Prony's Problem and the Restricted Isometric Property

Benedikt Diederichs, Armin Iske

Department of Mathematics, University of Hamburg

IM-Workshop
20.02.2018

Outline

Prony's Problem

Consider an exponential sum

$$f(x) = \sum_{j=1}^{M} c_j e^{2\pi i y_j x}.$$

- $Y^f = \{y_j \in [0, 1) \ : \ j = 1, \ldots, M\}$ are the frequencies of f
- $c_j \neq 0$ the corresponding coefficients
- M is called order of f

Prony's Problem

Consider an exponential sum

$$f(x) = \sum_{j=1}^{M} c_j e^{2\pi i y_j x}.$$

- $Y^f = \{y_j \in [0,1) \ : \ j = 1, \ldots, M\}$ are the frequencies of f
- $c_j \neq 0$ the corresponding coefficients
- M is called order of f

Prony's Problem

Given $2M$ samples $f(k)$, $k = 0, \ldots 2M-1$, determine f's frequencies and their coefficients.

Often, $2N$ samples are given, $N \geq M$ and M is unkown.

Sparse Model

Let

$$\mathcal{S} = \left\{ \sum_{j=1}^{M} c_j e^{2\pi i y_j x} \ : \ c_j \in \mathbb{C},\ y_j \in [0,1),\ M \in \mathbb{N} \right\}$$

and the sampling operator

$$\mathcal{P}_N : \mathcal{S} \to \mathbb{C}^{2N+1}, \quad \mathcal{P}_N(f) = (f(k))_{|k| \leq N}.$$

We denote the (unknown) ground truth by $\tilde{f}$. The inverse problem

$$\text{Find } f \in \mathcal{S} \ \text{ with } \ \mathcal{P}_N(f) = \mathcal{P}_N(\tilde{f})$$

has infinitely many solutions (for every N).

Model assumption: $\tilde{f}$ is the *sparsest* solution (always true if $N \geq M$).

Warning: $\mathcal{S}$ is not even a normed space!

Compressed Sensing

Want: *Stable recovery guarantees!* Quite similar to compressed sensing, so let us recall some ideas...

Reminder: Let $\tilde{x} \in \mathbb{R}^N$ be an s-sparse vector. Consider

$$\mathcal{A} : \mathbb{R}^N \to \mathbb{R}^M, \quad x \mapsto \mathcal{A}x = Ax$$

Now assume that $N \gg M$. Then

$$\text{Find } x \in \mathbb{R}^N \text{ with } \mathcal{A}x = \mathcal{A}\tilde{x}$$

has infinitely many solutions. But for special $\mathcal{A}$ we can recover $\tilde{x}$!

Compressed Sensing

Want: *Stable recovery guarantees!* Quite similar to compressed sensing, so let us recall some ideas...

Reminder: Let $\tilde{x} \in \mathbb{R}^N$ be an s-sparse vector. Consider

$$\mathcal{A} : \mathbb{R}^N \to \mathbb{R}^M, \quad x \mapsto \mathcal{A}x = Ax$$

Now assume that $N \gg M$. Then

$$\text{Find } x \in \mathbb{R}^N \text{ with } \mathcal{A}x = \mathcal{A}\tilde{x}$$

has infinitely many solutions. But for special $\mathcal{A}$ we can recover $\tilde{x}$!

Example: We pick $z_1, \ldots, z_N$ pairwise distinct and

$$A = \begin{pmatrix} 1 & \ldots & 1 \\ & \ddots & \\ z_1^{2s} & \ldots & z_N^{2s} \end{pmatrix} \in \mathbb{C}^{2s \times N}$$

If $x \neq y$ are s-sparse, $A(x - y) \neq 0$. But this is not very stable.

Compressed Sensing

Want: *Stable recovery guarantees!* Quite similar to compressed sensing, so let us recall some ideas...

Reminder: Let $\tilde{x} \in \mathbb{R}^N$ be an s-sparse vector. Consider

$$\mathcal{A} : \mathbb{R}^N \to \mathbb{R}^M, \quad x \mapsto \mathcal{A}x = Ax$$

Now assume that $N \gg M$. Then

$$\text{Find } x \in \mathbb{R}^N \text{ with } \mathcal{A}x = \mathcal{A}\tilde{x}$$

has infinitely many solutions. But for special $\mathcal{A}$ we can recover $\tilde{x}$!

Definition: A is said to have the *restricted isometric property* of order $2s$, if

$$(1-\delta)\|x\|_2^2 \le \|Ax\|_2^2 \le (1+\delta)\|x\|_2^2$$

for all $2s$-sparse $x \in \mathbb{R}^N$ and a constant $\delta < 1$.

Compressed Sensing: Stability

Now assume that $\tilde{x}$ is s-sparse and that A satisfies the RIP of order $2s$. Let x be s-sparse and

$$\mathcal{A}x = \mathcal{A}\tilde{x} + \varepsilon.$$

Then, because $x - \tilde{x}$ is $2s$-sparse,

$$\|x - \tilde{x}\|_2^2 \leq \frac{\|\varepsilon\|_2^2}{1-\delta}.$$

Conclusion: Every solution to the noisy problem satisfying our model assumption is close to the ground truth!

Compressed Sensing: Stability

Now assume that $\tilde{x}$ is s-sparse and that A satisfies the RIP of order $2s$. Let x be s-sparse and

$$\mathcal{A}x = \mathcal{A}\tilde{x} + \varepsilon.$$

Then, because $x - \tilde{x}$ is $2s$-sparse,

$$\|x - \tilde{x}\|_2^2 \leq \frac{\|\varepsilon\|_2^2}{1-\delta}.$$

Conclusion: Every solution to the noisy problem satisfying our model assumption is close to the ground truth!

If

$$x_{\min}, \tilde{x}_{\min} > \frac{\|\varepsilon\|_2}{\sqrt{1-\delta}}$$

we have $\operatorname{supp} x = \operatorname{supp} \tilde{x}$ and the error is due to solving an overdetermined linear system with a perturbed rhs.

RIP of $\mathcal{P}_N$

Question: Does our sampling operator $\mathcal{P}_N$ satisfies some kind of RIP?

Let $f \in \mathcal{S}$ be given by

$$f(x) = \sum_{j=1}^{M} c_j e^{2\pi i y_j \cdot x}.$$

Assume that $|y_j - y_k|_{\mathbb{T}} \geq q$. Here, we consider the *wrap-around distance*

$$|y - z|_{\mathbb{T}} := \min_{k \in \mathbb{Z}} |y - z - k|$$

Example:

$$|0.7 - 0.1|_{\mathbb{T}} = 0.4$$

We denote by q_f the largest q for which $|y_j - y_k|_{\mathbb{T}} \geq q$ and use the notation

$$\mathcal{S}_q = \{f \in \mathcal{S} \ : \ q_f \geq q\}.$$

RIP of $\mathcal{P}_N$

Question: Does our sampling operator $\mathcal{P}_N$ satisfies some kind of RIP?

Let $f \in \mathcal{S}$ be given by

$$f(x) = \sum_{j=1}^{M} c_j e^{2\pi i y_j \cdot x}.$$

Assume that $|y_j - y_k|_{\mathbb{T}} \geq q$. Then, if N is sufficiently large,

$$(1-\delta)\|c\|_2^2 \leq \|\mathcal{P}_N(f)\|_2^2 \leq (1+\delta)\|c\|_2^2.$$

RIP of $\mathcal{P}_N$

Question: Does our sampling operator $\mathcal{P}_N$ satisfies some kind of RIP?

Let $f \in \mathcal{S}$ be given by

$$f(x) = \sum_{j=1}^{M} c_j e^{2\pi i y_j \cdot x}.$$

Assume that $|y_j - y_k|_{\mathbb{T}} \geq q$. Then, if N is sufficiently large,

$$(1-\delta)\|c\|_2^2 \leq \|\mathcal{P}_N(f)\|_2^2 \leq (1+\delta)\|c\|_2^2.$$

The model assumption of q-separated frequencies + this property does not give rise to a stability guarantee!

Problem: $f, \tilde{f}$ q-separated does not say anything about the separation of $f - \tilde{f}$.

In this talk, we overcome this difficulty and prove stability.

Stable Reconstruction

Consider

$$\mathcal{S} = \left\{ \sum_{j=1}^{M} c_j e^{2\pi i y_j \cdot x} \;:\; c_j \in \mathbb{C},\; y_j \in [0,1),\; M \in \mathbb{N} \right\}$$

and the sampling operator

$$\mathcal{P}_N : \mathcal{S} \to \mathbb{C}^{2N+1}, \quad \mathcal{P}_N(f) = (f(k))_{|k| \le N}.$$

Want: If $f, g \in \mathcal{S}$ satisfy

$$\|\mathcal{P}_N(f-g)\|_2 \ll 1 \quad \Rightarrow \quad f, g \text{ are similar.}$$

Here, *similar* should imply

$$d_H(Y^f, Y^g) = \max\{\max_{y \in Y^f} \min_{y' \in Y^g} |y - y'|_{\mathbb{T}}, \max_{y' \in Y^g} \min_{y \in Y^f} |y - y'|_{\mathbb{T}}\} \ll 1.$$

Limits of Reconstruction

Several problems arise:

$$\|\mathcal{P}_N(\varepsilon e^{2\pi i y_j x})\|_2^2 \lesssim_N \varepsilon^2,$$

hence $\|\mathcal{P}_N(f-g)\|_2$ has to be small compared to the smallest coefficient of f, g. Modulus of smallest coefficient: $c_{\min}$.

Limits of Reconstruction

Several problems arise:

$$\|\mathcal{P}_N(\varepsilon e^{2\pi i y_j x})\|_2^2 \lesssim_N \varepsilon^2,$$

hence $\|\mathcal{P}_N(f-g)\|_2$ has to be small compared to the smallest coefficient of f, g. Modulus of smallest coefficient: $c_{\min}$.

$$\|\mathcal{P}_N(e^{2\pi i(y_j+\varepsilon)x} - e^{2\pi i y_j x})\|_2^2 = \mathcal{O}_N(\varepsilon^2) \quad \text{for} \quad \varepsilon \to 0,$$

thus "clustered" frequencies are problematic.

Limits of Reconstruction

Several problems arise:

$$\|\mathcal{P}_N(\varepsilon e^{2\pi i y_j x})\|_2^2 \lesssim_N \varepsilon^2,$$

hence $\|\mathcal{P}_N(f-g)\|_2$ has to be small compared to the smallest coefficient of f, g. Modulus of smallest coefficient: $c_{\min}$.

$$\|\mathcal{P}_N(e^{2\pi i(y_j+\varepsilon)x} - e^{2\pi i y_j x})\|_2^2 = \mathcal{O}_N(\varepsilon^2) \quad \text{for} \quad \varepsilon \to 0,$$

thus "clustered" frequencies are problematic. Actually, a much stronger result holds, which confirms the need for separation.

Theorem (Moitra, 2015, Informal)

If $N < (1-\varepsilon)/q$, then two exponential sums $f, g \in \mathcal{S}_q$ with k frequencies exists, where we need a noise level smaller than $2^{-\Omega(\varepsilon k)}$ to distinguish them.
Further, this result holds even true if we assume that we measure $f(k) + \eta_k$ where η_k are independent Gaussian random variables.

Main Theorem

Main Question

Given $f, g \in \mathcal{S}_q$ with $\|\mathcal{P}_N(f-g)\|_2^2 \ll 1$ and $c_{\min} \gtrsim 1$, are their frequencies Y^f and Y^g close?

Theorem (D., Iske 2017)

Let $f, g \in \mathcal{S}_{2q}$ be given, where $1/q \leq N$. If

$$\|\mathcal{P}_N(f-g)\|_2^2 = \sum_{k=-N}^{N} |f(k) - g(k)|^2 < N c_{\min}^2$$

we have that for any $y \in Y^f$ only one $n(y) \in Y^g$ with $|y - n(y)|_{\mathbb{T}} < q$ exists and

$$4N^4 c_{\min}^2 \|(d(y, Y^g))_y\|_{\ell^3(Y^f)}^3 + \frac{N}{2} \sum_{y \in Y^f} |c_y - c_{n(y)}|^2 \leq \|\mathcal{P}_N(f-g)\|_2^2.$$

Discussion

- The condition

$$\sum_{k=-N}^{N} |f(k) - g(k)|^2 < N c_{\min}^2$$

is, up to a constant, an optimal, necessary condition. Indeed, for $g(x) = f(x) + c_{\min} e^{2\pi ixy}$, we obtain

$$\sum_{k=-N}^{N} |f(k) - g(k)|^2 = (2N+1) c_{\min}^2 .$$

Discussion

- The condition

$$\sum_{k=-N}^{N} |f(k) - g(k)|^2 < Nc_{\min}^2$$

is, up to a constant, an optimal, necessary condition. Indeed, for $g(x) = f(x) + c_{\min} e^{2\pi i x y}$, we obtain

$$\sum_{k=-N}^{N} |f(k) - g(k)|^2 = (2N+1)c_{\min}^2.$$

- If we have only $f^n(k) = f(k) + noise(k)$, we can estimate $\|\mathcal{P}_N(f-g)\|_2^2$ by $\|\mathcal{P}_N(f^n - g)\|_2^2 +$ some noise term, depending on the noise model.
- What about the rate given?

Optimal Rates

$$4N^4 c_{\min}^2 \|(d(y, Y^g))_y\|_{\ell^3(Y^f)}^3 + \frac{N}{2}\sum_{y\in Y^f} |c_y - c_{n(y)}|^2 \le \|\mathcal{P}_N(f-g)\|_2^2.$$

The rate in c is optimal:

$$\sum_{k=-N}^{N} \left| ce^{2\pi iyk} - (c+c_2)e^{2\pi iyk} \right|^2 = (2N+1)|c_2|^2.$$

Optimal Rates

$$4N^4 c_{\min}^2 \|(d(y, Y^g))_y\|_{\ell^3(Y^f)}^3 + \frac{N}{2} \sum_{y \in Y^f} |c_y - c_{n(y)}|^2 \leq \|\mathcal{P}_N(f - g)\|_2^2.$$

The rate in $d(y, Y^g)$ not:

$$\sum_{k=-N}^{N} \left| e^{2\pi i y k} - e^{2\pi i (y+\varepsilon) k} \right|^2 \sim N^3 \varepsilon^2$$

Optimal Rates

$$4N^4 c_{\min}^2 \|(d(y, Y^g))_y\|_{\ell^3(Y^f)}^3 + \frac{N}{2} \sum_{y \in Y^f} |c_y - c_{n(y)}|^2 \leq \|\mathcal{P}_N(f-g)\|_2^2.$$

The rate in $d(y, Y^g)$ not:

$$\sum_{k=-N}^{N} \left| e^{2\pi i y k} - e^{2\pi i (y+\varepsilon) k} \right|^2 \sim N^3 \varepsilon^2$$

Corollary (D., Iske 2018)

Let $f, g \in \mathcal{S}_{2q}$ be given, where $2/q \leq N$.
If $\|\mathcal{P}_N(f-g)\|_2^2 \leq c_{\min}^2 N/8$ we have

$$\pi N^3 \|(d(y, Y^g))_y\|_{\ell^2(Y^f)}^2 \leq \|\mathcal{P}_N(f-g)\|_2^2$$

Numerical Example

We give a simple numerical example, where we choose

$$|Y^f| = 9, \qquad q_f \approx 0.1, \qquad c_y = 1 \;\; \forall y \in Y^f.$$

We take noisy samples, apply ESPRIT and obtain an exponential sum f^e.
Then we compare

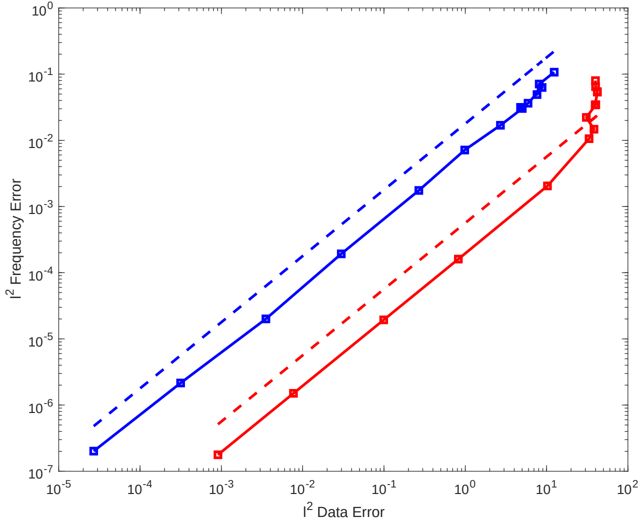
ℓ^2-Data Error : $$\left(\sum_{k=-N}^{N} |f(k) - f^e(k)|^2 \right)^{1/2}$$

ℓ^2-Frequency Error : $$\left(\sum_{y \in Y^{f^e}} (d(y, Y^f))^2 \right)^{1/2}$$

Error Indicator : $$4N^4 c_{\min}^2 \|(d(y, Y^g))_y\|_{\ell^3(Y^f)}^3 + \frac{N}{2} \sum_{y \in Y^f} |c_y - c_{n(y)}|^2$$

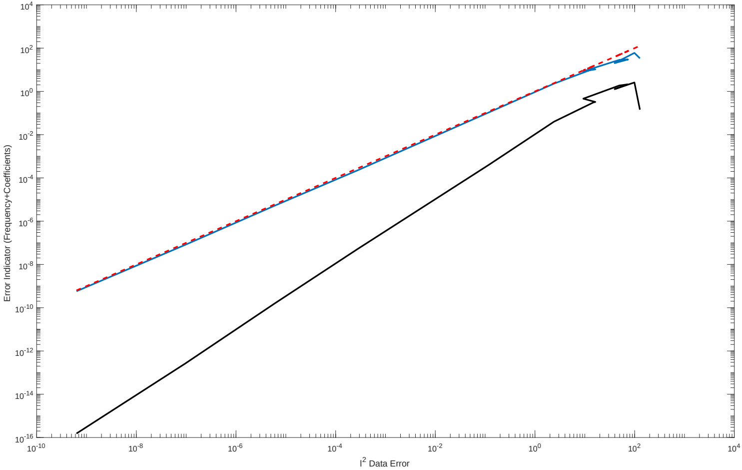
Replacing $f(k)$ by $f(k) + \text{noise}(k)$ chances almost nothing.

Observed Rate



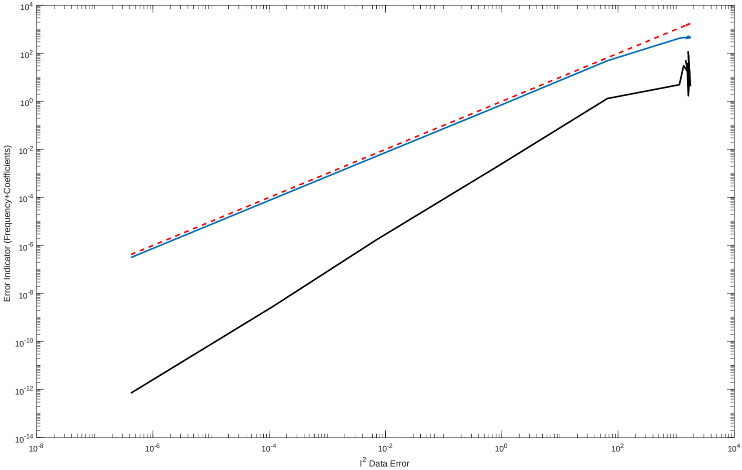
Blue: $N = 10$, Red: $N = 100$

Error Indicator



$N = 10$

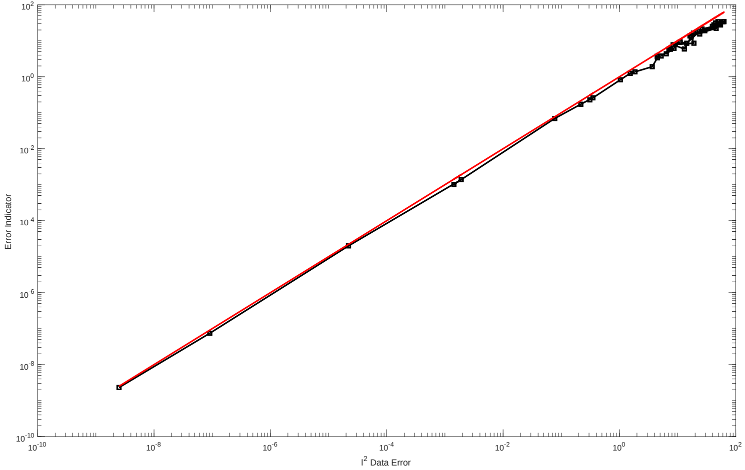
Error Indicator



$N = 100$

2nd Example

Now we pick a function $f \in \mathcal{S}_q$, perturb its frequencies, calculate least-squares coefficient and compare ℓ^2-error and our estimate.



Conclusion

- Prony's problem is a sparse reconstruction problem
- Using a 'well-separated frequency' prior, we have an analog to the Restricted Isometric Property
- This gives rise to conditional well-posedness and a-posteriori error estimates
- The given estimates are asymptotically optimal
- In higher dimensions, a similar result holds true, however more technical and not optimal

Conclusion

- Prony's problem is a sparse reconstruction problem
- Using a 'well-separated frequency' prior, we have an analog to the Restricted Isometric Property
- This gives rise to conditional well-posedness and a-posteriori error estimates
- The given estimates are asymptotically optimal
- In higher dimensions, a similar result holds true, however more technical and not optimal

Thank you for your attention!