Efficient First-Order Algorithms for Adaptive Signal Denoising

Dmitrii Ostrovskii[*] Zaid Harchaoui[†]

[*]INRIA Paris, Ecole Normale Supérieure

[†]University of Washington

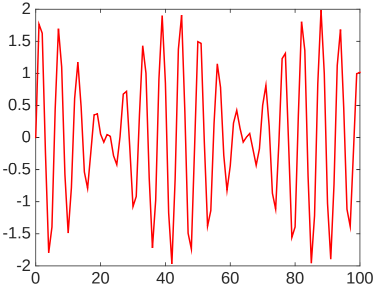
ICML 2018
Stockholm

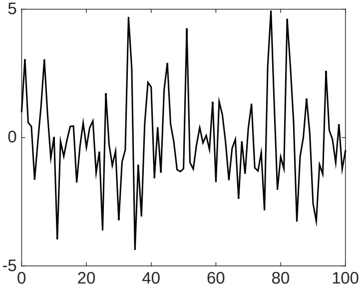
Signal denoising problem

Recover discrete-time **signal** $x = (x_\tau) \in \mathbb{C}^{2n+1}$ from noisy **observations**

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

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





Difficulty: unknown structure

Adaptive denoising: background*

Linear time-invariant estimator: convolution of y with filter $\varphi \in \mathbb{C}^{n+1}$:

$$\widehat{x}_t = [\varphi * y]_t := \sum_{0 \le \tau \le n} \varphi_\tau y_{t-\tau}, \quad 0 \le t \le n,$$

- Suppose x satisfies discrete ODE (sines, polynomials, exponentials):

$$P(\Delta)x \approx 0,$$

where $[\Delta x]_t := x_{t-1}$, and operator $P(\Delta) = \sum_{k=1}^{d} p_k \Delta^k$ is **unknown**.

- Then there **exists** φ^o with near-optimal risk and small ℓ_1-norm of Discrete Fourier transform $\mathcal{F}_n[\varphi^o]$:

$$\|\mathcal{F}_n[\varphi^o]\|_1 \le \frac{r}{\sqrt{n+1}}, \quad r = \text{poly}(\deg(P)).$$

Goal: construct **adaptive filter** $\widehat{\varphi} = \widehat{\varphi}(y)$ with similar properties to φ^o.

*[Juditsky and Nemirovski, 2009, 2010; Harchaoui et al., 2015; Ostrovsky et al., 2016]

Estimators

$$\text{minimize} \;\; \text{Res}_p(\varphi) := \left\| \mathcal{F}_{\boldsymbol{n}}[\boldsymbol{y} - \varphi * \boldsymbol{y}]_n^{2n} \right\|_p$$
$$\text{subject to} \;\; \varphi \in \Phi(r) := \left\{ \|\mathcal{F}_n[\varphi]\|_1 \leq \frac{r}{\sqrt{n+1}} \right\}.$$

Least Squares [Ostrovsky et al., 2016]:
$\boldsymbol{p} = \mathbf{2}$ ($\Rightarrow \ell_2$-loss guarantees)

Uniform Fit [Harchaoui et al., 2015]:
$\boldsymbol{p} = \boldsymbol{\infty}$ ($\Rightarrow \ell_\infty$-loss guarantees)

- ☺ **simple constraint:** proximal mapping computed in $O(n)$;
- ☺ **first-order oracle:** computed in $O(n \log n)$ by reducing to FFT;
- ☺ **low accuracy:** *are crude approximate solutions sufficient?*

First-order methods

Strategies

Fourier-domain: $u := \mathcal{F}_n[\varphi]$, $b = \mathcal{F}_n[[y]_n^{2n}]$, $\mathcal{A}u := \mathcal{F}_n[[y * \varphi]_n^{2n}]$.

Least Squares: **quadratic** problem on ℓ_1-ball:

$$\min_{\|u\|_1 \le \frac{r}{\sqrt{n+1}}} \|\mathcal{A}u - b\|_2^2.$$

- **Fast Gradient Method:** $O(1/T^2)$ convergence after T iterations.*

Uniform Fit: reduced to a **bilinear saddle-point** problem:

$$\min_{\|u\|_1 \le \frac{r}{\sqrt{n+1}}} \|\mathcal{A}u - b\|_\infty = \min_{\|u\|_1 \le \frac{r}{\sqrt{n+1}}} \max_{\|v\|_1 \le 1} \langle v, \mathcal{A}u \rangle - \langle v, b \rangle.$$

- **Mirror Prox:** $O(1/T)$ convergence after T iterations.*

☺ ℓ_1-adapted geometry, dual certificates, adaptive step, proximal terms.

*[Nesterov and Nemirovski, 2013; Juditsky and Nemirovski, 2011]

Statistical accuracy: theoretical result

Let $\|x\|_{n,p}$ be the "estimation norm" with the right scaling:

$$\|x\|_{n,p} = \left(\frac{1}{n+1} \sum_{t=n}^{2n} |x_t|^p \right)^{1/p}.$$

- **Exact solutions** [Harchaoui et al., 2015; Ostrovsky et al., 2016]:

$$\mathbb{P}\left\{ \|x - \widehat{\varphi}_{LS} * y\|_{n,2} \geq C\sigma r \sqrt{\frac{\log(n/\delta)}{n+1}} \right\} \leq \delta,$$

$$\mathbb{P}\left\{ \|x - \widehat{\varphi}_{UF} * y\|_{n,\infty} \geq C\sigma r^2 \sqrt{\frac{\log(n/\delta)}{n+1}} \right\} \leq \delta.$$

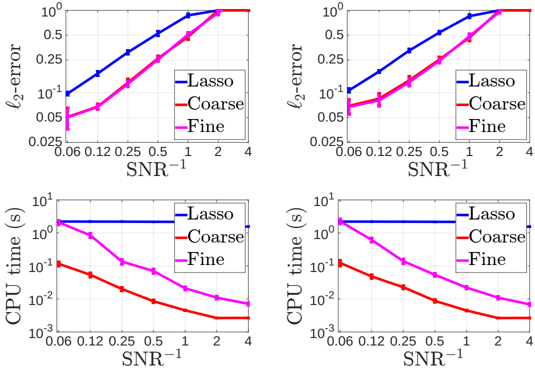
- We **extend** these results to **approximate solutions**:

Theorem A

Approximate solutions $\tilde{\varphi}$ with accuracy $\varepsilon_* = \boldsymbol{\sigma r}$ for Uniform Fit and $\varepsilon_* = \boldsymbol{\sigma^2 r^2}$ for Least Squares admit the same bounds as the exact ones.

Experiment: early stopping

Comparison of ℓ_2-loss and computation time in two scenarios: sum of sines with 4 random frequencies and 2 pairs of close frequencies (right)*.



- **Coarse:** crude Least Squares solution with accuracy $\varepsilon_* = \sigma^2 r^2$;
- **Fine:** near-optimal Least Squares solution with accuracy $0.01\varepsilon_*$;
- **Lasso:** 10-fold oversampled Lasso estimator [Bhaskar et al., 2013].

Code available at `https://github.com/ostrodmit/AlgoRec`

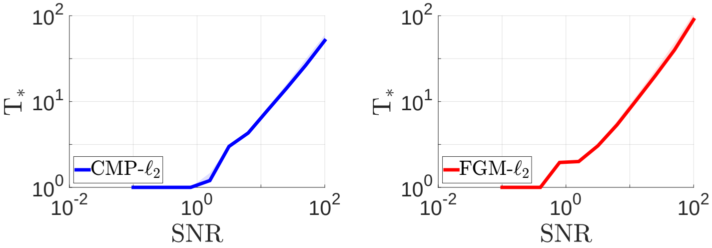
Algorithmic complexity

Theorem B

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

$$T_* = O\left(\text{PSNR} + 1\right)$$

steps of the corresponding algorithm.



Iteration at which accuracy ε_* is attained **experimentally** on the sum of sines with 4 random frequencies: Uniform Fit (left), Least Squares (right).

Thank you and see you at poster B#51

Where I will also show how to solve some non-smooth problems in $O(1/T^2)$.

References

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.

Harchaoui, Z., Juditsky, A., Nemirovski, A., and Ostrovsky, D. (2015). Adaptive recovery of signals by convex optimization. In *Proceedings of The 28th Conference on Learning Theory (COLT) 2015, Paris, France, July 3-6, 2015*, pages 929–955.

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.

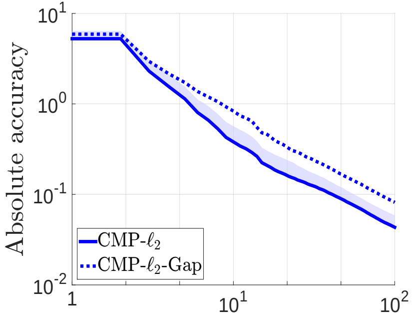
Juditsky, A. and Nemirovski, A. (2011). First-order methods for nonsmooth convex large-scale optimization, II: Utilizing problem structure. *Optimization for Machine Learning*, pages 149–183.

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

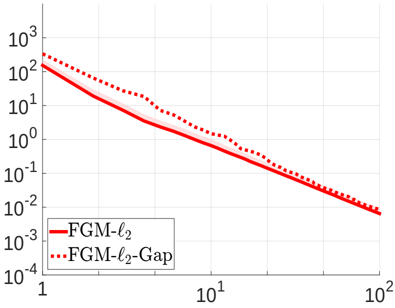
Ostrovsky, D., Harchaoui, Z., Juditsky, A., and Nemirovski, A. (2016). Structure-blind signal recovery. In *Advances in Neural Information Processing Systems*, pages 4817–4825.

Convergence: numerical experiment

Constrained uniform-fit (Mirror Prox)



Constrained least-squares (Fast Gradient Method)



Convergence of the residual (95% upper confidence bound) for a sum of $s = 4$ sinusoids with random frequencies and amplitudes, SNR $= 4$.

Dashed: online accuracy bounds via the dual certificate.