Near-optimal and tractable estimation of recurrent sequences

Dmitrii Ostrovskii

Georgia Tech

Research Horizons Seminar
November 5, 2025

Recurrent sequences

Let $\mathbb{C}^{\mathbb{N}}$ be the vector space of complex-valued sequences $x = (x_0, x_1, \dots)$.

Definition

$x \in \mathbb{C}^{\mathbb{N}}$ is **recurrent of order** s if it satisfies some equation of the form

$$x_t + \sum_{\tau \in [s]} p_\tau x_{t-\tau} = 0 \quad \forall t \geqslant s,$$

a.k.a., linear recurrence relation (LRR) of order s with coeffs $(1, p_1, \dots p_s)$.

- Whole sequence is defined by the first s initial values.

Recurrent sequences

Let $\mathbb{C}^{\mathbb{N}}$ be the vector space of complex-valued sequences $x = (x_0, x_1, \dots)$.

Definition

$x \in \mathbb{C}^{\mathbb{N}}$ is **recurrent of order** s if it satisfies some equation of the form

$$x_t + \sum_{\tau \in [s]} p_\tau x_{t-\tau} = 0 \quad \forall t \geqslant s,$$

a.k.a., linear recurrence relation (LRR) of order s with coeffs $(1, p_1, \dots p_s)$.

- Whole sequence is defined by the first s initial values.
- LRRs are in 1-1 correspondence with homogeneous linear ODEs with constant coeffs of the same order; LRR sol's are discretized ODE sol's.
- Description in terms of the **roots** of the characteristic polynomial

$$p(z) = 1 + \sum_{\tau \in [s]} p_\tau z^\tau = \prod_{k \in [s]} (1 - {z_k}^{-1} z).$$

Definition

A recurrent sequence is called **stationary** if $|z_k| = 1$ for all roots of $p(z)$.

Recurrent sequences as exponential polynomials

Let Δ be the **unit delay** operator on $\mathbb{C}^{\mathbb{N}}$, acting on $x = (x_0, x_1, x_2, \dots)$ as

$$\Delta x = (0, x_0, x_1, \dots), \quad \text{so that} \quad (\Delta x)_t = x_{t-1}, \quad t > 0.$$

- LRR with characteristic polynomial $p(z)$ can be expressed concisely as

$$p(\Delta)x = 0.$$

- Assuming the roots $z_1, z_2, \dots, z_s$ are distinct, its general solution is

$$\sum_{k \in [s]} c_k z_k^{-t}$$

– linear combination of exponential seqs with exponents $z_1^{-1}, \dots, z_s^{-1}$.
This is because each z_k^{-t} is killed by the factor $1 - z_k^{-1}\Delta$ within $p(\Delta)$.

- As in ODEs, **repeating** roots $\Rightarrow$ polynomial modulation $\Rightarrow$ **exp poly**.
E.g., solutions of $(1 - \Delta)^s x = 0$ are polynomials of degree $\leqslant s - 1$.

Recurrent sequences as exponential polynomials

Let Δ be the **unit delay** operator on $\mathbb{C}^{\mathbb{N}}$, acting on $x = (x_0, x_1, x_2, \dots)$ as

$$\Delta x = (0, x_0, x_1, \dots), \quad \text{so that} \quad (\Delta x)_t = x_{t-1}, \quad t > 0.$$

- LRR with characteristic polynomial $p(z)$ can be expressed concisely as

$$p(\Delta)x = 0.$$

- Assuming the roots $z_1, z_2, \dots, z_s$ are distinct, its general solution is

$$\sum_{k \in [s]} c_k z_k^{-t}$$

  – linear combination of exponential seqs with exponents $z_1^{-1}, \dots, z_s^{-1}$.
  This is because each z_k^{-t} is killed by the factor $1 - z_k^{-1}\Delta$ within $p(\Delta)$.
- **Stationary** seqs correspond to **harmonic oscillations** with s freqs

$$\sum_{k \in [s]} c_k e^{-i\omega_k t}$$

  if the roots are distinct. Repeating roots give polynomial modulation.

Estimation problem (Nemirovski, 2000)

- Assume that $x^* \in \mathbb{C}^{\mathbb{N}}$ satisfies some unknown LRR of order s

$$p(\Delta)x^* = 0.$$

- We observe x_t^*, for $0 \leqslant t \leqslant n$ with $n \geqslant s$, in complex Gaussian noise:

$$y_t = x_t^* + \sigma\xi_t, \quad |t| \leqslant n,$$

Here, $\xi_t \overset{\text{iid}}{\sim} \mathbb{C}\mathcal{N}(0,1)$, i.e. $\mathrm{Re}(\xi_t)$ and $\mathrm{Im}(\xi_t)$ are independent $\mathcal{N}(0,1)$.

How well can we estimate the observed sequence?

Estimation problem (Nemirovski, 2000)

- Assume that $x^* \in \mathbb{C}^{\mathbb{N}}$ satisfies some unknown LRR of order s

$$p(\Delta)x^* = 0.$$

- We observe x^*_t, for $0 \leqslant t \leqslant n$ with $n \geqslant s$, in complex Gaussian noise:

$$y_t = x^*_t + \sigma\xi_t, \quad |t| \leqslant n,$$

Here, $\xi_t \overset{\text{iid}}{\sim} \mathbb{C}\mathcal{N}(0,1)$, i.e. $\mathrm{Re}(\xi_t)$ and $\mathrm{Im}(\xi_t)$ are independent $\mathcal{N}(0,1)$.

How well can we estimate the observed sequence?

- Define the "empirically" rescaled ℓ_2-norm on $\mathbb{C}^{\mathbb{N}}$:

$$\|x\|^2_{n,2} := \frac{1}{n+1}\sum_{0\leqslant t\leqslant n} |x_t|^2.$$

- We want an estimate $\widehat{x} = \widehat{x}(y)$ with a guaranteed bound on the MSE:

$$\sup_{x^*\in\mathscr{X}_s} \mathbb{E}\|\widehat{x} - x^*\|^2_{n,2} \leqslant ?$$

where $\mathscr{X}_s$ is the class of all possible recurrent sequences of order s.

Known vs. unknown structure

- Let's speculate: assume that the LRR is **known**:
$$p(\Delta)x^* = 0.$$
Its solution set $X = X(p)$ is an s-dimensional subspace of $\mathbb{C}^{\mathbb{N}}$.
- We can estimate $x^* \in X$ via least-squares, a.k.a. projection estimator:
$$\Pi_X(y) := \operatorname*{argmin}_{x \in X} \|x - y\|_{n,2}^2.$$
It is the maximum likelihood estimator (MLE). Its worst-case MSE is
$$\sup_{x^* \in X} \mathbb{E}\,\|\Pi_X(y) - x^*\|_{n,2}^2 = \frac{\sigma^2}{n+1} s.$$
- Now, $\mathscr{X}_s$ is the **union** of all such subspaces:
$$\mathscr{X}_s = \bigcup\nolimits_{(p_1,\dots,p_s) \in \mathbb{C}^s} X(p).$$
Unfortunately, the corresponding maximum likelihood estimator
$$\mathrm{MLE}(y) = \operatorname*{argmin}_{x \in \mathscr{X}_s} \|x - y\|_{n,2}^2$$
is **intractable**, since $\mathscr{X}_s$ is a **nonconvex** set – and very much so.

Minimax δ-risk

Fix $\delta \in (0,1)$, say 0.05. *Worst-case δ-risk* of estimator $\widehat{x}(\cdot)$ over $\mathcal{X} \subseteq \mathbb{C}^{\mathbb{N}}$:

$$W_{n,\delta}(\widehat{x}(\cdot)|\mathcal{X}) := \min\left\{\varepsilon > 0 : \ \mathbb{P}\left(\|\widehat{x}(y) - x^*\|_{n,2}^2 > \varepsilon\right) \leqslant \delta \ \ \forall x^* \in \mathcal{X}\right\},$$

i.e., the worst-case, over $x^* \in \mathcal{X}$, $(1-\delta)$-percentile of the MSE of $\widehat{x}$.

Minimax δ-risk

Fix $\delta \in (0,1)$, say 0.05. *Worst-case δ-risk* of estimator $\widehat{x}(\cdot)$ over $\mathcal{X} \subseteq \mathbb{C}^{\mathbb{N}}$:

$$W_{n,\delta}(\widehat{x}(\cdot)|\mathcal{X}) := \min\left\{\varepsilon > 0 : \ \mathbb{P}\left(\|\widehat{x}(y) - x^*\|_{n,2}^2 > \varepsilon\right) \leqslant \delta \ \ \forall x^* \in \mathcal{X}\right\},$$

i.e., the worst-case, over $x^* \in \mathcal{X}$, $(1-\delta)$-percentile of the MSE of $\widehat{x}$.

Minimax δ-risk:

$$\mathrm{Risk}_{n,\delta}(\mathcal{X}) := \inf_{\widehat{x}(\cdot)} W_{n,\delta}(\widehat{x}|\mathcal{X}).$$

- For any subspace X with $\dim(X) = s$,

$$\mathrm{Risk}_{n,\delta}(X) \asymp \frac{\sigma^2}{n+1}(s + \log\delta^{-1}).$$

Questions, formalized

- What is the **minimax risk** $\mathrm{Risk}_{n,\delta}(\mathscr{X}_s)$?
- Is there a **tractable** near-optimal estimator?

Hierarchy of classes

Define the unit circle $\mathbb{T}$ and its discretization $\mathbb{T}_n := \{z \in \mathbb{C} : z^{n+1} = 1\}$.

- Define the set $\mathbb{T}^{\text{grid}}_{n,s} := \binom{\mathbb{T}_n}{s}$ of s-tuples from $\mathbb{T}_n$, and the larger set

$$\mathbb{T}^{\text{sep}}_{n,s} := \left\{(z_1, ..., z_s) \in \mathbb{T}^s : \ \text{dist}\,(z_{k'}, z_k) \geqslant \tfrac{2\pi}{n+1} \ \text{ for } \ k' \neq k \right\}.$$

  of $\frac{2\pi}{n+1}$-**separated** s-tuples from $\mathbb{T}$, where $\text{dist}(\cdot,\cdot)$ is the arc distance.
- Define the corresponding subspace

$$X(z_1, ..., z_s) := \left\{ x \in \mathbb{C}^{\mathbb{N}} : \ p(\Delta)x = 0 \text{ with } p(z) = \prod_{k \in [s]} (z - z_k) \right\},$$

  and $\mathscr{X}(\Omega) := \bigcup_{z_1,\dots,z_s \in \Omega} X(z_1, ..., z_s)$ the corresponding subclass of $\mathscr{X}_s$.

Hierarchy of classes

$$\underbrace{\mathscr{X}(\mathbb{T}^{\text{grid}}_{n,s})}_{\text{periodic}} \subset \underbrace{\mathscr{X}\left(\mathbb{T}^{\text{sep}}_{n,s}\right)}_{\text{quasiperiodic stationary}} \subset \underbrace{\mathscr{X}\left(\mathbb{T}^{s}\right)}_{\text{all stationary}} \subset \underbrace{\mathscr{X}\left(\mathbb{C}^{s}\right) = \mathscr{X}_s}_{\text{all recurrent}}.$$

$$\underbrace{\mathscr{X}(\mathbb{T}^{\text{grid}}_{n,s})}_{\text{periodic}} \subset \underbrace{\mathscr{X}(\mathbb{T}^{\text{sep}}_{n,s})}_{\text{quasiperiodic stationary}} \subset \underbrace{\mathscr{X}(\mathbb{T}^{s})}_{\text{all stationary}} \subset \underbrace{\mathscr{X}(\mathbb{C}^{s}) = \mathscr{X}_s}_{\text{all recurrent}}.$$

Discrete Fourier transform (DFT), unitary operator $\mathcal{F}_n : \mathbb{C}^{\mathbb{N}} \to \mathbb{C}^{n+1}$,

$$(\mathcal{F}_n u)_k = \frac{1}{\sqrt{n+1}} \sum_{t=0}^{n} u_t \chi_{n,k}^{-t} \quad \text{for} \;\; k \in \{0, 1, ..., n\},$$

where $\chi_{n,k} = \exp\left(\frac{i2\pi k}{n+1}\right)$ are the roots of unity, i.e. the nodes of $\mathbb{T}_n$.

- Any $x^* \in X(z_1, ..., z_s)$ with $(z_1, ..., z_s) \in \mathbb{T}^{\text{grid}}_{n,s}$ has sparse DFT:

$$\|\mathcal{F}_n x^*\|_0 \leqslant s.$$

- Moreover, $\mathcal{F}_n \xi$ has the same distribution as ξ, i.e. $(\mathcal{F}_n \xi)_k \overset{\text{iid}}{\sim} \mathbb{C}\mathcal{N}(0, 1)$.

Thus, estimation on $\mathscr{X}(\mathbb{T}^{\text{grid}}_{n,s})$ is equivalent to denoising a sparse vector:

$$\text{Risk}_{n,\delta}(\mathscr{X}_s) \geqslant \text{Risk}_{n,\delta}(\mathscr{X}(\mathbb{T}^{\text{grid}}_{n,s})) \asymp \tfrac{\sigma^2}{n+1}(s \log(en/s) + \log \delta^{-1}).$$

Quasiperiodic sequencies / Separated line spectra

$$\underbrace{\mathscr{X}(\mathbb{T}^{\text{grid}}_{n,s})}_{\text{periodic}} \subset \underbrace{\mathscr{X}\left(\mathbb{T}^{\text{sep}}_{n,s}\right)}_{\text{quasiperiodic stationary}} \subset \underbrace{\mathscr{X}\left(\mathbb{T}^{s}\right)}_{\text{all stationary}} \subset \underbrace{\mathscr{X}\left(\mathbb{C}^{s}\right) = \mathscr{X}_s}_{\text{all recurrent}}.$$

Spectral measure ν^* of $x^* \in X(z_1, ..., z_s)$ with $z_1 \neq ... \neq z_s \in \mathbb{T}$ is discrete.

- **Lasso analog** (Candès & Fernandez-Granda '14; Tang & Recht '14):

$$\hat{x} = \boldsymbol{A}(\hat{\nu}) \quad \text{where} \quad \hat{\nu} \in \operatorname*{Argmin}_{\nu \in \mathcal{L}^1(\mathbb{T})} \|y - \boldsymbol{A}(\nu)\|_{n,2}^2 + \lambda\|\nu\|_1$$

  and $\boldsymbol{A}(\nu) \in \mathbb{C}^{\mathbb{N}}$ is the sequence of moments of ν: $[\boldsymbol{A}(\nu)]_t = \int_{z \in \mathbb{T}} z^t d\nu(z)$.

- **RIP analog:** $\begin{pmatrix} 1 & \cdots & 1 \\ z_1 & \cdots & z_s \\ \vdots & & \vdots \\ z_1^n & \cdots & z_s^n \end{pmatrix}$ is nearly orthogonal if $(z_1, ..., z_s) \in \mathbb{T}^{\text{sep}}_{n,s}$.

$$\text{Risk}_{2n,\delta}(\mathscr{X}(\mathbb{T}^{\text{sep}}_{n,s})) \lesssim \tfrac{\sigma^2}{n+1}(s\log(en/s) + s\log\delta^{-1}).$$

- **Cannot** go beyond $\mathscr{X}(\mathbb{T}^{\text{sep}}_{n,s})$: RIP **fails** for $(z_1, ..., z_s) \in \mathbb{T}^{\text{sep}}_{N,s}$ with $N \geqslant n$.
  - No exact recovery on $\mathscr{X}(\mathbb{T}^{\text{sep}}_{n,s})$ from *noiseless* observations $x_0, ..., x_n$.

Intuition from the periodic case

- Estimation on $\mathscr{X}(\mathbb{T}^{\text{grid}}_{n,s})$ is equivalent to denoising of a sparse vector. Indeed, by the isometry of the DFT (a.k.a. Parseval's identity), one has
$$\|\widehat{x} - y\|_{n,2}^2 = \|\widehat{X} - Y\|_{n,2}^2$$
where $\widehat{X} = \mathcal{F}_n \widehat{x}$, and $Y = \widehat{F}_n y = X^* + \mathcal{F}_n \xi$ with **sparse** $X^* = \mathcal{F}_n x^*$.

- Thus, here we get a near-optimal estimator via **greedy approximation**:
$$\widehat{x} = \mathcal{F}_n^{-1} \widehat{X} \quad \text{where} \quad \widehat{X} = \operatorname*{argmin}_{\|X\|_0 \leqslant s} \|Y - X\|_{n,2}^2.$$

- In **soft thresholding** ("proto-Lasso"), we convexify the constraint $\|X\|_0 \leqslant s$ by replacing it first with $\|X\|_1 \leqslant \|X^*\|_1$, and then with the penalized version:
$$\widehat{X} = \operatorname*{argmin}_{X \in \mathbb{C}^{n+1}} \|Y - X\|_{n,2}^2 + \lambda \|X\|_1.$$

  However, this gives suboptimal confidence term $s \log \delta^{-1}$ Can we do better?

- Estimation on $\mathscr{X}(\mathbb{T}^{\text{grid}}_{n,s})$ is equivalent to denoising of a sparse vector. Indeed, by the isometry of the DFT (a.k.a. Parseval's identity), one has
$$\|\widehat{x} - y\|_{n,2}^2 = \|\widehat{X} - Y\|_{n,2}^2$$
where $\widehat{X} = \mathcal{F}_n \widehat{x}$, and $Y = \widehat{F}_n y = X^* + \mathcal{F}_n \xi$ with **sparse** $X^* = \mathcal{F}_n x^*$.

- Thus, here we get a near-optimal estimator via **greedy approximation**:
$$\widehat{x} = \mathcal{F}_n^{-1} \widehat{X} \quad \text{where} \quad \widehat{X} = \underset{\|X\|_0 \leqslant s}{\operatorname{argmin}} \|Y - X\|_{n,2}^2.$$

- **Yes!**[1] Reformulate the greedy estimator by focusing on **the support vector**:
$$\widehat{X} = \widehat{\Phi} \odot Y \quad \text{where} \quad \widehat{\Phi} = \underset{\|\Phi\|_0 \leqslant s,\, \|\Phi\|_\infty \leqslant 1}{\operatorname{argmin}} \|Y - \Phi \odot Y\|_{n,2}^2.$$
This works because the actual support vector $\Phi^* = \text{Supp}(X^*)$ satisfies both constraints. Then, replace $\|\Phi\|_0 \leqslant s$ with $\|\Phi\|_1 \leqslant s$ noting that $\|\Phi^*\|_1 = s$:
$$\widehat{x} = \mathcal{F}_n^{-1}(\widehat{\Phi} \odot Y) \quad \text{where} \quad \widehat{\Phi} = \underset{\|\Phi\|_1 \leqslant s,\, \|\Phi\|_\infty \leqslant 1}{\operatorname{argmin}} \|Y - \Phi \odot Y\|_{n,2}^2.$$

[1] To learn such wonders, take 7252 with me :)

Towards the right estimator

$$\widehat{x} = \mathcal{F}_n^{-1}(\widehat{\Phi} \odot Y) \quad \text{where} \quad \widehat{\Phi} = \underset{\|\Phi\|_1 \leqslant s,\, \|\Phi\|_\infty \leqslant 1}{\operatorname{argmin}} \|Y - \Phi \odot Y\|_{n,2}^2.$$

- DFT property: $\mathcal{F}_n^{-1}(\Phi \odot Y) = \varphi \circledast_n y$ for $\varphi = \frac{1}{\sqrt{n+1}}\mathcal{F}_n^{-1}\Phi$, where $\circledast_n$ is circular convolution modulo $n+1$. As such, the above estimator writes as

$$\widehat{x} = \widehat{\varphi} \circledast_n y \quad \text{where} \quad \widehat{\varphi} = \underset{\|\mathcal{F}_n\varphi\|_1 \leqslant \frac{s}{\sqrt{n+1}},\, \|\mathcal{F}_n\varphi\|_\infty \leqslant \frac{1}{\sqrt{n+1}}}{\operatorname{argmin}} \|y - \varphi \circledast_n y\|_{n,2}^2.$$

- Replace $\circledast$ with $*$ and **pray**.

Statistical guarantee

$$\widehat{\varphi} \in \operatorname*{Argmin}_{\varphi \in [\mathbb{C}^{\mathbb{N}}]_n} \left\{ \|y - \varphi * y\|_{n,2}^2 : \ \|\mathcal{F}_n \varphi\|_1 \leqslant \frac{c_\star s}{\sqrt{n+1}}, \quad \|\mathcal{F}_n \varphi\|_\infty \leqslant \frac{c_\star}{\sqrt{n+1}} \right\}.$$

Theorem (D.O. 2024)

As long as $n \gtrsim s$, the estimator $\widehat{x} = \widehat{\varphi} * y$, while tractable, guarantees that

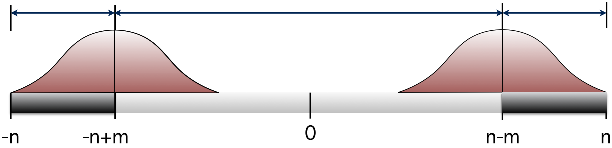
$$\mathrm{Risk}_{n,\delta}(\mathscr{X}_s) \lesssim \frac{\sigma^2}{n+1} \log(en/s)^3 \left(s \log(en/s) + \log \delta^{-1} \right).$$

Reproducing filters...

Let $\mathbb{C}_m^{\mathbb{Z}}$ be the space of sequences supported on $\{-m, ..., m\}$. For $\varphi \in \mathbb{C}_m^{\mathbb{Z}}$,

$$[\varphi(\Delta)x]_t = \sum_{|\tau| \leqslant m} \varphi_\tau x_{t-\tau}$$

is the convolution ("LTI filtering") of x with a filter φ of width m.

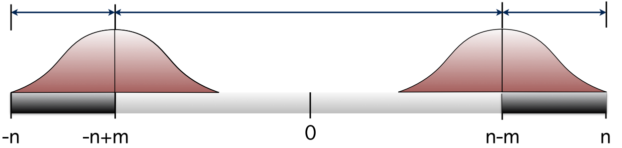


Definition

Filter $\varphi \in \mathbb{C}_m^{\mathbb{Z}}$ is **reproducing** on $X \subseteq \mathbb{C}^{\mathbb{Z}}$ if $\varphi(\Delta)x \equiv x$ for all $x \in X$.

- For any $z_1, \ldots, z_s \in \mathbb{C}$, the subspace $X(z_1, \ldots, z_s)$ is reproduced by $p \in \mathbb{C}_s^{\mathbb{Z}}$.

...as unbiased estimators



- If $\varphi \in \mathbb{C}_m^{\mathbb{Z}}$, then $[\varphi(\Delta)y]_t$ is a linear estimate of x_t^* from $y_{t-m}, \dots, y_{t+m}$.
- If φ is **reproducing** on X, then this estimate is **unbiased** over X:

$$\mathbb{E}[\varphi(\Delta)y]_t - x_t^* = [\varphi(\Delta)x^*]_t - x_t + \sigma\mathbb{E}(\varphi(\Delta)\xi)_t = 0.$$

Its MSE is controlled by $\|\phi\|_2$, namely $\mathbb{E}\|\varphi(\Delta)y - x^*\|_{n,2}^2 = \sigma^2\|\phi\|_2^2$.

Projector trick

Lemma (Juditsky, ca. 2016)

Let $m+1 \geqslant s$. Any shift-invariant X with $\dim(X) = s$ is reproduced by $\phi \in \mathbb{C}_m^{\mathbb{Z}}$:

$$\|\phi\|_2 \leqslant \sqrt{\frac{2s}{2m+1}}.$$

- Filter ϕ is constructed from the projector on X, hence the name.
- By Parseval, same ℓ_2-norm of the spectrum:

$$\|\mathcal{F}_m \phi\|_2 \leqslant \sqrt{\frac{2s}{2m+1}}.$$

- We verify the known minimax risk on a *fixed* shift-invariant subspace:

$$\mathrm{Risk}_{n,\delta}(X) \lesssim \frac{\sigma^2}{2n+1}(s + \log(\delta^{-1})).$$

Squaring trick (ℓ_2-to-ℓ_1 conversion)

Lemma (Nemirovski, 1990s)

The autoconvolution $\phi^2 \in \mathbb{C}_{2m}^{\mathbb{Z}}$ of a reproducing filter $\phi \in \mathbb{C}_m^{\mathbb{Z}}$ is reproducing, and

$$\|\mathcal{F}_{2m}\phi^2\|_2 \leqslant \|\mathcal{F}_{2m}\phi^2\|_1 = \sqrt{4m+1}\|\phi\|_2^2.$$

Proof:

1. $\mathrm{I} - \phi^2(\Delta) = (\mathrm{I} + \phi(\Delta))(\mathrm{I} - \phi(\Delta))$ erases $x \in X$ because $\mathrm{I} - \phi(\Delta)$ does so.
2. For the norm,

$$\|\mathcal{F}_{2m}\phi^2\|_1 = \frac{1}{\sqrt{4m+1}} \sum_{z \in \mathbb{T}_{2m}} \left|\phi^2(z)\right| = \sqrt{4m+1}\,\|\mathcal{F}_{2m}\phi\|_2^2 = \sqrt{4m+1}\,\|\phi\|_2^2. \quad \square$$

Corollary

Let $m+1 \geqslant s$. Any shift-invariant X with $\dim(X) = s$ is reproduced by $\varphi \in \mathbb{C}_{2m}^{\mathbb{Z}}$:

$$\|\mathcal{F}_{2m}\varphi\|_2 \leqslant \|\mathcal{F}_{2m}\varphi\|_1 \leqslant \frac{4s}{\sqrt{4m+1}}.$$

- Conversion of ℓ_2-norm to ℓ_1-norm with $\sqrt{s}$ inflation, "as if" under sparsity!
- Yet, ℓ_2-norm deteriorates: potentially, $\|\mathcal{F}_{2m}\varphi\|_2 \gg \|\mathcal{F}_m\phi\|_2$.

Oracle existence

Proposition

Let $m \geqslant s$. For any $z_1, ..., z_s \in \mathbb{C}$, all recurrent seqs in $X = X(z_1, \dots, z_s)$ are reproduced by some $\varphi^\star \in \mathbb{C}^{\mathbb{Z}}_{9m}$, depending only on X, such that:

$$\|\mathcal{F}_{9m}\varphi^\star\|_p \leqslant s^{1/p} \frac{3c_\star}{\sqrt{18m+1}} \quad \text{where} \;\; c_\star := 0.54\pi^2 + 1.$$

- **Conjecture:** worst case is **periodic**:

$$X = X(z_1, \dots, z_s) \quad \text{for} \quad (z_1, \dots, z_s) \in \mathbb{T}^{\text{grid}}_{n.s}.$$

Oracle existence

Proposition

Let $m \geqslant s$. For any $z_1, ..., z_s \in \mathbb{C}$, all recurrent seqs in $X = X(z_1, \dots, z_s)$ are reproduced by some $\varphi^\star \in \mathbb{C}^{\mathbb{Z}}_{9m}$, depending only on X, such that:

$$\|\mathcal{F}_{9m}\varphi^\star\|_p \leqslant s^{1/p} \frac{3c_\star}{\sqrt{18m+1}} \quad \text{where} \;\; c_\star := 0.54\pi^2 + 1.$$

- **Conjecture:** worst case is **periodic**:

$$X = X(z_1, \dots, z_s) \quad \text{for} \quad (z_1, \dots, z_s) \in \mathbb{T}^{\text{grid}}_{n.s}.$$

Corollary

As long as $n \gtrsim s$,

$$\text{Risk}_{n,\delta}(\mathscr{X}_s) \lesssim \frac{\sigma^2}{n+1} \log(en/s)^3 \left(s \log(en/s) + \log \delta^{-1}\right).$$

\begin{proof}

Oracle construction

Intuition:

- Let ϕ be "small" in ℓ_2. ϕ^2 is small in ℓ_1 but might be "large" in ℓ_2.
- Since $|\phi^2(z)| \gg |\phi(z)|$ requires that $|\phi^2(z)| \gg 1$, the only possible way for $\|\mathcal{F}_n \phi^2\|_2$ to be large is due to $z \in \mathbb{T}_n$ at which $|\phi(z)| \geqslant 1$.
- Can we correct $\phi^2(z)$ by renormalizing it at the bad frequencies?

Oracle construction

Intuition:

- Let ϕ be "small" in ℓ_2. ϕ^2 is small in ℓ_1 but might be "large" in ℓ_2.
- Since $|\phi^2(z)| \gg |\phi(z)|$ requires that $|\phi^2(z)| \gg 1$, the only possible way for $\|\mathcal{F}_n \phi^2\|_2$ to be large is due to $z \in \mathbb{T}_n$ at which $|\phi(z)| \geqslant 1$.
- Can we correct $\phi^2(z)$ by renormalizing it at the bad frequencies?

Construction:

1. Let $n = 9m$. Define the "approximate support" of $\phi \in \mathbb{C}_m^{\mathbb{Z}}$ on $\mathbb{T}_n$ as

$$\mathrm{Supp}_n(\phi) := \{z \in \mathbb{T}_n : |\phi(z)| \geqslant 1\}.$$

2. Let $\rho^\star \in \mathbb{C}_{5m}^{\mathbb{Z}}$ interpolate $\frac{1}{\phi^2(z)}$ on $\mathrm{Supp}_n(\phi)$ with minimal sup-norm on $\mathbb{T}$:

$$\rho^\star \in \operatorname*{Argmin}_{\rho \in \mathbb{C}_{5m}^{\mathbb{Z}}} \left\{ \|\rho\|_{\mathbb{T}} \ \text{ s.r. } \ \rho(z)\phi^2(z) = 1 \quad \forall z \in \mathrm{Supp}_n(\phi) \right\}.$$

3. Choose $\varphi^* \in \mathbb{C}_{9m}^{\mathbb{Z}}$ as

$$\varphi^\star(z) := \phi^2(z) + \rho^\star(z)(\phi^2(z) - \phi^4(z)).$$

Preparation

$$\mathsf{Supp}_n(\phi) := \{z \in \mathbb{T}_n : |\phi(z)| \geqslant 1\}.$$

$$\rho^\star \in \mathop{\mathrm{Argmin}}_{\rho \in \mathbb{C}_{5m}^{\mathbb{Z}}} \left\{ \|\rho\|_{\mathbb{T}} \ \text{ s.r. } \ \rho(z)\phi^2(z) = 1 \quad \forall z \in \mathsf{Supp}_n(\phi) \right\},$$

$$\mathsf{E}_{m,n}(\phi) := 1 \vee \|\rho^\star\|_{\mathbb{T}}.$$

$$\varphi^\star(z) := \phi^2(z) + \rho^\star(z)(\phi^2(z) - \phi^4(z)).$$

- $\varphi^\star$ is reproducing on X. Indeed: ϕ^2 is reproducing, and $1 - \phi^2$ divides $1 - \varphi^\star$:

$$1 - \varphi^\star = (1 - \phi^2)(1 - \rho\phi^2).$$

Lemma (Error bound on $\mathbb{T}_n$)

$$\|\phi^2\rho\|_{\mathbb{T}_n} \leqslant \mathsf{E}_{m,n}(\phi).$$

Proof:

- For $z \in \mathsf{Supp}_n(\phi)$, one has $|\rho^\star(z)\phi^2(z)| = 1$.
- For $z \in \mathbb{T}_n \setminus \mathsf{Supp}_n(\phi)$, $|\phi(z)| \leqslant 1$ and $|\rho^\star(z)\phi^2(z)| \leqslant |\rho^\star(z)| \leqslant \mathsf{E}_{m,n}(\phi)$. $\square$

Control of norms

Proposition 1

$$\|\mathcal{F}_n\varphi\|_1 \lesssim \frac{s\mathrm{E}_{m,n}(\phi)}{\sqrt{2n+1}}, \quad \|\mathcal{F}_n\varphi\|_\infty \leqslant \frac{3\mathrm{E}_{m,n}(\phi)}{\sqrt{2n+1}}.$$

Proof:

- Factor out ϕ^2 from $\varphi^\star$:

$$\varphi^\star = \phi^2(1+\rho^\star - \phi^2\rho^\star).$$

- For ℓ_1-norm,

$$\|\mathcal{F}_n\varphi^\star\|_1 = \frac{1}{\sqrt{2n+1}}\sum_{z\in\mathbb{T}_n}|\varphi(z)| \leqslant \|\mathcal{F}_n\phi^2\|_1\Big(1+\sup_{z\in\mathbb{T}_n}|\rho^\star(z)| + \sup_{z\in\mathbb{T}_n}|\rho^\star(z)\phi^2(z)|\Big)$$

$$\text{[Error Bound Lemma]} \quad \leqslant 3\mathrm{E}_{m,n}(\phi)\|\mathcal{F}_n\phi^2\|_1$$

$$\text{[Squaring]} \quad \lesssim \mathrm{E}_{m,n}(\phi)\frac{s}{\sqrt{2n+1}}.$$

- For ℓ_∞-norm, note that $\varphi^\star(z) = 1$ for all $z \in \mathrm{Supp}_n(\phi)$. On the other hand, for $z \in \mathbb{T}_n \setminus \mathrm{Supp}_n(\phi)$ one has $|\phi(z)| \leqslant 1$ by the definition of $\mathrm{Supp}_n(\phi)$, so

$$|\varphi^\star(z)| \leqslant |\phi^2(z)|(1+|\rho^\star(z)|+|\rho^\star(z)||\phi^2(z)|) \leqslant 1+2|\rho^\star(z)| \leqslant 3\mathrm{E}_{m,n}(\phi). \quad \square$$

Bounding $\mathrm{E}_{m,n}(\phi)$

Proposition 2

$$\mathrm{E}_{m,9m}(\phi) \leqslant 1.08\pi^2 + 2.$$

Proof:

$$\mathrm{E}_{m,n}(\phi) = \inf_{\rho \in \mathbb{C}_{5m}^{\mathbb{Z}}} \left\{ \|\rho\|_{\mathbb{T}} \ \text{ s.t. } \ \rho(z)\phi^2(z) = 1 \quad \forall z \in \mathrm{Supp}_n(\phi) \right\}.$$

Consider the Fejér interpolation polynomial on $\mathrm{Supp}_n(\phi)$,

$$\hat{\rho}(z) = \sum_{w \,\in\, \mathrm{Supp}_n(\phi)} \frac{1}{\phi^2(w)} \frac{\mathrm{Fej}_{5m}(z/w)}{5m+1},$$

where $\mathrm{Fej}_{5m} \in \mathbb{C}_{5m}^{\mathbb{Z}}$ is the Fejér kernel of width $5m$:

$$\mathrm{Fej}_{5m}(z) := \sum_{|\tau| \leqslant 5m} \left(1 - \frac{|\tau|}{5m+1}\right) z^{\tau}.$$

Note that $\mathrm{Fej}_{5m} \in \mathbb{C}_{5m}^{\mathbb{Z}}$ and $\mathrm{Fej}_{5m}(1) = 5m+1$, so $\hat{\rho}$ is feasible: $\mathrm{E}_{m,n}(\phi) \leqslant \|\hat{\rho}\|_{\mathbb{T}}$.

$$\hat{\rho}(z) \leqslant \sum_{w \,\in\, \mathrm{Supp}_n(\phi)} \frac{1}{|\phi(w)|^2} \frac{|\mathrm{Fej}_{5m}(z/w)|}{5m+1} \leqslant \sum_{w \,\in\, \mathbb{T}_n} \frac{|\mathrm{Fej}_{5m}(z/w)|}{5m+1} \leqslant 2 + \left(\frac{2n+1}{5m+1}\right)^2 \frac{\pi^2}{12}.$$

Perspectives

Future work:

- Deconvolution.
- Support estimation.
- Low-rank matrix recovery.
- Single- and multi-index models.

Fin!

Projector trick, cont'd

Lemma (Juditsky, ca. 2016)

Let $m + 1 \geqslant s$. For any $z_1, ..., z_s \in \mathbb{C}^s$, all recurrent sequences $x \in X(z_1, \dots, z_s)$ are reproduced by some $\phi \in \mathbb{C}_m^{\mathbb{Z}}$ such that

$$\|\phi\|_2^2 \leqslant \frac{2s}{2m+1}.$$

Proof:

1. Slices $(x_0, ..., x_m)$ of $x \in X$ form a subspace $X_m \subseteq \mathbb{C}^{m+1}$ with $\dim(X_m) \leqslant s$.
2. Hence, the projector $\Pi_m \in \mathbb{C}^{m+1} \to \mathbb{C}^{m+1}$ on X_m satisfies $\|\Pi_m\|_F^2 \leqslant s$, and
$$\|\pi^*\|_2^2 \leqslant \frac{s}{m+1} \leqslant \frac{2s}{2m+1}$$
for some row π^* of Π_m. Let $t_0 \in \{0, ..., m\}$ be the index of that row π^*.
3. On the other hand, the fact that π^* is row $\# t_0$ of the projector Π_m reads
$$x_{t_0} = \sum_{0 \leqslant \tau \leqslant m} \pi_\tau^* x_\tau = (\phi(\Delta)x)_{t_0}, \quad \forall x \in X,$$
where $\phi \in \mathbb{C}_m^{\mathbb{Z}}$ is constructed by shifting and zero-padding π^* appropriately.
4. By shift-invariance, this remains valid for $t \neq t_0$.

$\square$

Separated line spectra, made simple

$$\underbrace{\mathscr{X}(\mathbb{T}_n^{\langle s\rangle})}_{\text{grid spectra}} \subset \underbrace{\mathscr{X}(\mathbb{T}_{s,n})}_{\text{separated line spectra}} \subset \underbrace{\mathscr{X}(\mathbb{T}^s)}_{\text{arbitrary line spectra}} \subset \underbrace{\mathscr{X}(\mathbb{C}^s) = \mathscr{X}_s}_{\text{our problem}}.$$

- For $x \in X(z_1, ..., z_s)$ with distinct $z_1, ..., z_s \in \mathbb{T}$, we get $\mathcal{F}_n x$ by evaluating on $\mathbb{T}_n$ the convolution of a discrete measure supported on $\{z_1, ..., z_s\}$ with

$$\mathrm{Dir}_n(z) = \sum_{|t| \leqslant n} z^t, \quad z \in \mathbb{T}.$$

- If $\mathrm{dist}(z_1, z_2) \gtrsim \frac{4\pi}{2n+1}$, then $\theta^* = \mathcal{F}_n x^*$ is nearly sparse, so take $\widehat{x} = \mathcal{F}_n^\dagger \hat{\theta}$ with

$$\hat{\theta} \in \underset{\theta \in \mathbb{C}^{2n+1}}{\mathrm{Argmin}} \, \|y - \mathcal{F}_n^\dagger \theta\|_2^2 + \lambda \|\theta\|_1.$$

$$\mathrm{Risk}_{2n,\delta}(\mathscr{X}(\mathbb{T}_{s,n})) \lesssim \frac{\sigma^2}{2n+1} s \log(en\delta^{-1}).$$

- **Cannot** go beyond $\mathscr{X}(\mathbb{T}_{s,n})$.

Nonparametrics beyond smoothness

Differential inequalities of the form:

$$\mathcal{H}_{s,q,L} = \{f \in \mathrm{C}^s(\mathbb{R}) : \|\tfrac{d^s}{dt^s} f\|_{L_q} \leqslant L\}.$$

Smooth functions – those close to **polynomials** (Sobolev, Hölder, etc.)

Arbitrary differential inequalities:

$$\mathcal{H}_{\mathrm{P},q,L} = \{f \in \mathrm{C}^s(\mathbb{R}) : \|\mathrm{P}(\tfrac{d}{dt}) f\|_{L_q} \leqslant L\}.$$

Functions close to **exponential polynomials**, possibly very **nonsmooth**

- In classical nonparametrics, the minimax risk ($\asymp s$) on the subspace of polynomials controls the minimax rates on Sobolev, Hölder, etc. balls $\mathcal{H}_{s,q,L}$.
- For any fixed subspace with $\dim(s)$, the minimax risk is the same.
  Bias defined by $L \Rightarrow$ same bias-variance tradeoff & minimax rates on $\mathcal{H}_{\mathrm{P},q,L}$.
- If it turns out that the minimax risk on the whole union $\mathscr{X}_s$ is still $\asymp s$, then the minimax rates on $\mathcal{H}^*_{s,q,L} := \bigcup_{\deg(\mathrm{P})=s} \mathcal{H}_{\mathrm{P},q,L}$ are the same as on $\mathcal{H}_{s,q,L}$.