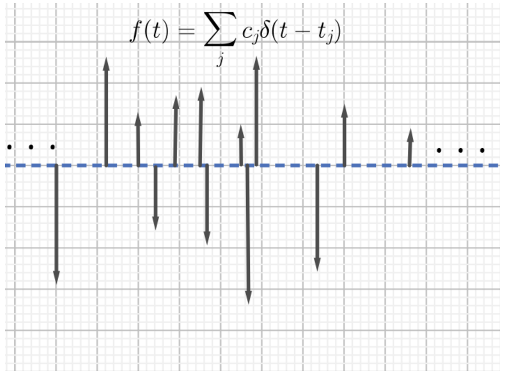
Topics in Inverse Problems and Super-Resolution

Dr. Dmitry Batenkov

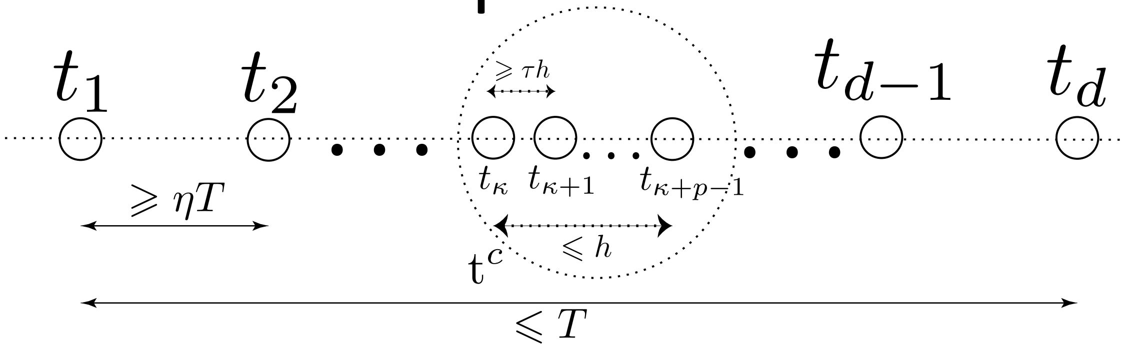
Fall 2022

Lecture 11

Parametric Super-Resolution Tractable algorithms



Recap: parametric super-resolution



- Model: $f(t) = \sum_{j=1}^{d} c_j \delta(t - t_j)$
- Continuous: $\hat{f}(\omega) = \sum_{j=1,\ldots,d} c_j \exp(2\pi i \omega t_j) + e(\omega)$, $|\omega| \leq \boldsymbol{\Omega}$, $|e(\omega)| \leq \boldsymbol{\epsilon}$
  - $\mathrm{SRF} := (\Omega h)^{-1}$
  - $|\delta t_j| \sim \frac{1}{\Omega} \mathrm{SRF}^{2p-2} \epsilon$ if $\epsilon < c \cdot \mathrm{SRF}^{2p-1}$
- Discrete: $\hat{f}(k) = \sum_{j=1}^{d} c_j \exp(2\pi i k t_j) + e(k)$, $k = 0, \ldots, N$, $|e(k)| \leq \epsilon$
  - $\mathrm{SRF} := (Nh)^{-1}$
  - ℓ_1 min: $\|\delta f\|_1 \sim \mathrm{SRF}^{2p} \epsilon$ if $c_j > 0$
  - $\|\delta f\|_2 \sim \mathrm{SRF}^{2p-1} \epsilon$

Today: 1D* SR algorithms

- Prony’s method
- Nonlinear Least Squares & Variable Projections
- Structured Low-Rank Approximation (“Cadzow desnoising”)
- (“single snapshot”) subspace methods
  - ESPRIT / (Modified) Matrix Pencil
  - Filter diagonalization
  - MUSIC

*Multivariate extensions exist for almost all of the above

Decimated Prony’s Method

$$\tilde{\tau}(k) = \sum_{j=1}^{d} c_j e^{2\pi i s_j k} + e(k)$$

$$0 < m \le |c_j| \le M < \infty$$
$$s_j := \boldsymbol{\lambda} t_j \bmod 1$$
$$|e(k)| \le \epsilon$$

- Construct $n \times n$ **Hankel** matrix

$$\widetilde{H}_d = [\tilde{\tau}(k+\ell)]_{k=0,1,\dots,d-1}^{\ell=0,1,\dots,d-1} = \begin{bmatrix} \tilde{\tau}(0) & \cdots & \tilde{\tau}(d-1) \\ \vdots & \ddots & \vdots \\ \tilde{\tau}(d-1) & \cdots & \tilde{\tau}(2d-2) \end{bmatrix}$$

- Find the unique solution $\tilde{\boldsymbol{q}}$ to $\widetilde{H}_d \tilde{\boldsymbol{q}} = \begin{bmatrix} \tilde{\tau}(d) \\ \vdots \\ \tilde{\tau}(2d-1) \end{bmatrix}$
- Construct $\widetilde{\mathrm{Q}}(x) = x^d + \sum_{j=0}^{d-1} \tilde{q}_j x^j$, let $\{\tilde{\rho}_j\}$ be its roots, $\tilde{s}_j = \frac{1}{2\pi} \angle \tilde{\rho}_j$
- Put $\tilde{\mathrm{z}}_j := e^{2\pi i \tilde{s}_j}$, then $\{\tilde{c}_j\}$ are given by

$$\begin{bmatrix} 1 & \cdots & 1 \\ \tilde{z}_1 & \cdots & \tilde{z}_n \\ \vdots & \ddots & \vdots \\ \tilde{z}_1^{d-1} & \cdots & \tilde{z}_n^{d-1} \end{bmatrix} \begin{bmatrix} \tilde{c}_1 \\ \vdots \\ \tilde{c}_n \end{bmatrix} = \begin{bmatrix} \tilde{\tau}(0) \\ \tilde{\tau}(1) \\ \vdots \\ \tilde{\tau}(d-1) \end{bmatrix}$$

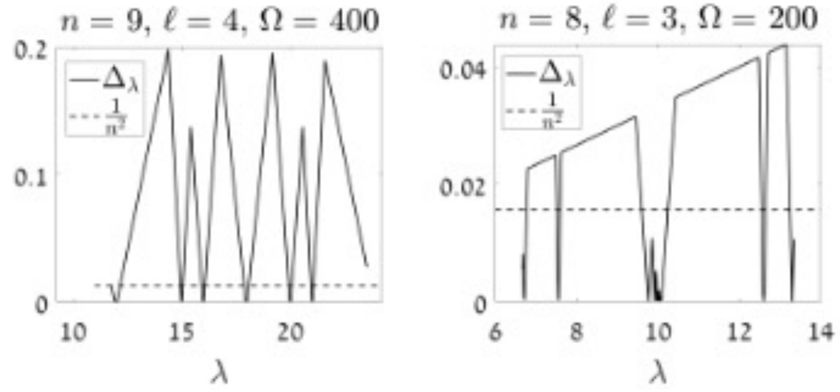
Finding λ (arXiv:2210.13329)

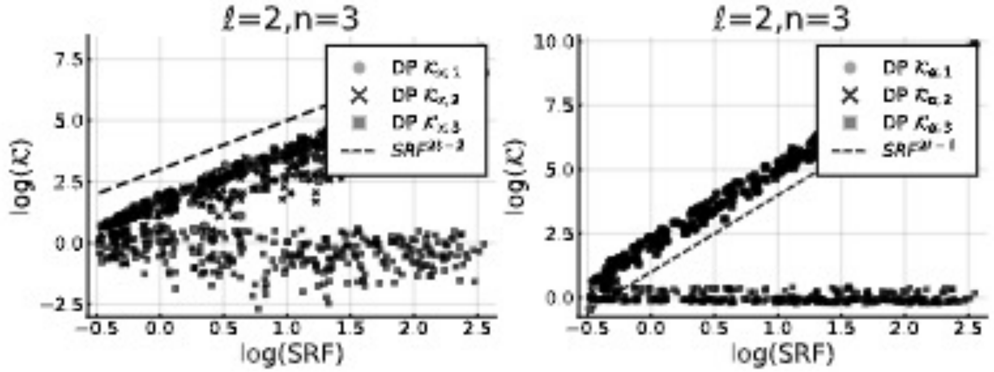
$$g(\omega) = \sum_{k=1}^{n} \alpha_k e^{2\pi \mathrm{j} x_k \omega} + e(\omega), \quad \omega \in [-\Omega, \Omega],$$

```
Algorithm 3.1: Decimated Prony Method
    Data: $N_\lambda, n, \Omega, \Delta, N_b > \frac{3}{\Delta}$
    Result: Estimates $\{\tilde{x}_k,\ \tilde{\alpha}_k\}_{k=1}^{n}$
for $\lambda \in G$ do
    $\tilde{m}^{(\lambda)} := \left\{\tilde{m}_k^{(\lambda)} = g(\lambda k)\right\}$
    $\{y_{\lambda,j}\} \longleftarrow \mathrm{Prony}\left(\tilde{m}^{(\lambda)}\right)$
    Compute $X_\lambda$ as in (2)
$X \longleftarrow \bigcup_{\lambda \in G} \left[X_\lambda \bigcap \left(G \times \left[-\frac{1}{2}, \frac{1}{2}\right]\right)\right]$
Compute $H$ - Histogram of $\{x : (\lambda, x) \in X\}$
    with $N_b$ bins. Set $\{B_k\}_{k=1}^{n} = \mathtt{ArgMax}(H, n)$
Compute $\Lambda$ as in (3) and $\lambda^* \leftarrow \max\{\lambda : \lambda \in \Lambda\}$
$\{\tilde{x}_k\}_{k=1}^{n} \longleftarrow \{x : (x, \lambda^*) \in X_{\lambda^*} \wedge x \in B_k\}$
Solve the Vandermonde system
    $$\left(e^{i\lambda^* k \tilde{x}_j}\right)_{k=0,\ldots,n-1}^{j=1,\ldots,n} \cdot \mathrm{col}\,\{\tilde{\alpha}_k\}_{k=1}^{n} = \mathrm{col}\left\{\tilde{m}_k^{(\lambda^*)}\right\}_{k=0}^{n-1}$$
    return the estimates $\{\tilde{x}_k,\ \tilde{\alpha}_k\}_{k=1}^{n}$
```

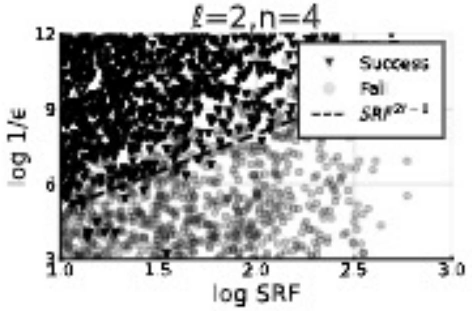
$$X_\lambda := \bigcup_{j=1,\ldots,n(\lambda)} \left\{ \left(\lambda, \frac{\tilde{y}_{\lambda,j} + k}{\lambda}\right) : k \in \mathbb{Z} \right\}.$$

$$\Lambda := \bigcap_{k=1}^{T} \{\lambda : (x, \lambda) \in X \wedge x \in B_k\}. \qquad (3)$$

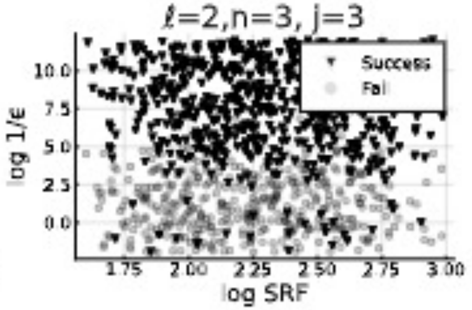




(a) Accuracy



(b) Cluster node



(c) Non-cluster node

Reminder: (linear) condition number

- Linear system: $Ax = b$, exact solution x_0
- Perturbed system: $(A + \Delta A)x = b + \Delta b$, $x_0 + \Delta x$ – exact solution
- Vector norm $\|v\|$, the induced matrix norm $\|A\| = \sup\limits_{\|v\|=1} \|Av\|$
- Set $\delta x := \frac{\|\Delta x\|}{\|x_0\|}, \delta A := \frac{\|\Delta A\|}{\|A\|}, \delta b := \frac{\|\Delta b\|}{\|b\|}$
- Condition number: $\kappa := \| A \| \, \|A^{-1}\|$
- <u>Lemma:</u>

$$\delta x \leq \frac{\kappa}{1 - \kappa \cdot \delta A} (\delta A + \delta b)$$

Prony’s method: Stability analysis

- Let $\tau, \boldsymbol{q}, Q, H_d$ denote the noiseless counterparts of $\tilde{\tau}, \tilde{\boldsymbol{q}}, \tilde{Q}$, $\tilde{H}_d$, $\rho_j = e^{2\pi i s_j}$
- Vandermonde factorization: $H_d = V_d(\boldsymbol{\rho}) \times C \times V_d^T(\boldsymbol{\rho})$ - check!
  - $\kappa(V_d) \lesssim \max_i \prod_{j \neq i} \left|\rho_i - \rho_j\right|^{-1} \lesssim \mathrm{SRF}^{p-1}$
  - $C = \mathrm{diag}\{c_1, \dots, c_d\}$

$s_j := \boldsymbol{\lambda} t_j \bmod 1$

$$V_d(\boldsymbol{\rho}) = \begin{bmatrix} 1 & \dots & 1 \\ \rho_1 & \ddots & \rho_d \\ \rho_1^2 & \ddots & \rho_d^2 \\ \vdots & \ddots & \vdots \\ \rho_1^{d-1} & \dots & \rho_d^{d-1} \end{bmatrix}$$

- $\kappa(H_d) \lesssim \frac{\kappa^2(V_d)M}{m}$, $\delta H_d \lesssim \epsilon$, $\delta\tau \lesssim \epsilon$
- $\left|\tilde{q}_j - q_j\right| \lesssim \mathrm{SRF}^{2p-2}\epsilon := \eta$ whenever $\epsilon \lesssim \mathrm{SRF}^{2-2p}$
- Stability of polynomial roots [1, p.38]: $\left|\tilde{\rho}_j - \rho_j\right| \lesssim \frac{\tilde{Q}(\rho_j)}{Q'(\rho_j)}\eta + O(\eta^2)$
  - Clearly $Q'(\rho_j) = \prod_{i \neq j}(\rho_j - \rho_i) \gtrsim \mathrm{SRF}^{1-p} \rightarrow |\tilde{s}_j - s_j| \lesssim \mathrm{SRF}^{3p-3}\epsilon$
  - Finally: $|\tilde{t}_j - t_j| \lesssim \frac{1}{\Omega}\mathrm{SRF}^{3p-3}\epsilon \leftarrow$ not optimal!!!
- Extra powers of SRF added for $|\tilde{c}_j - c_j|$

[1] J. Wilkinson, Rounding Errors in Algebraic Processes, Dover, New York, 1994.

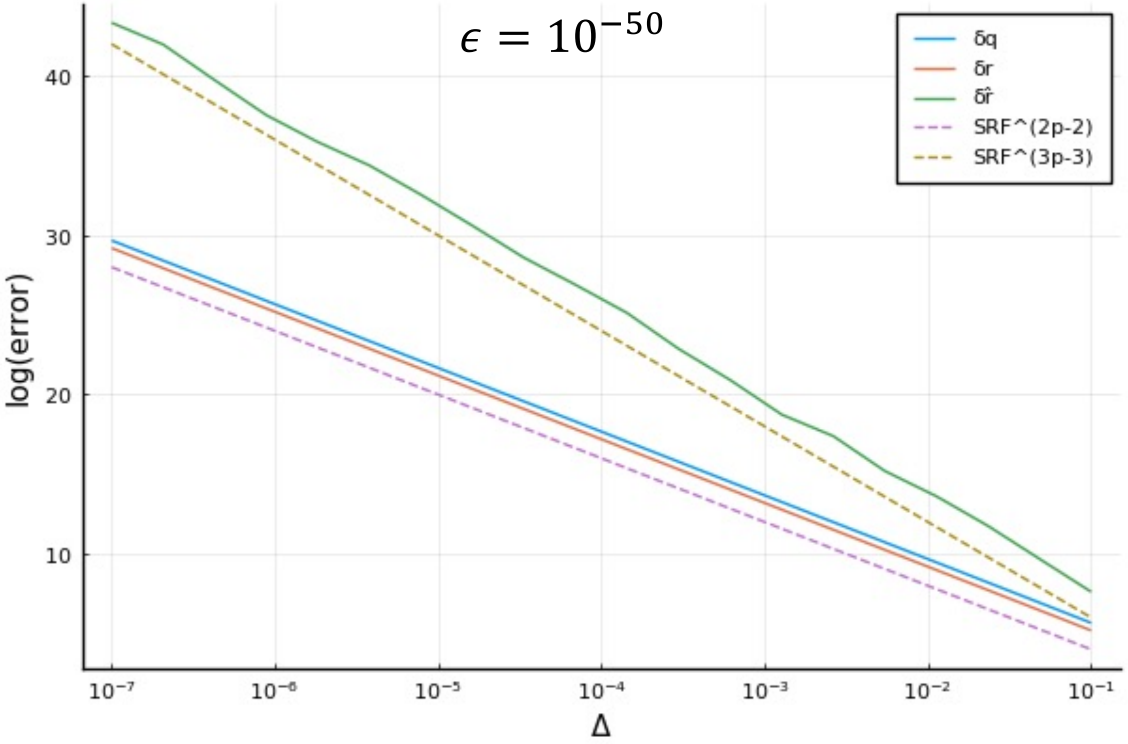
But…

- The analysis is not tight, since we know that Prony’s method computes F_λ^{-1} exactly
- Therefore, Prony + optimal decimation gives best accuracy, but we don’t know λ^*! (except when $p = d$)
- Q: Where is the problem in the analysis?

$\delta\boldsymbol{q} = \|\boldsymbol{q} - \tilde{\boldsymbol{q}}\|/\epsilon$ and $\delta\boldsymbol{r} = \|\boldsymbol{\rho} - \tilde{\boldsymbol{\rho}}\|/\epsilon$ have same scaling
➔ errors in the coefficients are not independent!

For comparison, we plot $\delta\hat{\boldsymbol{r}} = \|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}\|/\epsilon$ where $\{\hat{\rho}_j\}$ are the roots of a polynomial $\hat{Q}$ with <u>random</u> errors in the coefficients of same magnitude $\delta\boldsymbol{q}$

~~**Open question: explain this phenomenon!**~~



Prony’s method: accurate analysis

- Step #1: use Rouche’s theorem from complex analysis
  - If $|p(z)| > |q(z) - p(z)|, z \in \partial B_\rho(z_0)$ and $p(z_0) = 0$ (simple zero) then $q(z)$ has a simple zero inside $B_\rho(z_0)$ (proof: argument principle)
  - $p(z) \sim \prod(z - z_k)$
  - $q - p$ has some common factors in the coefficients
- Step #2:
  - compute $\widetilde{z_k} - z_k = A_k\epsilon + O(\epsilon^2)$
  - $\tilde{V}\boldsymbol{\delta c} = \tilde{V}^{-1}\boldsymbol{\delta m} - (I - \tilde{V}^{-1}V)\ \boldsymbol{c}$

R.Katz,N.Diab, DB, in prep.

Nonlinear Least Squares (NLS)

$$\boldsymbol{x} \coloneqq (\boldsymbol{c},\boldsymbol{t}) \in \mathbb{R}^{2d}, F_k(\boldsymbol{x}) = \sum_{j=1}^{d} c_j \exp\left(2\pi i k t_j\right), k = 0,\dots,N \text{ –} \underline{\textit{parametric model}}$$

$$g_k = F_k(\boldsymbol{x_0}) + e_k, |e_k| \le \epsilon \text{ – } \underline{\textit{noisy data}}$$

$$\boldsymbol{x}^{\boldsymbol{NLS}} = \operatorname*{argmin}_{\boldsymbol{x}} \sum_{k=0}^{N} r_k^*(\boldsymbol{x}) r_k(x), \textit{ where } r_k(\boldsymbol{x}) = g_k - F_k(\boldsymbol{x}) \textit{ are the } \underline{\textit{residuals}}$$

$$J_{\boldsymbol{F}}(\boldsymbol{x}) = \frac{\partial[\Re F_0\ \Im F_0,\dots,\Re F_N\ \Im F_N]^T}{\partial[c_1 t_1 \dots c_d t_d]}(\boldsymbol{x}) \in \mathbb{R}^{(2N+2)\times 2d}$$

- The NLS problem can be solved by iterative algorithms such as Gauss-Newton, Levenberg-Marquardt,...[1]
- NLS is equivalent to Maximum Likelihood Estimator with Gaussian noise in estimation theory
- Convergence to a local minimum can be guaranteed under sufficiently reasonable assumptions
- Problem 1: initialization needs to be sufficiently close to the global optimum
  - *Basin of attraction estimates are partially available [2], essentially depend on* $\sigma_{\min}(J_f)$
- Problem 2: Accuracy of the solution - can we estimate $\left\|\boldsymbol{x}^{\boldsymbol{NLS}} - \boldsymbol{x_0}\right\|$?

[1] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd edition. 2006.

[2] Y. Traonmilin and J.-F. Aujol, "The basins of attraction of the global minimizers of the non-convex sparse spike estimation problem," *Inverse Problems*, vol. 36, no. 4, p. 045003, 2020

Stability of nonlinear optimization problems [1]

- $f(\boldsymbol{x}, \epsilon)$ smooth, $\boldsymbol{x}^*(\epsilon) = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}, \epsilon)$, $\boldsymbol{x}^*(0) = \boldsymbol{x}_{\mathbf{0}}$
- $\boldsymbol{x}^*(\epsilon)$ local minimum when $\nabla_{\boldsymbol{x}} f(\boldsymbol{x}^*(\epsilon), \epsilon) \equiv \mathbf{0}$ in a neighb. of $\epsilon = 0$
- *Implicit function theorem*: $\nabla_{\boldsymbol{x}} f(\boldsymbol{x}, \epsilon) \coloneqq g(\boldsymbol{x}, \epsilon) = \mathbf{0}$. Sufficient conditions for existence of a solution which is also a minimum:
  - $g(\boldsymbol{x}_{\mathbf{0}}, 0) = 0$
  - $\nabla_{\boldsymbol{x}} g(\boldsymbol{x}, \epsilon)|_{\epsilon=0} = \nabla^2_{\boldsymbol{xx}} f(\boldsymbol{x}, \epsilon)|_{\epsilon=0} \succ 0$ (positive definite Hessian)
- Full differential: $0 \equiv \partial g(x^*, \epsilon) = \nabla_x g(x^*, \epsilon) \nabla_\epsilon x^*(\epsilon) + \nabla_\epsilon g(x^*, \epsilon)$
- $\nabla_\epsilon x^*(\epsilon) = -(\nabla^2_{xx} f(\boldsymbol{x}^*, \epsilon))^{-1} \nabla^2_{x\epsilon} f(x^*, \epsilon)$
- This gives a first order estimate $\|\boldsymbol{x}^*(\epsilon) - \boldsymbol{x}_0\| \approx \nabla_\epsilon \boldsymbol{x}^*(0) \cdot \epsilon$

[1] A. V. Fiacco, *Introduction to sensitivity and stability analysis in nonlinear programming*. Academic Press, 1983.

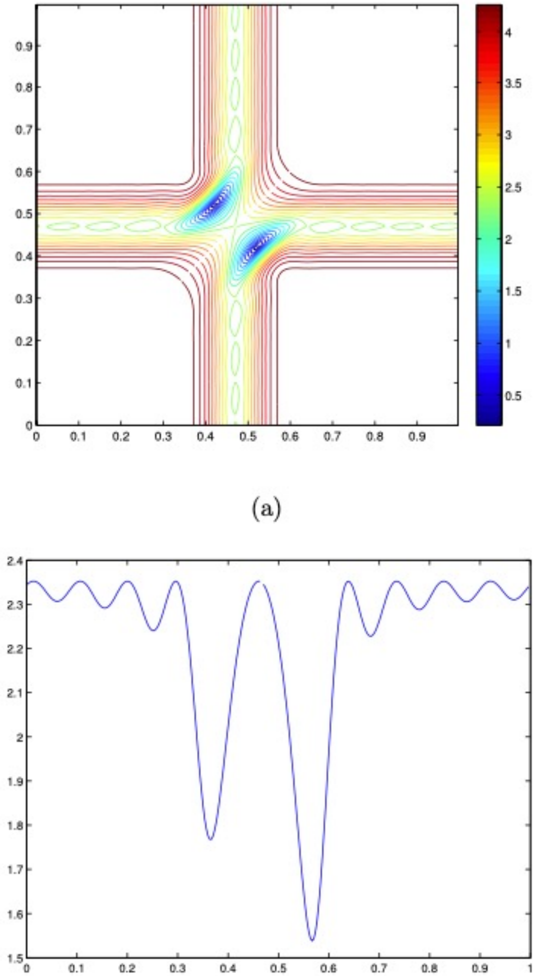
Back to least squares problems

- Residuals: $\boldsymbol{r}(\boldsymbol{x}, \epsilon) = \boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{g}(\epsilon)$, $\boldsymbol{F} = [f_0 \ldots f_N]$, $\boldsymbol{g} = \boldsymbol{F}(\boldsymbol{x_0}) + \boldsymbol{e}(\epsilon)$
- $f(\boldsymbol{x}, \epsilon) = \frac{1}{2}\|\boldsymbol{r}(\boldsymbol{x}, \epsilon)\|^2$
- Compute: $\nabla_{\boldsymbol{x}} f(\boldsymbol{x}, \epsilon) = J_{\boldsymbol{F}}^T(\boldsymbol{x})\boldsymbol{r}(\boldsymbol{x}, \epsilon)$, $\nabla_{xx}^2 f(\boldsymbol{x}, 0) = J_F^T(\boldsymbol{x}) J_F(\boldsymbol{x})$, $\nabla_{\boldsymbol{x}\epsilon}^2 f(\boldsymbol{x}, \epsilon) = -J_F^T(\boldsymbol{x})\nabla_\epsilon \boldsymbol{e}(\epsilon)$
- Substitute: $\nabla_\epsilon \boldsymbol{x}^*(0) = -(\nabla_{\boldsymbol{xx}}^2 f(\boldsymbol{x}^*, 0))^{-1}\nabla_{\boldsymbol{x}\epsilon}^2 f(\boldsymbol{x}^*, 0) = J_{\boldsymbol{F}}^\dagger(\boldsymbol{x_0})\nabla_\epsilon \boldsymbol{e}(0)$
    - $J_{\boldsymbol{F}}^\dagger := \left(J_{\boldsymbol{F}}^T J_{\boldsymbol{F}}\right)^{-1} J_{\boldsymbol{F}}^T$ - pseudo-inverse
- Stability at the global optimum is controlled by row norms of $J_{\boldsymbol{F}}^\dagger(\boldsymbol{x_0})$

NLS stability – open questions

$$\nabla_{\epsilon}\boldsymbol{x}^{*}(0) = J_{\boldsymbol{F}}^{\dagger}(\boldsymbol{x_0})\nabla_{\epsilon}\boldsymbol{e}(0)$$

- Estimate $\sigma_{\min}$ and row norms of $J_{\boldsymbol{F}}^{\dagger}(\boldsymbol{x_0})$ in the multi-cluster geometry
- Obtain quantitative estimates on the neighborhood of ϵ
- Structured perturbations: what if $\nabla_{\epsilon}\boldsymbol{e}(0) = J_{\boldsymbol{F}}(\boldsymbol{x_0}) + LOT$?



(a)

(b)

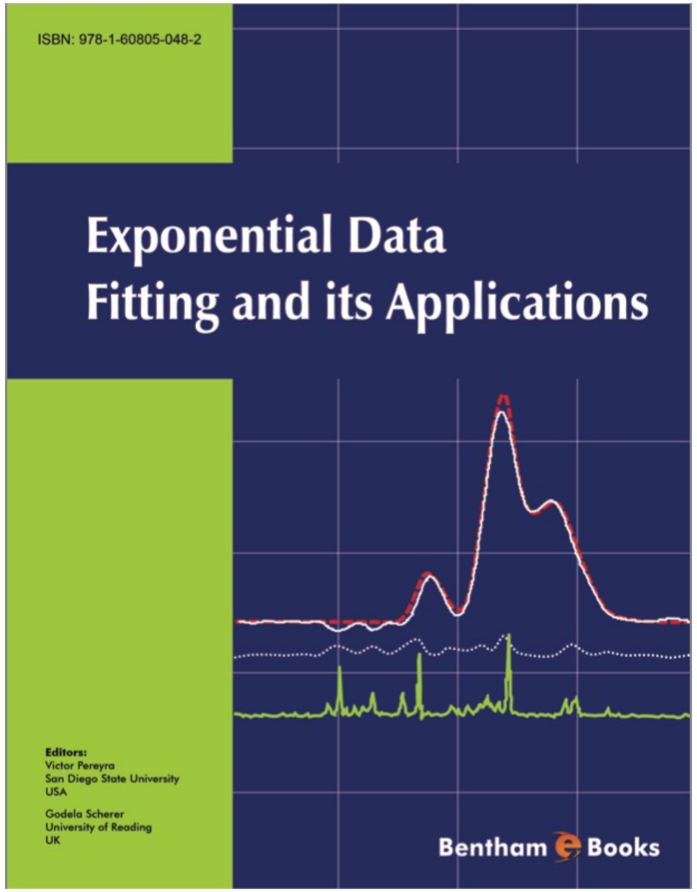
Multi-modality of the NLS cost [1]

[1] L. Condat and A. Hirabayashi, “Cadzow Denoising Upgraded: A New Projection Method for the Recovery of Dirac Pulses from Noisy Linear Measurements,” Mar. 2014.

VARiable PROjection (VARPRO)

$$\mathrm{x}^{\mathrm{NLS}} = \mathrm{argmin}_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{g}\|^2$$

- Note that in fact $\boldsymbol{F}(\boldsymbol{x} = (\boldsymbol{c}, \boldsymbol{t})) = V\left(e^{2\pi i \boldsymbol{t}}\right)\boldsymbol{c}$, where V is the Vandermonde matrix
- So $(\boldsymbol{c}^{NLS}, \boldsymbol{t}^{NLS}) = \arg\min_{\boldsymbol{t}} \min_{\boldsymbol{c}} \frac{1}{2}\|A(\boldsymbol{t})\boldsymbol{c} - \boldsymbol{g}\|^2$
- The inner min. is a linear LS! Its solution is $\boldsymbol{c}_{\boldsymbol{inner}}(\boldsymbol{t}) = A^{\dagger}(\boldsymbol{t})\boldsymbol{g}$
- We end up with $\boldsymbol{t}^{NLS} = \operatorname*{argmin}_{\boldsymbol{t}} \frac{1}{2}\left\|(A(\boldsymbol{t})A^{\dagger}(\boldsymbol{t}) - I)\boldsymbol{g}\right\|^2 = \operatorname*{argmin}_{\boldsymbol{t}} \mathrm{VP}(\boldsymbol{t}; \boldsymbol{g})$
    - $\boldsymbol{c}^{NLS} = A^{\dagger}(\boldsymbol{t}^{NLS})\, \boldsymbol{g}$
- This is a different nonlinear problem with the same global optimum
    - Better optimization landscape (smaller # of local minima)
    - Faster convergence
    - Efficient implementation utilizes explicit computation of J_{VP}
- Works for any separable $\boldsymbol{F}(\boldsymbol{x}) = A(\boldsymbol{t})\boldsymbol{c}$
- Stability properties in the SR regime are unknown.



ISBN: 978-1-60805-048-2

Exponential Data
Fitting and its Applications

Editors:
Victor Pereyra
San Diego State University
USA

Godela Scherer
University of Reading
UK

Bentham e Books

[1] L. Kaufman and V. Pereyra, "A Method for Separable Nonlinear Least Squares Problems with Separable Nonlinear Equality Constraints," *SIAM J. Numer. Anal.*, vol. 15, no. 1, pp. 12–20, Feb. 1978

[2] G. Golub and V. Pereyra, "Separable nonlinear least squares: the variable projection method and its applications," *Inverse problems*, vol. 19, no. 2, p. R1, 2003.

[3] V. Pereyra and G. Scherer, *Exponential Data Fitting and Its Applications*. Bentham Science Publishers, 2010.

Structured Low-Rank Approximation (“Cadzow denoising”)

$$\tau(k)=\sum_{j=1}^{d} c_j\rho_j^k \qquad H_{L,M}=[\tau(k+\ell)]_{k=0,1,\ldots,L}^{\ell=0,1,\ldots,M}=\begin{bmatrix}\tau(0) & \cdots & \tau(M)\\ \vdots & \ddots & \vdots\\ \tau(L) & \cdots & \tau(L+M)\end{bmatrix}$$

$L \geq d, M \geq d$

$\operatorname{rank} H_{L,M}=d$

$$(x-\rho_1)\cdots(x-\rho_n)=x^d+q_{d-1}x^{d-1}+\cdots+q_1x+q_0$$

Annihilation property: $\sum_{\ell=0}^{d} q_\ell\tau(k+\ell)=0, \qquad \forall k\in\mathbb{N}$

<u>Annihilation method</u>

1. Approximate $\widetilde{H}_{L,M}$ by a rank-d matrix A
2. Find $\boldsymbol{q}\in\ker(A)$, construct $q(z)=\sum_{j=0}^{M} q_j z^j$
3. Find d roots of $q(z)$ closest to the unit circle

- Problem: A is not Hankel anymore, so no “annihilation” as such
- We can “project” A to a set of Hankel matrices (by averaging anti-diagonals) but then it will not be of rank d
- Cadzow [1]: alternatively project onto rank-d (non-convex op.) and Hankel (linear op.) to “denoise” the matrix, after which annihilation will work
- This works in practice for low noise levels, but no proof of convergence yet
- [2] suggests an alternative impl. via proximal splitting (still no global convergence guarantees)
- Possible improvement: add the non-degeneracy condition of the <u>upper minor</u> – which we know to be necessary and sufficient for solvability

[1] J. A. Cadzow, “Total Least Squares, Matrix Enhancement, and Signal Processing,” *Digital Signal Processing*, vol. 4, no. 1, pp. 21–39, Jan. 1994, doi: 10.1006/dspr.1994.1003.

[2] L. Condat and A. Hirabayashi, “Cadzow Denoising Upgraded: A New Projection Method for the Recovery of Dirac Pulses from Noisy Linear Measurements,” Mar. 2014.

Subspace methods

$$\tau(k) = \sum_{j=1}^{d} c_j \rho_j^k$$

$$\rho_j = e^{2\pi i t_j}$$

$$H_{L,M} = [\tau(k+\ell)]_{k=0,1,\dots,L}^{\ell=0,1,\dots,M} = \begin{bmatrix} \tau(0) & \cdots & \tau(M) \\ \vdots & \ddots & \vdots \\ \tau(L) & \cdots & \tau(L+M) \end{bmatrix}$$

$$L \geq d, M \geq d$$

For simplicity assume $L \approx M$

$$V_L(\boldsymbol{\rho}) = \begin{bmatrix} 1 & \dots & 1 \\ \rho_1 & \ddots & \rho_d \\ \rho_1^2 & \ddots & \rho_d^2 \\ \vdots & \ddots & \vdots \\ \rho_1^L & \dots & \rho_d^L \end{bmatrix}$$

$$C = \mathrm{diag}\{c_1, \dots, c_d\}$$

Vandermonde decomposition: $H_{L,M} = V_L(\boldsymbol{\rho}) \times C \times V_M^T(\boldsymbol{\rho}) \rightarrow \mathrm{rank}\, H_{L,M} = d$

Singular Value Decomposition: $H_{L,M} = \begin{bmatrix} U_{H,d} & U_{H,d}^{\perp} \end{bmatrix} \times \begin{bmatrix} \Sigma_d & 0 \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} V_{H,d}^{*} \\ \left(V_{H,d}^{\perp}\right)^{*} \end{bmatrix}$

Signal subspace property: $\mathcal{R}\left(H_{L,M}\right) = \mathcal{R}\left(U_{H,d}\right) = \mathcal{R}(V_L(\boldsymbol{\rho}))$

Noise subspace property: $\left\| (U_{H,d}^{\perp})^{*} \times V_L(\boldsymbol{\rho}) \right\| = 0$

In the presence of noise, these become $\approx$, and taking $L \gg 2d$ should reduce the error

[1] P. Stoica and R. L. Moses, *Spectral analysis of signals (Chapter 4)*. Pearson/Prentice Hall, 2005.

EStimation of Parameters via Rotational Invariance Techniques (ESPRIT)

$$H_{L,M} = V_L(\boldsymbol{\rho})\times C\times V_M^T(\boldsymbol{\rho}) = \begin{bmatrix} U_{H,d} & U_{H,d}^{\perp}\end{bmatrix} \times \begin{bmatrix}\Sigma_d & 0\\ 0 & 0\end{bmatrix}\times\begin{bmatrix} V_{H,d}^{*} \\ \left(V_{H,d}^{\perp}\right)^{*}\end{bmatrix}$$

- Q: how to get $\{\rho_1, \dots \rho_d\}$ from $\mathcal{R}\left(U_{H,d}\right) = \mathcal{R}(V_L(\boldsymbol{\rho}))$?
- $V_L^{\uparrow} = \begin{bmatrix} 1 & \cdots & 1\\ \rho_1 & \ddots & \rho_d \\ \vdots & \ddots & \vdots \\ \rho_1^{L-1} & \cdots & \rho_d^{L-1}\end{bmatrix}$ (last row removed), $V_L^{\downarrow} = \begin{bmatrix} \rho_1 & \cdots & \rho_d\\ \rho_1^2 & \ddots & \rho_d^2 \\ \vdots & \ddots & \vdots \\ \rho_1^{L-1} & \cdots & \rho_d^{L-1}\end{bmatrix}$ (1st row removed), $\boldsymbol{\Gamma} \coloneqq \operatorname{diag}\{\rho_1, \dots, \rho_d\}$
- Rotational Invariance Property: $V_L^{\downarrow} = V_L^{\uparrow}\times\boldsymbol{\Gamma} \rightarrow \boldsymbol{\Gamma} = (V_L^{\uparrow})^{\dagger}V_L^{\downarrow}$
- $V_L = U_{H,d}R$ for an invertible $R \in \mathbb{C}^{d\times d}$ → $V_L^{\uparrow} = U_{H,d}^{\uparrow}R$, $V_L^{\downarrow} = U_{H,d}^{\downarrow}R$ → $\boldsymbol{\Gamma} = R^{-1}(U_{H,d}^{\uparrow})^{\dagger}U_{H,d}^{\downarrow}R$
- → $\{\rho_1, \dots, \rho_d\}$ are eigenvalues of $(U_{H,d}^{\uparrow})^{\dagger}U_{H,d}^{\downarrow}$

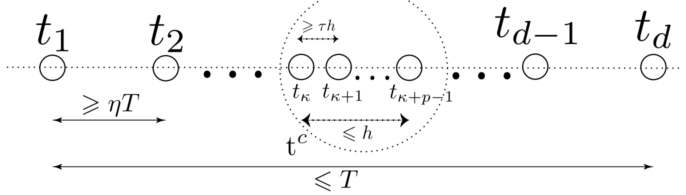
ESPRIT algorithm

1. Compute truncated SVD of $\widetilde{H}_{L,M}$, take $\widetilde{U}_{H,d}$
2. Compute d eigenvalues of $(\widetilde{U}_{H,d}^{\uparrow})^{\dagger}\widetilde{U}_{H,d}^{\downarrow}$

Matrix Pencil algorithm

1. $\widetilde{H}_{L,M}^{\uparrow} = U_1\Sigma_1V_1^H$, $\widetilde{H}_{L,M}^{\downarrow} = U_2\Sigma_2V_2^H$ (rank-d approximations)
2. Compute d eigenvalues of $\Sigma_2^{-1}U_2^HU_1\Sigma_1V_1^HV_2$

Signal subspace methods – analysis



$$H_{L,M} = V_L(\boldsymbol{\rho}) \times C \times V_M^T(\boldsymbol{\rho}) = \begin{bmatrix} U_{H,d} & U_{H,d}^{\perp} \end{bmatrix} \times \begin{bmatrix} \Sigma_d & 0 \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} V_{H,d}^* \\ \left(V_{H,d}^{\perp}\right)^* \end{bmatrix}$$

$$|\tau(k) - \tilde{\tau}(k)| \leq \epsilon$$

ESPRIT algorithm

1. Compute truncated SVD of $\widetilde{H}_{L,M}$, take $\widetilde{U}_{H,d}$
2. Compute d eigenvalues of $(\widetilde{U}_{H,d}^{\uparrow})^{\dagger}\widetilde{U}_{H,d}^{\downarrow}$

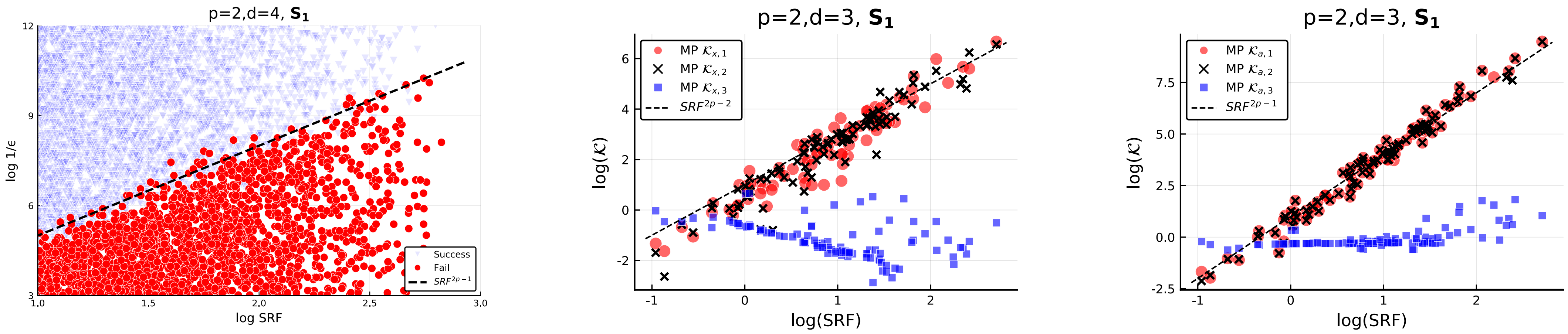
Matrix Pencil algorithm

1. $\widetilde{H}_{L,M}^{\uparrow} = U_1\Sigma_1 V_1^H$, $\widetilde{H}_{L,M}^{\downarrow} = U_2\Sigma_2 V_2^H$ (rank-d approximations)
2. Compute d eigenvalues of $\Sigma_2^{-1}U_2^H U_1 \Sigma_1 V_1^H V_2$

- Perturbation of $\widetilde{U}_{H,d}$ → related to $\sigma_{\min}(V_L)$
- Perturbation of eigenvalues: Gerschgorin's circle theorem [2, Ch.IV, Thm. 2.3]
  - $R^{-1}(\widetilde{U}_{H,d}^{\uparrow})^{\dagger}\widetilde{U}_{H,d}^{\downarrow}R = \Gamma + R^{-1}ER$ for some error matrix E
- Thm [1]: if $\epsilon \lesssim \frac{1}{\mathrm{M}}\mathrm{SRF}^{-4p-3}$ then $|\tilde{t}_j - t_j| \lesssim \mathrm{SRF}^{2p-2}\epsilon$ (for ESPRIT)
  - Min-max: $|\delta t_j| \sim \frac{1}{\mathrm{M}}\mathrm{SRF}^{2p-2}\epsilon$ if $\epsilon \lesssim \mathrm{SRF}^{2p-1}$ ← ESPRIT almost optimal
  - Experiments suggest $\epsilon \lesssim \mathrm{SRF}^{2p-1}$ is sufficient!
- Q: what is the true performance?

[1] W. Li, W. Liao, and A. Fannjiang, "Super-resolution limit of the ESPRIT algorithm," *IEEE Transactions on Information Theory*, 2020, 10/ggrnpw.
[2] Stewart, Gilbert W. "Matrix perturbation theory." (1990).

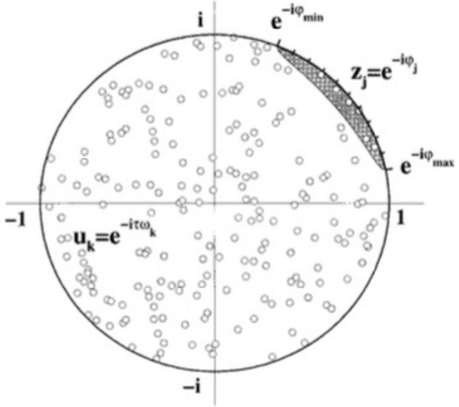
Are these algorithms optimal or not?



D. Batenkov, G. Goldman, and Y. Yomdin, “Super-resolution of near-colliding point sources,” *Information and Inference: A Journal of the IMA*, p. iaaa005, May 2020, doi: 10.1093/imaiai/iaaa005.

I suspect same problem as in the analysis of Prony’s method, i.e. using too general bounds, disregarding additional structure in the error terms

Filter Diagonalization Method (FDM)



- NMR/spectroscopy: want to recover t_j's only in a certain band
- Select large N and set $P_R := [e^{-2\pi i \frac{k}{N} j}]$ with $T_1 \le \frac{k}{N} \le T_2, j = 0, \dots, R$
- Key idea: if $t_j \notin [T_1, T_2]$ then $\left[1\ \rho_j\ \rho_j^2 \dots \rho_j^R\right] \times P_R \ll 1$

$$H_{L,M} = V_L(\boldsymbol{\rho}) \times C \times V_M^T(\boldsymbol{\rho})$$

Frequency-Selective Matrix Pencil algorithm (FDM)

1. *Input:* $\bar{\mathbf{d}} \le d, \widetilde{H}_{L,M}, T_1, T_2, N$
2. $\tilde{G}^{\uparrow}_{L,M} = P^T_{L-1}\widetilde{H}^{\uparrow}_{L,M}P_M, \tilde{G}^{\downarrow}_{L,M} = P^T_{L-1}\widetilde{H}^{\downarrow}_{L,M}P_M$
3. $\tilde{G}^{\uparrow}_{L,M} = U_1\Sigma_1V_1^H, \tilde{G}^{\downarrow}_{L,M} = U_2\Sigma_2V_2^H$ $(\mathrm{rank} - \bar{\mathbf{d}}$- approximations)
4. Compute $\bar{\mathbf{d}}$ eigenvalues of $\Sigma_2^{-1}U_2^H U_1\Sigma_1 V_1^H V_2$

No rigorous analysis of accuracy / optimality yet

[1] V. A. Mandelshtam and H. S. Taylor, "Harmonic inversion of time signals and its applications," *The Journal of Chemical Physics*, vol. 107, no. 17, pp. 6756–6769, Nov. 1997, doi: 10.1063/1.475324

[2] V. A. Mandelshtam, "FDM: the filter diagonalization method for data processing in NMR experiments," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 2, no. 38, pp. 159–196, 2001.

MUltipe SIgnal Classicication

$$H_{L,M} = V_L(\boldsymbol{\rho})\times C\times V_M^T(\boldsymbol{\rho}) = \begin{bmatrix} U_{H,d} & U_{H,d}^{\perp} \end{bmatrix}\times\begin{bmatrix}\Sigma_d & 0\\ 0 & 0\end{bmatrix}\times\begin{bmatrix} V_{H,d}^{*}\\ \left(V_{H,d}^{\perp}\right)^{*}\end{bmatrix}$$

<u>Vandermonde vector</u>: $\boldsymbol{v}(\rho) := [1\ \rho\ \rho^2\ \ldots\ \rho^L]$

<u>Noise subspace property:</u> $\left\| (U_{H,d}^{\perp})^{*}\times V_L(\boldsymbol{\rho})\right\| = 0$

<u>Claim</u>: $(U_{H,d}^{\perp})^{*}\times\boldsymbol{v}\left(e^{2\pi i t}\right) = \mathbf{0}$ if and only if $t \in \{t_1, \ldots, t_d\}$

Spectral MUSIC algorithm

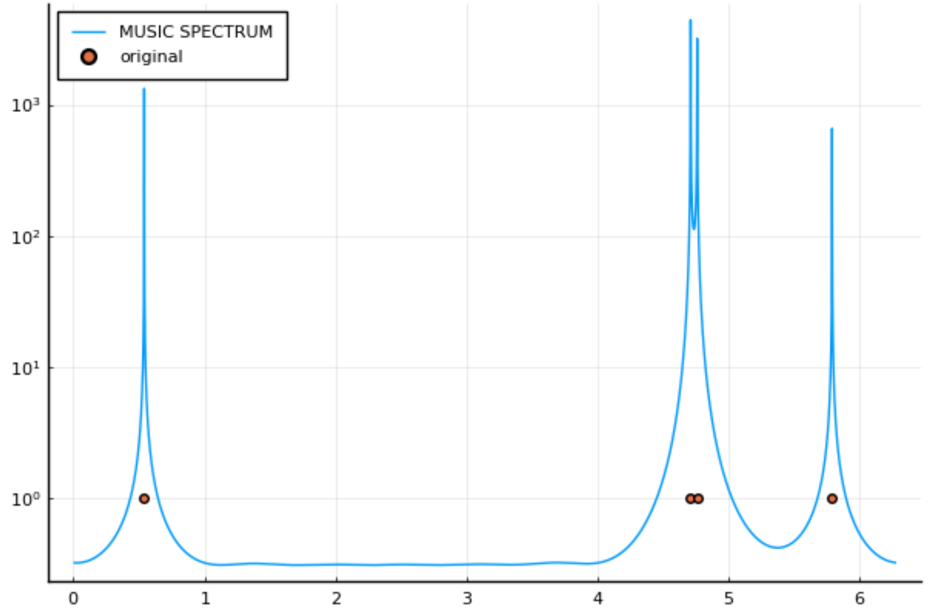
1. Compute truncated SVD of $\widetilde{H}_{L,M}$, take $\widetilde{U}_{H,d}^{\perp}$
2. Plot the "Imaging functional" $\widetilde{\mathcal{P}}(t) := \left\|(\widetilde{U}_{H,d}^{\perp})^{*}\times\boldsymbol{v}\left(e^{2\pi i t}\right)\right\|^{-1}$
3. Find d largest peaks of $\widetilde{\mathcal{P}}(t)$ (high computational complexity)

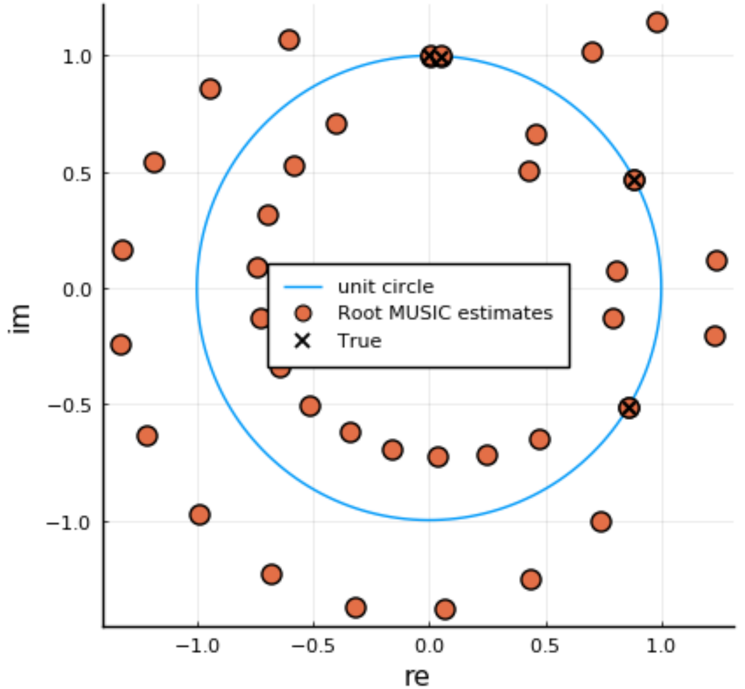
Root-MUSIC algorithm

1. Compute truncated SVD of $\widetilde{H}_{L,M}$, take $\widetilde{U}_{H,d}^{\perp}$
2. $\widetilde{G} := (\widetilde{U}_{H,d}^{\perp})^{*}\times\widetilde{U}_{H,d}^{\perp}$
3. Construct the Laurent polynomial $g(z) = v^T(z^{-1})\times\widetilde{G}\times\boldsymbol{v}(z)$

$$= \left\|(\widetilde{U}_{H,d}^{\perp})^{*}\times\boldsymbol{v}\left(e^{2\pi i t}\right)\right\|^{2}$$

4. Find d roots of $\mathrm{z}^{\mathrm{L}}g(z)$ closest to the unit circle <u>from inside</u>





MUSIC - Remarks

- Thm [1]: $\left\|\tilde{\mathcal{P}}(t)^{-1} - \mathcal{P}(t)^{-1}\right\|_{\infty} \lesssim \mathrm{SRF}^{2p-2}\epsilon$
  - But: no analysis of the peaks’ sharpness, so complete accuracy analysis still lacking
- The spectral algorithm provides a general approach to finding a range space of linear operators [2]
- Extensively used in imaging problems (maybe next lecture…)

[1] W. Li and W. Liao, “Stable super-resolution limit and smallest singular value of restricted Fourier matrices,” *Appl. Comput. Harm. Anal.*, vol. 51, 118–156,. 2021.
[2] M. Cheney, “The linear sampling method and the MUSIC algorithm,” *Inverse problems*, vol. 17, no. 4, p. 591, 2001.

There is a lot to do!