Near-optimal and tractable estimation under shift-invariance

Dmitrii Ostrovskii

Georgia Tech

ICCOPT 2025

Los Angeles

Toy example #1: polynomial regression

- Let $x_t^* = f(t)$, $t \in \mathbb{Z}$, where f is unknown polynomial, $\deg(f) = s - 1$.
- We observe $y_t = x_t^* + \sigma\xi_t$, where $\xi_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, for $t \in \{-n, ..., n\}$.
- Find $\widehat{x} = \widehat{x}(y)$ with small MSE $\frac{1}{2n+1}\sum_{|t|\leqslant n} |\widehat{x}_t - x_t^*|^2$ regardless of x^*.

Toy example #1: polynomial regression

- Let $x_t^* = f(t)$, $t \in \mathbb{Z}$, where f is unknown polynomial, $\deg(f) = s - 1$.
- We observe $y_t = x_t^* + \sigma\xi_t$, where $\xi_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, for $t \in \{-n, ..., n\}$.
- Find $\widehat{x} = \widehat{x}(y)$ with small MSE $\frac{1}{2n+1}\sum_{|t|\leqslant n} |\widehat{x}_t - x_t^*|^2$ regardless of x^*.

The assumption on x^* can be rephrased geometrically:

$$x^* \in X,$$

where X is a specific s-dimensional subspace of $\mathbb{C}^{\mathbb{Z}}$ – that of polynomials.

Toy example #1: polynomial regression

- Let $x_t^* = f(t)$, $t \in \mathbb{Z}$, where f is unknown polynomial, $\deg(f) = s - 1$.
- We observe $y_t = x_t^* + \sigma\xi_t$, where $\xi_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$, for $t \in \{-n, ..., n\}$.
- Find $\widehat{x} = \widehat{x}(y)$ with small MSE $\frac{1}{2n+1} \sum_{|t| \leqslant n} |\widehat{x}_t - x_t^*|^2$ regardless of x^*.

The assumption on x^* can be rephrased geometrically:

$$x^* \in X,$$

where X is a specific s-dimensional subspace of $\mathbb{C}^{\mathbb{Z}}$ – that of polynomials.

- Projection estimator $\widehat{x} = \Pi_X(y)$, a.k.a. linear least-squares, satisfies

$$\mathbb{E}\left[\frac{1}{2n+1} \sum_{|t| \leqslant n} |\widehat{x}_t - x_t^*|^2\right] \leqslant \frac{\sigma^2 s}{2n+1},$$

  for **any** $x^* \in X$. Cannot be improved by more than a constant factor.

- X is **shift-invariant**: $f(\cdot - 1)$ is a polynomial of the same degree as f.

Toy example #2: trigonometric polynomials

- Let $x_t^* = a_0 + \sum_{k=1}^{s} a_k \cos\left(\frac{2\pi kt}{T}\right) + b_k \sin\left(\frac{2\pi kt}{T}\right)$. Unknown a_k, b_k.
- $f(t)$ is a T-periodic function, spectrum lives on s lowest frequencies.

- Same conclusions hold, except that now $\dim(X) = 2s + 1$.
- In particular, X is again **shift-invariant**: $f(\cdot - 1)$ is T-periodic if f is.

- We can repeat this saga for any shift-invariant subspace X. However:

What if X is **unknown**?

- Of course, $\Pi_X[y]$ depends on X, and so is not available anymore

Adaptive estimation

Without knowing X, can we do almost as good as if we knew it?

Nemirovski's question

- Sequence x^* satisfies the unknown linear recurrence of order s, that is

$$\sum_{\tau=0}^{s} p_\tau x^*_{t-\tau} \equiv 0 \quad \forall t \in \mathbb{Z}.$$

- We observe x^* on the domain $\{-n, ..., n\}$ in Gaussian noise of level σ:

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

where $2n+1 \geqslant s$ and ξ is a sequence with i.i.d. entries $\xi_t \sim \mathbb{C}\mathcal{N}(0,1)$.

Nemirovski's question

- Sequence x^* satisfies the unknown linear recurrence of order s, that is

$$\sum_{\tau=0}^{s} p_\tau x^*_{t-\tau} \equiv 0 \quad \forall t \in \mathbb{Z}.$$

- We observe x^* on the domain $\{-n, \ldots, n\}$ in Gaussian noise of level σ:

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

where $2n+1 \geqslant s$ and ξ is a sequence with i.i.d. entries $\xi_t \sim \mathbb{C}\mathcal{N}(0,1)$.

Question

How well can we estimate x^* on this domain **without knowing** $p_1, \ldots, p_s$?

Why is it hard?

- The class $\mathscr{X}_s$ of sequences satisfying all linear recurrence relations

$$\sum_{\tau=0}^{s} p_\tau x_{t-\tau} \equiv 0$$

  is described by $2s$ params specifying $p_1, ..., p_s$ and initial conditions.
- Yet, $\mathscr{X}_s$ is extremely rich: **all exponential polynomials** of degree s.
- In particular, a **harmonic oscillation** regularly sampled on $[-n, n]$,

$$x_t = \sum_{1 \leqslant k \leqslant s} c_k e^{i\omega_k t}, \quad t \in \{-n, ..., n\},$$

  might itself resemble Gaussian noise for some frequencies $\omega_1, ..., \omega_s$.
  - We'll revisit this later, when discussing Super-Resolution.

Analysis perspective: difference equations

Let Δ be the unit shift (delay) operator on $\mathbb{C}^{\mathbb{Z}}$:

$$(\Delta x)_t = x_{t-1}, \quad t \in \mathbb{Z}.$$

- Linear recurrence relations are homogeneous difference eqs (ODiffEs):

$$\sum_{\tau=0}^{s} p_\tau x_{t-\tau} \equiv 0 \qquad \Longleftrightarrow \qquad p(\Delta)x \equiv 0$$

  where $p(z) := \sum_{\tau \in \mathbb{Z}} p_\tau z^\tau$ denotes the formal z-transform of $p \in \mathbb{C}^{\mathbb{Z}}$.

- The theory of such ODiffEs closely parallels that of continuous ODEs:

$$\mathrm{P}(\tfrac{d}{dt})f \equiv 0.$$

  Description via the *roots* $z_1, ..., z_s$ of characteristic polynomial $p(\cdot)$.

  - Stability: $|z| < 1$ for ODiffEs, $\mathrm{Re}(z) > 0$ for ODEs.
  - There is a 1-to-1 correspondence between p and P, such that solutions to ODE and ODiffE are pairwise related via discretization $x_t = f(t)$.

Geometric perspective: shift-invariant subspaces

Let Δ be the unit shift (delay) operator on $\mathbb{C}^{\mathbb{Z}}$:

$$(\Delta x)_t = x_{t-1}, \quad t \in \mathbb{Z}.$$

- Linear recurrence relations are homogeneous difference eqs (ODiffEs):

$$\sum_{\tau=0}^{s} p_\tau x_{t-\tau} \equiv 0 \qquad \Longleftrightarrow \qquad p(\Delta)x \equiv 0$$

  where $p(z) := \sum_{\tau \in \mathbb{Z}} p_\tau z^\tau$ denotes the formal z-transform of $p \in \mathbb{C}^{\mathbb{Z}}$.

- For any $p(z)$: $\deg(p) = s$, the solution set of the ODiffE $p(\Delta)x \equiv 0$ is a **shift-invariant** (i.e. Δ-invariant) s-**dimensional subspace** of $\mathbb{C}^{\mathbb{Z}}$.
  - Indeed: if x is such that $p(\Delta)x \equiv 0$, then $p(\Delta)\Delta x \equiv \Delta p(\Delta)x \equiv 0$.

Geometric perspective: shift-invariant subspaces

Let Δ be the unit shift (delay) operator on $\mathbb{C}^{\mathbb{Z}}$:

$$(\Delta x)_t = x_{t-1}, \quad t \in \mathbb{Z}.$$

- Linear recurrence relations are homogeneous difference eqs (ODiffEs):

$$\sum_{\tau=0}^{s} p_\tau x_{t-\tau} \equiv 0 \qquad \Longleftrightarrow \qquad p(\Delta)x \equiv 0$$

  where $p(z) := \sum_{\tau \in \mathbb{Z}} p_\tau z^\tau$ denotes the formal z-transform of $p \in \mathbb{C}^{\mathbb{Z}}$.

- For any $p(z)$: $\deg(p) = s$, the solution set of the ODiffE $p(\Delta)x \equiv 0$ is a **shift-invariant** (i.e. Δ-invariant) s-**dimensional subspace** of $\mathbb{C}^{\mathbb{Z}}$.
  - Indeed: if x is such that $p(\Delta)x \equiv 0$, then $p(\Delta)\Delta x \equiv \Delta p(\Delta) x \equiv 0$.

- Any shift-invariant X with $\dim(X) = s \Longleftrightarrow$ ODiffE with $\deg(p) = s$.
  - Proving it is a **great** exercise (Beurling '49, Halmos '61, Nikolskii '67).

$\mathscr{X}_s =$ union of **all** s-dimensional shift-invariant subspaces of $\mathbb{C}^{\mathbb{Z}}$.

Minimax risk

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

- $\|\widehat{x} - x^*\|_{n,2}^2$ is the mean-squred error (MSE) of an estimate $\widehat{x}$ of x^*.
- Fix a confidence level $1-\delta$. *Worst-case δ-risk* of $\widehat{x}(\cdot)$ over $\mathrm{X} \subseteq \mathbb{C}^{\mathbb{Z}}$:

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

i.e. uniform over $x^* \in \mathrm{X}$, tight $1-\delta$-confidence bound on MSE.

Minimax δ-risk:

$$\mathrm{Risk}_{n,\delta}^*(\mathrm{X}) := \inf_{\widehat{x}(\cdot):\mathbb{C}^{2n+1} \to X} \mathrm{Risk}_{n,\delta}(\widehat{x}|\mathrm{X}).$$

Question (formalized)

$$\mathrm{Risk}_{n,\delta}^*(\mathscr{X}_s) \asymp ?$$

Minimax risk

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

- $\|\widehat{x} - x^*\|_{n,2}^2$ is the mean-squred error (MSE) of an estimate $\widehat{x}$ of x^*.
- Fix a confidence level $1-\delta$. *Worst-case δ-risk* of $\widehat{x}(\cdot)$ over $\mathrm{X} \subseteq \mathbb{C}^{\mathbb{Z}}$:

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

i.e. uniform over $x^* \in \mathrm{X}$, tight $1-\delta$-confidence bound on MSE.

Minimax risk on a subspace

For any subspace X with $\dim(X) = s$, not necessarily a shift-invariant one,

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

Question (formalized)

$$\mathrm{Risk}_{n,\delta}^*(\mathscr{X}_s) \asymp ?$$

Classes of shift-invariant subspaces

Define the unit circle $\mathbb{T}$ and its discretization $\mathbb{T}_n := \{z \in \mathbb{C} : z^{2n+1} = 1\}$.

- Define the set $\mathbb{T}_n^{\langle s \rangle} := \binom{\mathbb{T}_n}{s}$ of s-tuples from $\mathbb{T}_n$, and the larger set

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

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

- Let $X(z_1, \dots, z_s) := \left\{ x \in \mathbb{C}^{\mathbb{Z}} : \ p(\Delta)x \equiv 0 \text{ with } p(z) = \prod_{k=1}^{s} (z - z_k) \right\}$,

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

Hierarchy of classes

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

Grid spectra

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

- **Discrete Fourier transform** is a unitary operator $\mathcal{F}_n : \mathbb{C}^{\mathbb{Z}} \to \mathbb{C}^{2n+1}$,

$$(\mathcal{F}_n u)_k = \frac{1}{\sqrt{2n+1}} \sum_{|t| \leqslant n} u_t \chi_{n,k}^{-t} \quad \text{for} \;\; k \in \{-n, ..., n\},$$

where $\chi_{n,k} = \exp\left(\frac{i2\pi k}{2n+1}\right)$ are the roots of unity, i.e. the nodes of $\mathbb{T}_n$.

- Any $x^* \in X(z_1, ..., z_s)$ with $(z_1, ..., z_s) \in \mathbb{T}_n^{\langle s \rangle}$ has s-**sparse** DFT $\mathcal{F}_n x^*$. 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 $\mathscr{X}(\mathbb{T}_n^{\langle s \rangle})$ is equivalent to denoising of a sparse vector:

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

- **Poll:** how to get the correct tail behavior with a tractable estimator?

Separated line spectra

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

Spectral signed 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):

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

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

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

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

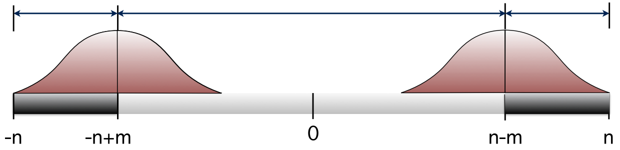
- **Cannot** go beyond $\mathscr{X}(\mathbb{T}_{s,n})$: RIP **fails** for $(z_1, ..., z_s) \in \mathbb{T}_{s,N}$ with $N \geqslant n$.
  - No exact recovery on $\mathscr{X}(\mathbb{T}_{s,n})$ from noiseless observations $x_{-n}, ..., x_n$.

Reproducing filters...

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

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

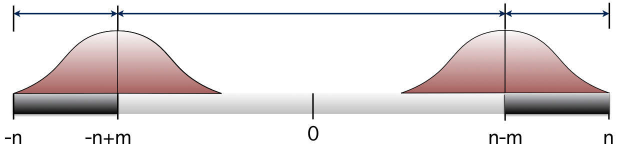
is the convolution ("LTI filtering") of x with a filter φ of width m.



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 $x \in X$.

- Any shift-invariant subspace X, $\dim(X) = s$ is reproduced by $p \in \mathbb{C}_s^{\mathbb{Z}}$.



- If $\varphi \in \mathbb{C}_m^{\mathbb{Z}}$, then $[\varphi(\Delta)y]_t$ is a linear estimate of x_t^* from $y_{t-m}, ..., y_{t+m}$.
- 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 $\phi^2 \in \mathbb{C}^{\mathbb{Z}}_{2m}$ of a reproducing filter $\phi \in \mathbb{C}^{\mathbb{Z}}_{m}$ 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,

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

Corollary

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

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

- Conversion of ℓ_2-norm to ℓ_1-norm with $\sqrt{s}$ inflation, "as if" under sparsity!

Result #1: Oracle inequality

$$\widehat{\varphi} \in \operatorname*{Argmin}_{\varphi \in \mathbb{C}_n^{\mathbb{Z}}} \left\{ \|\varphi(\Delta)y - y\|_{n,2}^2 : \ \|\mathcal{F}_n\varphi\|_1 \leqslant \frac{\mathrm{R}_1}{\sqrt{2n+1}}, \quad \|\mathcal{F}_n\varphi\|_\infty \leqslant \frac{\mathrm{R}_\infty}{\sqrt{2n+1}} \right\}.$$

Theorem 1 (O.'24)

Assume $x^* \in X$ where X has dimension s and is reproduced by $\varphi \in \mathbb{C}_n^{\mathbb{Z}}$ such that

$$\|\mathcal{F}_n\varphi\|_2 \leqslant \frac{\mathrm{R}_2}{\sqrt{2n+1}}, \quad \|\mathcal{F}_n\varphi\|_1 \leqslant \frac{\mathrm{R}_1}{\sqrt{2n+1}}, \quad \|\mathcal{F}_n\varphi\|_\infty \leqslant \frac{\mathrm{R}_\infty}{\sqrt{2n+1}}.$$

As long as $n \gtrsim s$, estimator $\widehat{x} = \widehat{\varphi}(\Delta)y$ with probability at least $1-\delta$ satisfies

$$\|\widehat{x} - x\|_{n,2}^2 \lesssim \frac{\sigma^2}{2n+1} \left(s + \mathrm{R}_2^2 + \log(2s)\, \mathrm{R}_1 \log(2n/s) + \log^2(2s)\, \mathrm{R}_\infty \log(\delta^{-1}) \right).$$

- Projector+Squaring: $\mathrm{R}_\infty \leqslant \mathrm{R}_2 \lesssim \mathrm{R}_1 \asymp s$, giving us $s^2 + s\log(n) + s\log(\delta^{-1})$.
- But ensuring $\mathrm{R}_\kappa \asymp s^{1/\kappa}$ for $\kappa \in \{1, 2, \infty\}$ would give us $s\log(n) + \log(\delta^{-1})$.
- It suffices to guarantee $\mathrm{R}_1 \lesssim s$ and $\mathrm{R}_\infty \lesssim 1$; then $\mathrm{R}_2^2 \lesssim \mathrm{R}_\infty \mathrm{R}_1 \leqslant s$ by Young.

Result #2: Oracle existence

Theorem 2 (O.'24)

Let $m+1 \geqslant s$. Any shift-invariant subspace X is reproduced by $\varphi^\star \in \mathbb{C}^{\mathbb{Z}}_{9m}$:

$$\|\varphi\|_2 \leqslant \frac{6\mathrm{c}_\star\sqrt{2s}}{\sqrt{18m+1}}, \quad \|\mathcal{F}_{9m}\varphi^\star\|_1 \leqslant \frac{36\mathrm{c}_\star s}{\sqrt{18m+1}}, \quad \|\mathcal{F}_{9m}\varphi^\star\|_\infty \leqslant \frac{2\mathrm{c}_\star}{\sqrt{18m+1}}$$

where $\mathrm{c}_\star := 1.08\pi^2 + 3.$

- Constant $\mathrm{c}_\star$ can be replaced with something like 3, using higher-order smoothing splines on $\mathbb{T}$ instead of the Fejér kernel (spline of order 2).
- **Conjecture:** worst case is **grid**: $X = X(z_1, \dots, z_s)$ for $(z_1, \dots, z_s) \in \mathbb{T}_n^{\langle s \rangle}$.

Result #2: Oracle existence

Theorem 2 (O.'24)

Let $m+1 \geqslant s$. Any shift-invariant subspace X is reproduced by $\varphi^\star \in \mathbb{C}_{9m}^{\mathbb{Z}}$:

$$\|\varphi\|_2 \leqslant \frac{6\mathsf{c}_\star\sqrt{2s}}{\sqrt{18m+1}}, \quad \|\mathcal{F}_{9m}\varphi^\star\|_1 \leqslant \frac{36\mathsf{c}_\star s}{\sqrt{18m+1}}, \quad \|\mathcal{F}_{9m}\varphi^\star\|_\infty \leqslant \frac{2\mathsf{c}_\star}{\sqrt{18m+1}}$$

where $\mathsf{c}_\star := 1.08\pi^2 + 3.$

- Constant $\mathsf{c}_\star$ can be replaced with something like 3, using higher-order smoothing splines on $\mathbb{T}$ instead of the Fejér kernel (spline of order 2).
- **Conjecture:** worst case is **grid**: $X = X(z_1, \dots, z_s)$ for $(z_1, \dots, z_s) \in \mathbb{T}_n^{\langle s \rangle}$.

Corollary

As long as $n \gtrsim s$,

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

\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?

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

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

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

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

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

- $\varphi^\star$ 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 $\mathbb{T}_n$)

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

Proof:

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

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,

$$\begin{aligned}
\|\mathcal{F}_n\varphi^\star\|_1 = \tfrac{1}{\sqrt{2n+1}} \sum_{z\in\mathbb{T}_n} |\varphi(z)| &\leqslant \|\mathcal{F}_n\phi^2\|_1 \Big(1 + \sup_{z\in\mathbb{T}_n} |\rho^\star(z)| + \sup_{z\in\mathbb{T}_n} |\rho^\star(z)\phi^2(z)|\Big) \\
\text{[Error Bound Lemma]} \quad &\leqslant 3\mathrm{E}_{m,n}(\phi)\|\mathcal{F}_n\phi^2\|_1 \\
\text{[Squaring]} \quad &\lesssim \mathrm{E}_{m,n}(\phi)\frac{s}{\sqrt{2n+1}}.
\end{aligned}$$

- 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

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

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 (ordinary and blind).
- Support estimation.
- Multi-index?

Fin!

Projector trick, cont'd

Lemma (Juditsky, ca. 2016)

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

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

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 $\|\Pi_m\|_{\mathrm{F}}^2 \leqslant s$, 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.

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

- 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

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

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

- **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 $\mathcal{H}_{s,q,L}$.