Lecture 11

- „Minimax" bounds for sparse SR
- Quantitative estimates on the neighb. ($\varepsilon \ll 1$)
- Stability of least squares
- Structural perturbations & extensions
- Open questions

Next time: overview & add. topics, info re. projects (november).

Last lecture: online quiz

Optimal SR, "mini-max", "worst-case error".

(local approach → „semi-local", „semi-global"

P_d (signal space) $= \{ (c,x),\ c \in \mathbb{C}^d,\ x \in \mathbb{R}^d \}$, $\|\cdot\|_\infty$ component wise

$x_1 < x_2 < \dots < x_d$

Ω: Bandlimit

data space: $L_\infty([-\Omega,\Omega]) \cap C^0 := \mathcal{S}$, $\|\cdot\| = \|\cdot\|_\infty$.

Will consider compact subsets of P_d, $\mathcal{U} \subset P_d$, ($\neq \emptyset$)

Prior information { min/max amplitudes; clustering, min. separation

Algorithm: $\mathcal{A} : \{ a : \mathcal{S} \to \mathcal{U} \}$ $\mathcal{A}(\mathcal{U},\Omega)$.

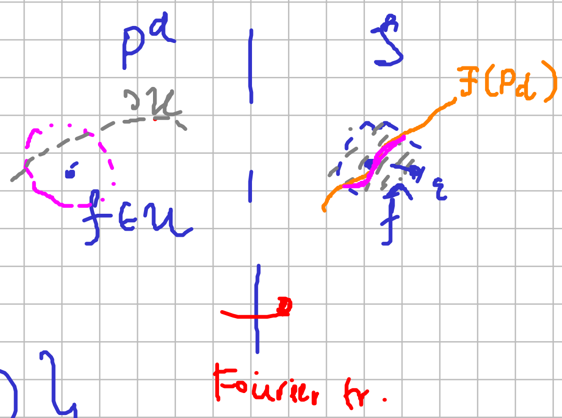
Min-max / worst case error:

$j = 1,\dots,d$: $\mathcal{E}_j^{(t)} = \mathcal{E}_j^{(t)}(\mathcal{U},\varepsilon,\Omega)$ $(\mathcal{E}_j^{(c)})$

$$:= \inf_{a \in \mathcal{A}} \sup_{\substack{f \in \mathcal{U} \\ e \in \mathcal{S} \\ \|e\| \le \varepsilon}} \left| t_j - (a(\hat{f} + e))_j^{(t)} \right|$$

$f = \sum_{j=1}^{d} c_j \delta(t - t_j)$

Error sets "preimages"



$$E(f,\varepsilon) := \{ g \in P_d : \|\hat{f}-\hat{g}\| \le \varepsilon \}$$

$$E_j^{(t)} = \{ (g)_j^{(t)} : g \in E(f,\varepsilon) \}$$

Proposition

$$\frac{1}{2} \sup_{f\in\mathcal{U}\cap E(f,\varepsilon)} \operatorname{diam} E_j^{(t)}(f,\varepsilon) \le \mathcal{E}_j^{(t)}(\mathcal{U}) \le \sup_{f\in\mathcal{U}} \operatorname{diam} E_j^{(t)}(f, 2\varepsilon)$$

Pf. RHS

"Oracle" $\quad y \in S$ given

Consider $\mathcal{B}(\varepsilon, y) = \{ f \in \mathcal{U} : \|\hat{f} - y\| \le \varepsilon \}$

(may be $\emptyset$)

$$\alpha_0 = \begin{cases} g \in \mathcal{B}(\varepsilon,y) & \text{if } \mathcal{B}(\varepsilon,y) \ne \emptyset \\ g_0 \in \mathcal{U} & \text{else.} \end{cases}$$

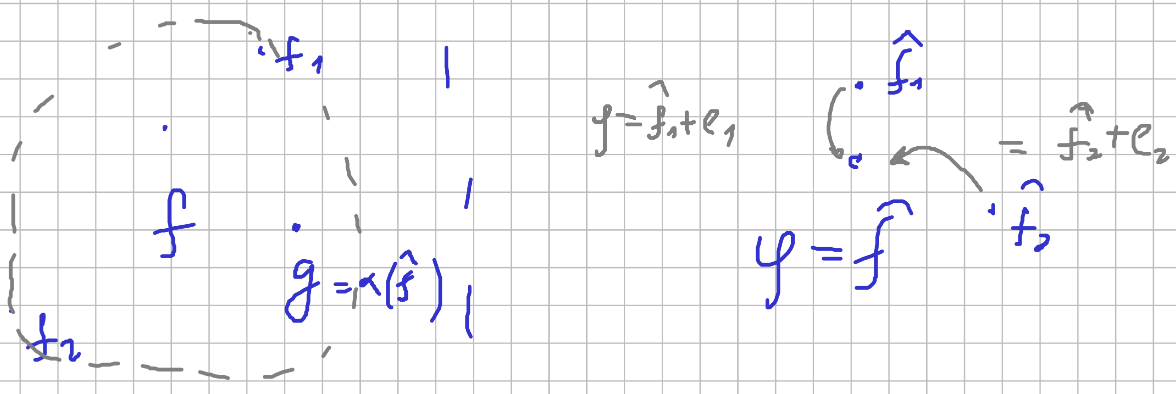
maybe can learn it?

$f\in\mathcal{U}$, $\|e\| \le \varepsilon$, $y = \hat{f} + e \Rightarrow f \in \mathcal{B}(\varepsilon, y)$

$f_0 = \alpha_0(y) \Rightarrow \|\hat{f}_0 - \hat{f}\| \le 2\varepsilon \Rightarrow f_0 \in E(f, 2\varepsilon)$

$\Rightarrow \mathcal{E}_j^{(t)} \le \sup_f \operatorname{diam} E_j^{(t)}(f, 2\varepsilon)$

LHS: $f \in \mathcal{U}$, $\alpha \in \mathcal{A}$, $g = \alpha(\hat{f})$



$\forall \varepsilon > 0$, let's take $f_1, f_2 \in E(f, \varepsilon)$,

$$|(f_1)_j^{(t)} - (f_2)_j^{(t)}| \geq \operatorname{diam} E(f, \varepsilon) - \varepsilon$$

$$\|\hat{f}_1 - y\|, \ \|\hat{f}_2 - y\| \leq \varepsilon$$

So α must make a "large" error either for f_1 or f_2

$$\varepsilon_j^{(t)} = \inf_{\alpha} \sup_{f, e} \left| (f)_j^{(t)} - \left(\alpha(\hat{f} + e)\right)_j^{(t)} \right|$$

$$\geq \inf_{\alpha} \max \{ |f_1 - g|, |f_2 - g| \}$$

proj

$$\geq \inf_{\alpha} \frac{1}{2} \{ |f_1 - g| + |f_2 - g| \}$$

$$\geq \frac{1}{2} |f_1 - f_2| \geq \frac{1}{2} \operatorname{diam} E(f, \varepsilon) - \frac{\varepsilon}{2}$$

$\varepsilon \to 0$.

Estimation of diam $E(f,\varepsilon)$.

Consider λ, "decimation parameter"

Prony map $\quad F_\lambda : P_d \to \mathbb{C}^{2d}, \qquad E^{(\lambda)}(f,\varepsilon) = F_\lambda^{-1}(B_\varepsilon(m))$ $\quad \mathbb{C}^{2d} \cup\!|$

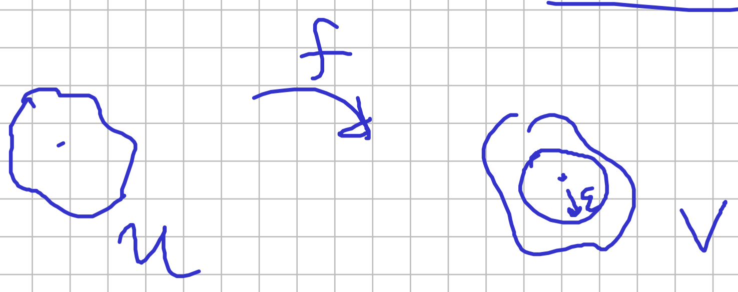
$$m = \Big[\sum_{j=1}^{d} c_j e^{i t_j \cdot \lambda \cdot k} \Big]_{k=0}^{2d-1}.$$

Easy to see that
$$\bigcap_{\lambda \in (0, \frac{\Omega}{2d-1})} E^{(\lambda)}(f,\varepsilon) \supseteq E(f,\varepsilon).$$

In particular, sufficient to estimate diam $E^{(\lambda)}(f,\varepsilon)$ (from above).
for an "optimal" λ^*. .

We've seen existence of $\lambda^* \asymp O(\Omega)$ $\Big|$ $\frac{1}{\lambda}\left(\frac{1}{\lambda\Delta}\right)^{2p-2}$

Additional ingredient: Inverse function theorem
<u>Quantitative</u>.



want to show existence of a large ball on which f^{-1} exists.

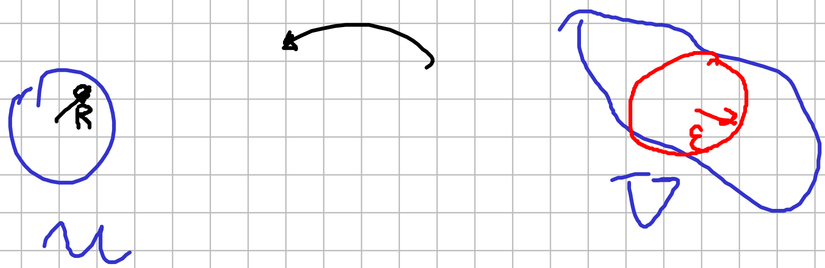
Classical IFT:

$f: B_1 \to B_2$, B_1, B_2 open, f holomorphic

then $|J_f(z_0)| \neq 0$ iff $\exists\, \mathcal{U} \subset B_1$, $V \subset B_2$ open neighborhoods

such that $f: \mathcal{U} \to V$ is biholomorphic
(f bijection, f^{-1} holomorphic)

Intuition:



Suppose $\|f^{-1}\| \leq A$

a ball of radius R will be mapped to a domain "of radius" $\leq A \cdot \varepsilon \leq R$

"safe" to take $\varepsilon \approx \frac{R}{A}$.

How to find f^{-1}?

Solve $y = f(x)$

$g(x) = y - f(x) = 0$

Newton's method

Can use but obtain non-optimal bounds.

Th: $U \subset \mathbb{C}^n$ open, $\forall z \in U$ $|J_f(z)| \neq 0$

Further, suppose U contains a rectangle of side lengths

$r = (r_1, \dots, r_n)$

$r_1 \dots r_{2d}$

r_j for node $j \approx \Delta$

for ampl. $\times \frac{\min \text{amp}}{2}$

$$Q_r(a) = \{ y : |y_i - a_i| \le r_i \}$$

$a \in \mathbb{C}^n$.

Assume also that $\forall z \in Q_r(a)$

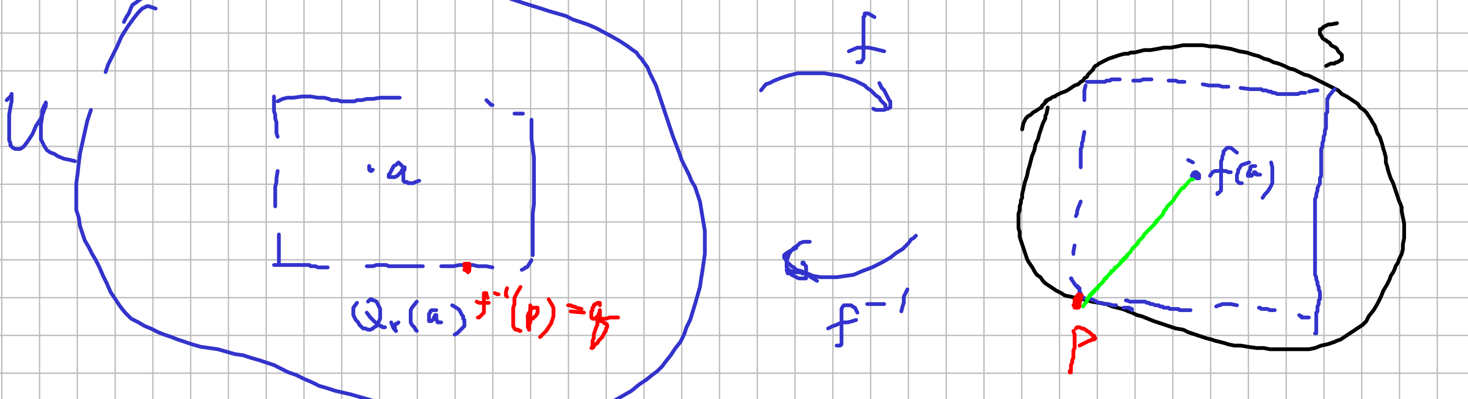
$$\sum_{j=1}^{n} \left| J_f^{-1}(z) \right|_{i,j} \le A_i$$

Then $f(U)$ contains $Q_{e=(e,\dots,e)}$, where

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

Pf: f biholomorphism between U and V.

and so homeomorphism $S = f(Q_r(a)) \to$ compact



Let $\tilde{Q}$ be the maximal cube centered at $f(a)$, $\tilde{Q} \subset S$.

$\tilde{Q} = Q_e$

$\Rightarrow \exists p \in \partial S \cap \partial \tilde{Q}$, $f^{-1}(p) \in \partial Q_r(a)$

$\Rightarrow \exists j$ s.t. $|q_j - a_j| = r_j$

Integrate f_j^{-1} from $f(a)$ to p.

Fundamental th. of calculus, mean value thm.

$$r_j = \underbrace{\left| f_j^{-1}(f(a)) - f_j^{-1}(p) \right|}_{a_j - q_j} = \left| \int_{f(a)}^{p} (D f_j^{-1})(y)\, dy \right|$$

$$\leq A_j \, \| p - f(a) \|$$

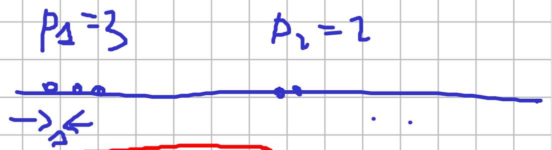
$$\Rightarrow \| p - f(a) \| \geq \frac{r_j}{A_j}.$$

□

$\left(\left| f^{-1}(a) - f^{-1}(b) \right|_j \leq A_j \| a - b \| . \right.$

$a, b \in Q_\varepsilon$

Back to clustered SR:



Th: for $\varepsilon < \varepsilon_{crit} \approx (\Omega \Delta)^{2p_{max} - 1}$

$\frac{\Delta}{A_j} = \frac{r_j}{A_j}$

(min-max) $\quad \mathcal{E}_j^{(t)} \leq A_j^{(t)} \cdot \varepsilon$

$$A_j^{(t)} = c \left(\frac{1}{\Omega \Delta} \right)^{2p_j - 2} \cdot \frac{1}{\Omega}$$

[Note: $p_j = 1$ we get "well-separated regime"].

Note: $\mathcal{U} = \{ m < |c_j| < M, \; |t_j - t_i| > \rho \;\; i,j$ not in same cluster $\; d, p_j, \; c = c(\rho, m, M) \}$

Fact: can show matching lower bounds for $\mathcal{E}_j^{(t)}$

Given $\mathcal{U}$, exhibit $f_1, f_2 \in \mathcal{U}$ $\quad |(f_1)_j^{(t)} - (f_2)_j^{(t)}| \geq A_j^{(t)} \varepsilon$

and $\|\hat{f}_1 - \hat{f}_2\| \leq \varepsilon$. ?

(divided differences, $f \sim \delta_\Delta^{(p-1)}(t)$)

[also can do: given f^{EPd}, find f' s.t. ...]

Need lower bounds for $\mathcal{F}^{-1}$

Open questions

1) Fine-tune inside $\mathcal{U}$, $c_j \in \mathbb{R}$ vs $c_j \in \mathbb{C}$, vs $c_j > 0$.

2) What exactly can we recover if $(\Omega\Delta)^{2p_{max}-1} < \mathcal{E} < (\Omega\Delta)^{2p'-1}$?
   - smaller clusters?
   - averages of big clusters?

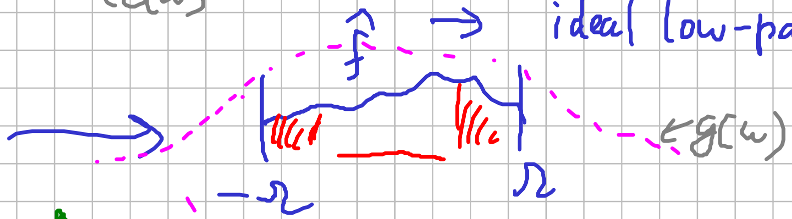
3) $d \to \infty$, $p_{max} \leq p$ need to find "local" algorithms

4) $d(\omega) = \left(\sum c_j e^{i\omega t_j} + e(\omega)\right) g(\omega)$, for some classes of filters

$\hat{f}$ → ideal low-pass filter

! ?

$f = \sum c_j \delta(t - t_j)$



$\int k(t-s) f(s)\, ds$

sinc, ...,

band pass

[In general, bad idea to just $\frac{d(\omega)}{g(\omega)}$

Algorithms (tractable or less so ...)
(polynomial time).

1) Try to find optimal λ^*.

2) "Sparsity-type". discretize the point grid N pts $[-\pi, \pi]$

M. Elad

R. Giryes

$c \in \mathbb{C}^M$, # non zeros $= d$, $\|c\|_0 \leq d$

fat matrix $\to Fc = m + e$

N measurements

compressed sensing.

3) Least squares. (Gauss). data fitting
(nonlinear) Machine learning

. (פדרי יוס ג׳ י אמ׳ר בק – אחר)

Task: compute $\min_x f(x, \varepsilon)$ ε "nuisance parameters"

s.t. $\begin{cases} g_i(x) \le 0 \\ h_i(x) = 0 \end{cases}$ x – "primary".

$$x^*(\varepsilon) = \arg\min_x f(x, \varepsilon)$$

L.S: $F(x) = \begin{bmatrix} f_1 \\ \vdots \\ f_p \end{bmatrix}$, model $\left[f_k = \sum c_j e^{i t_j k} \right]$

for simplicity $d(\varepsilon) = \begin{bmatrix} d_1 \\ \vdots \\ d_p \end{bmatrix}$ data (noise ε)

(assume everything sufficiently smooth). $d(0) = F(x_0)$

$$f(x, \varepsilon) = \frac{1}{2} \| F(x) - d(\varepsilon) \|^2$$

Question: estimate $\| x^* - x_0 \|$

Idea: consider $x^*(\varepsilon)$ as an implicit solution to

$(\nabla_x f(x^*(\varepsilon), \varepsilon) \equiv 0)$. $\underbrace{\nabla_x f(x, \varepsilon)}_{g} = 0$ (*)

Can use implicit f-n theorem,

$D_x g \neq 0 \Rightarrow D^2_{xx} f(x_0, 0)$

Minimum when $D^2_{xx} f(x_0, 0) \succ 0$. } sufficient conditions

$\nabla_x f(x_0, 0) = 0$

$Dg = 0$

$$x'(\varepsilon) = -\left(\nabla^2_{xx} f(x,\varepsilon)\right)^{-1} \nabla^2_{x\varepsilon} f(x,\varepsilon)$$

(can also do constrained min.)

For least squares: $J = \frac{\partial F}{\partial x}$

$$\nabla_x f = J^T (F-d)$$

$$\nabla^2_{xx} f(x,0) = J^T J$$

$$\nabla^2_{x\varepsilon} f(x,\varepsilon) = -J^T d'(\varepsilon)$$

perturbation model

Simple case: $d' \approx$ const $(d \approx c_1 + c_2 \varepsilon)$

$$x'(0) = J^+ d'(0).$$

Open questions:

1) Estimate $\|J^+_i\|_1$ in the multi-cluster setting

2) extend to finite ε.

3) Convergence speed, initialization. $x_{init} \approx x_0$

how close should this be?

(basin of attraction of the global optimum).