Structure-Blind Signal Recovery

Dmitry Ostrovsky* Zaid Harchaoui† Anatoli Juditsky* Arkadi Nemirovski ‡

*Université Grenoble Alpes †University of Washington ‡Georgia Tech

Problem

- **Discrete-time signal** $x = [x_{-T}, \dots, x_0, \dots, x_T]$ is observed with i.i.d. Gaussian noise:

$$y_\tau = x_\tau + \sigma\xi_\tau, \quad -T \le \tau \le T.$$

- **Full recovery**: recover the whole signal x, the loss is $\|\widehat{x} - x\|_2$.
- **Pointwise recovery**: recover one sample x_t at some $t \in \mathbb{Z}$, the loss is $|x_t - \widehat{x}_t|$.
- **Linear estimators**: $\widehat{x}_t^\varphi := \sum_\tau \varphi_\tau y_{t-\tau}$.

Structure-blind recovery

Theorem (Ibragimov and Khasminskii, 1984; Donoho, 1990)
Let a set $\mathcal{X} \subset \mathbb{R}^{2T+1}$ be compact, symmetric, and convex.

- *The minimax, over $x \in \mathcal{X}$, risk of recovering x_t is attained by a linear estimator φ^**
- *If $\mathcal{X}$ is known, φ^* can be found, along with its confidence interval, by cvx optim over $\mathcal{X}$.*

What if the set $\mathcal{X}$ is unknown? Can we "mimick" the "oracle" φ^*?

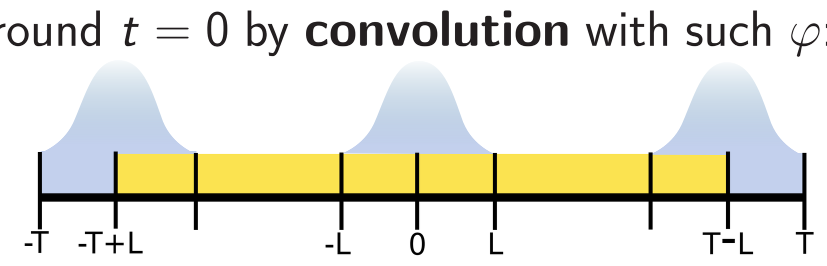
Example. Let $\mathcal{X}$ be a subspace, $\dim(\mathcal{X}) = s \ll T$.

- For sure there exists an estimator $\varphi = \varphi(\mathcal{X})$ with $|\widehat{x}_t^\varphi - x_t| = \mathcal{O}_\mathbb{P}(\sqrt{s/T})$.
- But we don't know $\mathcal{X}$ and hence cannot compute φ.

Linear time-invariant filtering

Write $\varphi \in \mathbf{B}_L$ if φ vanishes outside $[-L, L]$ for some $L \le T$.
Then one can estimate x_t in the central region around $t = 0$ by **convolution** with such φ:

$$\widehat{x}_t^\varphi = \sum_{|\tau| \le L} \varphi_\tau y_{t-\tau} = [\varphi * y]_t, \quad |t| \le T - L.$$



Recoverable signals

For the sake of simplicity, let us assume that we want to estimate x_0.

Definition. x is $(\mathbf{T}, \boldsymbol{\rho})$**-recoverable** at $t = 0$ if there is a filter $\varphi^{\text{oracle}} \in \mathbf{B}_L$, $L = \mathcal{O}(T)$, which has an $\mathcal{O}(1/\sqrt{T})$ error of recovering x_τ in the $T - L$ neighbourhood of $t = 0$:

$$\mathbb{E}^{1/2}|x_\tau - [\varphi^{\text{oracle}} * y]_\tau|^2 \le \frac{\sigma\rho}{\sqrt{T}}, \quad |\tau| \le T - L,$$

As a consequence, in this neighbourhood w.h.p. one has $\|x - \varphi^{\text{oracle}} * y\|_2 \le C\sigma\rho$.

- By simple algebra this is equivalent, up to a small constant, to:
  1. *small ℓ_2-norm of the oracle:* $\|\varphi^{\text{oracle}}\|_2 \le \frac{\rho}{\sqrt{T}}$;
  2. *reproduction of the signal:* $|x_\tau - [\varphi^{\text{oracle}} * x]_\tau| \le \frac{\sigma\rho}{\sqrt{T}}$, $\quad |\tau| \le T - L$.
- Of course, we can introduce the same assumption for any $t \in \mathbb{Z}$.

Example: estimation of smooth curves

Consider the problem of estimating a smooth function $f : [0,1] \to \mathbb{R}$

$$y_\tau = f(\tau/n) + \sigma\xi_\tau, \quad \tau = -n, \dots, n, \ \xi \sim \mathcal{N}(0, I_n).$$

The classical kernel estimator $\widehat{f}_t$ of $f(t)$ with bandwidth h is

$$\widehat{f}(t) = \sum_{|\tau| \le n} \frac{1}{2nh} K\left(\frac{t - \tau/n}{h}\right) y_i,$$

and $K(t) : [-1, 1] \to \mathbb{R}$ is a kernel such that

$$\int_{-1}^{1} K(t)dt = 1, \quad \int_{-1}^{1} K^2(t)dt = \rho^2 < \infty.$$

Let $x_\tau = f(\tau/n)$, $\tau = -n, \dots, n$, and let $T = [nh]$. Then, the kernel estimator can be rewritten for $|t| \le n - T$:

$$\widehat{x}_t = \widehat{f}(t/n) = (\varphi * y)_t, \quad \varphi_\tau = \frac{1}{T} K\left(\frac{\tau}{T}\right), \ \tau = -T, \dots, T.$$

Note that for big enough T the ℓ_2-norm of φ satisfies $\|\varphi\|_2 \sim \frac{\rho}{\sqrt{T}}$, and if bandwidth h is "properly chosen", the bias of the estimator is $\frac{\sigma\rho}{\sqrt{T}}$.

Main assumption: approximate shift-invariance

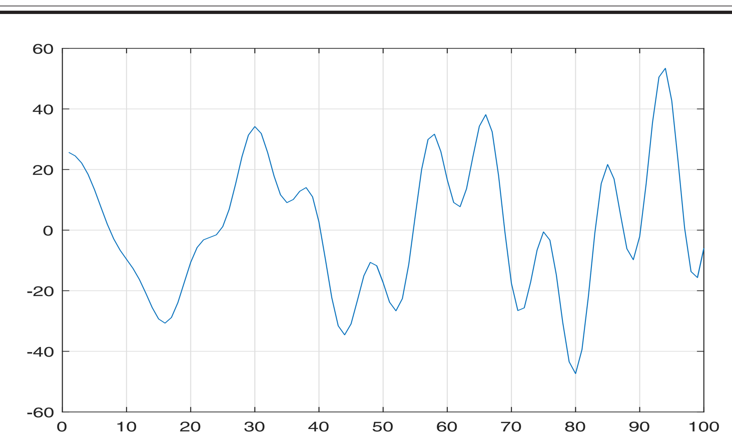
A set $\mathcal{S}$ of signals on $\mathbb{Z}$ is called shift-invariant if it is preserved under the shift $x_t \mapsto x_{t-1}$.

Approximate Shift-Invariance Assumption (ASIA) at $t = 0$:

x is close to an (unknown) shift-invariant subspace $\mathcal{S} \subset \mathbb{C}^\mathbb{Z}$ of dimension $s \le T$.
Specifically, $x = x^\mathcal{S} + \delta$, where $x^\mathcal{S} \in \mathcal{S}$ and $\left\|[\delta]_{-T}^{T}\right\|_2 \le \varkappa\sigma$.

SIA $\Leftrightarrow$ exp. polynomials. x satisfying SIA can be approximated by an exponential polynomial

$$p_\tau = \sum_{k=1}^{s} c_k \tau^{r_k} e^{i\omega_k \tau}$$



An exponential polynomial with $s = 2$

SIA $\Rightarrow$ recoverable. x satisfying SIA is (T, ρ)-recoverable at $t = 0$ with $\rho = (1 + \varkappa)\sqrt{s}$.

Example: sum of harmonic oscillations

Sum of s harmonic oscillations $x_\tau = \sum_{k=1}^{s} c_k e^{\imath\omega_k\tau}$ satisfies ASIA with $\varkappa = 0 \Rightarrow \rho = \sqrt{s}$.

What we want

Estimator $\widehat{\varphi}$ with the following properties:

- error of $\widehat{\varphi}$ of the same order as for φ^{oracle};
- $\widehat{\varphi}$ can be efficiently computed.

Discrete Fourier transform

Let F_m be the $(2m+1) \times (2m+1)$ DFT matrix:

$$[F_m]_{jk} = \frac{1}{\sqrt{2m+1}} \exp\left(\frac{2\pi\imath jk}{2m+1}\right).$$

For $x = [x_{-m}; \dots; x_m]$ consider the Fourier norms

$$\|x\|_p^F := \|F_m x\|_p.$$

Estimator: spectral regularization

$\widehat{x} := \widehat{\varphi} * y$ where $\widehat{\varphi} \in \mathbf{B}_{2L}$ is an optimal solution to:

$$\min_{\varphi \in \mathbf{B}_{2L}} \|y - \varphi * y\|_2^2 + \lambda\|\varphi\|_1^F, \qquad (\text{Pen-}\ell_2)$$

- Favourable structure: efficiently solved by Mirror Prox or Accelerated Gradient Descent.

Main result

Theorem 1 Under ASIA at $t = 0$, estimator (Pen-ℓ_2) with $\lambda \asymp \sigma^2\sqrt{T}\log(T)$ achieves

$$|x_0 - [\widehat{\varphi} * y]_0| \le \frac{C\sigma\rho}{\sqrt{T}}(\rho^2 + \rho\sqrt{\log T}).$$

$$\|x - \widehat{\varphi} * y\|_2 \le C\sigma\rho(\rho + \sqrt{\log T}).$$

Discussion I: uniform fit estimator

Another estimator studied in [J. & N., 2009], and its ρ-adaptive version in [H. et al., 2015]

$$\min_{\varphi \in \mathbf{B}_{2L}} \|y - \varphi * y\|_\infty^F$$

$$\text{such that} \quad \|\varphi\|_1^F \le \frac{\rho^2}{\sqrt{T}}.$$

Theorem (H. et al., 2009). If x is (T, ρ)-recoverable at $t = 0$, $\widehat{\varphi}$ as above satisfies

$$|x_0 - [\widehat{\varphi} * y]_0| \le \frac{C\sigma\rho}{\sqrt{T}}(\rho^3\sqrt{\log T}).$$

$$\|x - \widehat{\varphi} * y\|_2 \le C\sigma\rho(\rho^3\sqrt{\log T}).$$

- Higher adaptation price but ASIA is not needed, recoverability is enough.

Discussion II: sum of harmonic oscillations

For a sum of s harmonic oscillations, we obtain, with no frequency separation assumption,

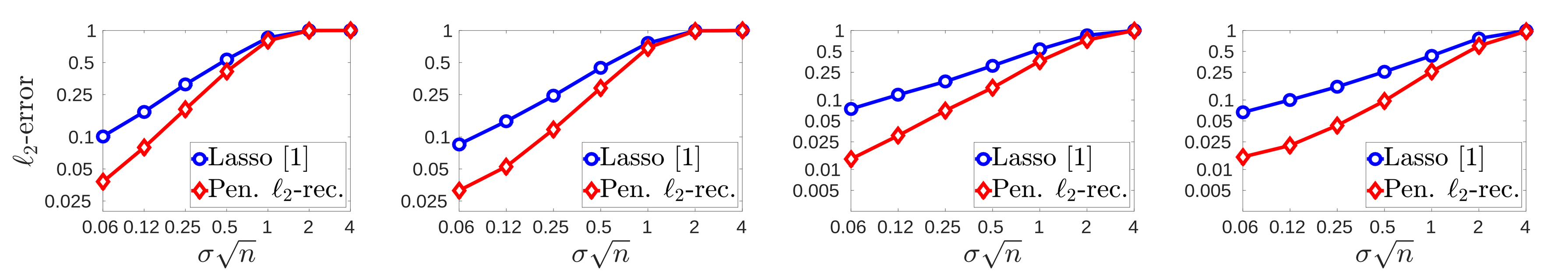
$$\mathbb{E}^{1/2}\|\widehat{x} - x\|_2^2 \le \tilde{O}\left(\frac{\sigma s}{\sqrt{T}}\right)$$

- Assuming frequency separation, s can be replaced by $\sqrt{s}$ for the Lasso estimator of [Tang et al.]
- Our result is more general: it admits exponential polynomials instead of harmonic oscillations with separated frequencies.

Numerical experiments

We present the results of Monte-Carlo experiments on simulated data, comparing the penalized ℓ_2-estimator with the Lasso [Tang et al.].

- Harmonic oscillations with random frequencies (*RandomFreq*) and random pairs of close frequencies (*CoherentFreq*) in 1-D and 2-D.



Signal and image denoising in scenarios *RandomFreq*, *CoherentFreq*, *RandomFreq-2D*, *CoherentFreq-2D*.

- In scenario *DimensionReduction* we consider the single-index regression model:

$$f(t/n) = g(\theta^T t/n), \quad g(\cdot) \in \mathcal{S}_\beta^1(1). \qquad (1)$$

where $\mathcal{S}_\beta^1(1) = \{g : \mathbb{R} \to \mathbb{R}, \|g^{(\beta)}(\cdot)\|_2 \le 1\}$ is the Sobolev ball of smooth periodic functions on $[0,1]$, and the unknown structure is formalized as the direction θ.

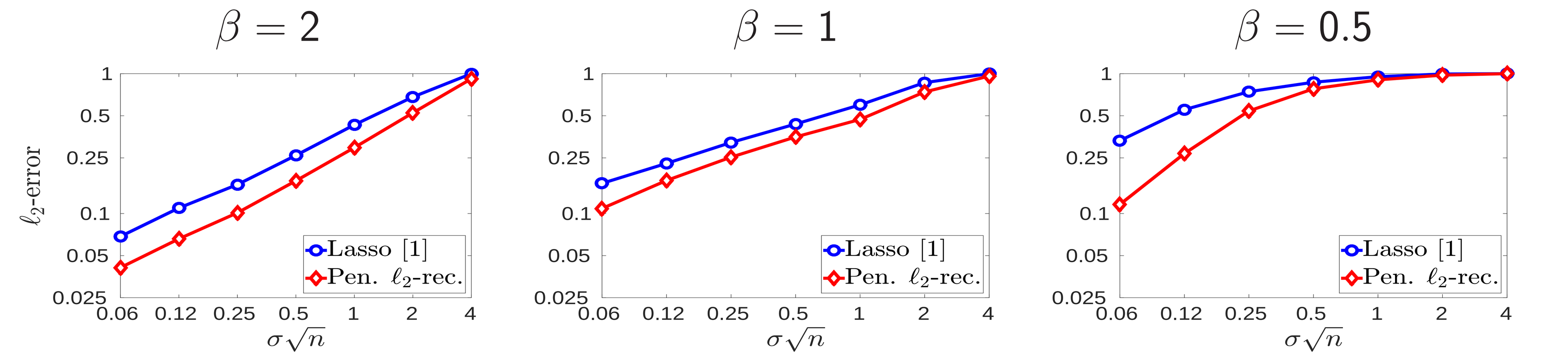


Image denoising in *DimensionReduction* scenario; smoothness decreases left to right.

Bandwidth adaptation

- For each point t of the grid, and for each bandwidth $\{T_0 = 1, T_1 = 2, \dots, T_K = 2^K\}$, compute a solution $\widehat{\varphi}_{T_k,t}$ of (Pen-ℓ_2).
- Compute $\widehat{x}_t[T_k] = [\widehat{\varphi}_{T_k,t} * y]_t$ and choose the "best" among them via Lepski's algorithm.
- To reduce the numerical cost, instead of proceeding point-wise, one can use block-wise update of filters, using the ℓ_2-bound of Theorem 1.