A Regularized Linear Dynamical System Framework for Multivariate Time Series Analysis

Zitao Liu and Milos Hauskrecht
{ztliu, milos}@cs.pitt.edu
Department of Computer Science, University of Pittsburgh

Introduction

Multivariate time series (MTS) analysis is an important statistical tool to study the behavior of time dependent data and to forecast its future values depending on the history of variations in the data.

Challenges in MTS Modeling

- A large number of MTS collected in the real-world problems have a relatively short span.
- The number of MTS instances available is often limited.

Overfitting Problem!

Linear Dynamical Systems

The Linear Dynamical System (LDS) is arguably the most commonly used MTS model for real-world application.

$$\mathbf{z}_t = A\mathbf{z}_{t-1} + \boldsymbol{\epsilon}_t; \quad \mathbf{y}_t = C\mathbf{z}_t + \boldsymbol{\zeta}_t$$

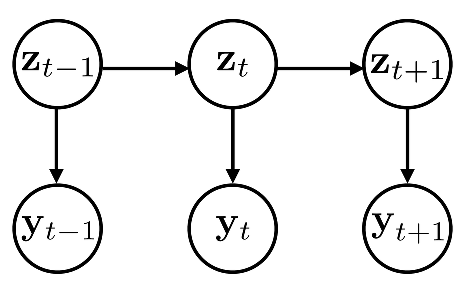


Figure 1: Graphical illustration of an LDS.

- Real-valued MTS observations $\{\mathbf{y}_t \in \mathbb{R}^n\}_{t=1}^T$.
- Hidden states $\{\mathbf{z}_t \in \mathbb{R}^d\}_{t=1}^T$.
- Transition matrix $A \in \mathbb{R}^{d\times d}$.
- Emission matrix $C \in R^{n\times d}$.
- $\{\boldsymbol{\epsilon}_t\}_{t=1}^T \sim \mathcal{N}(\mathbf{0}, Q)$ and $\{\boldsymbol{\zeta}_t\}_{t=1}^T \sim \mathcal{N}(\mathbf{0}, R)$.
- Initial state $\mathbf{z}_1 \sim \mathcal{N}(\boldsymbol{\xi}, \Psi)$.

The complete set of the LDS parameters is $\Omega = \{A, C, Q, R, \boldsymbol{\xi}, \Psi\}$.

Question when an LDS is learned from MTS

Learning an LDS model from **short-span low-sample** MTS datasets gives rise to numerous important questions:

? *How many hidden states are needed to represent the highly dependent MTS well?*

? *How do we prevent the overfit of the model parameters when the training size is small?*

Goal

In this work we address the above issues by presenting a regularized LDS framework (rLDS) which

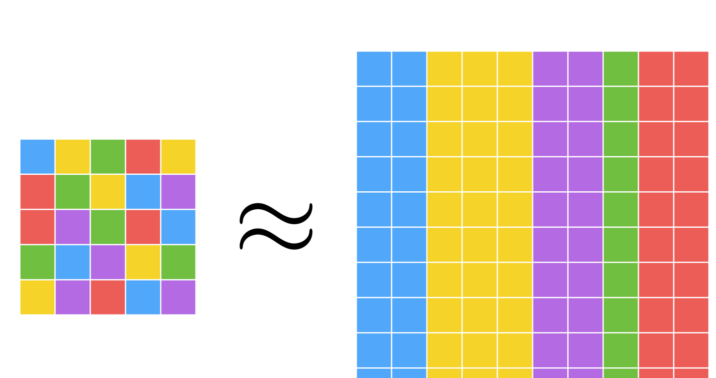
✓ recovers the intrinsic dimensionality of MTS.
✓ prevents model overfitting given short MTS datasets.
✓ supports accurate MTS forecasting.

The Regularized LDS Framework

Intuition

Recover the intrinsic dimensionality of MTS by ***minimizing the rank of the transition matrix rather than the state space size***.

***A* is a *d* x *d* matrix**



rLDS Framework

In rLDS, the LDS has a large implicit state space but a low-rank transition matrix. The rLDS recovers the intrinsic dimensionality of MTS by using ***the rank of transition matrix rather than the state space size***.

In order to achieve the low-rank property, we introduce a prior, i.e., $p(A)$ for the hidden state transition matrix *A*. The log joint probability distribution for our rLDS is:

$$\log\Big(p(\mathbf{z},\mathbf{y},A)\Big) = \log p(\mathbf{z}_1) + \sum_{t=1}^T p(\mathbf{y}_t|\mathbf{z}_t) + \sum_{t=2}^T \log p(\mathbf{z}_t|\mathbf{z}_{t-1},A) + \log p(A)$$

Learning

E-step(Inference)

The E-step requires computing the expected log likelihood of the log joint probability with respect to the hidden state distribution, i.e., $\mathcal{Q} = \mathbb{E}_{\mathbf{z}}[\log p(\mathbf{z},\mathbf{y},A|\Omega)]$, which depends on 3 sufficient statistics $\mathbb{E}[\mathbf{z}_t|\mathbf{y}]$, $\mathbb{E}[\mathbf{z}_t\mathbf{z}_t'|\mathbf{y}]$ and $\mathbb{E}[\mathbf{z}_t\mathbf{z}_{t-1}'|\mathbf{y}]$. Here we follow the standard Kalman smoother backward algorithm to compute them [1].

M-step(Learning)

In the M-step, we try to find Ω that maximizes the likelihood lower bound $\mathcal{Q}$.

O1: Optimization of A

In each iteration in the M-step, we need to maximize $\mathbb{E}_{\mathbf{z}}\Big[\sum_{t=2}^T \log p(\mathbf{z}_t|\mathbf{z}_{t-1},A)\Big] + \log p(A)$ with respect to *A*, which is equivalent to

$$\min_A \underbrace{\frac{1}{2}\sum_{t=2}^T \mathbb{E}_{\mathbf{z}}\Big[(\mathbf{z}_t - A\mathbf{z}_{t-1})'Q^{-1}(\mathbf{z}_t - A\mathbf{z}_{t-1})\Big] - \log p(A)}_{g(A)}$$

I1: rLDS$_{\mathcal{G}}$ with multivariate Laplacian priors

$$\log p(A|\lambda_1,\lambda_3) = -\lambda_1\sum_{i=1}^d \|A_i\|_2 - \frac{\lambda_3}{2}\|A\|_F^2 + const,$$

and the objective function we want to optimize becomes:

$$\min_A g(A) + \frac{\lambda_3}{2}\|A\|_F^2 + \lambda_1\sum_{i=1}^d \|A_i\|_2$$

I2: rLDS$_{\mathcal{R}}$ with a nuclear norm prior

$$\log p(A|\lambda_2,\lambda_3) = -\lambda_2\|A\|_* - \frac{\lambda_3}{2}\|A\|_F^2 + const,$$

and the objective function we want to optimize becomes:

$$\min_A g(A) + \frac{\lambda_3}{2}\|A\|_F^2 + \lambda_2\|A\|_*$$

O2: Optimization of $\Omega\backslash A = \{C, R, Q, \boldsymbol{\xi}, \Psi\}$

Each of these parameters is estimated by taking the corresponding derivative of $\mathcal{Q} = \mathbb{E}_{\mathbf{z}}[\log p(\mathbf{z},\mathbf{y},A|\Omega)]$, setting it to zero, and by solving it analytically [1].

$$C^{(k+1)} = \Big(\sum_{t=1}^T \mathbf{y}_t\mathbb{E}[\mathbf{z}_t|\mathbf{y}]'\Big)\Big(\sum_{t=1}^T \mathbb{E}[\mathbf{z}_t\mathbf{z}_t'|\mathbf{y}]\Big)^{-1}$$

$$R^{(k+1)} = \frac{1}{T}\sum_{t=1}^T (\mathbf{y}_t\mathbf{y}_t' - C^{(k+1)}\mathbb{E}[\mathbf{z}_t|\mathbf{y}]\mathbf{y}_t')$$

$$Q^{(k+1)} = \frac{1}{T-1}\Big(\sum_{t=2}^T \mathbb{E}[\mathbf{z}_t\mathbf{z}_t'|\mathbf{y}] - A^{(k+1)}\sum_{t=2}^T \mathbb{E}[\mathbf{z}_t\mathbf{z}_{t-1}'|\mathbf{y}]\Big)$$

$$\boldsymbol{\xi}^{(k+1)} = \mathbb{E}[\mathbf{z}_1|\mathbf{y}]$$

$$\Psi^{(k+1)} = \mathbb{E}[\mathbf{z}_1\mathbf{z}_1'|\mathbf{y}] - \mathbb{E}[\mathbf{z}_1|\mathbf{y}]\mathbb{E}[\mathbf{z}_1|\mathbf{y}]'$$

Experiments

Evaluation Metrics

$$\text{Average-MAPE} = \frac{1}{nT}\sum_{i=1}^n\sum_{j=1}^T |1 - \hat{y}_{ij}/y_{ij}| \times 100\%$$

where $|\cdot|$ denotes the absolute value; y_{ij} and $\hat{y}_{ij}$ are the *j*th true and predicted observations from time series *i*. *n* is the number of time series and *T* is the length of a MTS.

Baselines

- LDS learned using the standard EM algorithm (EM).
- Subspace identification algorithm (SubspaceID).
- Stable linear dynamical system (StableLDS).

Recovery of Intrinsic Dimensionality

We generate a synthetic MTS dataset of length $T = 200$ using a 20-state LDS with zero-mean, 0.01 variance Gaussian innovations. We uniformly randomly generate the transition matrix *A* and emission matrix *C*. The transition matrix *A* is truncated to an exact 10-rank matrix.

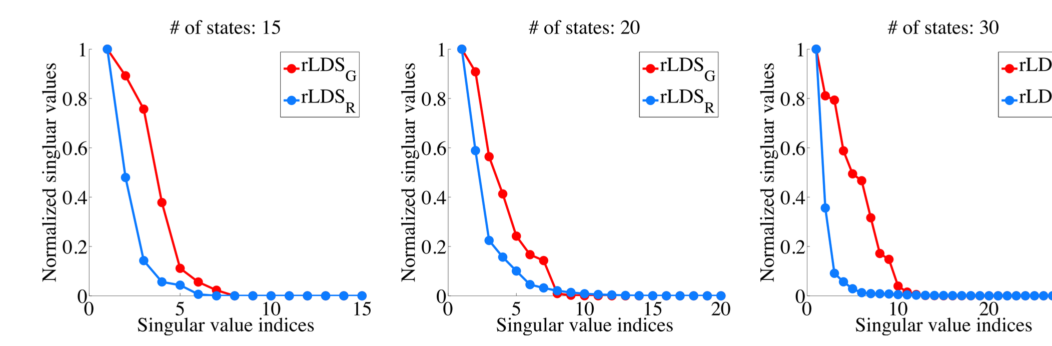


Figure 2: State space recovery on a synthetic dataset.

Overfitting Phenomena

Two real-world datasets:

- PB: Production and billing figures data [2]. The data is a bi-variate time series of length $T = 100$.
- CL: A MTS clinical data obtained from electronic health records of 500 post-surgical cardiac patients in PCP database [3].

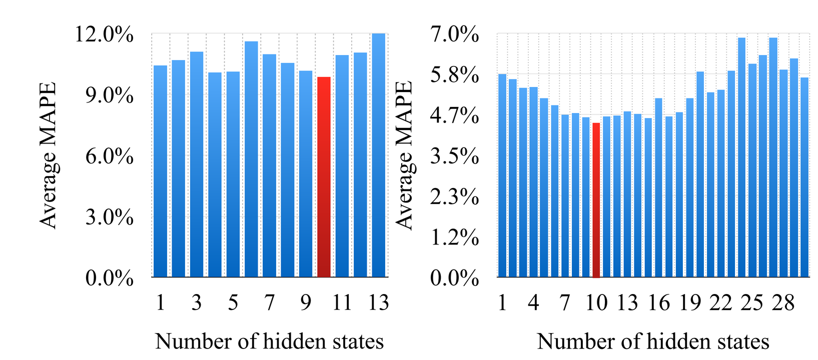


Figure 3: Overfitting Phenomena on both PB and CL datasets.

Sparsification Effects

State space recovery results on both PB and CL datasets are shown in Figure 4 and Figure 4.

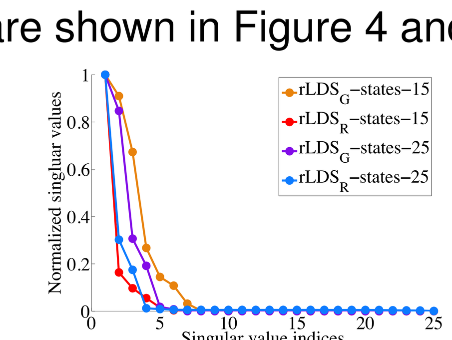


Figure 4: PB dataset.

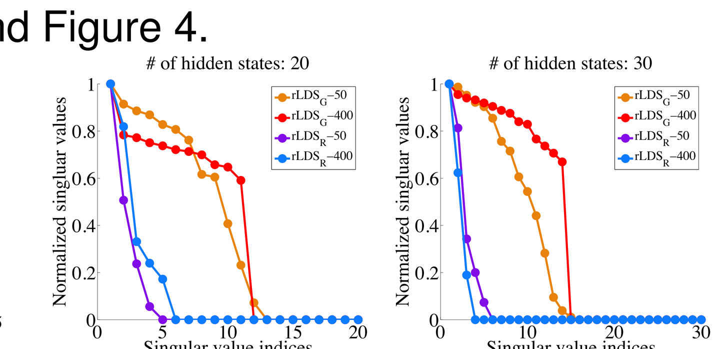


Figure 5: CL dataset.

Prediction Performance

In order to gain a more comprehensive insight into rLDS's prediction abilities, we explored different initial state space sizes. The results of these experiments on CL dateset are summarized in Table 1.

Table 1: Average-MAPE results on CL dataset with different training sizes.

	Training Size: 50			Training Size: 400		
# of states	10	20	30	10	20	30
EM	6.28	17.24	23.98	**4.43**	5.91	5.72
SubspaceID	6.55	6.99	7.44	6.10	6.16	6.27
StableLDS	6.54	6.99	7.40	6.10	6.16	6.27
rLDS$_{\mathcal{G}}$	4.98	4.97	**4.86**	4.51	**4.25**	**4.35**
rLDS$_{\mathcal{R}}$	**4.65**	**4.95**	5.01	4.65	4.46	4.67

Conclusion

In this paper, we presented a regularized LDS learning framework for MTS modeling, whose advantages are: (1) it automatically seeks the intrinsic state dimensionality; (2) it is robust in preventing model overfitting even for a small amount of MTS data; and (3) it is able to make accurate MTS prediction.

Acknowledgement

The work in this paper was supported by grants R01GM088224 and R01LM010019 from the NIH. Its content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Reference

[1] Ghahramani, Zoubin, and Geoffrey E. Hinton. Parameter estimation for linear dynamical systems. Technical Report CRG-TR-96-2, University of Totronto, Dept. of Computer Science, 1996.
[2] `http://www.stat.wisc.edu/~reinsel/emtsa-data/prod-bill`.
[3] Hauskrecht, Milos, et al. "Conditional outlier detection for clinical alerting." AMIA Annual Symposium Proceedings. Vol. 2010. American Medical Informatics Association, 2010.