Nonlinear Sampling Recovery for Multivariate Function Classes

Tino Ullrich

Faculty of Mathematics
Chemnitz University of Technology

SIGMA 2024 Workshop
Luminy, France

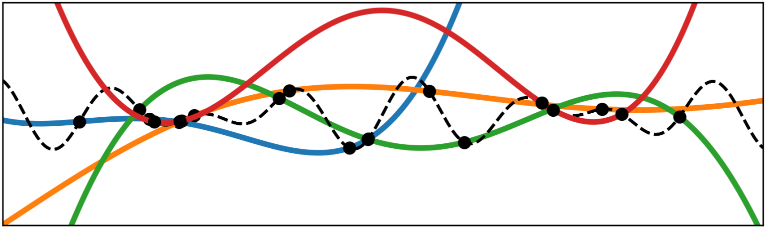
Joint work with...

Thomas Jahn (KU Eichstätt-Ingolstadt)
Felix Voigtlaender (KU Eichstätt-Ingolstadt)

recently published as

- Sampling numbers of smoothness classes via ℓ_1-minimization, **J. Complexity**, 79, 2023

Sampling widths



$$\varrho_m(\mathcal{F})_X := \inf_{t_1,\ldots,t_m\in\Omega} \inf_{R:\mathbb{C}^m\to X} \sup_{\|f\|_{\mathcal{F}}\leq 1} \|f - R(f(t_1),\ldots,f(t_m))\|_X .$$

sampling width = **minimal worst-case** error for **optimal** *standard information*

- Bartel, Cohen, Dai, Dolbeault, Dũng, Heinrich, Kämmerer, Krieg, Nagel, Novak, Schäfer, Sickel, Temlyakov, Triebel, M. Ullrich, T. Ullrich, Voigtlaender, Vybíral, Wojtaszczyk, Woźniakowski, ...

Sampling widths vs. best n-term approximation

$$\varrho_{\lceil Cn \log(n)^4 \rceil}(\mathcal{F})_{L_2} \leq \tilde{C}\sigma_n(\mathcal{F}, \mathcal{B})_{L_\infty}$$

Approximate using $\geq Cn\log(n)^4$ **samples** of $f \in \mathcal{F}$, error measured in L_2

approximate $f \in \mathcal{F}$ by **linear combinations of n basis elements** of $\mathcal{B}$, error measured in L_∞

- Simplified version of main result
- Example: best-m-term trig. approximation
- Constant C and $\tilde{C}$ are under control!
- Quantity on the right-hand side has been studied intensively in various scenarios

Sparse (non-linear) approximation

- Best n-term approximation wrt. a dictionary $\mathcal{B} = (\varphi_j)_j$
- X quasi-Banach space, $f \in X$

$$\begin{aligned}\sigma_s(f,\mathcal{B})_X &= \inf\left\{\Big\|f - \sum_{j\in\Lambda}\lambda_j\varphi_j\Big\|_X \;:\; |\Lambda| \le n, \lambda_j \in \mathbb{C}, \varphi_j \in \mathcal{B}\right\}\\ &= \inf_{g\in\Sigma_n}\|f-g\|_X\end{aligned}$$

- Best n-term widths

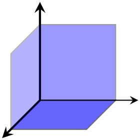
$$\sigma_n(\mathcal{F},\mathcal{B})_X := \sup_{f\in\mathcal{F}}\sigma_n(f,\mathcal{B})_X$$

- $\sigma_1 \ge \sigma_2 \ge ... \ge \sigma_n \ge ...$
- See also **Pietsch 1981**

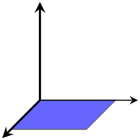
Non-linear vs. linear approximation

$V_J :=$ linear combinations of basis elements with coeff. in J

$\Sigma_n :=$ linear combinations of n dictionary elements



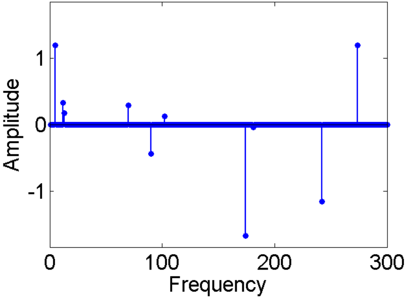
Σ_2

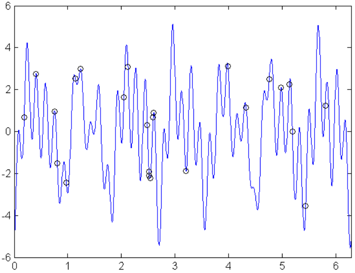


$V_{\{1,2\}}$

$$\sigma_n(\mathcal{F},\mathcal{B})_X := \sup_{\|f\|_{\mathcal{F}}\leq 1} \inf_{g\in\Sigma_n} \|f-g\|_X \quad , \quad E_J(\mathcal{F})_X := \sup_{\|f\|_{\mathcal{F}}\leq 1} \inf_{g\in V_J} \|f-g\|_X$$

Sparse trigonometric polynomials





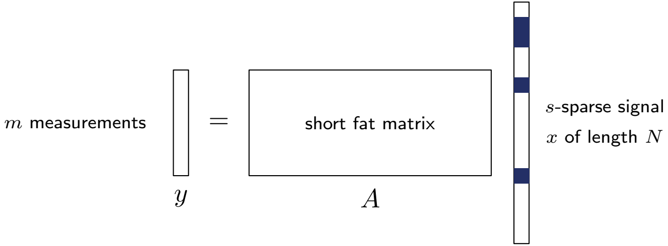
$$f(x) = \sum_{k \in \Gamma} c_k e^{2\pi i k t} \quad , \quad |\Gamma| = N$$

Sparsity s: Only few frequencies are "active", i.e.,

$$|\{k \, : \, c_k \neq 0\}| \leq s$$

Goal: Reconstruct f from samples $y^T = (f(t_1), ..., f(t_m))$ where $m << N$.

Basis pursuit denoising - ℓ_1-minimization



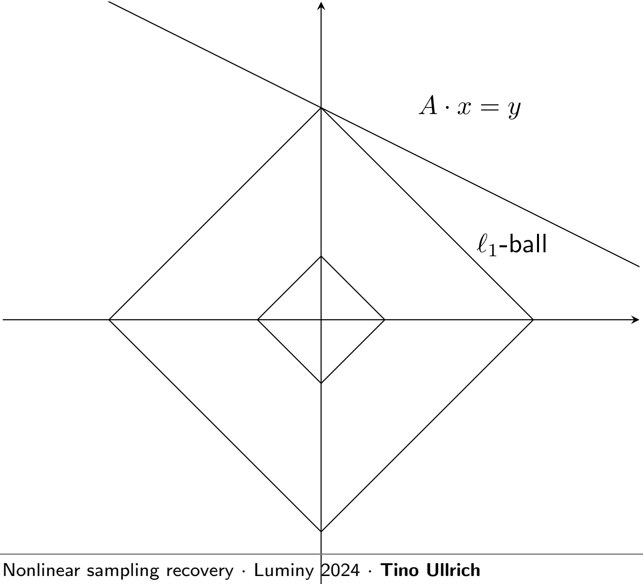
Perturbed samples: $\tilde{y}_i = y_i + \delta = A \cdot x + \delta$, A satisfies **RIP** of order s

Basis pursuit denoising: $\min_{x \in \mathbb{C}^N} \|x\|_1$ subject to $\|y - Ax\|_2 \leq \delta\sqrt{m}$

$$\left\|x - x^{\#}\right\|_{\ell^2} \leq \frac{C_1}{\sqrt{s}} \sigma_s(x)_1 + C_2 \delta.$$

Foucart, Rauhut '13: A mathematical introduction to CS

Illustration



RIP for bounded orthonormal systems

- $\mathcal{B} := (\phi_j)_{j\in[N]} \subset L_2(\mu)$ bounded orthonormal system, i.e. $\|\varphi_j\|_{L_\infty} \leq K$
- Number of samples

$$m \geq C \cdot K^2 \cdot s \cdot \log(s)^3 \cdot \log(N)$$

- $t_1, \ldots, t_m \overset{iid}{\sim} \mu$
- Then, for $A = (\phi_j(t_\ell))_{\ell\in[m], j\in[N]}$, the matrix $\frac{1}{\sqrt{m}}A$ has RIP(s).

  Sparse signals can be recovered robustly using ℓ^1-minimization for the measurements given by A.

- **Candes, Tao, Donoho, Foucart, Rauhut ...**
- **Bourgain '14, Haviv, Regev '17**: $\log(s)^2$ for Fourier basis

Minimal number of samples

Lemma (Foucart, Pajor, Rauhut, T. Ullrich 2010)

Let $0 < p \leq 1$ *and* $N, m, s \in \mathbb{N}$. *If* $A \in \mathbb{R}^{m \times N}$ *is a matrix such that every* $2s$*-sparse vector is exactly recovered by* ℓ_1*-minimization. Then*

$$m \geq cs \log\left(\frac{N}{4s}\right),$$

where $c := 1/\log 9 \approx 0.455$.

Corollary: Sharp behavior of Gelfand widths for ℓ_p with $0 < p \leq 1$ in ℓ_2

$$c_m(\ell_p, \ell_2) \asymp \left(\frac{\log(eN/m)}{m}\right)^{1/p-1/2}$$

Recovery with high probability

Theorem [Jahn, T. Ullrich, Voigtlaender '23]

Let $\mathcal{F} \hookrightarrow L_\infty$ and $P : L_\infty \to L_\infty$ a (J, J^*)-quasi-projection.
Put

$$\eta := 2\|P\|_{\infty\to\infty} \cdot \sigma_n(\mathcal{F})_{L_\infty} + (1 + \|P\|_{\infty\to\infty}) \cdot E_J(\mathcal{F})_{L_\infty}$$

and $N := |J^*|$. Drawing at least

$$m := \left\lceil CK^2\kappa \cdot n \cdot \log(n)^3 \cdot \log(N) \right\rceil$$

nodes $t_1, \ldots, t_m \overset{iid}{\sim} \mu$, then, with prob. $\geq 1 - N^{-\gamma \log(n)^3}$

$$\sup_{\|f\|_{\mathcal{F}} \leq 1} \|f - R_\eta(f(t_1), \ldots, f(t_m))\|_{L_2} \leq \widetilde{C}\eta$$

with universal constants $C, \widetilde{C}, \gamma > 0$.
The approximant $R_\eta(f(t_1), \ldots, f(t_m))$ is contained in V_{J^*}.

Sampling widths

Corollary

$$\varrho_{\lceil Cn\log(n)^3\log(M)\rceil}(\mathcal{F})_{L_2} \leq \widetilde{C}\Big(\sigma_n(\mathcal{F},\mathcal{B})_{L_\infty} + E_{\{0,\dots,M\}}(\mathcal{F})_{L_\infty}\Big) \tag{1}$$

For trigonometric polynomials improvement (due to improved RIP):

Corollary (Trigonometric system)

$$\varrho_{\lceil Cd\log(d+1)n\log(n)^2\log(M)\rceil}(\mathcal{F})_{L_2} \leq \widetilde{C}\Big(\sigma_n(\mathcal{F},\mathcal{T}^d)_{L_\infty} + E_{[-M,M]^d\cap\mathbb{Z}^d}(\mathcal{F})_{L_\infty}\Big)\,. \tag{2}$$

- Constant $\tilde{C}$ is universal and absolute, i.e., not depending on d

- $\mathbb{T}^d$...d-torus represented by $[0,1)^d$
- $I_j = I_{j_1} \times \cdots \times I_{j_d}$ dyadic frequency block, where $I_0 = \{-1,0,1\}$ and

$$I_n = \{k \in \mathbb{Z} \;:\; 2^{n-1} < |k| \le 2^n\}$$

- Sobolev spaces mixed smoothness $r > 0$, integrability $1 < p < \infty$

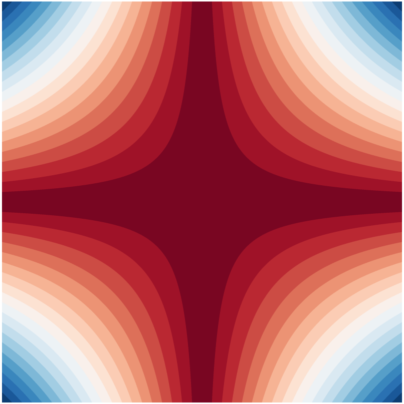
$$\mathbf{W}^r_p(\mathbb{T}^d) := \Big\{ f \in L_p(\mathbb{T}^d) : \Big\| \Big(\sum_{j \in \mathbb{N}_0^d} 2^{r|j|_1 2} \Big| \sum_{k \in I_j} \hat{f}(k) \exp(i2\pi k \cdot x) \Big|^2 \Big)^{1/2} \Big\|_p \le 1 \Big\}$$

- Besov spaces mixed smoothness $r > 0$, fine index $0 < \theta \le \infty$

$$\mathbf{B}^r_{p,\theta}(\mathbb{T}^d) := \Big\{ f \in L_p(\mathbb{T}^d) : \Big(\sum_{j \in \mathbb{N}_0^d} 2^{r|j|_1 \theta} \Big\| \sum_{k \in I_j} \hat{f}(k) \exp(i2\pi k \cdot x) \Big\|_p^\theta \Big)^{1/\theta} \le 1 \Big\}$$

- $r > 1/p$ embedding into $C(\mathbb{T}^d)$
- **Amanov, Nikolskij, Temlyakov, Schmeisser, Triebel . . .**

Hyperbolic cross projection



- Mixed Sobolev regularity

$$\|f\|^2_{\mathbf{W}^r_2} \asymp \sum_{\mathbf{k}\in\mathbb{Z}^d} |\hat{f}(\mathbf{k})|^2 \prod_{i=1}^{d} (1+|k_i|^2)^r$$

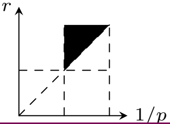
- Hyperbolic cross projection

$$P_{\mathcal{H}_n} f := \sum_{\mathbf{k}\in\mathcal{H}_n} \hat{f}(\mathbf{k}) e^{2\pi i \mathbf{k}\cdot\mathbf{x}}$$

- **Error:** $\|f - P_{\mathcal{H}_n} f\|_{L_2} \lesssim n^{-r}$
- **Cost:** $m := \sharp$ grid points in $\mathcal{H}_n$
- **Rate:** $m^{-r}(\log m)^{(d-1)r}$

Sampling recovery for Sobolev spaces

Let $1 < p < 2$ and $r > \frac{1}{p}$.



Corollary

$$\varrho_{\lceil Cd\log(d+1)n\log(n)^3\rceil}(\mathbf{W}^r_p(\mathbb{T}^d))_{L_2} \le \sigma_n(\mathbf{W}^r_p, \mathcal{T}^d)_{L_\infty}$$
$$\lesssim \left(\frac{\log(n)^{d-1}}{n}\right)^{r-\frac{1}{p}+\frac{1}{2}} \log(n)^{\frac{1}{2}-(d-1)(\frac{1}{p}-\frac{1}{2})}$$

Compare to $\varrho_n^{\mathrm{lin}}(\mathbf{W}^r_p(\mathbb{T}^d))_{L_2} \asymp \left(\frac{\log(n)^{d-1}}{n}\right)^{r-\frac{1}{p}+\frac{1}{2}}$ (Sparse grids, least squares)

Worse in the logarithm if d is large!

Effect not present for isotropic spaces, see **Heinrich '09**

Very recent results by **Feng Dai** and **V.N. Temlyakov**

improve the bound for $\varrho_n(\mathbf{W}^r_p(\mathbb{T}^d))_{L_2}$ by $\sqrt{\log n}$

Tractability

- Consequence of general Theorem: **Moeller, Stasyuk, T. Ullrich '24** considered the space $\mathbf{B}^r_{p,\theta}(\mathbb{T}^d)$, for $2 < p < \infty, 0 < \theta \le 1$ and $r = 1/\theta - 1/2$

$$\varrho_{\lceil cd^2(\log^2 d)n(\log^3 n)\rceil}(\mathbf{B}^r_{p,\theta}(\mathbb{T}^d))_{L_2} \le C_{p,\theta} d^{3/2} n^{-r} \log(dn)^{1/2} \tag{3}$$

- Compare with corresponding results for Gelfand widths in **Dirksen, T. Ullrich '18**

$$n^{-r} \lesssim c_n(\mathbf{B}^r_{p,\theta})_{L_2} \lesssim n^{-r} (\log\log n)^{r+1}$$

Information complexity vs. computational cost

- Nonlinear recovery error in terms of samples is sometimes smaller by the factor $n^{-1/2}$

Good Basis pursuit denoising needs a search space $\Lambda = [-M, M]^d$, where its size (dimension) $N = (2M+1)^d$ enters only logarithmically in the number of samples

$$m \geq C \cdot K^2 \cdot s \cdot \log(s)^3 \cdot \log(N)$$

Bad A matrix vector multiplication needs $(2M)^d$ flops and hence the **computational cost** of the recovery algorithm increases dramatically!

- J. Bourgain: An improved estimate in the restricted isometry problem
  in: Geometric Aspects of Functional Analysis, p. 65–70, 2014.
- F. Dai, V. Temylakov: Random points are good for universal discretization,
  J. Math. Anal. Appl, 529(1), 2024
- D. Dũng, V. Temlyakov, T. Ullrich: Hyperbolic Cross Approximation,
  Birkhäuser/Springer, Cham, 2018.
- S. Foucart, H. Rauhut: A Mathematical Introduction to Compressive Sensing,
  Birkhäuser/Springer, New York, NY, 2013.
- I. Haviv, O. Regev: The restricted isometry property of subsampled Fourier matrices
  in: Geometric Aspects of Functional Analysis, p. 163–179, 2017.
- T. Jahn, T. Ullrich, F. Voigtlaender: Sampling numbers of smoothness classes via ℓ_1-mi
  in: J. Complexity, 79, 2023.
- D. Krieg: Tractability of sampling numbers on unweighted function classes.
  in: Proc. Amer. Math. Soc. 11, 115–125. 2024
- E. Novak, H. Woźniakowski: Tractability of Multivariate Problems. Vol. 1: Linear Inforr
  European Mathematical Society, Zürich, 2008.
- H. Rauhut, R. Ward: Sparse Legendre expansions via ℓ_1-minimization,
  J. Approx. Theory 164(5):517–533, 2012.

Thank you for your attention!

Control the quasi-projection

Let $\kappa, n \in \mathbb{N}$, $J, J^* \subset I$ and $\tau > 0$.

A linear operator $P : L_2 \to L_2$ is called a $(\kappa, n, J, J^*, \tau)$ **quasi-projection** if

- $P(\Sigma_n) \subset \Sigma_{\kappa n}$,
- $Pf = f$ for all $f \in V_J$,
- $Pf \in V_{J^*}$ for all $f \in L_2$,
- $P : L_\infty \to L_\infty$ is well-defined and $\|P\|_{L_\infty \to L_\infty} \leq \tau$.

Filbir, Temistoclacis 2004: If $\mathcal{B}$ is the Fourier basis or an OPS, take de la Vallée Poussin operators.