Super-Resolution

Yuxin Chen

Princeton University, Fall 2020

Outline

- Classical methods for parameter estimation
  - Polynomial method: Prony's method
  - Subspace method: MUSIC
  - Matrix pencil algorithm
- Optimization-based methods
  - Basis mismatch
  - Atomic norm minimization
  - Connections to low-rank matrix completion

Parameter estimation

Model: a signal is mixture of r modes

$$x[t] = \sum_{i=1}^{r} d_i \psi(t; f_i), \qquad t \in \mathbb{Z}$$

- d_i : amplitudes
- f_i : frequencies
- ψ: (known) modal function, e.g. $\psi(t, f_i) = e^{j2\pi t f_i}$
- r: model order
- $2r$ unknown parameters: $\{d_i\}$ and $\{f_i\}$

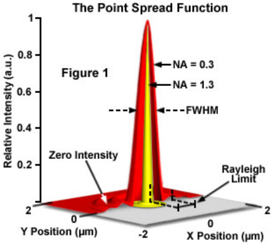
Application: super-resolution imaging

Consider a time signal (a dual problem)

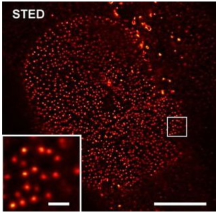
$$z(t) = \sum_{i=1}^{r} d_i \delta(t - t_i)$$

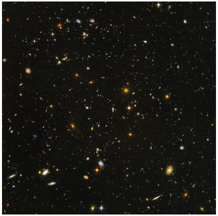
- Resolution is limited by point spread function $h(t)$ of imaging system

$$x(t) = z(t) * h(t)$$



point spread function $h(t)$





Application: super-resolution imaging

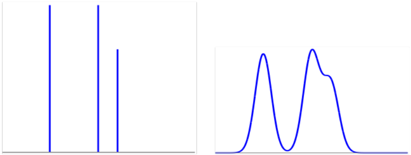
time domain: $$x(t) = z(t) * h(t) = \sum_{i=1}^{r} d_i \, h(t - t_i)$$

spectral domain: $$\hat{x}(f) = \hat{z}(f)\,\hat{h}(f) = \sum_{i=1}^{r} d_i \underbrace{\hat{h}(f)}_{\text{known}} e^{j2\pi f t_i}$$

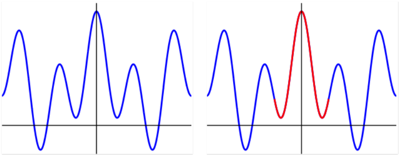
$\Longrightarrow$ observed data: $$\frac{\hat{x}(f)}{\hat{h}(f)} = \sum_{i=1}^{r} d_i \underbrace{e^{j2\pi f t_i}}_{\psi(f;t_i)}, \qquad \forall f : \hat{h}(f) \neq 0$$

$h(t)$ is usually band-limited (suppress high-frequency components)

Application: super-resolution imaging



(a) highly resolved signal $z(t)$; (b) low-pass version $x(t)$



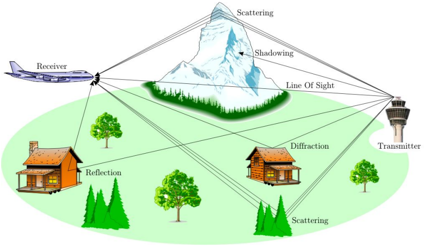
(c) Fourier transform $\hat{z}(f)$; (d) (red) observed spectrum $\hat{x}(f)$

Fig. credit: Candes, Fernandez-Granda '14

Super-resolution: extrapolate high-end spectrum (fine scale details) from low-end spectrum (low-resolution data)

Application: multipath communication channels

In wireless communications, transmitted signals typically reach receiver by multiple paths, due to reflection from objects (e.g. buildings).



Suppose $h(t)$ is transmitted signal, then received signal is

$$x(t) = \sum_{i=1}^{r} d_i \, h(t - t_i) \qquad (t_i : \text{ delay in } i^{\text{th}} \text{ path})$$

$\rightarrow$ same as super-resolution model

Basic model

- **Signal model:** a mixture of sinusoids at r distinct frequencies

$$x[t] = \sum_{i=1}^{r} d_i e^{j2\pi t f_i}$$

where $f_i \in [0,1)$: frequencies; d_i : amplitudes
  - *Continuous dictionary*: f_i can assume *ANY* value in $[0,1)$

- **Observed data:**

$$\boldsymbol{x} = \left[x[0], \cdots, x[n-1]\right]^\top$$

- **Goal:** retrieve frequencies / recover signal (also called harmonic retrieval)

Matrix / vector representation

Alternatively, observed data can be written as

$$\boldsymbol{x} = \boldsymbol{V}_{n\times r}\,\boldsymbol{d} \tag{15.1}$$

where $\boldsymbol{d} = [d_1, \cdots, d_r]^\top$;

$$\boldsymbol{V}_{n\times r} := \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ z_1 & z_2 & z_3 & \cdots & z_r \\ z_1^2 & z_2^2 & z_3^2 & \cdots & z_r^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_1^{n-1} & z_2^{n-1} & z_3^{n-1} & \cdots & z_r^{n-1} \end{bmatrix} \quad \text{(Vandermonde matrix)}$$

with $z_i = e^{j2\pi f_i}$.

Polynomial method: Prony's method

Prony's method

- A *polynomial method* proposed by Gaspard Riche de Prony in 1795
- **Key idea:** construct an annililating filter + polynomial root finding

Annihilating filter

- Define a filter by (Z-transform or characteristic polynomial)

$$G(z) = \sum_{l=0}^{r} g_l z^{-l} = \prod_{l=1}^{r} (1 - z_l z^{-1})$$

whose roots are $\{z_l = e^{j2\pi f_l} \mid 1 \le l \le r\}$

- $G(z)$ is called annihilating filter since it annihilates $x[k]$, i.e.

$$q[k] := \underbrace{g_k * x[k]}_{\text{convolution}} = 0 \qquad (15.2)$$

Proof:

$$q[k] = \sum_{i=0}^{r} g_i x[k-i] = \sum_{i=0}^{r} \sum_{l=1}^{r} g_i d_l z_l^{k-i}$$

$$= \sum_{l=1}^{r} d_l z_l^{k} \Big(\underbrace{\sum_{i=0}^{r} g_i z_l^{-i}}_{=0} \Big) = 0$$

Annihilating filter

Equivalently, one can write (15.2) as

$$\boldsymbol{X}_{\mathrm{e}} \boldsymbol{g} = \boldsymbol{0}, \tag{15.3}$$

where $\boldsymbol{g} = [g_r, \cdots, g_0]^\top$ and

$$\boldsymbol{X}_{\mathrm{e}} := \underbrace{\begin{pmatrix} x[0] & x[1] & x[2] & \cdots & x[r] \\ x[1] & x[2] & x[3] & \cdots & x[r+1] \\ x[2] & x[3] & x[4] & \cdots & x[r+2] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x[n-r-1] & x[n-r] & \cdots & \cdots & x[n-1] \end{pmatrix}}_{\text{Hankel matrix}} \in \mathbb{C}^{(n-r)\times(r+1)} \tag{15.4}$$

Thus, we can obtain coefficients $\{g_i\}$ by solving linear system (15.3).

A crucial decomposition

Vandermonde decomposition

$$\boldsymbol{X}_{\mathrm{e}} = \boldsymbol{V}_{(n-r)\times r}\,\mathrm{diag}(\boldsymbol{d})\,\boldsymbol{V}_{(r+1)\times r}^{\top} \tag{15.5}$$

Implications: if $n \geq 2r$ and $d_i \neq 0$, then

- $\mathsf{rank}(\boldsymbol{X}_{\mathrm{e}}) = \mathsf{rank}(\boldsymbol{V}_{(n-r)\times r}) = \mathsf{rank}(\boldsymbol{V}_{(r+1)\times r}) = r$
- $\mathrm{null}(\boldsymbol{X}_{\mathrm{e}})$ is 1-dimensional $\iff$ nonzero solution to $\boldsymbol{X}_{\mathrm{e}}\boldsymbol{g} = \boldsymbol{0}$ is unique

A crucial decomposition

Vandermonde decomposition

$$\boldsymbol{X}_{\mathrm{e}} = \boldsymbol{V}_{(n-r)\times r}\,\mathrm{diag}(\boldsymbol{d})\,\boldsymbol{V}_{(r+1)\times r}^{\top} \tag{15.5}$$

Proof: For any i and j,

$$\begin{aligned}
[\,\boldsymbol{X}_{\mathrm{e}}\,]_{i,j} = x[i+j-2] = \sum_{l=1}^{r} d_l z_l^{i+j-2} = \sum_{l=1}^{r} z_l^{i-1} d_l z_l^{j-1} \\
= \left(\boldsymbol{V}_{(n-r)\times r}\right)_{i,:} \mathrm{diag}(\boldsymbol{d}) \left(\boldsymbol{V}_{(r+1)\times r}\right)_{j,:}^{\top}
\end{aligned}$$

Prony's method

Algorithm 15.1 Prony's method

1. Find $\boldsymbol{g} = [g_r, \cdots, g_0]^\top \neq \mathbf{0}$ that solves $\boldsymbol{X}_{\mathrm{e}}\boldsymbol{g} = \mathbf{0}$
2. Compute r roots $\{z_l \mid 1 \leq l \leq r\}$ of $G(z) = \sum_{l=0}^{r} g_l z^{-l}$
3. Calculate f_l via $z_l = e^{j2\pi f_l}$

Drawback

- Root-finding for polynomials becomes difficult for large r
- Numerically unstable in the presence of noise

Subspace method: MUSIC

MUltiple SIgnal Classification (MUSIC)

Consider a (slightly more general) Hankel matrix

$$\boldsymbol{X}_{\mathrm{e}} = \begin{pmatrix} x[0] & x[1] & x[2] & \cdots & x[k] \\ x[1] & x[2] & x[3] & \cdots & x[k+1] \\ x[2] & x[3] & x[4] & \cdots & x[k+2] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x[n-k-1] & x[n-k] & \cdots & \cdots & x[n-1] \end{pmatrix} \in \mathbb{C}^{(n-k)\times(k+1)}$$

where $r \leq k \leq n-r$ (note that $k=r$ in Prony's method).

- $\mathrm{null}(\boldsymbol{X}_{\mathrm{e}})$ might span multiple dimensions

MUltiple SIgnal Classification (MUSIC)

- Generalize Prony's method by computing $\{\boldsymbol{v}_i \mid 1 \leq i \leq k-r+1\}$ that forms orthonormal basis for $\mathrm{null}(\boldsymbol{X}_{\mathrm{e}})$

- Let $\boldsymbol{z}(f) := \begin{bmatrix} 1 \\ e^{j2\pi f} \\ \vdots \\ e^{j2\pi k f} \end{bmatrix}$, then it follows from Vandermonde decomposition that

$$\boldsymbol{z}(f_l)^\top \boldsymbol{v}_i = 0, \qquad 1 \leq i \leq k-r+1,\ 1 \leq l \leq r$$

- Thus, $\{f_l\}$ are peaks in pseudospectrum

$$S(f) := \frac{1}{\sum_{i=1}^{k-r+1} |\boldsymbol{z}(f)^\top \boldsymbol{v}_i|^2}$$

MUSIC algorithm

Algorithm 15.2 MUSIC

1. Compute orthonormal basis $\{\boldsymbol{v}_i \mid 1 \leq i \leq k-r+1\}$ for $\mathrm{null}(\boldsymbol{X}_{\mathrm{e}})$
2. Return r largest peaks of $S(f) := \frac{1}{\sum_{i=1}^{k-r+1} |\boldsymbol{z}(f)^\top \boldsymbol{v}_i|^2}$, where $\boldsymbol{z}(f) := [1, e^{j2\pi f}, \cdots, e^{j2\pi k f}]^\top$

Matrix pencil algorithm

Matrix pencil

Definition 15.1

A *linear matrix pencil* $\boldsymbol{M}(\lambda)$ is defined as a combination

$$\boldsymbol{M}(\lambda) = \boldsymbol{M}_1 - \lambda \boldsymbol{M}_2$$

of 2 matrices $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$, where $\lambda \in \mathbb{C}$.

Matrix pencil algorithm

Key idea:

- Create 2 closely related Hankel matrices $\boldsymbol{X}_{\mathrm{e},1}$ and $\boldsymbol{X}_{\mathrm{e},2}$
- Recover $\{f_l\}$ via generalized eigenvalues of matrix pencil

$$\boldsymbol{X}_{\mathrm{e},2} - \lambda \boldsymbol{X}_{\mathrm{e},1}$$

(i.e. all $\lambda \in \mathbb{C}$ obeying $\det(\boldsymbol{X}_{\mathrm{e},2} - \lambda \boldsymbol{X}_{\mathrm{e},1}) = 0$)

Matrix pencil algorithm

Construct 2 Hankel matrices (which differ only by 1 row)

$$\boldsymbol{X}_{\mathrm{e},1} := \begin{pmatrix} x[0] & x[1] & x[2] & \cdots & x[k-1] \\ x[1] & x[2] & x[3] & \cdots & x[k] \\ x[2] & x[3] & x[4] & \cdots & x[k+1] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x[n-k-1] & x[n-k] & \cdots & \cdots & x[n-2] \end{pmatrix} \in \mathbb{C}^{(n-k)\times k}$$

$$\boldsymbol{X}_{\mathrm{e},2} := \begin{pmatrix} x[1] & x[2] & x[3] & \cdots & x[k] \\ x[2] & x[3] & x[4] & \cdots & x[k+1] \\ x[3] & x[4] & x[5] & \cdots & x[k+2] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x[n-k] & x[n-k+1] & \cdots & \cdots & x[n-1] \end{pmatrix} \in \mathbb{C}^{(n-k)\times k}$$

where k is pencil parameter

Matrix pencil algorithm

Fact 15.2

Similar to (15.5), one has

$$\begin{aligned}
\boldsymbol{X}_{\mathrm{e},1} &= \boldsymbol{V}_{(n-k)\times r}\,\mathrm{diag}(\boldsymbol{d})\,\boldsymbol{V}_{k\times r}^{\top}\\
\boldsymbol{X}_{\mathrm{e},2} &= \boldsymbol{V}_{(n-k)\times r}\,\mathrm{diag}(\boldsymbol{d})\,\mathrm{diag}\left(\boldsymbol{z}\right)\,\boldsymbol{V}_{k\times r}^{\top}
\end{aligned}$$

where $\boldsymbol{z} = \left[e^{j2\pi f_1}, \cdots, e^{j2\pi f_r}\right]^{\top}$.

Proof: The result for $\boldsymbol{X}_{\mathrm{e},1}$ follows from (15.5). Regarding $\boldsymbol{X}_{\mathrm{e},2}$,

$$\begin{aligned}
\left[\boldsymbol{X}_{\mathrm{e},2}\right]_{i,j} = x[i+j-1] = \sum_{l=1}^{r} d_l z_l^{i+j-1} &= \sum_{l=1}^{r} z_l^{i-1}(d_l z_l) z_l^{j-1}\\
&= \left(\boldsymbol{V}_{(n-k)\times r}\right)_{i,:}\mathrm{diag}(\boldsymbol{d})\,\mathrm{diag}(\boldsymbol{z})\left(\boldsymbol{V}_{k\times r}\right)_{j,:}^{\top}
\end{aligned}$$

Matrix pencil algorithm

Based on Fact 15.2,

$$\boldsymbol{X}_{\mathrm{e},2} - \lambda \boldsymbol{X}_{\mathrm{e},1} = \boldsymbol{V}_{(n-k)\times r}\, \mathrm{diag}(\boldsymbol{d})\, (\mathrm{diag}\,(\boldsymbol{z}) - \lambda \boldsymbol{I})\, \boldsymbol{V}_{k\times r}^{\top}$$

(Exercise) If $r \leq k \leq n - r$, then

- $\mathrm{rank}(\boldsymbol{V}_{(n-k)\times r}) = \mathrm{rank}(\boldsymbol{V}_{k\times r}) = r$
- Generalized eigenvalues of $\boldsymbol{X}_{\mathrm{e},2} - \lambda \boldsymbol{X}_{\mathrm{e},1}$ are $\{z_l = e^{j2\pi f_l}\}$, which can be obtained by finding eigenvalues of $\boldsymbol{X}_{\mathrm{e},1}^{\dagger} \boldsymbol{X}_{\mathrm{e},2}$

Matrix pencil algorithm

Algorithm 15.3 Matrix pencil algorithm

1. Compute all eigenvalues $\{\lambda_l\}$ of $\boldsymbol{X}_{\mathrm{e},1}^{\dagger}\boldsymbol{X}_{\mathrm{e},2}$
2. Calculate f_l via $\lambda_l = e^{j2\pi f_l}$

Basis mismatch issue

Optimization methods for super resolution?

Recall our representation in (15.1):

$$\boldsymbol{x} = \boldsymbol{V}_{n\times r}\,\boldsymbol{d} \tag{15.6}$$

- **Challenge:** both $\boldsymbol{V}_{n\times r}$ and $\boldsymbol{d}$ are unknown

Alternatively, one can view (15.6) as sparse representation over a continuous dictionary $\{\boldsymbol{z}(f) \mid 0 \le f < 1\}$, where $\boldsymbol{z}(f) = [1, e^{j2\pi f}, \cdots, e^{j2\pi(n-1)f}]^\top$

Optimization methods for super resolution?

One strategy:

- Convert nonlinear representation into linear system via discretization at desired resolution:

$$\text{(assume)} \qquad \boldsymbol{x} = \underbrace{\boldsymbol{\Psi}}_{n\times p \text{ partial DFT matrix}} \boldsymbol{\beta}$$

  - representation over a discrete frequency set $\{0, \frac{1}{p}, \cdots, \frac{p-1}{p}\}$
  - gridding resolution: $1/p$

- Solve ℓ_1 minimization:

$$\text{minimize}_{\boldsymbol{\beta}\in\mathbb{C}^p} \ \|\boldsymbol{\beta}\|_1 \quad \text{s.t. } \boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{\beta}$$

Discretization destroys sparsity

Suppose $n = p$, and recall

$$\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{\beta} = \boldsymbol{V}_{n\times r}\boldsymbol{d}$$

$$\Longrightarrow \qquad \boldsymbol{\beta} = \boldsymbol{\Psi}^{-1}\boldsymbol{V}_{n\times r}\boldsymbol{d}$$

Ideally, if $\boldsymbol{\Psi}^{-1}\boldsymbol{V}_{n\times r} \approx$ submatrix of $\boldsymbol{I}$, then sparsity is preserved.

Discretization destroys sparsity

Suppose $n = p$, and recall

$$\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{\beta} = \boldsymbol{V}_{n\times r}\boldsymbol{d}$$

$$\Longrightarrow \qquad \boldsymbol{\beta} = \boldsymbol{\Psi}^{-1}\boldsymbol{V}_{n\times r}\boldsymbol{d}$$

Simple calculation gives

$$\boldsymbol{\Psi}^{-1}\boldsymbol{V}_{n\times r} = \begin{bmatrix} D(\delta_0) & D(\delta_1) & \cdots & D(\delta_r) \\ D(\delta_0 - \frac{1}{p}) & D(\delta_1 - \frac{1}{p}) & \cdots & D(\delta_r - \frac{1}{p}) \\ \vdots & \vdots & \ddots & \vdots \\ D(\delta_0 - \frac{p-1}{p}) & D(\delta_1 - \frac{p-1}{p}) & \cdots & D(\delta_r - \frac{p-1}{p}) \end{bmatrix}$$

where f_i is mismatched to grid $\{0, \frac{1}{p}, \cdots, \frac{p-1}{p}\}$ by δ_i, and

$$D(f) := \frac{1}{p}\sum_{l=0}^{p-1} e^{j2\pi lf} = \frac{1}{p}e^{j\pi f(p-1)}\underbrace{\frac{\sin(\pi f p)}{\sin(\pi f)}}_{\text{heavy tail}} \quad \text{(Dirichlet kernel)}$$

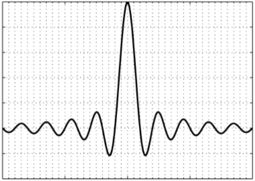
Discretization destroys sparsity

Suppose $n = p$, and recall

$$\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{\beta} = \boldsymbol{V}_{n\times r}\boldsymbol{d}$$

$$\Longrightarrow \qquad \boldsymbol{\beta} = \boldsymbol{\Psi}^{-1}\boldsymbol{V}_{n\times r}\boldsymbol{d}$$

Slow decay / spectral leakage of Dirichlet kernel



If $\delta_i = 0$ (no mismatch), $\boldsymbol{\Psi}^{-1}\boldsymbol{V}_{n\times r}$ = submatrix of $\boldsymbol{I}$

$\Longrightarrow \quad \boldsymbol{\Psi}^{-1}\boldsymbol{V}_{n\times r}\boldsymbol{d}$ is sparse

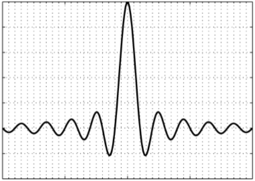
Discretization destroys sparsity

Suppose $n = p$, and recall

$$\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{\beta} = \boldsymbol{V}_{n\times r}\boldsymbol{d}$$

$$\Longrightarrow \qquad \boldsymbol{\beta} = \boldsymbol{\Psi}^{-1}\boldsymbol{V}_{n\times r}\boldsymbol{d}$$

Slow decay / spectral leakage of Dirichlet kernel



If $\delta_i \neq 0$ (e.g. randomly generated), $\boldsymbol{\Psi}^{-1}\boldsymbol{V}_{n\times r}$ may be far from submatrix of $\boldsymbol{I}$

$$\Longrightarrow \qquad \boldsymbol{\Psi}^{-1}\boldsymbol{V}_{n\times r}\boldsymbol{d} \text{ may be incompressible}$$

- Finer gridding does not help!

Mismatch of DFT basis

Loss of sparsity after discretization due to basis mismatch

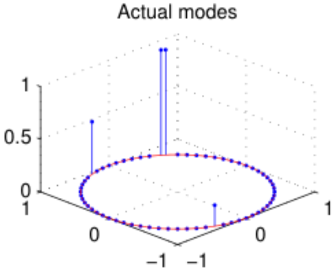


Fig. credit: Chi, Pezeshki, Scharf, Calderbank '10

Optimization-based methods

Atomic set

Consider a set of atoms $\mathcal{A} = \{\psi(\nu) : \nu \in S\}$

- S can be finite, countable, or even continuous
- **Examples:**
  - standard basis vectors (used in compressed sensing)
  - rank-one matrices (used in low-rank matrix recovery)
  - line spectral atoms

$$a(f,\phi) := \underbrace{e^{j\phi}}_{\text{global phase}} [1, e^{j2\pi f}, \cdots, e^{j2\pi(n-1)f}]^\top, \; f \in [0,1), \; \phi \in [0, 2\pi)$$

Atomic norm

Definition 15.3 (Atomic norm, Chandrasekaran, Recht, Parrilo, Willsky '10)

Atomic norm of any $\boldsymbol{x}$ is defined as

$$\|\boldsymbol{x}\|_{\mathcal{A}} := \inf \left\{ \|\boldsymbol{d}\|_1 : \boldsymbol{x} = \sum_k d_k \psi(\nu_k) \right\} = \inf \left\{ t > 0 : \boldsymbol{x} \in t \operatorname{conv}(\mathcal{A}) \right\}$$

- Generalization of ℓ_1 norm for vectors

SDP representation of atomic norm

Consider set of line spectral atoms
$\mathcal{A} := \left\{ a(f,\phi) := e^{j\phi} \cdot [1, e^{j2\pi f}, \cdots, e^{j2\pi(n-1)f}]^\top \;\middle|\; f \in [0,1), \phi \in [0, 2\pi) \right\}$,
then

$$\|\boldsymbol{x}\|_{\mathcal{A}} = \inf_{d_k \geq 0,\ \phi_k \in [0,2\pi),\ f_k \in [0,1)} \left\{ \sum\nolimits_k d_k \mid \boldsymbol{x} = \sum\nolimits_k d_k a(f_k, \phi_k) \right\}$$

Lemma 15.4 (Tang, Bhaskar, Shah, Recht '13)

For any $\boldsymbol{x} \in \mathbb{C}^n$,

$$\|\boldsymbol{x}\|_{\mathcal{A}} = \inf \left\{ \frac{1}{2n} \mathrm{Tr}\left(\mathrm{Toeplitz}(\boldsymbol{u})\right) + \frac{1}{2} t \;\middle|\; \begin{bmatrix} \mathrm{Toeplitz}(\boldsymbol{u}) & \boldsymbol{x} \\ \boldsymbol{x}^* & t \end{bmatrix} \succeq \mathbf{0} \right\} \tag{15.7}$$

Vandermonde decomposition of PSD Toeplitz matrices

Lemma 15.5

Any Toeplitz matrix $\boldsymbol{P} \succeq \boldsymbol{0}$ *can be represented as*

$$\boldsymbol{P} = \boldsymbol{V} \mathrm{diag}(\boldsymbol{d}) \boldsymbol{V}^*,$$

where $\boldsymbol{V} := [a(f_1, 0), \cdots, a(f_r, 0)]$, $d_i \geq 0$, *and* $r = \mathrm{rank}(\boldsymbol{P})$

- Vandermonde decomposition can be computed efficiently via root finding

Proof of Lemma 15.4

Let SDP($\boldsymbol{x}$) be value of RHS of (15.7).

1. **Show that** SDP($\boldsymbol{x}$) $\leq \|\boldsymbol{x}\|_{\mathcal{A}}$**.**

- Suppose $\boldsymbol{x} = \sum_k d_k a(f_k, \phi_k)$ for $d_k \geq 0$. Picking $\boldsymbol{u} = \sum_k d_k a(f_k, 0)$ and $t = \sum_k d_k$ gives (exercise)

$$\mathrm{Toeplitz}(\boldsymbol{u}) = \sum_k d_k a(f_k, 0) a^*(f_k, 0) = \sum_k d_k a(f_k, \phi_k) a^*(f_k, \phi_k)$$

$$\Rightarrow \quad \begin{bmatrix} \mathrm{Toeplitz}(\boldsymbol{u}) & \boldsymbol{x} \\ \boldsymbol{x}^* & t \end{bmatrix} = \sum_k d_k \begin{bmatrix} a(f_k, \phi_k) \\ 1 \end{bmatrix} \begin{bmatrix} a(f_k, \phi_k) \\ 1 \end{bmatrix}^* \succeq \mathbf{0}$$

- Given that $\frac{1}{n}\mathrm{Tr}(\mathrm{Toeplitz}(\boldsymbol{u})) = t = \sum_k d_k$, one has

$$\mathsf{SDP}(\boldsymbol{x}) \leq \sum\nolimits_k d_k.$$

Since this holds for any decomposition of $\boldsymbol{x}$, we conclude this part.

Proof of Lemma 15.4

2. **Show that $\|\boldsymbol{x}\|_{\mathcal{A}} \leq \mathsf{SDP}(\boldsymbol{x})$.**

   i) Suppose for some $\boldsymbol{u}$,

$$\begin{bmatrix} \mathrm{Toeplitz}(\boldsymbol{u}) & \boldsymbol{x} \\ \boldsymbol{x}^* & t \end{bmatrix} \succeq \boldsymbol{0}. \tag{15.8}$$

   Lemma 15.5 suggests Vandermonde decomposition

$$\mathrm{Toeplitz}(\boldsymbol{u}) = \boldsymbol{V}\mathrm{diag}(\boldsymbol{d})\boldsymbol{V}^* = \sum_k d_k a(f_k, 0) a^*(f_k, 0).$$

   This together with the fact $\|a(f_k, 0)\| = \sqrt{n}$ gives

$$\frac{1}{n}\mathrm{Tr}\left(\mathrm{Toeplitz}(\boldsymbol{u})\right) = \sum_k d_k.$$

Proof of Lemma 15.4

2. **Show that $\|\boldsymbol{x}\|_{\mathcal{A}} \leq \mathsf{SDP}(\boldsymbol{x})$.**

   ii) It follows from (15.8) that $\boldsymbol{x} \in \mathrm{range}(\boldsymbol{V})$, i.e.

   $$\boldsymbol{x} = \sum_k w_k a(f_k, 0) = \boldsymbol{V}\boldsymbol{w}$$

   for some $\boldsymbol{w}$. By Schur complement lemma,

   $$\boldsymbol{V}\mathrm{diag}(\boldsymbol{d})\boldsymbol{V}^* \succeq \frac{1}{t}\boldsymbol{x}\boldsymbol{x}^* = \frac{1}{t}\boldsymbol{V}\boldsymbol{w}\boldsymbol{w}^*\boldsymbol{V}^*.$$

   Let $\boldsymbol{q}$ be any vector s.t. $\boldsymbol{V}^*\boldsymbol{q} = \mathrm{sign}(\boldsymbol{w})$. Then

   $$\sum\nolimits_k d_k = \boldsymbol{q}^*\boldsymbol{V}\mathrm{diag}(\boldsymbol{d})\boldsymbol{V}^*\boldsymbol{q} \succeq \frac{1}{t}\boldsymbol{q}^*\boldsymbol{V}\boldsymbol{w}\boldsymbol{w}^*\boldsymbol{V}^*\boldsymbol{q} = \frac{1}{t}\left(\sum\nolimits_k |w_k|\right)^2$$

   $$\Rightarrow \quad t\sum\nolimits_k d_k \geq \left(\sum\nolimits_k |w_k|\right)^2$$

   $$\overset{\text{AM-GM inequality}}{\Longrightarrow} \frac{1}{2n}\mathrm{Tr}\left(\mathrm{Toeplitz}(\boldsymbol{u})\right) + \frac{1}{2}t \geq \sqrt{t\sum\nolimits_k d_k} \geq \sum\nolimits_k |w_k| \geq \|\boldsymbol{x}\|_{\mathcal{A}}$$

Atomic norm minimization

$$\begin{aligned} \text{minimize}_{\boldsymbol{z}\in\mathbb{C}^n} \quad & \|\boldsymbol{z}\|_{\mathcal{A}} \\ \text{s.t.} \quad & z_i = x_i, \quad i \in T \quad \text{(observation set)} \end{aligned}$$

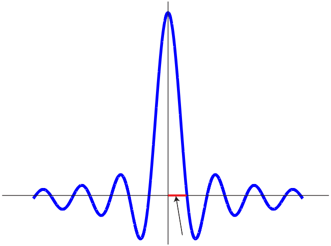
$$\Updownarrow$$

$$\begin{aligned} \text{minimize}_{\boldsymbol{z}\in\mathbb{C}^n} \quad & \frac{1}{2n}\mathrm{Tr}\left(\text{Toeplitz}(\boldsymbol{u})\right) + \frac{1}{2}t \\ \text{s.t.} \quad & z_i = x_i, \quad i \in T \\ & \begin{bmatrix} \mathrm{Toeplitz}(\boldsymbol{u}) & \boldsymbol{z} \\ \boldsymbol{z}^* & t \end{bmatrix} \succeq \boldsymbol{0} \end{aligned}$$

Key metrics

Minimum separation Δ of $\{f_l \mid 1 \le l \le r\}$ is

$$\Delta := \min_{i \neq l} |f_i - f_l|$$



Rayleigh resolution limit: $\lambda_{\mathrm{c}} = \frac{2}{n-1}$

Performance guarantees for super resolution

Suppose $T = \{-\frac{n-1}{2}, \cdots, \frac{n-1}{2}\}$

Theorem 15.6 (Candes, Fernandez-Granda '14)

Suppose that

- **Separation condition:** $\Delta \geq \frac{4}{n-1} = 2\lambda_{\mathrm{c}}$;

Then atomic norm (or total-variation) minimization is exact.

- A deterministic result
- Can recover at most $n/4$ spikes from n consecutive samples
- Does not depend on amplitudes / phases of spikes
- There is no separation requirement if all d_i are positive

Fundamental resolution limits

If $T = \{-\frac{n-1}{2}, \cdots, \frac{n-1}{2}\}$, we cannot go below Rayleigh limit λ_{c}.

Theorem 15.7 (Moitra '15)

If $\Delta < \frac{2}{n-1} = \lambda_{\mathrm{c}}$, then no estimator can distinguish between a particular pair of Δ-separated signals even under exponentially small noise.

Compressed sensing off the grid

Suppose T is random subset of $\{0, \cdots, N-1\}$ of cardinality n

— Extend compressed sensing to continuous domain

Theorem 15.8 (Tang, Bhaskar, Shah, Recht '13)

Suppose that

- **Random sign:** $\mathrm{sign}(d_i)$ *are i.i.d. and random;*
- **Separation condition:** $\Delta \geq \frac{4}{N-1}$*;*
- **Sample size:** $n \gtrsim \max\{r \log r \log N, \log^2 N\}$*.*

Then atomic norm minimization is exact with high prob.

- Random sampling improves resolution limits ($\frac{4}{n-1}$ vs. $\frac{4}{N-1}$)

Connection to low-rank matrix completion

Recall Hankel matrix

$$\boldsymbol{X}_{\mathrm{e}} := \begin{pmatrix} x[0] & x[1] & x[2] & \cdots & x[k] \\ x[1] & x[2] & x[3] & \cdots & x[k+1] \\ x[2] & x[3] & x[4] & \cdots & x[k+2] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x[n-k-1] & x[n-k] & \cdots & \cdots & x[n-1] \end{pmatrix}$$

$$= \boldsymbol{V}_{(n-k)\times r}\,\mathrm{diag}(\boldsymbol{d})\,\boldsymbol{V}_{(k+1)\times r}^{\top} \quad \text{(Vandermonde decomposition)}$$

- rank $(\boldsymbol{X}_{\mathrm{e}}) \leq r$
- Spectral sparsity $\iff$ low rank

Recovery via Hankel matrix completion

Enhanced Matrix Completion (EMaC):

$$\begin{aligned} \underset{\boldsymbol{z}\in\mathbb{C}^n}{\text{minimize}} \quad & \|\boldsymbol{Z}_{\mathsf{e}}\|_* \\ \text{s.t.} \quad & z_i = x_i, \quad i \in T \end{aligned}$$

When T is random subset of $\{0, \cdots, N-1\}$:

- Coherence measure is closely related to separation condition (Liao & Fannjiang '16)
- Similar performance guarantees as atomic norm minimization (Chen, Chi, Goldsmith '14)

Extension to 2D frequencies

Signal model: a mixture of 2D sinusoids at r distinct frequencies

$$x[\boldsymbol{t}] = \sum_{i=1}^{r} d_i e^{j2\pi\langle \boldsymbol{t}, \boldsymbol{f}_i \rangle}$$

where $\boldsymbol{f}_i \in [0,1)^2$: frequencies; d_i : amplitudes

- Multi-dimensional model: $\boldsymbol{f}_i$ can assume *ANY* value in $[0,1)^2$

Vandermonde decomposition

$$\boldsymbol{X} = [x(t_1, t_2)]_{0 \le t_1 < n_1, 0 \le t_2 < n_2}$$

Vandermonde decomposition:

$$\boldsymbol{X} = \boldsymbol{Y} \cdot \mathrm{diag}(\boldsymbol{d}) \cdot \boldsymbol{Z}^\top.$$

where

$$\boldsymbol{Y} := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ y_1 & y_2 & \cdots & y_r \\ \vdots & \vdots & \vdots & \vdots \\ y_1^{n_1-1} & y_2^{n_1-1} & \cdots & y_r^{n_1-1} \end{bmatrix}, \boldsymbol{Z} := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_r \\ \vdots & \vdots & \vdots & \vdots \\ z_1^{n_2-1} & z_2^{n_2-1} & \cdots & z_r^{n_2-1} \end{bmatrix}$$

with $y_i = \exp(j2\pi f_{1i}), \quad z_i = \exp(j2\pi f_{2i})$.

Multi-fold Hankel matrix (Hua '92)

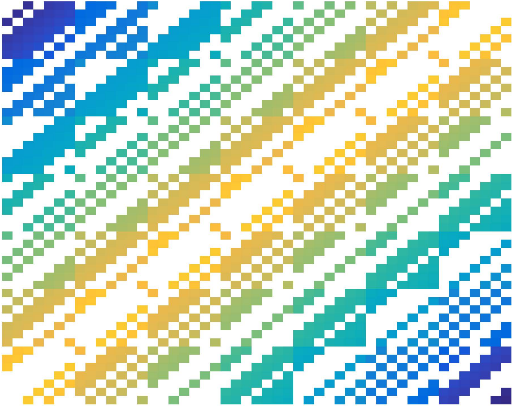
An **enhanced form** $\boldsymbol{X}_{\mathsf{e}}$: $k_1 \times (n_1 - k_1 + 1)$ block Hankel matrix

$$\boldsymbol{X}_{\mathsf{e}} = \begin{bmatrix} \boldsymbol{X}_0 & \boldsymbol{X}_1 & \cdots & \boldsymbol{X}_{n_1-k_1} \\ \boldsymbol{X}_1 & \boldsymbol{X}_2 & \cdots & \boldsymbol{X}_{n_1-k_1+1} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{X}_{k_1-1} & \boldsymbol{X}_{k_1} & \cdots & \boldsymbol{X}_{n_1-1} \end{bmatrix},$$

where each block is $k_2 \times (n_2 - k_2 + 1)$ Hankel matrix:

$$\boldsymbol{X}_l = \begin{bmatrix} x_{l,0} & x_{l,1} & \cdots & x_{l,n_2-k_2} \\ x_{l,1} & x_{l,2} & \cdots & x_{l,n_2-k_2+1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{l,k_2-1} & x_{l,k_2} & \cdots & x_{l,n_2-1} \end{bmatrix}.$$

Multi-fold Hankel matrix (Hua '92)



$\boldsymbol{X}_{\mathrm{e}}$

Low-rank structure of enhanced matrix

- Enhanced matrix can be decomposed as

$$\boldsymbol{X}_{\mathsf{e}} = \begin{bmatrix} \boldsymbol{Z}_{\mathsf{L}} \\ \boldsymbol{Z}_{\mathsf{L}} \boldsymbol{Y}_{\mathsf{d}} \\ \vdots \\ \boldsymbol{Z}_{\mathsf{L}} \boldsymbol{Y}_{\mathsf{d}}^{k_1 - 1} \end{bmatrix} \mathrm{diag}(\boldsymbol{d}) \left[\boldsymbol{Z}_{\mathsf{R}}, \boldsymbol{Y}_{\mathsf{d}} \boldsymbol{Z}_{\mathsf{R}}, \cdots, \boldsymbol{Y}_{\mathsf{d}}^{n_1 - k_1} \boldsymbol{Z}_{\mathsf{R}} \right],$$

  - $\boldsymbol{Z}_{\mathsf{L}}$ and $\boldsymbol{Z}_{\mathsf{R}}$ are Vandermonde matrices specified by $z_1, \ldots, z_r$
  - $\boldsymbol{Y}_{\mathsf{d}} = \mathsf{diag}\left[y_1, y_2, \cdots, y_r\right]$

- Low-rank: $\mathsf{rank}\left(\boldsymbol{X}_{\mathsf{e}}\right) \leq r$

Recovery via Hankel matrix completion

Enhanced Matrix Completion (EMaC):

$$\begin{aligned} \underset{\boldsymbol{z}\in\mathbb{C}^n}{\text{minimize}} \quad & \|\boldsymbol{Z}_{\mathsf{e}}\|_* \\ \text{s.t.} \quad & z_{i,j} = x_{i,j}, \quad (i,j) \in T \end{aligned}$$

- Can be easily extended to higher-dimensional frequency models

Reference

- "*Tutorial: convex optimization techniques for super-resolution parameter estimation*," Y. Chi, G. Tang, *ICASSP*, 2016.
- "*Sampling theory: beyond bandlimited systems*," Y. Eldar, *Cambridge University Press*, 2015.
- "*Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise*," Y. Hua, T. Sarkar, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1990.
- "*Multiple emitter location and signal parameter estimation*," R. Schmidt, *IEEE transactions on antennas and propagation*, 1986.
- "*Estimating two-dimensional frequencies by matrix enhancement and matrix pencil*," Y. Hua, *IEEE Transactions on Signal Processing*, 1992.
- "*ESPRIT-estimation of signal parameters via rotational invariance techniques*," R. Roy, T. Kailath, *IEEE Transactions on acoustics, speech, and signal processing*, 1989.

Reference

- "*Superresolution via sparsity constraints*," D. Donoho, *SIAM journal on mathematical analysis*, 1992.
- "*Sparse nonnegative solution of underdetermined linear equations by linear programming*," D. Donoho, J. Tanner, *Proceedings of the National Academy of Sciences*, 2005.
- "*Sensitivity to basis mismatch in compressed sensing*," Y. Chi, L. Scharf, A. Pezeshki, R. Calderbank, *IEEE Transactions on Signal Processing*, 2011.
- "*The convex geometry of linear inverse problems*," V. Chandrasekaran, B. Recht, P. Parrilo, A. Willsky, *Foundations of Computational mathematics*, 2012.
- "*Towards a mathematical theory of super-resolution*," E. Candes, C. Fernandez-Granda, *Communications on Pure and Applied Mathematics*, 2014.

Reference

- "*Compressed sensing off the grid*," G. Tang, B. Bhaskar, P. Shah, B. Recht, *IEEE Transactions on Information Theory*, 2013.
- "*Robust spectral compressed sensing via structured matrix completion*," Y. Chen, Y. Chi, *IEEE Transactions on Information Theory*, 2014.
- "*Super-resolution, extremal functions and the condition number of Vandermonde matrices*," A. Moitra, *Symposium on Theory of Computing*, 2015.
- "*MUSIC for single-snapshot spectral estimation: Stability and super-resolution*," W. Liao, A. Fannjiang, *Applied and Computational Harmonic Analysis*, 2016.