Topics in Inverse Problems and Super-Resolution

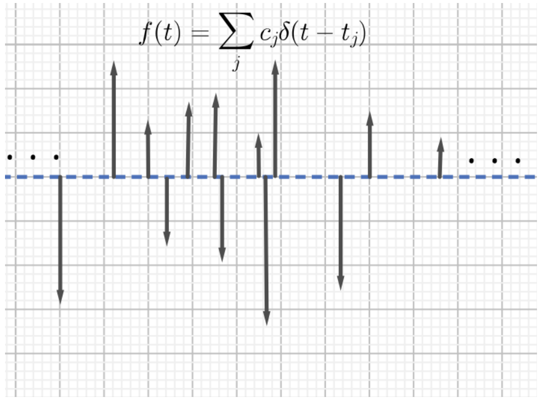
Dr. Dmitry Batenkov

Fall 2022

Lecture 10

Parametric Super-Resolution Min-Max bounds

Recap: parametric super-resolution



- Model: $m(k) = \sum_{j=1}^{d} c_j \exp(2\pi i k t_j) + e(k), k = 0, \dots, N$
  - $t_j \in [0,1)$ to avoid aliasing
- Square case: $N = 2d - 1$
  - Problem generically solvable
  - Prony's method (1795): direct "algebraic" solution method
- 1D extensions
  - $m(k) \sim \sum_j f_j(k) \exp(ikt_j)$ for some parametric family $\{f_j\}$
  - $\mathfrak{D}(k)m(k) = 0$ for some difference operator $\mathfrak{D}(k)$ with poly. coefficients
  - $f = \sum_\lambda c_\lambda v_\lambda,\ Av_\lambda = \lambda v_\lambda,\ m(k) = F(A^k f)$
- N-D extensions
  - $m(\boldsymbol{k}) \sim \sum_j c_j \exp(i\langle \boldsymbol{k}, \boldsymbol{t_j} \rangle)$
  - Algebraically defined domains

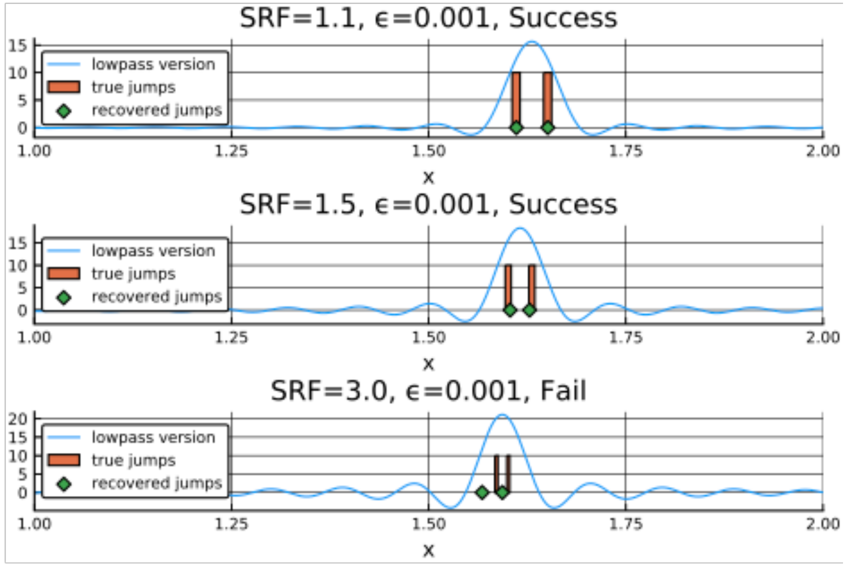
Existence/uniqueness?

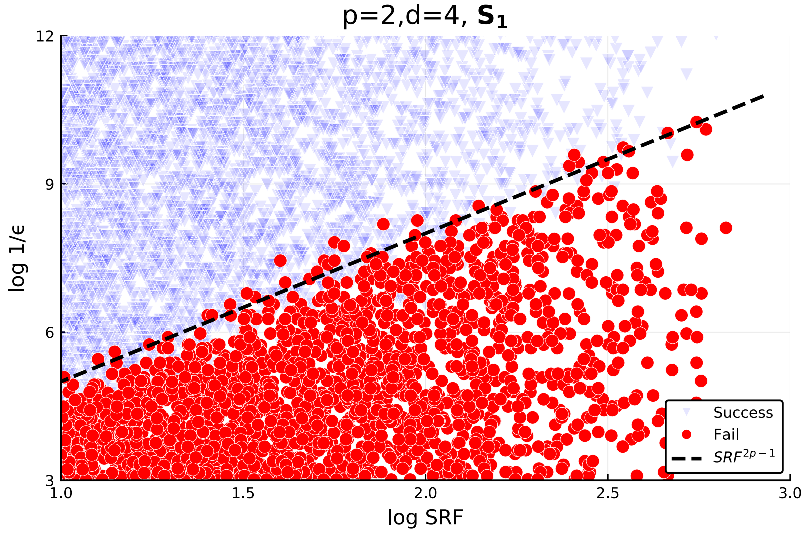
Stability?

Today: stability analysis

$$\Delta := \min_{j \neq k} |x_j - x_k|$$

$$\mathrm{SRF} := \frac{1}{\Delta\Omega} = \frac{\text{Rayleigh length}}{\Delta}$$





<u>Main reference:</u>

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.

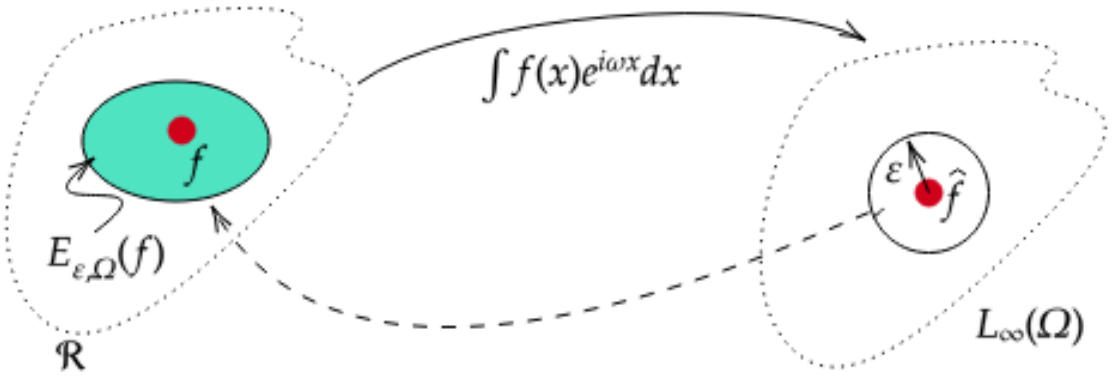
Setup

- Data: $\hat{f}(\omega) = \sum_{j=1,\ldots,d} c_j \exp(2\pi i \omega t_j) + e(\omega)$, $|\omega| \le \mathbf{\Omega}$, $|e(\omega)| \le \boldsymbol{\epsilon}$
- Model space: $\mathcal{P}_d = \{(\boldsymbol{c}, \boldsymbol{t}) : \boldsymbol{c} \in (\mathbb{C} \setminus \{0\})^d,\ \boldsymbol{t} \in \mathbb{R}^d, 0 < t_1 < \cdots < t_d < 1\}$
- Prior information: a compact subset $\emptyset \ne U(m, M, \Delta, \ldots) \subset \mathcal{P}_d$
  - $0 < m \le |c_j| \le M < \infty$
  - Additional "geometric" constraints - *clustering*
- $\mathcal{A} = \mathcal{A}(U, \Omega) = \{\alpha : L^\infty[-\Omega, \Omega] \cap C^0 \to U\}$ – admissible algorithms
- Projections onto components: $P_j^{(\boldsymbol{t})} : \mathcal{P}_d \to \mathbb{R}$, $P_j^{(\boldsymbol{c})} : \mathcal{P}_d \to \mathbb{C}$
- Uniform min-max error:

$$\mathcal{E}_j^{(\boldsymbol{t})}(U, \Omega, \epsilon) := \inf_{\alpha \in \mathcal{A}} \sup_{f \in U} \sup_{\|e\|_\infty < \epsilon} \left| P_j^{(\boldsymbol{t})} f - P_j^{(\boldsymbol{t})} \{\alpha(\hat{f} + e)\} \right|$$

$$\mathcal{E}_j^{(\boldsymbol{c})}(U, \Omega, \epsilon) := \inf_{\alpha \in \mathcal{A}} \sup_{f \in U} \sup_{\|e\|_\infty < \epsilon} \left| P_j^{(\boldsymbol{c})} f - P_j^{(\boldsymbol{c})} \{\alpha(\hat{f} + e)\} \right|$$

“Error sets” and min-max error



$$\mathcal{E}_j^{(t)}(U,\Omega,\epsilon) := \inf_{\alpha\in\mathcal{A}} \sup_{f\in U} \sup_{\|e\|_\infty<\epsilon} \left| t_j - P_j^{(t)}\{\alpha(\hat{f}+e)\} \right|$$

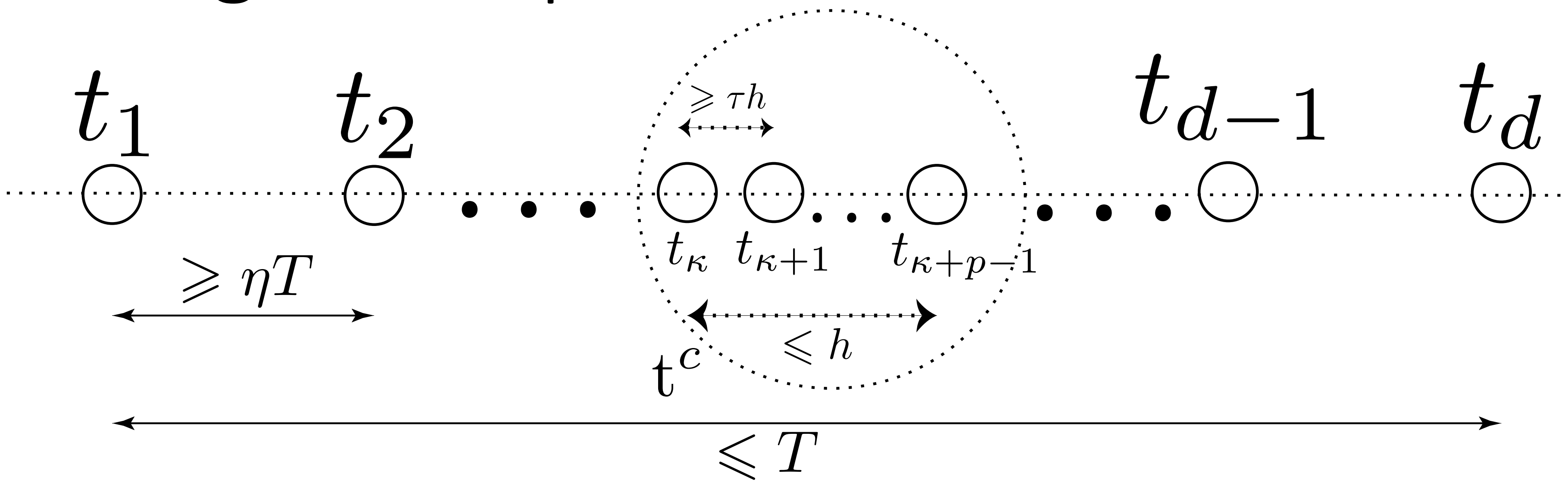
- $E_{\epsilon,\Omega}(f) := \{g \in \mathcal{P}_d\colon \|\hat{f}-\hat{g}\|_\infty \le \epsilon\}$
- $E^{(t)}_{\epsilon,\Omega,j}(f) := P_j^{(t)} E_{\epsilon,\Omega}(f) = \{P_j^{(t)} g\colon\ g \in E_{\epsilon,\Omega}(f)\}$

Proposition: $$\frac{1}{2}\sup_{f\in U\cap E_\epsilon(f)} \operatorname{diam} E^{(t)}_{\epsilon,\Omega,j}(f) \le \mathcal{E}_j^{(t)}(U,\Omega,\epsilon) \le \sup_{f\in U} \operatorname{diam} E^{(t)}_{2\epsilon,\Omega,j}(f)$$

Proof outline:

- RHS: consider an “oracle” $\alpha_0\colon f \mapsto \{\text{any element} \in E_{\epsilon,\Omega}(f)$ if not empty, otherwise some $f_0 \in U\}$
- LHS: consider $f_1, f_2 \in E_{\epsilon,\Omega}(f)$ “far away from each other”, then any α must make large error on either f_1 or f_2

Clustering assumptions



- Overall extent $T = 1$
- Single cluster t^c of multiplicity $p \geq 2$ and extent h
- “Rigidity” parameters $0 < \eta, \tau \leq 1$ – assumed to be fixed
- $U = U(m, M, p, d, h, \tau, \eta)$
- <u>Super-resolution regime:</u> $\frac{c_1}{\eta} \leq \Omega \leq \frac{c_2}{h}$ where c_1, c_2 depend on d, p, m, M only

Upper bounds on diam $E^{(t)}_{\epsilon,\Omega,j}(f)$

$$E_{\epsilon,\Omega}(f) := \{g \in \mathcal{P}_d : \|\hat{f} - \hat{g}\|_{L^\infty[-\Omega,\Omega]} \le \epsilon\}$$

- For $\lambda \in \left(0, \frac{\Omega}{2d-1}\right)$ ("blowup" parameter) define $F_\lambda : \mathcal{P}_d \to \mathbb{C}^{2d}$ by

$$F_\lambda : (\boldsymbol{c}, \boldsymbol{t}) \mapsto \left[\sum_{j=1}^{d} c_j \exp(2\pi i \lambda t_j k) = \hat{f}(\lambda k)\right]_{k=0,1,\ldots,2d-1}$$

- This is the square **Prony problem …** but with "new" nodes $s_j := \lambda t_j \bmod 1$
- $A^{(\lambda)}_{\epsilon,\Omega}(f) := F_\lambda^{-1}\{\boldsymbol{m} \in \mathbb{C}^{2d} : \boldsymbol{m} = F_\lambda(f) + \boldsymbol{e}, \|\boldsymbol{e}\|_\infty \le \epsilon\}$ (F_λ^{-1} is the local inverse)
- Step 1: there exists a large "admissible" set $\Lambda = \Lambda(\Omega, U) \subset \left(0, \frac{\Omega}{2d-1}\right)$ s.t. $\forall \lambda \in \Lambda$ we can effectively bound diam $A^{(\lambda)}_{\epsilon,\Omega}(f)$
  - For *small enough* $\epsilon < \epsilon(\Omega, U)$ and all $f \in U$ we have

$$\text{diam } A^{(\lambda)}_{\epsilon,\Omega}(f) \le C(U, \Omega) \cdot \epsilon$$

Quantitative Inverse Function Theorem

  - In fact, provide bounds for diam $P^{(t)}_j\, A^{(\lambda)}_{\epsilon,\Omega}(f)$ and diam $P^{(c)}_j\, A^{(\lambda)}_{\epsilon,\Omega}(f)$
  - Different scalings depending on whether $t_j \in t^c$ or not
- Step 2: $\exists \lambda^* \in \Lambda$ s.t. $\forall f \in U$: $E_{\epsilon,\Omega}(f) \subset A^{(\lambda^*)}_{\epsilon,\Omega}(f)$
  - Very technical because of the globally multi-valued nature of the "full inverse" $F^{-1}_{\lambda,\text{global}}$

Existential Inverse Function Theorem

- Let $f: B_1 \to B_2$ holomorphic, $B_1, B_2 \subset \mathbb{C}^n$ open
- $J_f = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{z}} \in \mathbb{C}^{n \times n}$ - Jacobian matrix
- Then $\det J_f(z_0) \neq 0$ iff $\exists U \subset B_1, V \subset B_2$ open s.t. $f: U \to V$ is a bijection and f^{-1} is holomorphic.
- In that case, $\frac{\partial \boldsymbol{f}^{-1}}{\partial \boldsymbol{v}}(\boldsymbol{v}) = J_f^{-1}(f^{-1}(\boldsymbol{v}))$ for all $\boldsymbol{v} \in V$.
- Q: how large can V be?
- Intuitively, this depends on $\|J_f^{-1}\|$:
  - a ball of diameter ϵ in B_2 is mapped to a domain of diameter $\leq \|J_f^{-1}\|\epsilon$ in B_1
  - Can we take $\epsilon \approx \frac{\operatorname{diam} B_1}{\|J_f^{-1}\|}$??

Quantitative Inverse Function Theorem (QIFT)

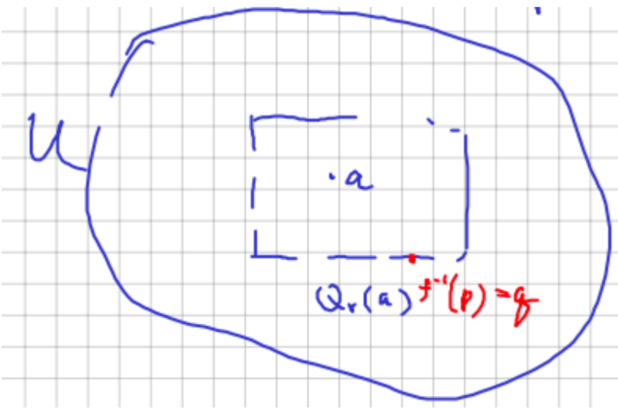
- Assumptions:
  - $f: U \subset \mathbb{C}^n \to \mathbb{C}^n$ holomorphic, injective (U open), $\forall z \in U: \det J_f(z) \neq 0$
  - U contains a <u>rectangle</u> of side lengths $\boldsymbol{r} = (r_1, \dots, r_n)$

$$Q_{\boldsymbol{r}}(\boldsymbol{a}) := \{\boldsymbol{y}: |y_i - a_i| \leq r_i, \forall i = 1, \dots, n\} \subset U$$

  - $\forall \boldsymbol{z} \in Q_r(\boldsymbol{a})$

$$\kappa_i(f; \boldsymbol{z}) := \sum_{j=1}^{n} \left|J_f^{-1}(\boldsymbol{z})\right|_{i,j} \leq A_i, \qquad i = 1, \dots, n.$$

Component-wise condition numbers



- Then
  - $f(U) =: V$ contains the cube $\tilde{Q} = Q_{\boldsymbol{e}}(f(\boldsymbol{a}))$,$\boldsymbol{e} = (e, \dots, e)$ where

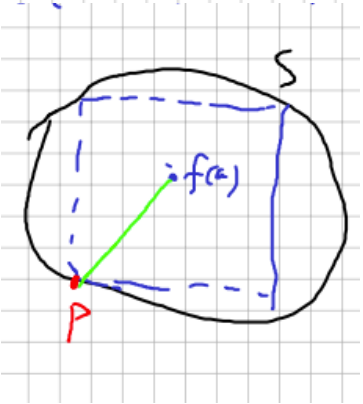
$$e := \min\left(\frac{r_1}{A_1}, \dots, \frac{r_n}{A_n}\right)$$

  - f^{-1} is holomorphic in an open nbhd. of $\tilde{Q}$, and

$$\forall \boldsymbol{v_1}, \boldsymbol{v_2} \in \tilde{Q}: \ |f^{-1}(\boldsymbol{v_1})_i - f^{-1}(\boldsymbol{v_2})_i| \leq A_i \|\boldsymbol{v_1} - \boldsymbol{v_2}\|_\infty$$

- Proof main idea: take $\boldsymbol{p}$ on the boundary of $S := f\big(Q_r(a)\big)$, on a cube of max. radius, then

$$r_i = \left|f_i^{-1}\big(f(\boldsymbol{a})\big) - f_i^{-1}(\boldsymbol{p})\right| = \left|\int_{f(\boldsymbol{a})}^{\boldsymbol{p}} \left(J_f^{-1}\right)_i\right| \leq A_i \|\boldsymbol{p} - f(\boldsymbol{a})\|_\infty$$



“Admissible” set Λ

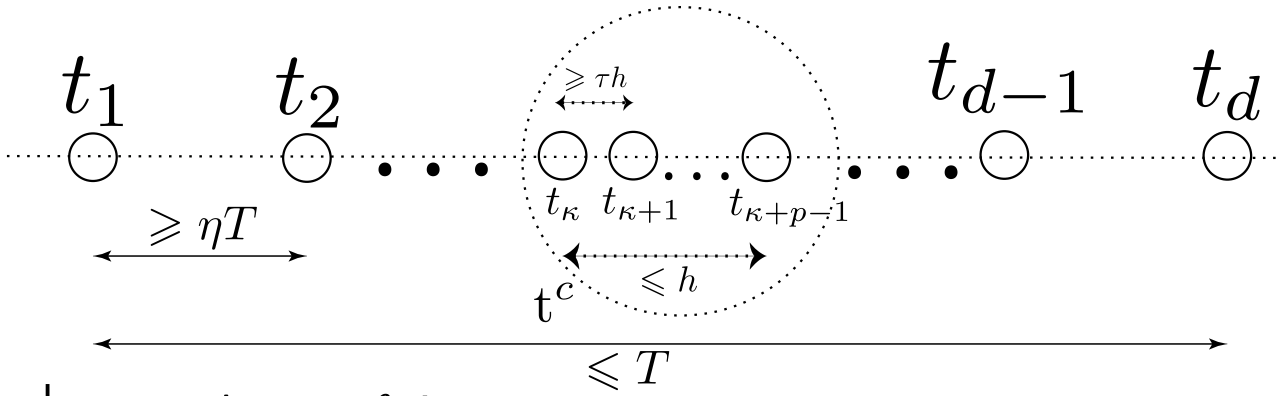
$$F_\lambda : (\boldsymbol{c}, \boldsymbol{t}) \mapsto \left[\sum_{j=1}^{d} c_j \exp(2\pi i s_j k)\right]_{k=0,1,\dots,2d-1}, \quad s_j := \lambda t_j \bmod 1$$

$$A_{\epsilon,\Omega}^{(\lambda)}(f) := F_\lambda^{-1}\{\boldsymbol{m} \in \mathbb{C}^{2d} : \boldsymbol{m} = F_\lambda(f) + \boldsymbol{e}, \|\boldsymbol{e}\|_\infty \le \epsilon\}$$

$$\kappa_i(F_\lambda; \boldsymbol{z}) := \sum_{j=1}^{2d} \left|J_{F_\lambda}^{-1}(\boldsymbol{z})\right|_{i,j}, \qquad i = 1, \dots, 2d.$$

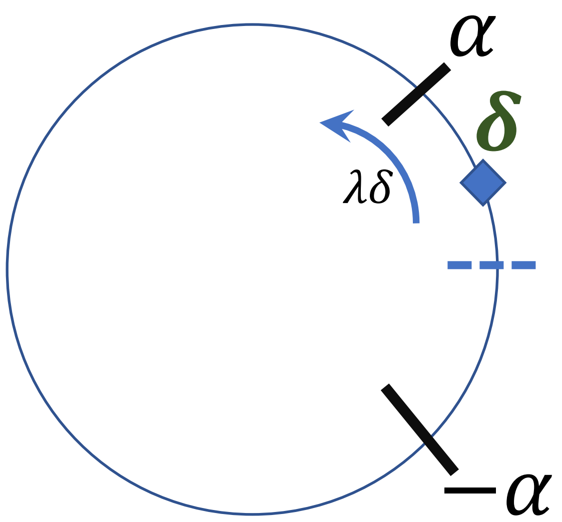
- Fact: $\kappa_i(F_\lambda) \sim \prod_{j \ne i} |s_i - s_j|^{-2}$ (will show next)
- Therefore: choose $\lambda \in \left(\frac{1}{2} \cdot \frac{\Omega}{2d-1}, \frac{\Omega}{2d-1}\right) = O(\Omega)$ s.t. $s_j := \lambda t_j \bmod 1$ are well-separated
  - This will imply well-conditioning of F_λ and a tight upper bound on diam $A_{\epsilon,\Omega}^{(\lambda)}(f)$
- $\Lambda = \Lambda(\Omega, \boldsymbol{t})$ is the subset of λ’s for which the following two conditions hold
  - $|s_\ell - s_j|_{\text{mod } 1} \ge \lambda \tau h = O(\Omega h)$ for $t_\ell, t_j \in \mathrm{t}^c$ - true for all λ
  - $|s_\ell - s_j|_{\text{mod } 1} \ge \frac{1}{d^2}$ for all other cases – true only for certain λ ($\frac{1}{d^2}$ is an artifact of proof)
- Lemma: $m(\Lambda) \ge \text{const}(d, \eta) \cdot \Omega$ ($m(\cdot)$ is the Lebesgue measure)
  - In fact Λ is a union of finite # of intervals

The existence of admissible λ



$s_j := \lambda t_j \bmod 1$

- Task: choose $\lambda \in I_0 := \left(\frac{1}{2} \cdot \frac{\Omega}{2d-1}, \frac{\Omega}{2d-1}\right)$ s.t. $\left|s_\ell - s_j\right|_{\text{mod } 1}$ is large $\forall \ell, j$
- If $p = d$ then any λ as above will suffice
- Otherwise, may get collisions between $t_j \notin \mathrm{t}^c$ and another node
- Key idea: collisions occur only for a small subset of λ's
- Consider $t, y \in \mathrm{t}$, put $\boldsymbol{\delta} := |t - y|_{\text{mod } 1} \geq \eta$
- $\beta(\alpha) := m\{\lambda \in I_0 \colon |\lambda\boldsymbol{\delta}|_{\text{mod } 1} \leq \alpha\}$ for $\alpha < \frac{1}{2}$
- $\lambda \in I$ with $|I| = \frac{1}{\boldsymbol{\delta}}$ → $\lambda\boldsymbol{\delta}$ traverses the circle once
- → $\frac{m\{\lambda \colon \lambda\boldsymbol{\delta} \in [-\alpha, \alpha]\}}{|I|} = 2\alpha$ → $\beta(\alpha) \leq 2|I|\alpha$
- Put $\alpha := \frac{1}{2d^2}$
- $m\{\lambda \in I_0 \colon \exists (t, y)\ s.t.\ |\lambda(t - y)|_{\text{mod } 1} \leq \alpha\} \leq \binom{d}{2}\beta(\alpha) \leq \frac{1}{2}|I|$



Estimating $\kappa_i(F_\lambda)$

- Consider $F: \{c_j, \rho_j\}_{j=1}^{d} \mapsto \left[\sum_{j=1}^{d} c_j \rho_j^k\right]_{k=0}^{2d-1}$
- Check: $J_F((\boldsymbol{c}, \boldsymbol{\rho})) = \frac{\partial[F_0 \ldots F_{2d-1}]^T}{\partial[c_1\ \rho_1 \ldots c_d\ \rho_d]} = U(\boldsymbol{\rho}) \times C(\boldsymbol{c})$ where
  - $U(\boldsymbol{\rho}) = \begin{bmatrix} 1 & 0 & \cdots & 1 & 0 \\ \rho_1 & 1 & \ddots & \rho_d & 1 \\ \rho_1^2 & 2\rho_1 & \ddots & \rho_d^2 & 2\rho_d \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_1^{2d-1} & (2d-1)\rho_1^{2d-2} & \cdots & \rho_d^{2d-1} & (2d-1)\rho_d^{2d-2} \end{bmatrix}$ - confluent Vandermonde matrix
  - $C(\boldsymbol{c}) = \mathrm{diag}\{1, c_1, 1, c_2, \ldots, 1, c_d\}$
- Since $J_F^{-1} = C^{-1}U^{-1}$, $\kappa_i(F) \asymp \nu_i(U^{-1})$ where $\nu_i(A) = \sum_{j=1}^{2d} |A_{i,j}|$
- To estimate $\nu_i(U^{-1})$ we use its connection to Hermite interpolation

Estimating $\nu_i(U^{-1})$

$$U(\boldsymbol{\rho}) = \begin{bmatrix} 1 & 0 & \dots & 1 & 0 \\ \rho_1 & 1 & \ddots & \rho_d & 1 \\ \rho_1^2 & 2\rho_1 & \ddots & \rho_d^2 & 2\rho_d \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_1^{2d-1} & (2d-1)\rho_1^{2d-2} & \dots & \rho_d^{2d-1} & (2d-1)\rho_d^{2d-2} \end{bmatrix}$$

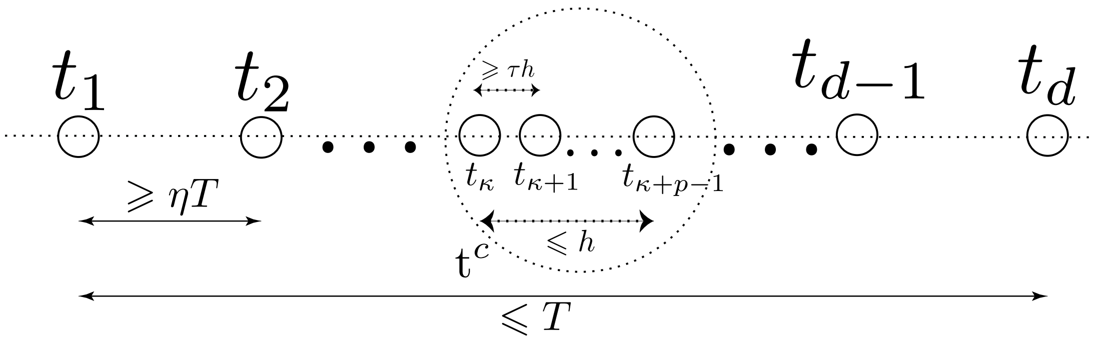
- Let $p(z) = (z-\xi_1)\dots(z-\xi_d) = \sum_{j=1}^{d} p_j z^j$ be any polynomial, then $U^T(\rho)\boldsymbol{p} = [p(\rho_1)\ p'(\rho_1) \dots p(\rho_d)\ \ p'(\rho_d)\ ]^T = p(\boldsymbol{\rho})^T$
  - $\boldsymbol{p} = (U^T)^{-1} p(\boldsymbol{\rho})^T$.
- Substitute for $p(z)$ the fundamental polynomials of (Hermite) interpolation
  - $h_i(\rho_j) = \delta_{ij}, h_i'(\rho_j) = 0$ for $1 \le i,j \le d$ → $\boldsymbol{h_i}$ = $2i-1$-st row of U^{-1}
  - $g_i'(\rho_j) = \delta_{ij}, g_i(\rho_j) = 0$ for $1 \le i,j \le d$ → $\boldsymbol{g_i}$ = $2i$-th row of U^{-1}
- Lemma [1]: for any $p(z)$ as above, $\sum_{j=0}^{d} |p_j| \le \prod_{i=1}^{d}(1+|\xi_i|)$
- From basic numerical analysis: set $\ell_i(z) = \frac{\prod_{j\ne i}(z-\rho_j)}{\prod_{i\ne j}(\rho_i-\rho_j)}$, then
  - $h_i(z) = \ell_i^2(z)(1-2\ell_i'(\rho_i)(z-\rho_i))$
  - $g_i(z) = (z-\rho_i)\ell_i^2(z)$

All zeros + leading coefficients are known

- Example: single cluster, $|\rho_i - \rho_j| \approx \Delta \ll 1$
  - $\nu_{2i-1}(U^{-1}) \approx \Delta^{1-2d}$
  - $\nu_{2i}(U^{-1}) \approx \Delta^{2-2d}$

[1] W. Gautschi, “On inverses of Vandermonde and confluent Vandermonde matrices,” *Numerische Mathematik*, vol. 4, no. 1, pp. 117–123, 1962.

Back to our business…



- Theorem: $\frac{c_1}{\eta T} \leq \Omega \leq \frac{c_2}{h}$, then with $\mathrm{SRF} \coloneqq (\Omega h)^{-1}$ and $\epsilon < c_3 \mathrm{SRF}^{1-2p}$

$$\mathcal{E}_j^{(t)}(U,\Omega,\epsilon) \asymp \frac{\epsilon}{\Omega} \times \begin{cases} \mathrm{SRF}^{2p-2}, & t_j \in \mathrm{t}^c \\ 1 & t_j \notin \mathrm{t}^c \end{cases}$$

$$\mathcal{E}_j^{(c)}(U,\Omega,\epsilon) \asymp \epsilon \times \begin{cases} \mathrm{SRF}^{2p-1}, & t_j \in \mathrm{t}^c \\ 1 & t_j \notin \mathrm{t}^c \end{cases}$$

- Corollary: $\mathrm{SRF} \asymp \left(\frac{1}{\epsilon}\right)^{\frac{1}{2p-1}}$
- Upper bound: choose $\lambda^* \in \Lambda$, estimate $\kappa_i(F_{\lambda^*})$ and apply QIFT.
- Lower bound: given $f \in U$, construct $f_1 \in U$ with $m_k(f_1) = m_k(f)$ for $k = 0, \ldots, 2d-2$ but $m_{2d-1}(f_1) = m_{2d-1}(f) + \epsilon$
  - Needs an additional argument from [1]
  - Works only if nodes of f are real

[1] A. Akinshin, G. Goldman, and Y. Yomdin, "Geometry of error amplification in solving the Prony system with near-colliding nodes," *Math. Comp.*, 90(327) , 267–302, 2021, 10.1090/mcom/3571.

Summary – SR stability

1. ℓ_1 min: $\|\delta f\|_1 \sim \mathrm{SRF}^{2p}\epsilon$ if $c_j > 0$
2. "Off-grid" (today): $|\delta t_j| \sim \frac{1}{\Omega}\mathrm{SRF}^{2p-2}\epsilon$ if $\epsilon < c \cdot \mathrm{SRF}^{2p-1}$
3. "On-grid": $\|\delta f\|_2 \sim \mathrm{SRF}^{2p-1}\epsilon$ [1,2,3]
   - In [3] a somewhat different model was considered
   - Similar techniques for $\sigma_{\min,s,\boldsymbol{p}}(F_n)$ in [1], completely different in [2]

- Next time: tractable algorithms for 2/3

[1] **DB**, L.Demanet, G. Goldman, and Y. Yomdin, "Conditioning of Partial Nonuniform Fourier Matrices with Clustered Nodes," *SIAM J. Matrix Anal. Appl.*, 44(1), 199–220, 2020, 10/ggjwzb.

[2] W. Li and W. Liao, "Stable super-resolution limit and smallest singular value of restricted Fourier matrices," *Appl. Comput. Harm. Anal.*, 51, 118–156,. 2021, 10.1016/j.acha.2020.10.004.

[3] D. L. Donoho, "Superresolution via sparsity constraints," *SIAM Journal on Mathematical Analysis*, vol. 23, no. 5, pp. 1309–1331, 1992.

Open research questions

- What if $\mathrm{c}_3\mathrm{SRF}^{-2\mathrm{p}+1} \leq \epsilon \leq c_4\mathrm{SRF}^{-2\tilde{p}+1}$?
- Finding optimal λ? (in progress…)
- Arbitrary complex ρ_j?
- What happens with the constants as $d \to \infty$ but p remains bounded?
- $\hat{f}(\omega) \sim \hat{k}(\omega) \sum_j c_j \exp(2\pi i \omega t_j)$ - dependence on supp $\hat{k}$?
- $\hat{f}(\omega) \sim \sum_j c_j(\omega) \exp\left(2\pi i \omega t_j\right)$ where $c_j' \ll c_j$?
- $\hat{f}(\omega) \sim \sum_j c_j(\omega) \exp(2\pi i \varphi_j(\omega))$ where $\varphi_j" \ll \varphi_j'$?
- …
- …