Finite Sample System Identification: Improved Rates and the Role of Regularization

Yue Sun
University of Washington

Joint work with:
Samet Oymak (UC Riverside), Maryam Fazel (Univ. of Washington)

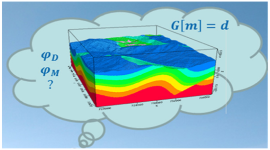
2nd Annual Conference on
Learning for Dynamics and Control
June 12, 2020

Low-rank Hankel matrices and applications

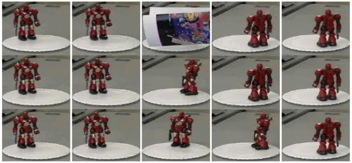
input-output LTI system ID

https://www.mathworks.com/products/sysid.html

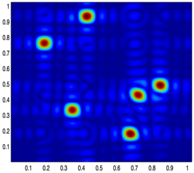
shape from moments estimation, tomography, geophysical inversion [Elad et al '04]



video inpainting [Ding et al, '07]



Super-resolution [Chen & Chi '14]



The role of regularization

Input-output system identification

- Classical: matrix pencil method; subspace system ID
  Ho and Kálmán 1966; Van Overschee and De Moor 1995,...
- More recent: regularization with Hankel nuclear norm
  Fazel et al. 2013; Hansson, Liu, and Vandenberghe 2012; Liu, Hansson, and Vandenberghe 2013; Verhaegen and Hansson 2016; Blomberg 2016

Goal: statistical bounds: sample complexity, error rates

Related work

Goal: statistical bounds: sample complexity, error rates

Only recently explored for (unregularized) least squares:
state observation: Simchowitz, Mania, et al. 2018; Sarkar and Rakhlin 2019; Faradonbeh, Tewari, and Michailidis 2018; Dean et al. 2017; Mania, Tu, and Recht 2019; Sattar and Oymak 2020; Foster, Rakhlin, and Sarkar 2020
stable/unstable, random/designed input, sys id/control, linear/nonlinear
input-output system: Oymak and Ozay 2018; Sarkar, Rakhlin, and Dahleh 2019; Tu et al. 2017; Simchowitz, Boczar, and Recht 2019; Hazan, Singh, and Zhang 2017; Hazan, Lee, et al. 2018; Tsiamis and Pappas 2019; Simchowitz, Singh, and Hazan 2020
strictly/marginally stable, random/designed input, prediction/sysid/control, filtering

Related work

Goal: statistical bounds: sample complexity, error rates

Only recently explored for (unregularized) least squares:
state observation: Simchowitz, Mania, et al. 2018; Sarkar and Rakhlin 2019; Faradonbeh, Tewari, and Michailidis 2018; Dean et al. 2017; Mania, Tu, and Recht 2019; Sattar and Oymak 2020; Foster, Rakhlin, and Sarkar 2020
stable/unstable, random/designed input, sys id/control, linear/nonlinear
input-output system: Oymak and Ozay 2018; Sarkar, Rakhlin, and Dahleh 2019; Tu et al. 2017; Simchowitz, Boczar, and Recht 2019; Hazan, Singh, and Zhang 2017; Hazan, Lee, et al. 2018; Tsiamis and Pappas 2019; Simchowitz, Singh, and Hazan 2020
strictly/marginally stable, random/designed input, prediction/sysid/control, filtering

Cai et al. 2016 regularization for super-resolution, with partial results

This talk: towards guarantees for regularization

Order-R single-input single-output system

Let u, x, y, ξ be the system input, state, output, noise. There exists state-space model of order R

$$\begin{aligned} x(t+1) &= Ax(t) + bu(t) \\ y(t) &= cx(t) + \xi(t) \end{aligned}$$

where $A \in \mathbb{R}^{R \times R}$, $b \in \mathbb{R}^{R \times 1}$, $c \in \mathbb{R}^{1 \times R}$.
Impulse response $h_i = cA^{i-1}b,\ i = 1, \ldots,$

$$R = \operatorname{rank} \begin{pmatrix} h_1 & h_2 & \cdots & h_n \\ h_2 & & \cdots & \\ \vdots & \vdots & & \vdots \\ h_n & & \cdots & h_{2n-1} \end{pmatrix} = \mathcal{H}(h),\ \forall n \geq R$$

extends to block-Hankel (e.g. Fazel et al. 2013; Liu, Hansson, and Vandenberghe 2013)

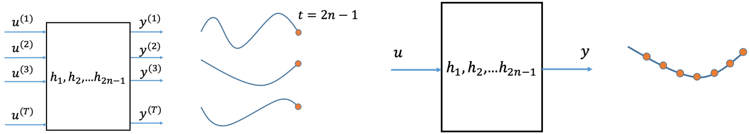
Data acquisition

The input-output mapping can also be

$$y = u * h + \xi$$

Denote **snr** $= \frac{\boldsymbol{E}(\|u\|^2)}{2n-1} \; / \; \boldsymbol{E}(\xi^2)$

Two data aqcuisition models: (a) Multi-rollout (left), and (b) single rollout (right). Sample complexity T = # of dots



Nuclear norm regularization for low order system recovery

Consider the optimization problem

$$\min_{h' \in \mathbb{R}^{2n-1}} \underbrace{\|y - Uh'\|_2^2}_{\text{squared loss}} + \lambda \underbrace{\|\mathcal{H}(h')\|_*}_{\text{regularization}} \tag{1}$$

- $\mathcal{H}(h)$ denotes the $n \times n$ Hankel matrix

$$\mathcal{H}(h) := \begin{bmatrix} h_1 & h_2 & \dots & h_{n-1} & h_n \\ h_2 & h_3 & \dots & h_n & h_{n+1} \\ \dots & \dots & \dots & \dots & \dots \\ h_n & h_{n+1} & \dots & h_{2n-2} & h_{2n-1} \end{bmatrix}$$

- $\lambda = 0$, least squares; $\lambda > 0$, regularization.
- Regularizer encourages low rank Hankel structure

Which error metrics?

$$\hat{h} = \operatorname{argmin}_{h'} \underbrace{\|y - Uh'\|_2^2}_{\text{squared loss}} + \lambda \underbrace{\|\mathcal{H}(h')\|_*}_{\text{regularization}}$$

How useful is $\hat{h}$?

Which error metrics?

$$\hat{h} = \operatorname{argmin}_{h'} \underbrace{\|y - Uh'\|_2^2}_{\text{squared loss}} + \lambda \underbrace{\|\mathcal{H}(h')\|_*}_{\text{regularization}}$$

How useful is $\hat{h}$?

- ℓ_2 error: $\|h - \hat{h}\|_2$
- Hankel spectral error: $\|\mathcal{H}(h) - \mathcal{H}(\hat{h})\|$
  - Useful for model selection & finding realization
  - Hankel spectral error upper bounds ℓ_2 error, and we'll prove they are on the same order.

Spectral analysis - Nuclear norm

Theorem

Let the dimension of Hankel be $n \times n$, system order be R and number of samples be T.

$$\|\mathcal{H}(h) - \mathcal{H}(\hat{h})\| \lesssim \begin{cases} \sqrt{\frac{n}{\mathbf{snr} \times T}} & \text{if} \quad T \gtrsim \min(R^2, n) \\ \sqrt{\frac{Rn}{\mathbf{snr} \times T}} & \text{if} \quad R \lesssim T \lesssim \min(R^2, n) \end{cases}$$

(Second line extended from Cai et al. 2016)

Remark:

- Minimum sample complexity is $R < n$.
  Number of observed samples is smaller than the dimension of variables.

Spectral analysis - Nuclear norm

Theorem

Let the dimension of Hankel be $n \times n$, system order be R and number of samples be T.

$$\|\mathcal{H}(h) - \mathcal{H}(\hat{h})\| \lesssim \begin{cases} \sqrt{\frac{n}{\mathbf{snr} \times T}} & \text{if} \quad T \gtrsim \min(R^2, n) \\ \sqrt{\frac{Rn}{\mathbf{snr} \times T}} & \text{if} \quad R \lesssim T \lesssim \min(R^2, n) \end{cases}$$

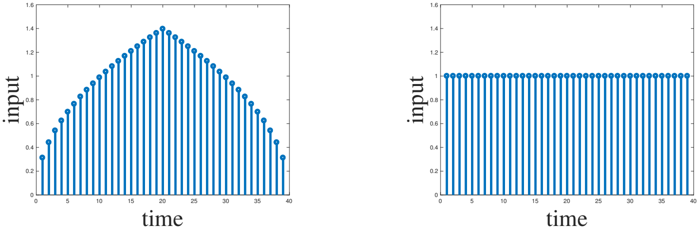
(Second line extended from Cai et al. 2016)

Remark:

- Minimum sample complexity is $R < n$.
  Number of observed samples is smaller than the dimension of variables.
- Oymak and Ozay 2018 obtains ℓ_2 bound $\|h - \hat{h}\|_2 \lesssim \sqrt{\frac{n}{\mathbf{snr} \times T}}$ in unregularized least squares algorithm. Our bound sharpens with enough samples.

Optimal sample complexity due to shaped input

- It is thought that i.i.d. random input is typically proper for system id in least squares works. (Oymak and Ozay 2018; Sarkar, Rakhlin, and Dahleh 2019; Simchowitz, Boczar, and Recht 2019; Hazan, Singh, and Zhang 2017; Hazan, Lee, et al. 2018 etc.)
- With regularization, we have to use a shaped input for optimal sample complexity.



(a) Shaped input, recovery is guaranteed when $T \approx R$; (b) our result, i.i.d input, deterministic recovery failure $T \approx n^{1/6}$.

Optimal spectral bounds for unregularized least-squares

- Oymak and Ozay 2018 provides naive spectral error estimates.
- Sarkar, Rakhlin, and Dahleh 2019 provides suboptimal sample sizes.
- Can we get both?

Optimal spectral bounds for unregularized least-squares

- Oymak and Ozay 2018 provides naive spectral error estimates.
- Sarkar, Rakhlin, and Dahleh 2019 provides suboptimal sample sizes.
- Can we get both?

Theorem

Let the dimension of Hankel be $n \times n$ and number of samples be T. Let $\hat{h}_{LS}$ be the least squares estimate. Suppose $T \gtrsim n$, then The Hankel error obeys

$$\|\mathcal{H}(\hat{h}_{LS}) - \mathcal{H}(h)\| \lesssim \sqrt{\frac{n}{\mathbf{snr} \times T}}$$

Optimal spectral bounds for unregularized least-squares

- Oymak and Ozay 2018 provides naive spectral error estimates.
- Sarkar, Rakhlin, and Dahleh 2019 provides suboptimal sample sizes.
- Can we get both?

Theorem

Let the dimension of Hankel be $n \times n$ and number of samples be T. Let $\hat{h}_{LS}$ be the least squares estimate. Suppose $T \gtrsim n$, then The Hankel error obeys

$$\|\mathcal{H}(\hat{h}_{LS}) - \mathcal{H}(h)\| \lesssim \sqrt{\frac{n}{\textbf{snr} \times T}}$$

Takeaway: Spectral error is as good as the impulse response ℓ_2 error i.e.

$$\|\mathcal{H}(\hat{h} - h)\| \propto \|\hat{h} - h\|_2$$

Algorithmic comparison

Table 1: Hankel matrix is $n \times n$, system order is R, number of samples is T. Noise level $\sigma = 1/\sqrt{\mathbf{snr}}$. LS-IR stands for regressing impulse response with output by LS, LS-Hankel stands for regressing Hankel with output by LS.

Paper	This work	This work	[OO '18]	[SRD '19]
Sample complexity	R^2	n	n	n^2
Method	Nuc-norm	LS-IR	LS-IR	LS-Hankel
IR ℓ_2 Error	$\sigma\sqrt{n/T}$	$\sigma\sqrt{n/T}$	$\sigma\sqrt{n/T}$	$(1+\sigma)\sqrt{n/T}$
Hankel Spectral Error	$\sigma\sqrt{n/T}$	$\sigma\sqrt{n/T}$	$\sigma n/\sqrt{T}$	$(1+\sigma)\sqrt{n/T}$

Algorithmic comparison

Table 1: Hankel matrix is $n \times n$, system order is R, number of samples is T. Noise level $\sigma = 1/\sqrt{\textbf{snr}}$. LS-IR stands for regressing impulse response with output by LS, LS-Hankel stands for regressing Hankel with output by LS.

Paper	This work	This work	[OO '18]	[SRD '19]
Sample complexity	R^2	n	n	n^2
Method	Nuc-norm	LS-IR	LS-IR	LS-Hankel
IR ℓ_2 Error	$\sigma\sqrt{n/T}$	$\sigma\sqrt{n/T}$	$\sigma\sqrt{n/T}$	$(1+\sigma)\sqrt{n/T}$
Hankel Spectral Error	$\sigma\sqrt{n/T}$	$\sigma\sqrt{n/T}$	$\sigma n/\sqrt{T}$	$(1+\sigma)\sqrt{n/T}$

- Optimal guarantee for LS-IR

Algorithmic comparison

Table 1: Hankel matrix is $n \times n$, system order is R, number of samples is T. Noise level $\sigma = 1/\sqrt{\mathbf{snr}}$. LS-IR stands for regressing impulse response with output by LS, LS-Hankel stands for regressing Hankel with output by LS.

Paper	This work	This work	[OO '18]	[SRD '19]
Sample complexity	R^2	n	n	n^2
Method	Nuc-norm	LS-IR	LS-IR	LS-Hankel
IR ℓ_2 Error	$\sigma\sqrt{n/T}$	$\sigma\sqrt{n/T}$	$\sigma\sqrt{n/T}$	$(1+\sigma)\sqrt{n/T}$
Hankel Spectral Error	$\sigma\sqrt{n/T}$	$\sigma\sqrt{n/T}$	$\sigma n/\sqrt{T}$	$(1+\sigma)\sqrt{n/T}$

- Optimal guarantee for LS-IR
- Good error guarantees hold for $T \gtrsim R^2$ regime.

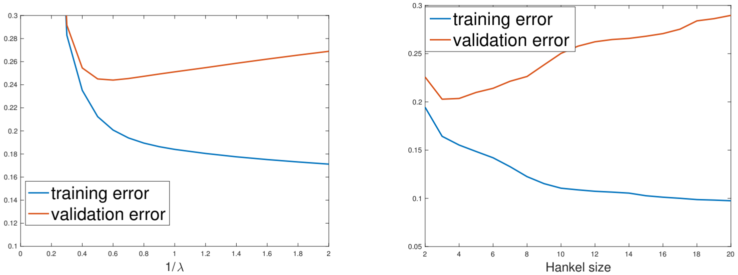
Algorithmic comparison

Table 1: Hankel matrix is $n \times n$, system order is R, number of samples is T. Noise level $\sigma = 1/\sqrt{\mathbf{snr}}$. LS-IR stands for regressing impulse response with output by LS, LS-Hankel stands for regressing Hankel with output by LS.

Paper	This work	This work	[OO '18]	[SRD '19]
Sample complexity	R^2	n	n	n^2
Method	Nuc-norm	LS-IR	LS-IR	LS-Hankel
IR ℓ_2 Error	$\sigma\sqrt{n/T}$	$\sigma\sqrt{n/T}$	$\sigma\sqrt{n/T}$	$(1+\sigma)\sqrt{n/T}$
Hankel Spectral Error	$\sigma\sqrt{n/T}$	$\sigma\sqrt{n/T}$	$\sigma n/\sqrt{T}$	$(1+\sigma)\sqrt{n/T}$

- Optimal guarantee for LS-IR
- Good error guarantees hold for $T \gtrsim R^2$ regime.
  - Question: Can Nuc-norm error bounds be improved to $\sigma\sqrt{R/T}$ (independent of n)?

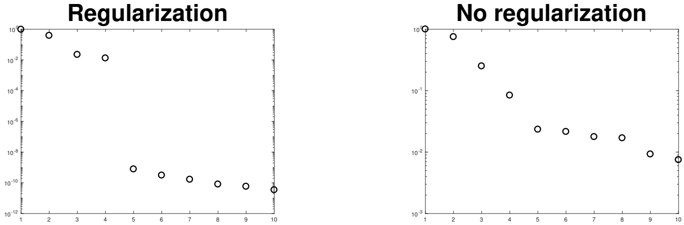
Experiments: regularization and least squares, DaISy dataset[1]



System ID for CD player arm data, assuming $n = 10$. Training data size $=$ 200 and validation data size $=$ 600. (a) training and validation error of different λ, (b) training and validation error of different Hankel size n.

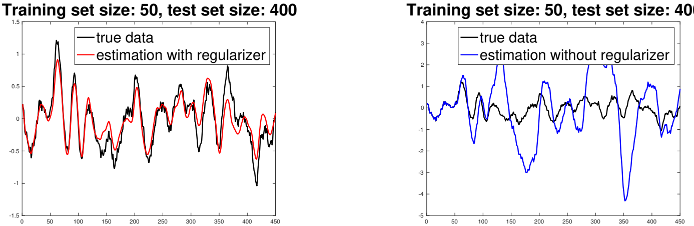
[1] De Moor et al. 1997

Experiments: regularization and least squares, DaISy dataset



CD player arm data, $n = 10$, normalized singular value of (a) regularized (b) unregularized Hankel. We can see the low rank structure of recoverer Hankel matrix from regularization method.

Experiments: regularization and least squares, DaISy dataset



CD player arm data, (a) regularized (b) unregularized prediction, number of samples is 50 when Hankel matrix is 10×10. 400 output samples for validation.

Conclusions

- Nuclear norm regularization is practical but poorly understood
  - Simplifies model selection
  - Less sensitive to hyperparameter tuning
- New estimation error guarantees
  - Improved guarantees for nuclear norm
  - Optimal Hankel spectral error bound for LS

Future: Better understanding the empirical behavior: spectral gap, sensitivity to regularization parameters, etc.
Can error bounds/statistical rates be improved?

References I

Blomberg, Niclas (2016). "On nuclear norm minimization in system identification". PhD thesis. KTH Royal Institute of Technology.

Cai, Jian-Feng et al. (2016). "Robust recovery of complex exponential signals from random Gaussian projections via low rank Hankel matrix reconstruction". In: *Applied and computational harmonic analysis* 41.2, pp. 470–490.

De Moor, Bart et al. (1997). "DAISY: A database for identification of systems". In: *JOURNAL A* 38, pp. 4–5.

Dean, Sarah et al. (2017). "On the sample complexity of the linear quadratic regulator". In: *arXiv preprint arXiv:1710.01688*.

Faradonbeh, Mohamad Kazem Shirani, Ambuj Tewari, and George Michailidis (2018). "Finite time identification in unstable linear systems". In: *Automatica* 96, pp. 342–353.

References II

Fazel, Maryam et al. (2013). "Hankel matrix rank minimization with applications to system identification and realization". In: *SIAM Journal on Matrix Analysis and Applications* 34.3, pp. 946–977.

Foster, Dylan J, Alexander Rakhlin, and Tuhin Sarkar (2020). "Learning nonlinear dynamical systems from a single trajectory". In: *arXiv preprint arXiv:2004.14681*.

Hansson, Anders, Zhang Liu, and Lieven Vandenberghe (2012). "Subspace system identification via weighted nuclear norm optimization". In: *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*. IEEE, pp. 3439–3444.

Hazan, Elad, Holden Lee, et al. (2018). "Spectral filtering for general linear dynamical systems". In: *Advances in Neural Information Processing Systems*, pp. 4634–4643.

Hazan, Elad, Karan Singh, and Cyril Zhang (2017). "Learning linear dynamical systems via spectral filtering". In: *Advances in Neural Information Processing Systems*, pp. 6705–6715.

References III

Ho, BL and Rudolf E Kálmán (1966). "Effective construction of linear state-variable models from input/output functions". In: *at-Automatisierungstechnik* 14.1-12, pp. 545–548.

Liu, Zhang, Anders Hansson, and Lieven Vandenberghe (2013). "Nuclear norm system identification with missing inputs and outputs". In: *Systems & Control Letters* 62.8, pp. 605–612.

Mania, Horia, Stephen Tu, and Benjamin Recht (2019). "Certainty equivalent control of LQR is efficient". In: *arXiv preprint arXiv:1902.07826.*

Oymak, Samet and Necmiye Ozay (2018). "Non-asymptotic identification of lti systems from a single trajectory". In: *arXiv preprint arXiv:1806.05722.*

Sarkar, Tuhin and Alexander Rakhlin (2019). "Near optimal finite time identification of arbitrary linear dynamical systems". In: *arXiv preprint arXiv:1812.01251.*

References IV

- Sarkar, Tuhin, Alexander Rakhlin, and Munther A Dahleh (2019). "Finite-Time System Identification for Partially Observed LTI Systems of Unknown Order". In: *arXiv preprint arXiv:1902.01848*.
- Sattar, Yahya and Samet Oymak (2020). "Non-asymptotic and accurate learning of nonlinear dynamical systems". In: *arXiv preprint arXiv:2002.08538*.
- Simchowitz, Max, Ross Boczar, and Benjamin Recht (2019). "Learning Linear Dynamical Systems with Semi-Parametric Least Squares". In: *arXiv preprint arXiv:1902.00768*.
- Simchowitz, Max, Horia Mania, et al. (2018). "Learning Without Mixing: Towards A Sharp Analysis of Linear System Identification". In: *Conference On Learning Theory*, pp. 439–473.
- Simchowitz, Max, Karan Singh, and Elad Hazan (2020). "Improper learning for non-stochastic control". In: *arXiv preprint arXiv:2001.09254*.

References V

- Tsiamis, Anastasios and George J Pappas (2019). "Finite Sample Analysis of Stochastic System Identification". In: *arXiv preprint arXiv:1903.09122*.
- Tu, Stephen et al. (2017). "Non-Asymptotic Analysis of Robust Control from Coarse-Grained Identification". In: *arXiv preprint arXiv:1707.04791*.
- Van Overschee, Peter and Bart De Moor (1995). "A unifying theorem for three subspace system identification algorithms". In: *Automatica* 31.12, pp. 1853–1864.
- Verhaegen, Michel and Anders Hansson (2016). "N2SID: Nuclear norm subspace identification of innovation models". In: *Automatica* 72, pp. 57–63.