Fessler, J. A. (2019).
Probabilistic PCA for heteroscedastic data.
In *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 26–30.

Hong, D., Gilman, K., Balzano, L., and Fessler, J. A. (2021).
HePPCAT: probabilistic PCA for data with heteroscedastic noise.
IEEE Transactions on Signal Processing.

Hong, D., Yang, F., Fessler, J. A., and Balzano, L. (2023).
Optimally weighted pca for high-dimensional heteroscedastic data.
SIAM Journal on Mathematics of Data Science, 5(1):222–250.

Hotelling, H. (1933).
Analysis of a complex of statistical variables into principal components.
Journal of educational psychology, 24(6):417.

Jeffers, J. N. R. (1967).
Two case studies in the application of principal component analysis.
Journal of the Royal Statistical Society. Series C (Applied Statistics), 16(3):225–236.

References VII

Kennedy, R., Balzano, L., Wright, S. J., and Taylor, C. J. (2016).
Online algorithms for factorization-based structure from motion.
Computer Vision and Image Understanding.

Kennedy, R., Taylor, C. J., and Balzano, L. (2014).
Online completion of ill-conditioned low-rank matrices.
In *Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on*, pages 507–511. IEEE.

Kileel, J., Klock, T., and M Pereira, J. (2021).
Landscape analysis of an improved power method for tensor decomposition.
Advances in Neural Information Processing Systems, 34:6253–6265.

Kwon, S. M., Zhang, Z., Song, D., Balzano, L., and Qu, Q. (2024).
Efficient compression of overparameterized deep models through low-dimensional learning dynamics.
In *Proceedings of AI Stats*.

Li, Y., Balzano, L., Needell, D., and Lyu, H. (2023).
Convergence and complexity of block majorization-minimization for constrained block-riemannian optimization.
arXiv preprint arXiv:2312.10330.

Li, Y., Balzano, L., Needell, D., and Lyu, H. (2024).
Convergence and complexity guarantee for inexact first-order riemannian optimization algorithms.
In *International Conference on Machine Learning*, pages 27376–27398. PMLR.

References VIII

Ling, S. and Strohmer, T. (2019).
Regularized gradient descent: a non-convex recipe for fast joint blind deconvolution and demixing.
Information and Inference: A Journal of the IMA, 8(1):1–49.

Lipor, J. and Balzano, L. (2017).
Leveraging union of subspace structure to improve constrained clustering.
In *Proceedings of the International Conference on Machine Learning (ICML)*.

Lipor, J., Hong, D., Tan, Y. S., and Balzano, L. (2021).
Subspace clustering using ensembles of k-subspaces.
Information and Inference: A Journal of the IMA, 10(1):73–107.

Liu, H., Wang, P., Huang, L., Qu, Q., and Balzano, L. (2024).
Symmetric matrix completion with relu sampling.
In *Forty-first International Conference on Machine Learning*.

Lyu, H., Needell, D., and Balzano, L. (2020).
Online matrix factorization for markovian data and applications to network dictionary learning.
Journal of Machine Learning Research, 21(251):1–49.

Meyers, J. and Meneveau, C. (2012).
Optimal turbine spacing in fully developed wind farm boundary layers.
Wind energy, 15(2):305–317.

References IX

Newton, R., Du, Z., Balzano, L., and Seiler, P. (2023).
Manifold optimization for data driven reduced-order modeling.
In *59th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1–6.

Ongie, G., Pimentel-Alarcón, D., Balzano, L., Willett, R., and Nowak, R. D. (2021).
Tensor methods for nonlinear matrix completion.
SIAM Journal on Mathematics of Data Science, 3(1):253–279.

Ongie, G., Willett, R., Nowak, R. D., and Balzano, L. (2017).
Algebraic variety models for high-rank matrix completion.
In *International Conference for Machine Learning (ICML)*.

Pearson, K. (1901).
On lines and planes of closest fit to systems of points in space.
The London, Edinburgh, and Dublin philosophical magazine and journal of science, 2(11):559–572.

Ritchie, A., Balzano, L., and Scott, C. (2020).
Supervised pca: A multiobjective approach.
arXiv preprint arXiv:2011.05309.

Schmidt, E. (1907).
Zur theorie der linearen und nicht linearen integralgleichungen. i teil. entwicklung willkürlichen funkionen nach system vorgeschriebener.
Mathematische Annalen, 63:433–476.

References X

Spearman, C. (1904).
"general intelligence," objectively determined and measured.
The American Journal of Psychology, 15(2):201–292.

Stewart, G. (2011).
Fredholm, hilbert, schmidt: Three fundamental papers on integral equations.
Available at www.cs.umd.edu/~stewart/FHS.pdf.

Stewart, G. W. (1993).
On the early history of the singular value decomposition.
SIAM review, 35(4):551–566.

Sun, J., Qu, Q., and Wright, J. (2017a).
Complete dictionary recovery over the sphere i: Overview and the geometric picture.
IEEE Transactions on Information Theory, 63(2):853–884.

Sun, J., Qu, Q., and Wright, J. (2017b).
Complete dictionary recovery over the sphere ii: Recovery by riemannian trust-region method.
IEEE Transactions on Information Theory, 63(2):885–914.

Udell, M. and Townsend, A. (2019).
Why are big data matrices approximately low rank?
SIAM Journal on Mathematics of Data Science, 1(1):144–160.

References XI

Wang, P., Liu, H., So, A. M.-C., and Balzano, L. (2022).
Convergence and recovery guarantees of the k-subspaces method for subspace clustering.
In *International Conference on Machine Learning*, pages 22884–22918. PMLR.

Willems, J. C., Rapisarda, P., Markovsky, I., and De Moor, B. L. (2005).
A note on persistency of excitation.
Systems & Control Letters, 54(4):325–329.

Wright, J., Ganesh, A., Rao, S., Peng, Y., and Ma, Y. (2009).
Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization.
Advances in neural information processing systems, 22:2080–2088.

Xiao, P. and Balzano, L. (2016).
Online sparse and orthogonal subspace estimation from partial information.
In *Proceedings of the Allerton conference on Communication, Control, and Computing*.

Yang, D., Tang, G., and Wakin, M. B. (2016).
Super-resolution of complex exponentials from modulations with unknown waveforms.
IEEE Transactions on Information Theory, 62(10):5809–5830.

Yaras, C., Wang, P., Balzano, L., and Qu, Q. (2024).
Compressible dynamics in deep overparameterized low-rank learning & adaptation.
In *Proceedings of ICML*.

References XII

Yaras, C., Wang, P., Hu, W., Zhu, Z., Balzano, L., and Qu, Q. (2023).
The law of parsimony in gradient descent for learning deep linear networks.
arXiv preprint arXiv:2306.01154.

Yaras, C., Wang, P., Zhu, Z., Balzano, L., and Qu, Q. (2022).
Neural collapse with normalized features: A geometric analysis over the riemannian manifold.
Advances in neural information processing systems, 35:11547–11560.

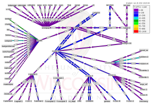
Zhang, D. and Balzano, L. (2016).
Global convergence of a grassmannian gradient descent algorithm for subspace estimation.
In *Proceedings of Artificial Intelligence and Statistics*.

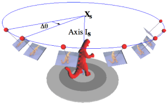
Zhang, D. and Balzano, L. (2018).
Convergence of a grassmannian gradient descent algorithm for subspace estimation from undersampled data.
Submitted to Foundations on Computational Mathematics.
Preprint available at `https://arxiv.org/abs/1610.00199`.

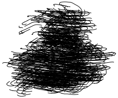
Example Applications I

1. Missing Data SVD (Low-Rank Matrix Completion):
   P_Ω projects onto the coordinates $\Omega \subset \{1, \dots, n\}$.

$$\underset{U \in \mathbb{R}^{n \times d}, W \in \mathbb{R}^{N \times d}}{\text{minimize}} \quad \|P_\Omega(UW^T - Y)\|_F^2 \quad \text{s.t. } U \in \mathcal{G}(n, d) \quad (7)$$







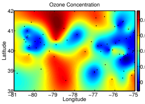
Example Applications II

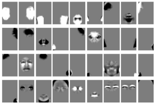
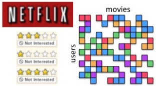
2. Robust SVD:

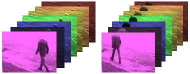
$$\underset{U\in\mathbb{R}^{n\times d}, W\in\mathbb{R}^{N\times d}}{\text{minimize}} \quad \|P_\Omega(UW^T - Y)\|_1 \quad \text{s.t.} \quad U \in \mathcal{G}(n,d) \tag{8}$$

3. Sparse SVD:

$$\underset{U\in\mathbb{R}^{n\times d}, W\in\mathbb{R}^{N\times d}}{\text{minimize}} \quad \|P_\Omega(UW^T - Y)\|_F^2 \text{ s.t. } \|U\|_1 \le \tau, \;\; U \in \mathcal{G} \tag{9}$$





Example Applications III

4. Heteroscedastic PCA:

$$\underset{U \in \mathbb{R}^{n \times d}}{\text{minimize}} \quad \ell(U, Y) \text{subject to} \quad U \in \mathcal{G}(n, d) \tag{10}$$

where $\ell()$ is the negative log likelihood for heteroscedastic PCA.

5. Dictionary learning / Non-neg Matrix Factorization:

$$\begin{aligned} \underset{U \in \mathbb{R}^{n \times m}, W \in \mathbb{R}^{N \times m} : m \geq n}{\text{minimize}} \quad & \|UW^T - Y\|_F^2 \\ \text{subject to} \quad & \|W\|_1 \leq \tau \\ & W_{ij} \geq 0 \end{aligned} \tag{11}$$

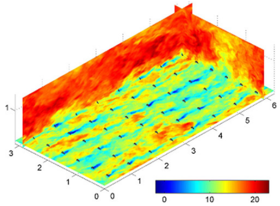
Example Applications IV

6. Calibration SVD:

$$\begin{aligned} \underset{U\in\mathbb{R}^{n\times d}, W\in\mathbb{R}^{N\times d}, g}{\text{minimize}} \quad & \|P_\Omega\left(g(UW^T - Y)\right)\|_F^2 \\ \text{subject to} \quad & g \text{ is } L-\text{Lipschitz and monotonic} \\ & U \in \mathcal{G}(n,d) \end{aligned} \tag{12}$$

7. Data-Driven ROM

Collect input-output data from a (potentially nonlinear) discrete-time dynamical system, and find a system matrices for a low-rank linear system that best fits the data.



Wind farm simulation with 6.3 million states [Meyers and Meneveau, 2012].

Example Applications V

$$\underset{Q,F,G,H,D}{\text{minimize}} \quad \left\| \begin{bmatrix} X_1 \\ Y_0 \end{bmatrix} - \begin{bmatrix} Q & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} F & G \\ H & D \end{bmatrix} \begin{bmatrix} Q^\top & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} X_0 \\ U_0 \end{bmatrix} \right\|_F^2 \tag{13}$$

$$\text{subject to} \quad Q \in \mathcal{G}(n, r)$$

Astronomical Imaging

Courtesy of `https://noirlab.edu/public/news/noao1901/`

Imaging the night sky is done with a variety of instrumentation whose images are subject to nightly conditions.

- moonlight
- haze, clouds, other particulate matter
- imaging at the edge of the field of view

The measurements are noisier on some nights or for some locations in the sky.

New Directions: Maximum Likelihood

$$y_i = U\Theta z_i + \eta_i \varepsilon_i$$

Suppose z_i and ε_i are all iid normal $\mathcal{N}(0,1)$. Then the maximum likelihood estimator for U, Θ is given as

$$\max_{\substack{U \in \mathcal{G}(d,k) \\ \theta_i, i=1,\dots,k}} \frac{n}{2} \sum_{i=1}^{n} \sum_{j=1}^{k} \log\left(\frac{1}{\eta_i^2} - \frac{1}{\eta_i^2} \frac{\theta_j^2}{\theta_j^2 + \eta_i^2}\right) + \frac{1}{2}\left(\sum_{i=1}^{n} y_i^T U \Gamma_i U^T y_i\right) .$$

where Γ_i is a $k \times k$ diagonal matrix with entries $\frac{1}{\eta_i^2}\frac{\theta_j^2}{\theta_j^2+\eta_i^2}$. Given θ_i (and therefore Γ_i) then the maximum likelihood U is identified by

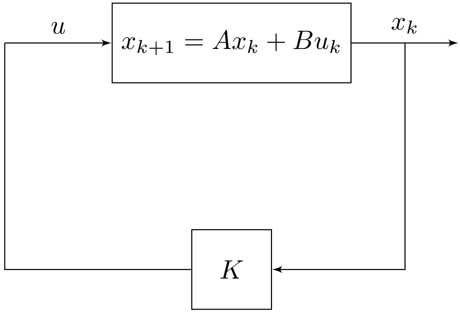
$$\max_{U \in \mathcal{G}(d,k)} \sum_{i=1}^{n} y_i^T U \Gamma_i U^T y_i \ .$$

Linear-Quadratic Regulator

Denote $\mathrm{LQR}(A, B, P, R)$ as the state-feedback problem

$$\begin{aligned} \min_{K} \quad & \sum_{k=0}^{\infty} \left(x_k^\top P x_k + u_k^\top R u_k \right) \\ \text{subject to} \quad & x_{k+1} = A\, x_k + B\, u_k, \\ & u_k = -K\, x_k. \end{aligned}$$

When (A, B) are known, we can already solve this exactly, but the computation scales with $\mathcal{O}(n_x^3)$.



Optimality of POD

Theorem 2 (Informal)

For a lifted low-rank LQR problem with sufficiently exciting input, the POD controller found with the data-driven LQR pipeline is equivalent to the optimal LQR controller for the full-order system.

Proof outline:

1. Prove POD identifies the correct subspace using persistently exciting assumption.
2. Prove equivalence of $K_f^\star = \mathrm{LQR}(A, B, P, R)$ and POD controller using low-rank assumptions.

Perturbations to the Low-Rank Assumption

Given parameters

Low-order system	$F \in \mathbb{R}^{r \times r}$	$G \in \mathbb{R}^{r \times n_u}$
Perturbations	$F_\perp \in \mathbb{R}^{(n_x - r) \times (n_x - r)}$	$G_\perp \in \mathbb{R}^{(n_x - r) \times n_u}$
Orthogonal matrix	$\begin{bmatrix} Q & Q_\perp \end{bmatrix} \in \mathbb{R}^{n_x \times n_x}$,	$\begin{bmatrix} Q & Q_\perp \end{bmatrix}^\top \begin{bmatrix} Q & Q_\perp \end{bmatrix} = I$
LQR Costs	$R = I_{n_u}$	$P = QP_rQ^T$

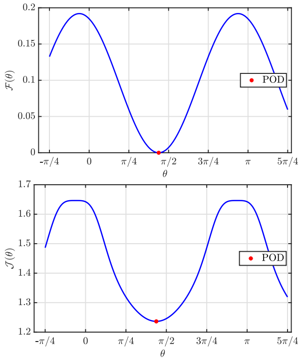
we generate an ϵ-perturbed lifted low-order system defined by

$$A = QFQ^\top + \epsilon Q_\perp F_\perp Q_\perp^\top$$
$$B = QG + \epsilon Q_\perp G_\perp$$

to evaluate the impact of breaking the low-rank system matrices assumption.

Small Exactly Low-Rank Example

- Problem setup
  - Generate a random $Q \in \mathbb{R}^{2\times 1}$, $F \in \mathbb{R}$, $G \in \mathbb{R}$ and construct lifted low-rank system as $A = QFQ^\top$, $B = QG$, $P = QQ^\top$, and $R = 1$.
  - Simulate and gather snapshots for $n_s = 4$ steps.
  - Define $\hat{Q}(\theta) = [\sin(\theta), \cos(\theta)]^\top$.
- POD controller is optimal for
  - Least-squares ROM cost (upper)
  - End-to-end data-driven ROM LQR cost (lower)



LQR Solution

Lemma 3 (Section 3.3 of [Anderson and Moore, 2007])

When (A, B) is stabilizable and (A, Q) is detectable, there is a unique solution $P_f^\star \succeq 0$ to the following DARE:

$$P - A^\top PA - Q - A^\top PB\,(R + B^\top PB)^{-1}\,B^\top PA = 0 \qquad (14)$$

such that

$$K_f^\star := (R + B^\top P_f^\star B)^{-1} B^\top P_f^\star A \qquad (15)$$

stabilizes the linear discrete time system (3) *and minimizes the quadratic cost in* (20) *with $\mathcal{J}(K_f^\star, x_0) = x_0^\top P_f^\star x_0$ for all initial conditions.* LQR Slides

LQR Intermediate Results II

Proposition 4

Consider discrete LTI system matrices (A, B) and LQR cost matrices (Q, R) that satisfy LLR Assumptions. Denote the Riccati solution and optimal controller for LQR(A, B, P, R) as $(P_f^\star, K_f^\star)$. Let $\hat{Q} \in \mathbb{R}^{n_x \times r}$ be any matrix with orthonormal columns such that $\mathrm{span}(V) = \mathrm{span}(\hat{Q})$. Define $\hat{F} := \hat{Q}^\top A \hat{Q}$, $\hat{G} := \hat{Q}^\top B$, and $\hat{Q}_r := \hat{Q}^\top P \hat{Q}$. Then there exists a unique stabilizing solution, that we denote $(\hat{P}_r^\star, \hat{K}_r^\star)$, to the reduced-order LQR$(\hat{F}, \hat{G}, \hat{P}_r, R)$. Moreover,

$$P_f^\star = \hat{Q}\hat{P}_r^\star\hat{Q}^\top, \quad K_f^\star = \hat{K}_r^\star\hat{Q}^\top. \tag{16}$$

LQR Intermediate Results III

Lemma 5 (extension of [Willems et al., 2005, Cor. 2])

Consider discrete LTI system matrices (A, B) that satisfy LLR Assumptions with input $\{u_k\}_{k=0}^{n_s-1}$ persistently exciting of order $r+1$. Then

1. $\mathrm{rank}(X_0) = r$, *and*
2. $\mathrm{rank}\left(\begin{bmatrix} X_0 \\ U_0 \end{bmatrix}\right) = r + n_u$

LQR Intermediate Results I

Proposition 6

Consider discrete LTI system matrices (A, B) that satisfy LLR Assumptions with input $\{u_k\}_{k=0}^{n_s-1}$ persistently exciting of order $r+1$.
Let $\hat{Q}$ be the POD projection matrix and $(\hat{F}, \hat{G})$ be the resulting reduced-order model matrices. Then

$$\mathrm{span}(V) = \mathrm{span}(\hat{Q}) \tag{17}$$

and

$$\hat{F} = T^{-1}FT\,,\ \hat{G} = T^{-1}G \tag{18}$$

for the orthogonal matrix $T := V^\top \hat{Q}$.

LQR Theorem Formal Statement

Theorem 7 (LQR Theorem Slide)

Consider discrete LTI system matrices (A, B) *and LQR*(A, B, P, R) *problem that satisfy Lifted Low-Rank Assumptions with input* $\{u_k\}_{k=0}^{n_s-1}$ *persistently exciting of order* $r+1$. *Then, the POD controller* $\hat{K}_f$ *is the optimal LQR controller for the full-order system, i.e.* $\hat{K}_f = K_f^\star$.