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

\[
y_\tau = x_\tau + \sigma \xi_\tau, \quad \tau = -n, \ldots, n,
\]

where \( \xi_\tau \) are i.i.d. standard Gaussian random variables.

Difficulty: unknown structure
Adaptive denoising: background

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

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

- 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 \( \varphi^o \) with near-optimal risk and small \( \ell_1 \)-norm of Discrete Fourier transform \( \mathcal{F}_n[\varphi^o] \):
  \[
  \| \mathcal{F}_n[\varphi^o] \|_1 \leq \frac{r}{\sqrt{n+1}}, \quad r = \text{poly}(\deg(P)).
  \]

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

*Juditsky and Nemirovski, 2009, 2010; Harchaoui et al., 2015; Ostrovsky et al., 2016*
minimize $\text{Res}_p(\varphi) := \|\mathcal{F}_n[y - \varphi * y]_n^{2n}\|_p$

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]:

$p = 2 \ (\Rightarrow \ell_2\text{-loss guarantees})$

**Uniform Fit** [Harchaoui et al., 2015]:

$p = \infty \ (\Rightarrow \ell_\infty\text{-loss guarantees})$

- 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?
Strategies

Fourier-domain: \( u := \mathcal{F}_n[\varphi], \quad b = \mathcal{F}_n[y]_n, \quad Au := \mathcal{F}_n[y \ast \varphi]_n \).

**Least Squares:** quadratic problem on \( \ell_1 \)-ball:

\[
\min_{\|u\|_1 \leq \frac{r}{\sqrt{n+1}}} \|Au - b\|_2^2.
\]

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

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

\[
\min_{\|u\|_1 \leq \frac{r}{\sqrt{n+1}}} \|Au - b\|_\infty = \min_{\|u\|_1 \leq \frac{r}{\sqrt{n+1}}} \max_{\|v\|_1 \leq 1} \langle v, Au \rangle - \langle v, b \rangle.
\]

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

\( \ell_1 \)-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]:

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

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

**Theorem A**

Approximate solutions $\hat{\varphi}$ with accuracy $\varepsilon_* = \sigma r$ for Uniform Fit and $\varepsilon_* = \sigma^2 r^2$ for Least Squares admit the same bounds as the exact ones.
Comparison of $\ell_2$-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 $\varepsilon_*$, in each case it is sufficient to perform

$$T_* = O(SNR + 1)$$

steps of the corresponding algorithm.

Iteration at which accuracy $\varepsilon_*$ 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 #51

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


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.