Adaptive Recovery of Signals by Convex Optimization

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

*NYU, INRIA †Univ. Grenoble Alpes ‡Georgia Tech

Filtering problem

- **Data:** Samples of signal $x \in \mathcal{X} \subset \mathbb{R}^N$, corrupted with i.i.d. noise $\xi_t \in \mathcal{N}(0,1)$:
$$y_t = x_t + \sigma\xi_t, \quad t = 1, \dots, N.$$
- **Goal:** Estimate x_N via N previous observations.

Motivation: minimaxity of linear estimators

Consider *linear estimators* $\widehat{x}_N^{\varphi} = \sum_{\tau=0}^{N-1} \varphi_\tau y_{N-\tau} = [\varphi * y]_N, \quad \varphi \in \mathbb{R}^N.$

Theorem (Ibragimov and Khasminskii '1984, Donoho '1990)
Let $\mathcal{X}$ be convex, compact and symmetric. Then for many common loss functions $\ell(\cdot,\cdot)$
$$\min_{\varphi} \max_{x\in\mathcal{X}} \mathbb{E}\,\ell(\widehat{x}_N^{\varphi}, x_N) \le C \min_{\widehat{x}_N} \max_{x\in\mathcal{X}} \mathbb{E}\,\ell(\widehat{x}_N, x_N),$$
where C is a small absolute constant; in particular, it does not depend on $\mathcal{X}$.

- If $\mathcal{X}$ is known, minimax estimator is found by convex programming.
- Finding it is computationally easy for a well-structured set (subspace, ellipsoid).

Sparse recovery perspective

Can we adapt to the (linear) minimax estimator if $\mathcal{X}$ is unknown?

For example, let $\mathcal{X} = \bigcup_{\omega\in\Omega} X_\omega$, where all X_ω are subspaces with $\dim(X_\omega) = p \ll N$.
- For each X_ω there exist (its own) minimax estimator $\widehat{x}_N^\omega$ with error $C_p/\sqrt{N}$.
- We don't know which X_ω does x come from, and hence the actual minimax estimator.

Main assumption

There exists a **linear time-invariant (LTI) oracle** – a filter which recovers the last $O(N)$ samples with pointwise error $O_{\mathbb{P}}(1/\sqrt{N})$.

Assumption (ρ) Let $N = 4n$. For each $x \in \mathcal{X}$, there exists $\varphi^{\text{oracle}} \in \mathbb{R}^n$ such that
$$\left\|\varphi^{\text{oracle}}\right\|_2 \le \frac{\rho}{\sqrt{n}}$$
$$\left\| \left[x - \varphi^{\text{oracle}} * x\right]_n^N \right\|_\infty \le O(1)\frac{\rho}{\sqrt{n}}\sigma.$$

Corollary. φ^{oracle} has error
$$\left|x_t - [\varphi^{oracle} * y]_t\right| = O_{\mathbb{P}}(1)\frac{\rho}{\sqrt{n}}\sigma, \quad n \le t \le N.$$

Example: line spectral signal

$\mathcal{X} = \bigcup_{\omega\in\Omega} X_\omega$, where each X_ω is spanned by p harmonic oscillations: $x_t = \sum_{k=1}^p A_k e^{\imath\omega_k t}$
- Assumption (ρ) holds with $\rho(p) = O(p^{3/2}\ln^{1/2} p)$. [Juditsky & Nemirovski '2013]

What we dream of

Estimator $\widehat{\varphi}$ with the following properties:
- error close to that of (unknown) φ^{oracle},
- $\widehat{\varphi}$ is found by convex optimization.

Discrete Fourier transform

Let $F_N \in \mathbb{R}^{N\times N}$ be the matrix of unitary DFT
$$F_N = \frac{1}{\sqrt{N}}\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{\frac{2\pi\imath}{N}} & \cdots & e^{\frac{2\pi\imath(N-1)}{N}} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & e^{\frac{2\pi\imath(N-1)}{N}} & \cdots & e^{\frac{2\pi\imath(N-1)^2}{N}} \end{pmatrix},$$
and for $x \in \mathbb{R}^N$ consider semi-norms $\|x\|_p^* = \|F_N \cdot x\|_p$ in the frequency domain.

Idea of construction, I

- For any (data-dependent) filter $\widehat{\varphi} \in \mathbb{R}^n$ decompose the estimation error:
$$\left|[x - \widehat{\varphi} * y]_N\right| \le \left|[x - \widehat{\varphi} * x]_N\right| + \left|[\widehat{\varphi} * \xi]_N\right| \le \left|[x - \widehat{\varphi} * x]_N\right| + O_{\mathbb{P}}(\sigma\sqrt{\log n})\,\|\widehat{\varphi}\|_1^*$$
- For the "bias" term, we get, using any fixed filter φ°:
$$\left|[x - \widehat{\varphi} * x]_N\right| \le \left|[(1-\widehat{\varphi}) * (1-\varphi^\circ) * x]_N\right| + \left|[\varphi^\circ * (1-\widehat{\varphi}) * x]_N\right|$$
$$\le (1 + \|\widehat{\varphi}\|_1)\left\|[x - \varphi^\circ * x]_{N-n+1}^N\right\|_\infty + \|\varphi^\circ\|_1^* \left\|[(1-\widehat{\varphi}) * x]_{N-n+1}^N\right\|_\infty^*$$
- Furthermore,
$$\left\|[(1-\widehat{\varphi}) * x]_{N-n+1}^N\right\|_\infty^* \le \left\|[(1-\widehat{\varphi}) * y]_{N-n+1}^N\right\|_\infty^* + \sigma\left\|[(1-\widehat{\varphi}) * \xi]_{N-n+1}^N\right\|_\infty^*$$
$$\le \left\|[(1-\widehat{\varphi}) * y]_{N-n+1}^N\right\|_\infty^* + O_{\mathbb{P}}(\sigma\sqrt{\log n})\,(1 + \|\widehat{\varphi}\|_1)$$
- Let us choose $\widehat{\varphi}$ such that
$$\left\|[(1-\widehat{\varphi}) * y]_{N-n+1}^N\right\|_\infty^* \le \left\|[(1-\varphi^\circ) * y]_{N-n+1}^N\right\|_\infty^*$$
$$\le \left\|[(1-\varphi^\circ) * x]_{N-n+1}^N\right\|_\infty^* + \sigma\left\|[(1-\varphi^\circ) * \xi]_{N-n+1}^N\right\|_\infty^*$$
$$\le \sqrt{n}\left\|[(1-\varphi^\circ) * x]_{N-n+1}^N\right\|_\infty + O_{\mathbb{P}}(\sigma\sqrt{\log n})\,(1 + \|\varphi^\circ\|_1)$$

Idea of construction, II

Back to the estimation error:
$$\left|x_N - [\widehat{\varphi} * y]_N\right| \le O_{\mathbb{P}}(\sigma\sqrt{\log n})\,(1 + \|\widehat{\varphi}\|_1 + \|\varphi^\circ\|_1)\,(\|\widehat{\varphi}\|_1^* + \|\varphi^\circ\|_1^*)$$
$$+ (1 + \|\widehat{\varphi}\|_1 + \sqrt{n}\,\|\varphi^\circ\|_1^*)\left\|[x - \varphi^\circ * x]_{N-n+1}^N\right\|_\infty$$

Thus we need a filter φ° with stronger properties than those of φ^{oracle}, namely
$$\left\|[x - \varphi^\circ * x]_{N-n+1}^N\right\|_\infty \sim \sigma/\sqrt{n}, \qquad \text{(small bias, as for LTI oracle)}$$
$$\|\varphi^\circ\|_1^* \sim 1/\sqrt{n}. \qquad \text{(small } \ell_1^*\text{-norm)}$$

Auto-convolution trick

It turns out that such φ° exists 'automatically' as soon as there exists φ^{oracle}.

Lemma *For* $\varphi^\circ = \left(\varphi^{\text{oracle}} * \varphi^{\text{oracle}}\right) \in \mathbb{R}^{2n-1}$ *it holds:*
$$\left\|[x - \varphi^\circ * x]_{N-2n+2}^N\right\|_\infty \le O(1)\frac{\rho^2}{\sqrt{2n-1}}\sigma, \qquad \|\varphi^\circ\|_2 \le \|\varphi^\circ\|_1^* \le 2\frac{\rho^2}{\sqrt{2n-1}}.$$

It remains only to state a 'suitable' optimization problem for which φ° is feasible w.h.p.

Estimator

- For desired confidence level α, find an optimal solution $(\widehat{\varphi}, \widehat{r})$ of the program
$$\min r, \text{ subject to}$$
$$\left\|[y - \varphi * y]_{N-2n+2}^N\right\|_\infty^* \le 2\sigma(r+1)\sqrt{\ln\left(\tfrac{2n-1}{\alpha}\right)},$$
$$\|\varphi\|_1^* \le \tfrac{r}{\sqrt{2n-1}}, \quad \varphi \in \mathbb{C}^{2n-1}.$$
- Then build the estimate
$$\widehat{x}_N = [\widehat{\varphi} * y]_N.$$

Well-structured SOCP (may be approximated by an LP). Solved by a first-order method.

High-probability bound

Theorem *With probability at least* α,
$$|\widehat{x}_N - x_N| \le O(1)\frac{\rho^4\sqrt{\ln(N/\alpha)}}{\sqrt{N}}\sigma.$$

- Cost of adaptation is $\rho^3\sqrt{\ln N}$. Further reduced to $\rho^2\sqrt{\ln N}$ if ρ is known in advance.
- Bayesian lower bound gives the cost $\rho\sqrt{\ln N}$.

Prediction setting

Given noisy observations of $x_1, .., x_N$, the goal now is to estimate $x_{N+\mathbf{h}}$ for the *forecasting horizon* $\mathbf{h} \in \mathbb{Z}_+$.
- We modify Assumption (ρ) to state the existence of **h**-*predictive* LTI oracle.
- The estimator and the corresponding bound are readily given.

Pointwise recovery

Given noisy signal $\mathbf{y} = (y_1, .., y_N)$, we may construct a recovery $\widehat{\mathbf{x}}$ by separately solving N filtering problems, obtaining the bound
$$\mathbb{E}^{1/2}\,\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2 \le \rho^4\sqrt{\frac{\ln N}{N}}\sigma.$$

Application to line spectral estimation

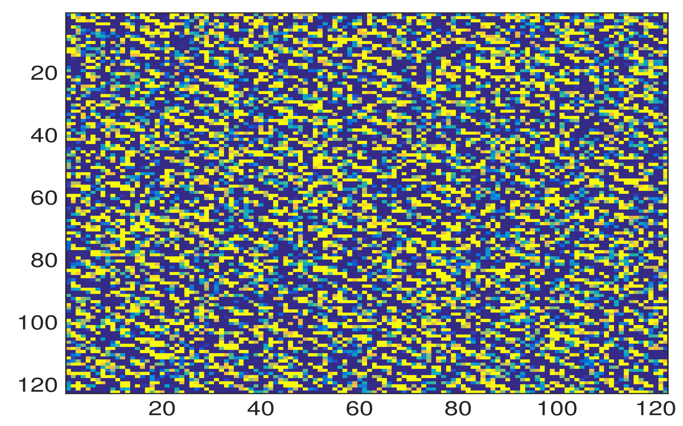
For a line spectral signal, we obtain the bound
$$\mathbb{E}^{1/2}\,\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2 \le O(\rho^6\ln^3 p)\sqrt{\frac{\ln N}{N}}\sigma$$

- We 'almost' recover the state-of-the-art bound of [Tang et al. '2013], without $O(1/N)$ frequency separation assumption.

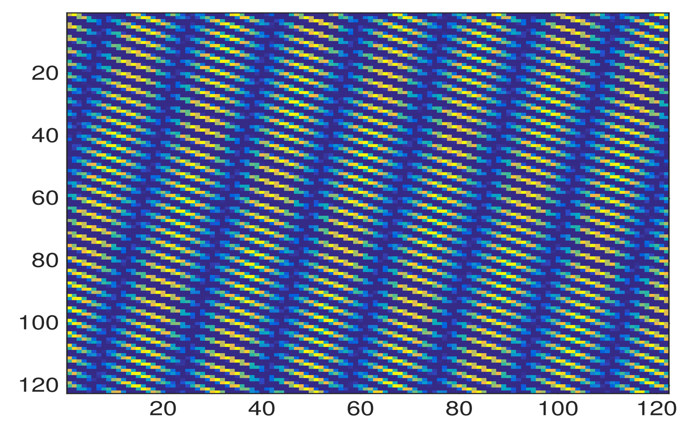
Numerical demonstration

Prediction of a 2-d signal – sum of 2 sinusoids with unit amplitudes and random frequencies. SNR = −3 dB, $h = N = 12$.

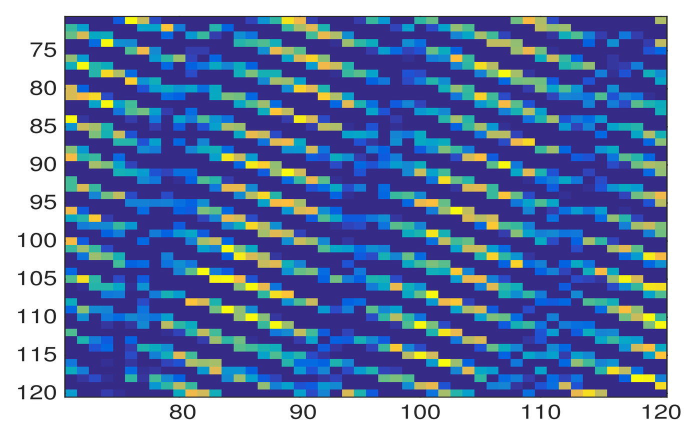
a) noisy signal



b) true signal



c) prediction in the target zone (magnified)



d) true signal in the target zone (magnified)