Adaptive Signal Recovery by Convex Optimization

Dmitrii Ostrovskii

CWI, Amsterdam
19 April 2018

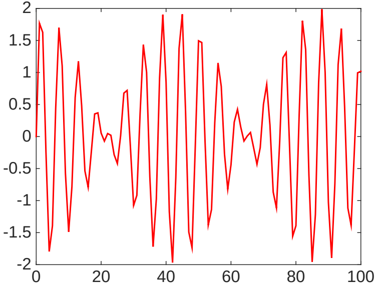
Communauté
UNIVERSITÉ Grenoble Alpes

Signal denoising problem

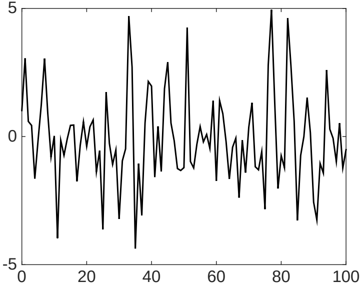
Recover complex **signal** $x = (x_\tau)$, $\tau = -n, ..., n$, from noisy observations

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

where ξ_τ are i.i.d. standard complex Gaussian random variables.



Signal



Observations

- **Assumption:** signal has **unknown shift-invariant structure**.

Preliminaries

- **Finite-dimensional spaces and norms:**
  - $\mathbb{C}_n(\mathbb{Z}) = \{x = (x_\tau)_{\tau\in\mathbb{Z}} : x_\tau = 0 \text{ whenever } |\tau| > n\}$;
  - ℓ_p-norms restricted to $\mathbb{C}_n(\mathbb{Z})$:

$$\|x\|_p = \left(\sum_{|\tau|\le n} |x_\tau|^p\right)^{\frac{1}{p}};$$

  - Scaled ℓ_p-norms:

$$\|x\|_{n,p} = \frac{1}{(2n+1)^{1/p}}\|x\|_p.$$

- **Loss:**
  - $\ell(\widehat{x}, x) = |\widehat{x}_0 - x_0|$ – pointwise loss;
  - $\ell(\widehat{x}, x) = \|\widehat{x} - x\|_{n,2}$ – ℓ_2-loss.
- **Risk:**
  - $R(\widehat{x}, x) = [\mathbf{E}\ell(\widehat{x}, x)^2]^{\frac{1}{2}}$;
  - $R_\delta(\widehat{x}, x) = \min\{r \ge 0 : \ \ell(\widehat{x}, x) \le r \text{ with probability} \ge 1-\delta\}$.

Adaptive estimation: disclaimer

Classical approach

Given a set $\mathcal{X}$ containing x, look for a **near-minimax**, over $\mathcal{X}$, estimator $\widehat{x}^o$. One can often assume that $\widehat{x}^o$ is linear in y (e.g. for pointwise loss)*.

If $\mathcal{X}$ is **unknown**, $\widehat{x}^o$ becomes an unavailable **linear oracle**. Mimic it!

Oracle approach

Knowing that there exists a linear oracle $\widehat{x}^o$ with small risk $R(\widehat{x}^o, x)$, construct an adaptive estimator $\widehat{x} = \widehat{x}(y)$ satisfying an **oracle inequality**:

$$R(\widehat{x}, x) \le P \cdot R(\widehat{x}^o, x) + \mathtt{Rem}, \quad \mathtt{Rem} \ll R(\widehat{x}^o, x).$$

x, $\widehat{x}^o$ can change but P and Rem must be uniformly bounded over $(\widehat{x}^o, x)$.

- $P =$ "price of adaptation". Inequalities with $P = 1$ are called **sharp***.

*[Ibragimov and Khasminskii, 1984; Donoho et al., 1990], *[Tsybakov, 2008]

Classical example: unknown smoothness

Let x be a regularly sampled function:

$$x_t = f(t/N), \quad t = -N, ..., N,$$

where $f : [-1, 1] \to \mathbb{R}$ has weak derivative $D^s f$ of order $s \geq 1$ on $[-1, 1]$, and belongs to a Sobolev ($q = 2$) or Hölder ($q = \infty$) smoothness class:*

$$\mathcal{F}_{s,L} = \{f(\cdot) : \|D^s f\|_{\mathcal{L}_q} \leq L\}.$$

- **Linear oracle:** kernel estimator with properly chosen bandwidth h:

$$\widehat{f}(t/N) = \frac{1}{2hN+1} \sum_{|\tau| \leq hN} K\left(\frac{\tau}{hN}\right) y_{t-\tau}, \quad |t| \leq N - hN.$$

- **Adaptive bandwidth selection*:** Lepski's method, Stein's method, ...

*[Adams and Fournier, 2003; Brown et al., 1996; Watson, 1964; Nadaraya, 1964; Tsybakov, 2008; Johnstone, 2011], *[Lepski, 1991; Lepski et al., 1997, 2015; Goldenshluger et al., 2011]

Recoverable signals

- We consider **convolution-type** (or time-invariant) estimators

$$\widehat{x}_t = [\varphi * y]_t := \sum_{\tau \in \mathbb{Z}} \varphi_\tau y_{t-\tau},$$

where $*$ is discrete **convolution**, and $\varphi \in \mathbb{C}_n(\mathbb{Z})$ is called **a filter**.

Definition*

A signal x is (n, ρ)-**recoverable** if there exists $\phi^o \in \mathbb{C}_n(\mathbb{Z})$ which satisfies

$$\left(\mathbf{E}\,|x_t - [\phi^o * y]_t|^2\right)^{1/2} \le \frac{\sigma\rho}{\sqrt{2n+1}}, \quad |t| \le 3n.$$

- **Consequence:** small ℓ_2-risk: $[\mathbf{E}\,\|x - \phi^o * y\|_{3n,2}^2]^{1/2} \le \frac{\sigma\rho}{\sqrt{2n+1}}$.

*[Juditsky and Nemirovski, 2009; Nemirovski, 1991; Goldenshluger and Nemirovski, 1997]

Adaptive signal recovery: main questions

Goal

Assuming that x is (n, ρ)-recoverable, construct an **adaptive filter** $\widehat{\varphi} = \widehat{\varphi}(y)$ such that the pointwise or ℓ_2-risk of $\widehat{x} = \widehat{\varphi} * y$ is close to $\frac{\sigma\rho}{\sqrt{2n+1}}$.

Main questions:

- Can we adapt to the oracle?
  - **Yes**, but we must pay the price polynomial in ρ;
- Can $\widehat{\varphi}$ be efficiently computed?
  - **Yes**, by solving a well-structured convex optimization problem.
- Do recoverable signals with small ρ exist?
  - **Yes:** when the signal **belongs to shift-invariant subspace** $\mathcal{S} \subset \mathbb{C}(\mathbb{Z})$, $\dim(\mathcal{S}) = s$, we have "nice" bounds on $\rho = \rho(s)$.

Adaptive estimators and their analysis

Main idea

- "Bias-variance decomposition"

$$\underbrace{x_t - [\phi^o * y]_t}_{\text{total error}} = \underbrace{x_t - [\phi^o * x]_t}_{\text{bias}} + \underbrace{\sigma[\phi^o * \xi]_t}_{\text{stochastic error}} .$$

- (n, ρ)-recoverability implies

$$|x_t - [\phi^o * x]_t| \le \frac{\sigma\rho}{\sqrt{2n+1}}, \ |t| \le 3n, \quad \text{and} \quad \|\phi^o\|_2 \le \frac{\rho}{\sqrt{2n+1}}.$$

- Unitary **Discrete Fourier transform** operator $\mathcal{F}_n : \mathbb{C}_n(\mathbb{Z}) \to \mathbb{C}_n(\mathbb{Z})$.

Look at the Fourier transforms

Estimate x via $\widehat{x} = \widehat{\varphi} * y$, where $\widehat{\varphi} = \widehat{\varphi}(y) \in \mathbb{C}_{2n}(\mathbb{Z})$ minimizes the Fourier-domain residual $\|\mathcal{F}_{2n}[y - \varphi * y]\|_p$ while keeping $\|\mathcal{F}_{2n}[\varphi]\|_1$ small.

Motivation: new oracle

Oracle with small ℓ_1-norm of DFT*

If x is (n,ρ)-recoverable, then there exists a $\varphi^o \in \mathbb{C}_{2n}(\mathbb{Z})$ s.t. for $R = 2\rho^2$,

$$|x_t - [\varphi^o * x]_t| \le \frac{C\sigma R}{\sqrt{4n+1}}, \; |t| \le 2n, \quad \|\mathcal{F}_{2n}[\varphi^o]\|_1 \le \frac{R}{\sqrt{4n+1}}.$$

Proof. **1º**. Consider $\varphi^o = \phi^o * \phi^o \in \mathbb{C}_{2n}(\mathbb{Z})$. On one hand, for $|t| \le 2n$,

$$\begin{aligned}|x_t - [\varphi^o * x]_t| &= |x_t - [\phi^o * x]_t| + |[\phi^o * (x - \phi^o * x)]_t| \\ &\le (1 + \|\phi^o\|_1) \max_{|\tau| \le 3n} |x_\tau - [\phi^o * x]_\tau| \le \frac{\sigma\rho(1+\rho)}{\sqrt{2n+1}}.\end{aligned}$$

2º. On the other hand, we get

$$\|\mathcal{F}_{2n}[\varphi^o]\|_1 = \frac{4n+1}{\sqrt{4n+1}} \|\mathcal{F}_{2n}[\phi^o]\|_2^2 = \sqrt{4n+1}\|\mathcal{F}_n[\phi^o]\|_2^2 \le \frac{2\rho^2}{\sqrt{4n+1}}.$$

■

*[Juditsky and Nemirovski, 2009]

Uniform-fit estimators

- **Constrained uniform-fit estimator***:

$$\widehat{\varphi} \in \operatorname*{Argmin}_{\varphi \in \mathbb{C}_n(\mathbb{Z})} \left\{ \|\mathcal{F}_n[y - \varphi * y]\|_{\infty} : \; \|\mathcal{F}_n[\varphi]\|_1 \le \frac{\overline{R}}{\sqrt{2n+1}} \right\}. \quad \text{(CUF)}$$

- **Penalized estimator**: for some $\lambda \ge 0$,

$$\widehat{\varphi} \in \operatorname*{Argmin}_{\varphi \in \mathbb{C}_n(\mathbb{Z})} \left\{ \|\mathcal{F}_n[y - \varphi * y]\|_{\infty} + \sigma\lambda\sqrt{2n+1}\|\mathcal{F}_n[\varphi]\|_1 \right\}. \quad \text{(PUF)}$$

Pointwise upper bound for uniform-fit estimators

Let x be $(\lceil \frac{n}{2} \rceil, \rho)$-recoverable. Let $\overline{R} = 2\rho^2$ for the constrained estimator, and $\lambda = 2\sqrt{\log[(2n+1)/\delta]}$ for the penalized one, then w.p. $\ge 1 - \delta$,

$$|x_0 - [\widehat{\varphi} * y]_0| \le \frac{C\sigma\rho^4\sqrt{\log[(2n+1)/\delta]}}{\sqrt{2n+1}}.$$

High price of adaptation: $O(\rho^3\sqrt{\log n})$.

*[Juditsky and Nemirovski, 2009]

Analysis of uniform-fit estimators

Let $\widehat{\varphi}$ be an optimal solution to (CUF) with $\overline{R} = R$, and let

$$\Theta_n(\zeta) = \|\mathcal{F}_n[\zeta]\|_\infty = O(\sqrt{\log n}) \quad \text{w.h.p.}$$

1º. Already in the first step, we see why the new oracle is useful:

$$\begin{aligned}
|[x - \widehat{\varphi} * y]_0| &\le \sigma |[\widehat{\varphi} * \zeta]_0| + |[x - \widehat{\varphi} * x]_0| \\
&\le \sigma \|\mathcal{F}_n[\widehat{\varphi}]\|_1 \|\mathcal{F}_n[\zeta]\|_\infty + |[x - \widehat{\varphi} * x]_0| \qquad \text{[Young's ineq.]} \\
&\le \frac{\sigma \Theta_n(\zeta) R}{\sqrt{2n+1}} + |[x - \widehat{\varphi} * x]_0|. \qquad \text{[Feasibility of } \widehat{\varphi}\text{]}
\end{aligned}$$

2º. To control $|[x - \widehat{\varphi} * x]_0|$, we can add & subtract convolution with φ^o:

$$\begin{aligned}
|x_0 - [\widehat{\varphi} * x]_0| &\le |[\varphi^o * (x - \widehat{\varphi} * x)]_0| + |[(1 - \widehat{\varphi}) * (x - \varphi^o * x)]_0| \\
&\le \|\mathcal{F}_n[\varphi^o]\|_1 \|\mathcal{F}_n[x - \widehat{\varphi} * x]\|_\infty + (1 + \|\widehat{\varphi}\|_1) \|[x - \varphi^o * x]\|_\infty \\
&\le \frac{R}{\sqrt{2n+1}} \|\mathcal{F}_n[x - \widehat{\varphi} * x]\|_\infty + \frac{CR(1+R)}{\sqrt{2n+1}}.
\end{aligned}$$

Analysis of uniform-fit estimators, cont.

3°. It remains to control $\|\mathcal{F}_n[x - \widehat{\varphi} * x]\|_\infty$ which can be done as follows:

$$
\begin{aligned}
\|\mathcal{F}_n[x - \widehat{\varphi} * x]\|_\infty &\le \|\mathcal{F}_n[y - \widehat{\varphi} * y]\|_\infty + \sigma\|\mathcal{F}_n[\zeta - \widehat{\varphi} * \zeta]\|_\infty \\
&\le \|\mathcal{F}_n[y - \widehat{\varphi} * y]\|_\infty + \sigma(1 + \|\widehat{\varphi}\|_1)\Theta_n(\zeta) \\
&\le \|\mathcal{F}_n[y - \varphi^o * y]\|_\infty + \sigma(1 + \|\widehat{\varphi}\|_1)\Theta_n(\zeta) \\
&\qquad\qquad\qquad\qquad\qquad\qquad [\text{Feas. of } \varphi^o] \\
&\le \|\mathcal{F}_n[x - \varphi^o * x]\|_\infty + 2\sigma(1 + R)\Theta_n(\zeta).
\end{aligned}
$$

4°. Finally, note that

$$
\begin{aligned}
\|\mathcal{F}_n[x - \varphi^o * x]\|_\infty &\le \|\mathcal{F}_n[x - \varphi^o * x]\|_2 \\
&= \|[x - \varphi^o * x]\|_2 \qquad\qquad [\text{Parseval's identity}] \\
&\le \sqrt{2n+1}\,\|x - \varphi^o * x\|_\infty \le \sigma C R.
\end{aligned}
$$

Collecting the above, we obtain a bound dominated by $\frac{\sigma CR(1+R)\Theta_n(\zeta)}{\sqrt{2n+1}}$. ■

Limit of performance

Proposition: pointwise lower bound

For any integer $n \geq 2$, $\alpha < 1/4$, and ρ satisfying $1 \leq \rho \leq n^{\alpha}$, one can point out a family of signals $\mathcal{X}_{n,\rho} \in \mathbb{C}_{2n}(\mathbb{Z})$ such that

- any signal in $\mathcal{X}_{n,\rho}$ is (n,ρ)-recoverable;
- for any estimate $\widehat{x}_0$ of x_0 from observations $y \in \mathbb{C}_{2n}(\mathbb{Z})$, one can find $x \in \mathcal{X}_{n,\rho}$ satisfying

$$\mathbb{P}\left\{ |x_0 - \widehat{x}_0| \geq \frac{c\sigma\rho^2\sqrt{(1-4\alpha)\log n}}{\sqrt{2n+1}} \right\} \geq 1/8.$$

Conclusion: there is a gap ρ^2 between upper and lower bounds.

- To bridge it (and encompass ℓ_2-loss), we introduce new estimators.

Least-squares estimators

- **Constrained formulation:**

$$\widehat{\varphi} \in \operatorname*{Argmin}_{\varphi \in \mathbb{C}_n(\mathbb{Z})} \left\{ \|\mathcal{F}_n[y - \varphi * y]\|_2 : \|\mathcal{F}_n[\varphi]\|_1 \le \frac{\overline{R}}{\sqrt{2n+1}} \right\}; \quad \text{(CLS)}$$

- **Penalized formulations:** $\langle \ldots \rangle$.

For the analysis, we have to restrict the set of signals, introducing **shift-invariant subspaces** (s.-i.s.)

Definition. A linear subspace $\mathcal{S} \subseteq \mathbb{C}_\infty(\mathbb{Z})$ is called **shift-invariant** if it is an invariant subspace of the unit lag operator $[\Delta x]_t = x_{t-1}$.

Oracle inequality for ℓ_2-loss

Theorem: sharp ℓ_2-oracle inequality for least-squares estimators

Suppose that x belongs to some s.-i.s. $\mathcal{S}$, and let φ^o be feasible in (CLS):

$$\|\mathcal{F}_n[\varphi^o]\|_1 \le \frac{\overline{R}}{\sqrt{2n+1}}.$$

For any $\delta \in (0,1]$, an optimal solution $\widehat{\varphi}$ to (CLS) w.p. $\ge 1-\delta$ satisfies

$$\|x - \widehat{\varphi} * y\|_{n,2} \le \|x - \varphi^o * y\|_{n,2} + \frac{C\sigma}{\sqrt{2n+1}} \sqrt{\overline{R} \log\left(\frac{2n+1}{\delta}\right) + \dim(\mathcal{S})}.$$

Consequence. Suppose that x is $(\lceil \frac{n}{2} \rceil, \rho)$-recoverable, and let $\overline{R} = 2\rho^2$. Then, $\varphi^o = \phi^o * \phi^o$ satisfies $\|x - \varphi^o * y\|_{n,2} = O\left(\frac{\sigma\rho^2}{\sqrt{2n+1}}\right)$, whence

$$\|x - \widehat{\varphi} * y\|_{n,2} = O\left(\frac{\sigma\left(\boldsymbol{\rho}^2 + \rho\sqrt{\log n} + \sqrt{\dim(\mathcal{S})}\right)}{\sqrt{2n+1}}\right).$$

Sketch of the proof of ℓ_2-oracle inequality

Control of the cross-term

$$\widehat{\varphi} \in \underset{\varphi \in \mathbb{C}_n(\mathbb{Z})}{\mathrm{Argmin}} \left\{ \|y - \varphi * y\|_2^2 : \ \|\mathcal{F}_n[\varphi]\|_1 \le \frac{\overline{R}}{\sqrt{2n+1}} \right\}.$$

- φ^o is **feasible**, so that $\|y - \widehat{\varphi} * y\|_2^2 \le \|y - \varphi^o * y\|_2^2$.
- Expand the squares:

$$\|x - \widehat{\varphi} * y\|_2^2 = \|x - \varphi^o * y\|_2^2 + 2\sigma^2 \mathbf{Re}\langle \xi, \widehat{\varphi} * \xi \rangle + [\ldots]$$

- **Heuristic:** replace convolution in $\langle \xi, \widehat{\varphi} * \xi \rangle_n$ with the cyclic one $\circledast$:

$$\begin{aligned}
\langle \xi, \widehat{\varphi} \circledast \xi \rangle &= \langle \mathcal{F}_n[\xi], \mathcal{F}_n[\widehat{\varphi} \circledast \xi] \rangle && \text{[Parseval]} \\
&= \sqrt{2n+1} \langle \mathcal{F}_n[\xi], \mathcal{F}_n[\widehat{\varphi}] \odot \mathcal{F}_n[\xi] \rangle && \text{[Diagonalization]} \\
&\le \sqrt{2n+1} \|\mathcal{F}_n[\xi]\|_\infty^2 \|\mathcal{F}_n[\widehat{\varphi}]\|_1 && \textbf{[Young]} \\
&\le C\overline{R} \log\left(\frac{2n+1}{\delta}\right) && \text{with probability at least } 1-\delta.
\end{aligned}$$

- **Rigorous argument**: represent $\langle \xi, \varphi * \xi \rangle_n$ as a random process indexed by φ, and control its maximum on ℓ_1-ball. ■

Error decomposition

$$\|x - \widehat{\varphi} * y\|_2^2 \leq \|x - \varphi^o * y\|_2^2 + 2\sigma \mathbf{Re}\langle \xi, x - \varphi^o * y \rangle - 2\sigma \mathbf{Re}\langle \xi, x - \widehat{\varphi} * y \rangle.$$

$\widehat{\varphi}$-cross-term poses the main difficulty. It can be decomposed as:

$$\langle \xi, x - \widehat{\varphi} * y \rangle = \langle \Pi_{\mathcal{S}} \xi, x - \widehat{\varphi} * y \rangle + \sigma \langle \Pi_{\mathcal{S}}^{\perp} \xi, \widehat{\varphi} * \xi \rangle + \langle \xi, \Pi_{\mathcal{S}}^{\perp} [x - \widehat{\varphi} * x] \rangle,$$

where $\Pi_{\mathcal{S}}$ is the projector onto $\mathcal{S}$.

- **For the first term**, we use Cauchy-Schwarz + χ^2-deviation bound:

$$\mathbf{Re} \langle \Pi_{\mathcal{S}} \xi, x - \widehat{\varphi} * y \rangle \leq \|x - \widehat{\varphi} * y\|_2 \left[\sqrt{2 \dim(\mathcal{S})} + \sqrt{2 \log\left(\frac{1}{\delta}\right)} \right].$$

- **The second term** $\langle \Pi_{\mathcal{S}}^{\perp} \xi, \widehat{\varphi} * \xi \rangle$ is bounded similarly to $\langle \xi, \widehat{\varphi} * \xi \rangle$.
- **The third term** vanishes due to the shift-invariance of $\mathcal{S}$:

$$\Pi_{\mathcal{S}}^{\perp} [x - \widehat{\varphi} * x] \equiv [\Pi_{\mathcal{S}}^{\perp} x] - \widehat{\varphi} * [\Pi_{\mathcal{S}}^{\perp} x] \equiv 0.$$

Summary

We summarize the risk multiplier for $\frac{\sigma}{\sqrt{2n+1}}$ (up to a constant factor):

	Pointwise loss	ℓ_2-loss
Oracle	$\boldsymbol{\rho}$	$\boldsymbol{\rho}$
(Adaptive) lower bound	$\rho^2\sqrt{\log n}$	$\rho\sqrt{\log n}$ *
(Adaptive) upper bound	$\rho^4\sqrt{\log n}$	$\rho^2 + \rho\sqrt{\log n} + \sqrt{\dim(\mathcal{S})}$

In fact, one can also control the pointwise loss for least-squares estimators, so that $\rho^4\sqrt{\log n}$ can be replaced with $\rho^3 + \rho^2\sqrt{\log n} + \rho\sqrt{\dim(\mathcal{S})}$.

*Obtained via a simple argument from the corresponding pointwise bound.

Application:
Recovery from an unknown shift-invariant subspace

Shift-invariant subspaces

Assume that $x \in \mathcal{S} \subset \mathbb{C}_\infty(\mathbb{Z})$, a shift-invariant subspace with $\dim(\mathcal{S}) = s$.

Equivalent formulations:

- x satisfies a homogeneous **difference equation** of order $s = \dim(\mathcal{S})$,

  $$[\mathrm{P}(\Delta)x]_t \equiv 0, \quad t \in \mathbb{Z},$$

  where $\Delta : [\Delta x]_t = x_{t-1}$ is the lag operator, and $\mathrm{P}(z)$ is a polynomial with $\deg(\mathrm{P}) = s$.

- x is an **exponential polynomial** of order s: for some $r \leq s$,

  $$x_t = \sum_{k=1}^{r} q_k(t) e^{\lambda_k t}, \quad \lambda_k \in \mathbb{C},$$

  where $\deg(q_k) - 1$ is the multiplicity of the root $z_k = e^{\lambda_k}$ of $\mathrm{P}(z)$.

Unknown shift-invariant structure of x is encoded by $\mathcal{S}$, or equivalently, P.

Recoverability for shift-invariant subspaces

Signals from shift-invariant subspaces admit oracle filters with $\rho = \rho(s)$.

Theorem

Let $x \in \mathcal{S}$ where $\mathcal{S}$ is a shift-invariant subspace, $\dim(\mathcal{S}) = s$. Then, for any $n \geq s$ there exists a filter $\phi^o \in \mathbb{C}_n(\mathbb{Z})$ which satisfies

$$x_t - [\phi^o * x]_t \equiv 0 \quad \text{and} \quad \|\phi^o\|_2 \leq \sqrt{\frac{s}{2n+1}}.$$

- Lower bound $\rho(s) = \Omega(\sqrt{s})$ for $\phi^o = \phi^o(\mathcal{S})$ from parametric theory.
- The result can be extended to signals close to $\mathcal{S}$ in $\|\cdot\|_p$-norm, encompassing **general differential inequalities***:

$$\|\mathrm{P}(D)f\|_{\mathcal{L}_p} \leq L, \quad \deg(\mathrm{P}) \leq s.$$

*[Juditsky and Nemirovski, 2010]

Recoverability for shift-invariant subspaces (cont.)

One-sided filters: $\phi^o \in \mathbb{C}_n^+(\mathbb{Z}) = \{\varphi \in \mathbb{C}_n(\mathbb{Z}) : \ \varphi_\tau = 0 \ \text{ for } \ \tau < 0\}.$

- In this case, we consider **"generalized harmonic oscillations"**:

$$x_t = \sum_{k=1}^{r \le s} q_k(t) e^{i\omega_k t}, \quad \omega_k \in [0, 2\pi).$$

We improve over the state-of-the art bound*

$$\|\phi^o\|_2 \le \sqrt{\frac{C\mathbf{s^3}\log(s+1)}{n+1}}:$$

Theorem

Under the premise of the previous theorem, there exists $\phi^o \in \mathbb{C}_n^+(\mathbb{Z})$:

$$x_t - [\phi^o * x]_t \equiv 0 \quad \text{and} \quad \|\phi^o\|_2 \le \sqrt{\frac{C\mathbf{s^2}\log(ns+1)}{n+1}}.$$

*[Juditsky and Nemirovski, 2013]

Recovery in ℓ_2-loss on the whole domain

Goal: recover an ordinary harmonic oscillation on the whole $[-n, n]$:

$$x_\tau = \sum_{k=1}^{s} C_k e^{i\omega_k \tau}, \quad \omega_k \in [0, 2\pi).$$

- **Atomic Soft Thresholding***: requires frequency separation by $\frac{2\pi}{2n+1}$.
- **One-sided recovery:** ℓ_2-oracle inequality + one-sided oracles.
- **Two-zone recovery:** ℓ_2-oracle inequality + two-sided oracle in the center + one-sided oracles in the border zones of size $n/(s \log n)$.

	Arbitrary frequencies	Separated frequencies
AST	$O(n^{-1/4})$ – slow rate	$\frac{\sigma}{\sqrt{n}} \cdot (s \log n)^{1/2}$ – optimal
One-sided recovery	$\frac{\sigma}{\sqrt{n}} \cdot s^2 \log n$	$\frac{\sigma}{\sqrt{n}} \cdot [s + (s \log n)^{1/2}]$
Two-zone recovery	$\frac{\sigma}{\sqrt{n}} \cdot s^{3/2} \log n$	$\frac{\sigma}{\sqrt{n}} \cdot [s + (s \log n)^{1/2}]$

*[Bhaskar et al., 2013; Tang et al., 2013]

Algorithmic implementation

Optimization problem

$$\min_{\varphi \in \Phi(r)} \{F(\varphi) + \text{Pen}(\varphi)\},$$

where $F(\varphi) = \begin{cases} \|\mathcal{F}_n[y - y * \varphi]\|_\infty & \text{for uniform-fit recovery,} \\ \|\mathcal{F}_n[y - y * \varphi]\|_2^2 & \text{for least-squares recovery,} \end{cases}$

$$\text{Pen}(\varphi) := \mu \|\mathcal{F}_n[\varphi]\|_1, \quad \text{and} \quad \Phi(r) := \{\varphi \in \mathbb{C}_n(\mathbb{Z}) : \|\mathcal{F}_n[\varphi]\|_1 \le r\}.$$

- **Simple constraint / penalization** after changing variables to $\mathcal{F}_n[\varphi]$.
- **Large scale:** n up to 10^4 in signal processing and 10^6-10^9 in imaging.
- **(Sub-)gradient** of $F(\varphi)$ in $\mathcal{O}(n \log n)$ via FFT and elementwise ops.
- **Low accuracy:** approximate solutions with medium accuracy in the objective are sufficient (more precisely later).

First-order proximal methods

Strategies

Least-squares recovery

$$\min_{\varphi \in \Phi(r)} \left\{ \|\mathcal{F}_n[y - y * \varphi]\|_2^2 + \mathrm{Pen}(\varphi) \right\}.$$

- Composite objective with Lipschitz continuous gradient $\nabla F(\varphi)$;
  - **Nesterov's Fast Gradient Method**, $O(1/k^2)$ convergence.

Uniform-fit recovery

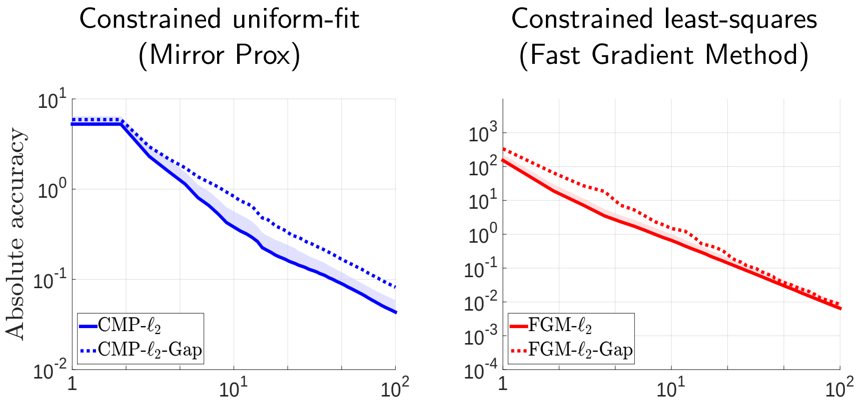
$$\begin{aligned} &\min_{\varphi \in \Phi(r)} \{\|\mathcal{F}_n[y - y * \varphi]\|_\infty + \mathrm{Pen}(\varphi)\} \\ &= \min_{\varphi \in \Phi(r)} \max_{\psi \in \Phi(1)} \{\langle \psi, y - y * \varphi \rangle + \mathrm{Pen}(\varphi)\}. \end{aligned}$$

- Convex-concave saddle-point problem, smooth part is bilinear;
  - **Composite Mirror Prox**, $O(1/k)$ convergence.

- Non-Euclidean prox (ℓ_1/ℓ_2-norm), accuracy certificates, adaptive stepsize.

[Nesterov and Nemirovski, 2013; Juditsky and Nemirovski, 2011a,b; Nemirovski et al., 2010]

Convergence



Convergence of the residual (95% upper confidence bound) for harmonic oscillations with $s = 4$ random frequencies, observed with SNR $= 4$.

Dashed: online accuracy bounds via the accuracy certificate technique.

Statistical accuracy: theoretical results

- Recalling the statistical analysis of the adaptive estimators, we get:

Theorem

Approximate solutions $\tilde{\varphi}$ with objective accuracy $\varepsilon_* = \sigma\rho^2$ for uniform fit, or $\varepsilon_* = \sigma^2\rho^4$ for least-squares fit, admit the same statistical guarantees as the exact solutions (up to a constant).

- Combining this with the usual guarantees for CMP and FGM, $\langle \dots \rangle$

Corollary

To reach the threshold accuracy ε_*, in each case it is sufficient to perform

$$T_* = O(\|\mathcal{F}_n[y]\|_\infty / \sigma)$$

iterations of the suitable first-order algorithm (CMP or FGM).

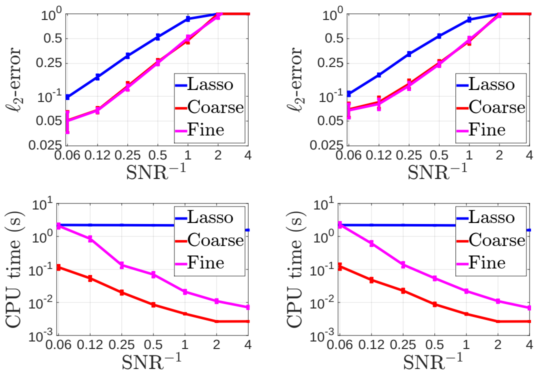


Figure. Comparison of (CLS) with an $\sigma\rho^2$-accurate solution (Coarse), $0.01\sigma\rho^2$-accurate solution (Fine), and the oversampled Lasso estimator*.

Two signal generation scenarios are compared: 4 random frequencies on $[0, 2\pi]$ (left) and 2 random pairs of $\frac{0.2\pi}{n}$-close frequencies (right).

Bhaskar et al. [2013]

Statistical accuracy: T_* experiment

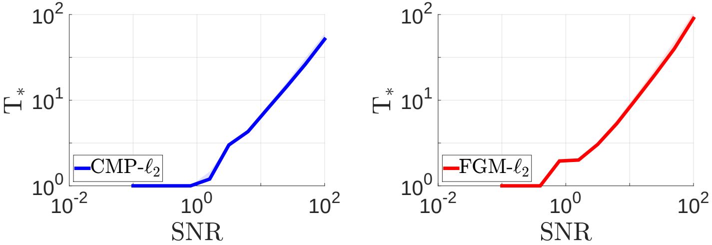


Figure. Iteration at which accuracy ε_* is attained *experimentally* for (CUF), left, and (CLS), right (signal with 4 random frequencies).

Constrained least-squares: phase transition

Constrained least-squares can be recast as (non-squared) ℓ_2-minimization:

$$\min_{\varphi \in \Phi(r)} \mathrm{Res}_2(\varphi) := \|\mathcal{F}_n[y - y * \varphi]\|_2.$$

Objective is non-smooth but can be minimized at rate $O(1/k^2)$ by FGM:

- Indeed, after k iterations of FGM applied to the "squared" problem,

  $$\mathrm{Res}_2^2(\varphi^k) - \mathrm{Res}_2^2(\tilde{\varphi}^*) \le \frac{Q}{k^2},$$

  where φ^* is any minimizer of $\mathrm{Res}_2^2(\cdot)$ on $\Phi(r)$, and Q is a constant.
- Since $t \to t^2$ is monotone on $t \ge 0$, φ_* also minimizes $\mathrm{Res}_2(\cdot)$.
- By the difference-of-squares formula,

  $$\mathrm{Res}_2(\tilde{\varphi}^k) - \mathrm{Res}_2(\varphi^*) \le \frac{Q}{(\mathrm{Res}_2(\tilde{\varphi}_k) + \mathrm{Res}_2(\varphi^*))k^2} \le \frac{Q}{2\mathrm{Res}_2(\varphi^*)k^2}$$

  (Note that this requires the "non-ideal" fit: $\mathrm{Res}_2(\varphi^*) > 0$.)

Constrained least-squares: phase transition (cont.)

- We also have the usual $O(1/k)$ rate as in "Nesterov's smoothing":

$$\mathrm{Res}_2(\tilde{\varphi}^k) \le \sqrt{\mathrm{Res}_2^2(\varphi^*) + \frac{Q}{k^2}} \le \mathrm{Res}_2(\varphi^*) + \frac{\sqrt{Q}}{k}.$$

- To summarize,

$$\mathrm{Res}_2(\tilde{\varphi}^k) - \mathrm{Res}_2(\varphi^*) \le \min\left(\frac{\sqrt{Q}}{k}, \frac{Q}{2\mathrm{Res}_2(\varphi^*)k^2}\right),$$

i.e. there is an "elbow" at $k \approx \frac{\sqrt{Q}}{2\mathrm{Res}_2(\varphi_*)}$. Confirmed empirically:

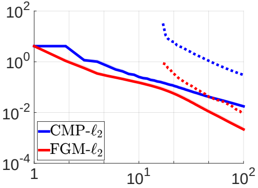


Figure. Relative accuracy vs. iteration for (CLS) with non-squared residual solved with Mirror Prox and FGM (2 pairs of close frequencies, SNR = 4).

Conclusions and perspectives

Conclusions

- We construct adaptive estimators for signals with unknown shift-invariant structure;
- We prove statistical bounds on the pointwise and ℓ_2-loss of the new estimators, and compare them with lower bounds.
- We provide efficient algorithmic implementation for the estimators.
- As an application, we address the problem of signal recovery from a shift-invariant subspace without frequency separation assumptions.

Perspectives

- **GPU implementation:** gradient computations are reduced to FFT.
- **Generalization to indirect observations:**

$$y_\tau = [a * x]_\tau + \sigma\xi_\tau,$$

where $a \in \mathbb{C}_m(\mathbb{Z})$ is a known filter.

  - **Applications:** inverse PDEs[1], fluorescence microscopy[2], exoplanet detection[3], ...
  - **Challenge:** adaptation to the "mutual coherency" of a and x.

- **Signal recovery on graphs**[4]: other domains than $\mathbb{Z}$.
  - **Applications:** social network analysis, sensor networks,...
  - **Challenge:** no FFT, difficult to work in the Fourier domain.

[1] [Cavalier et al., 2002], [2] [Waters, 2009; Bissantz et al., 2015], [3] [Fischer et al., 2015; Kim et al., 2017], [4] [Sandryhaila and Moura, 2013]

Thank you for your attention!

Publications and preprints:

- Z. Harchaoui, A. Juditsky, A. Nemirovski, D.O.
  Adaptive Signal Recovery by Convex Optimization. *COLT 2015.*
- D.O., Z. Harchaoui, A. Juditsky, A. Nemirovski.
  Structure-Blind Signal Recovery. *NIPS 2016*.
  extended version: `arXiv:1607.05712`.
- D.O, Z. Harchaoui.
  Efficient First-Order Algorithms for Adaptive Signal Denoising.
  Submitted to *ICML 2018*.
- D.O, Z. Harchaoui, A. Juditsky, A. Nemirovski.
  Adaptive Signal Recovery: an Overview. *In preparation*.
- D.O, Z. Harchaoui, A. Juditsky, A. Nemirovski.
  Adaptive Signal Deconvolution by Convex Optimization. *In preparation*.

References I

Adams, R. A. and Fournier, J. J. (2003). *Sobolev spaces*, volume 140. Academic press.

Bhaskar, B., Tang, G., and Recht, B. (2013). Atomic norm denoising with applications to line spectral estimation. *IEEE Trans. Signal Processing*, 61(23):5987–5999.

Bickel, P., Ritov, Y., and Tsybakov, A. (2009). Simultaneous analysis of Lasso and Dantzig selector. *Ann. Statist.*, 37(4):1705–1732.

Bissantz, K., Bissantz, N., and Proksch, K. (2015). *Monitoring of Significant Changes Over Time in Fluorescence Microscopy Imaging of Living Cells*. Universitätsbibliothek Dortmund.

Brown, L. D., Low, M. G., et al. (1996). Asymptotic equivalence of nonparametric regression and white noise. *The Annals of Statistics*, 24(6):2384–2398.

Bühlmann, P. and Van De Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.

References II

Cavalier, L., Golubev, G., Picard, D., Tsybakov, A., et al. (2002). Oracle inequalities for inverse problems. *The Annals of Statistics*, 30(3):843–874.

Donoho, D. L., Liu, R. C., and MacGibbon, B. (1990). Minimax risk over hyperrectangles, and implications. *The Annals of Statistics*, pages 1416–1437.

Fischer, D. A., Howard, A. W., Laughlin, G. P., Macintosh, B., Mahadevan, S., Sahlmann, J., and Yee, J. C. (2015). Exoplanet detection techniques. *arXiv preprint arXiv:1505.06869*.

Goldenshluger, A., Lepski, O., et al. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39(3):1608–1632.

Goldenshluger, A. and Nemirovski, A. (1997). Adaptive de-noising of signals satisfying differential inequalities. *IEEE Transactions on Information Theory*, 43(3):872–889.

Ibragimov, I. and Khasminskii, R. (1984). Nonparametric estimation of the value of a linear functional in gaussian white noise. *Theor. Probab. & Appl.*, 29:1–32.

Johnstone, I. (2011). Gaussian estimation: sequence and multiresolution models.

Juditsky, A. and Nemirovski, A. (2009). Nonparametric denoising of signals with unknown local structure, I: Oracle inequalities. *Appl. & Comput. Harmon. Anal.*, 27(2):157–179.

Juditsky, A. and Nemirovski, A. (2010). Nonparametric denoising signals of unknown local structure, II: Nonparametric function recovery. *Appl. & Comput. Harmon. Anal.*, 29(3):354–367.

Juditsky, A. and Nemirovski, A. (2011a). First-order methods for nonsmooth convex large-scale optimization, I: general purpose methods. *Optimization for Machine Learning*, pages 121–148.

Juditsky, A. and Nemirovski, A. (2011b). First-order methods for nonsmooth convex large-scale optimization, II: utilizing problem structure. *Optimization for Machine Learning*, pages 149–183.

Juditsky, A. and Nemirovski, A. (2013). On detecting harmonic oscillations. *Bernoulli*, 23(2):1134–1165.

Juditsky, A. and Nemirovski, A. (2017). Near-optimality of linear recovery from indirect observations. *arXiv preprint arXiv:1704.00835*.

References IV

Kim, T. H., Lee, K. M., Schölkopf, B., and Hirsch, M. (2017). Online video deblurring via dynamic temporal blending network. In *IEEE International Conference on Computer Vision (ICCV 2017)*.

Laurent, B. and Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.*, 28(5):1302–1338.

Lepski, O. (1991). On a problem of adaptive estimation in Gaussian white noise. *Theory of Probability & Its Applications*, 35(3):454–466.

Lepski, O. et al. (2015). Adaptive estimation over anisotropic functional classes via oracle approach. *The Annals of Statistics*, 43(3):1178–1242.

Lepski, O., Mammen, E., and Spokoiny, V. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *The Annals of Statistics*, pages 929–947.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142.

Nemirovski, A. (1991). On non-parametric estimation of functions satisfying differential inequalities.

References V

Nemirovski, A., Onn, S., and Rothblum, U. (2010). Accuracy certificates for computational problems with convex structure. *Mathematics of Operations Research*, 35(1):52–78.

Nesterov, Y. and Nemirovski, A. (2013). On first-order algorithms for ℓ_1/nuclear norm minimization. *Acta Numerica*, 22:509–575.

Sandryhaila, A. and Moura, J. M. (2013). Discrete signal processing on graphs. *IEEE transactions on signal processing*, 61(7):1644–1656.

Tang, G., Bhaskar, B., and Recht, B. (2013). Near minimax line spectral estimation. In *Information Sciences and Systems (CISS), 2013 47th Annual Conference on*, pages 1–6. IEEE.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 58(1):267–288.

Tsybakov, A. (2008). *Introduction to Nonparametric Estimation*. Springer.

Waters, J. C. (2009). Accuracy and precision in quantitative fluorescence microscopy. *The Journal of Cell Biology*, 185(7):1135–1148.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 359–372.