Sparse polynomial approximation for surrogate modeling on networks

Giuseppe Alessio D'Inverno
LMO, Université Paris-Saclay

Northernmost GraphML Group
05/06/2026

université PARIS-SACLAY

Collaborators

- Simone Brugiapaglia, Concordia University
- Iacopo Iacopini, NU London
- Kylian Ajavon, Concordia University

Motivation

- Differential models on complex systems are at the core of many biological phenomena

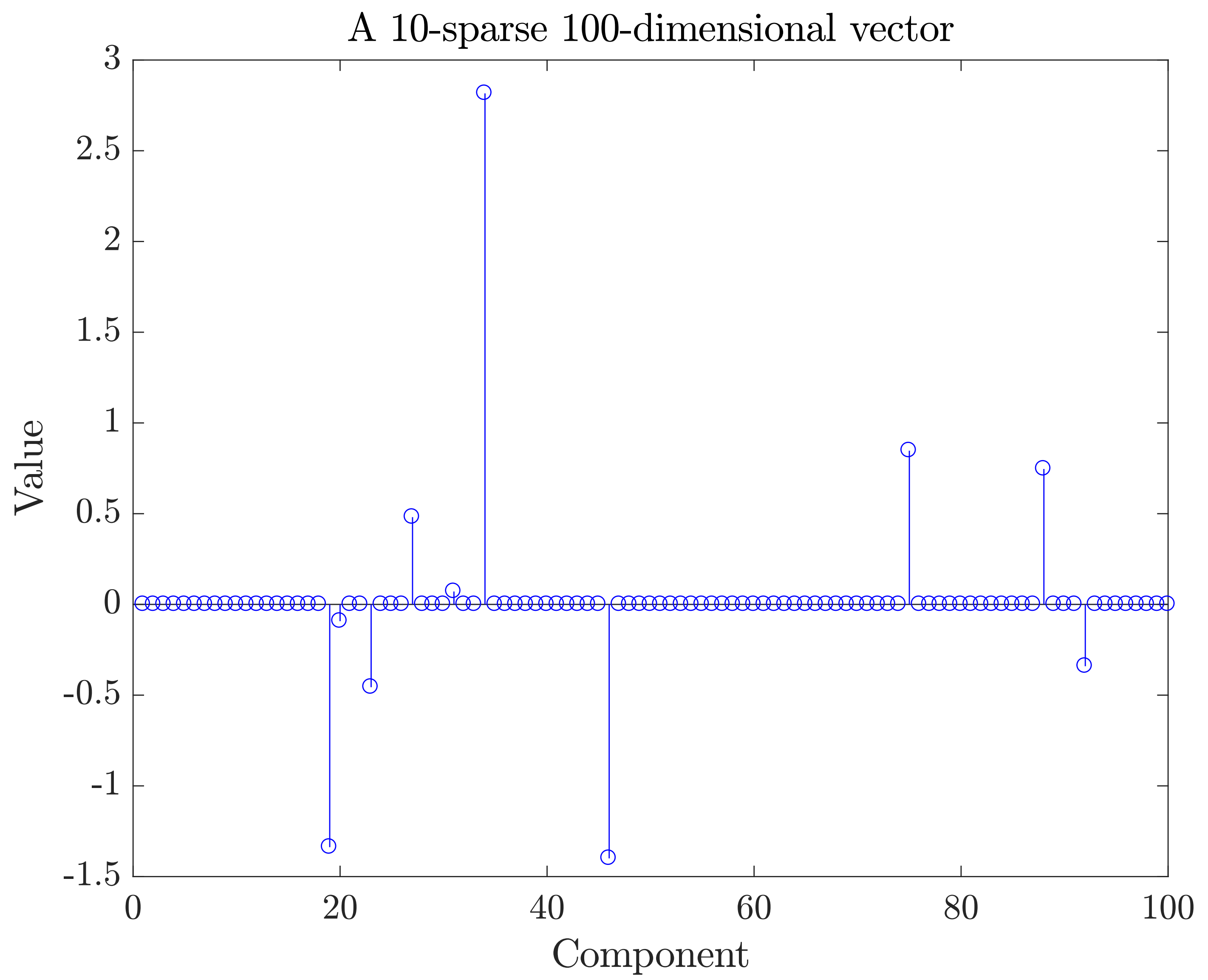
- However, running simulations of such models for wide ranges of system parameters requires prohibitive computational time

- Deep learning is empirically convenient in terms of accuracy and inference time - but often lacks of ***theoretical guarantees***

→ alternative: theoretically-grounded nonlinear methods

- Guiding principle of this talk: **sparsity**

Sparsity



- **Sparsity** is measured by the ℓ^0-norm

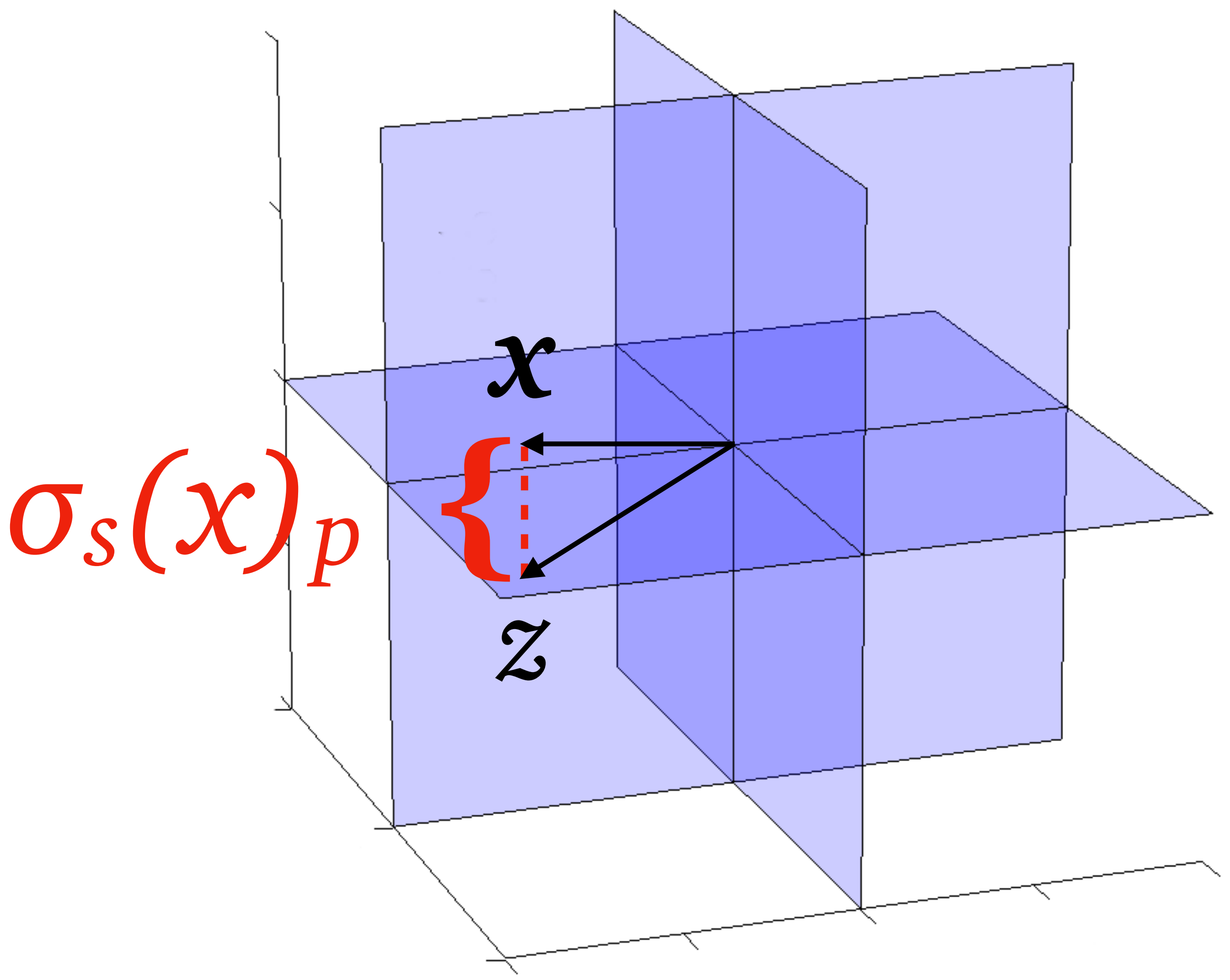
$$\|\mathbf{x}\|_0 = |\{j : x_j \neq 0\}|.$$

- A vector $\mathbf{x}$ is ***s*-sparse** if $\|\mathbf{x}\|_0 \leq s$.

- The best s-term approximation error measures the **compressibility** of a vector

$$\sigma_s(\mathbf{x})_p = \min_{\mathbf{z}:\ \|\mathbf{z}\|_0 \leq s} \|\mathbf{z} - \mathbf{x}\|_p$$

- Typically, we consider signals $\mathbf{\Psi x}$ with $\mathbf{x}$ sparse or compressible



2-sparse vectors in $\mathbb{R}^3$

*This slide is courtesy of Simone Brugiapaglia

Sparse polynomial approximation

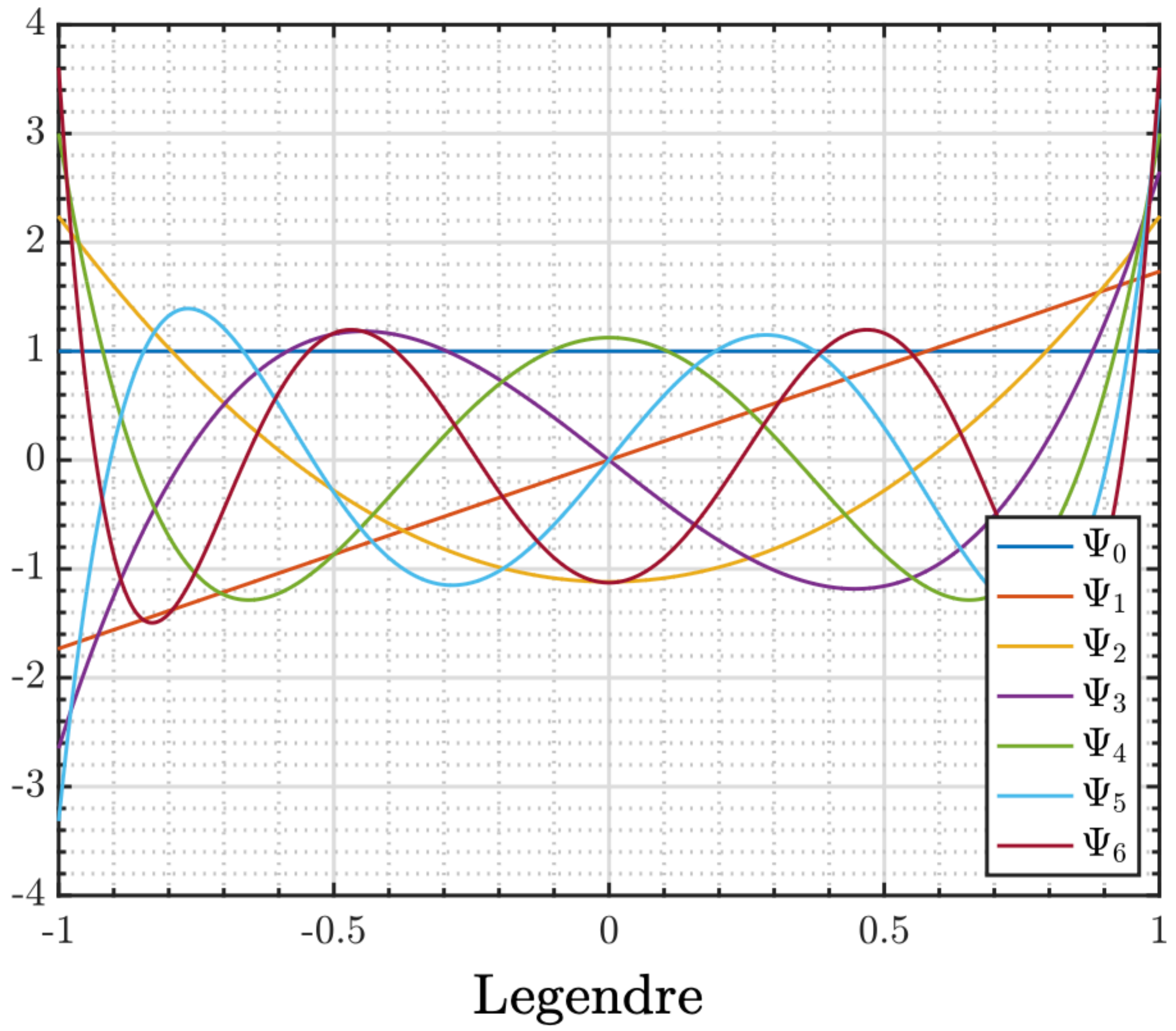
- **Problem:** Approximate $f : [-1,1]^d \to \mathbb{R}$ from $f(x_1), \ldots, f(x_m)$
- Typical applications: Surrogate modeling, uncertainty quantification
- **Goal:** compute $\hat{f} = \sum_{\nu \in \mathbb{N}_0^d} \hat{c}_\nu \Psi_\nu \approx f$ with $\|\hat{\mathbf{c}}\|_0 = \#\{\hat{c}_\nu \neq 0\}$;

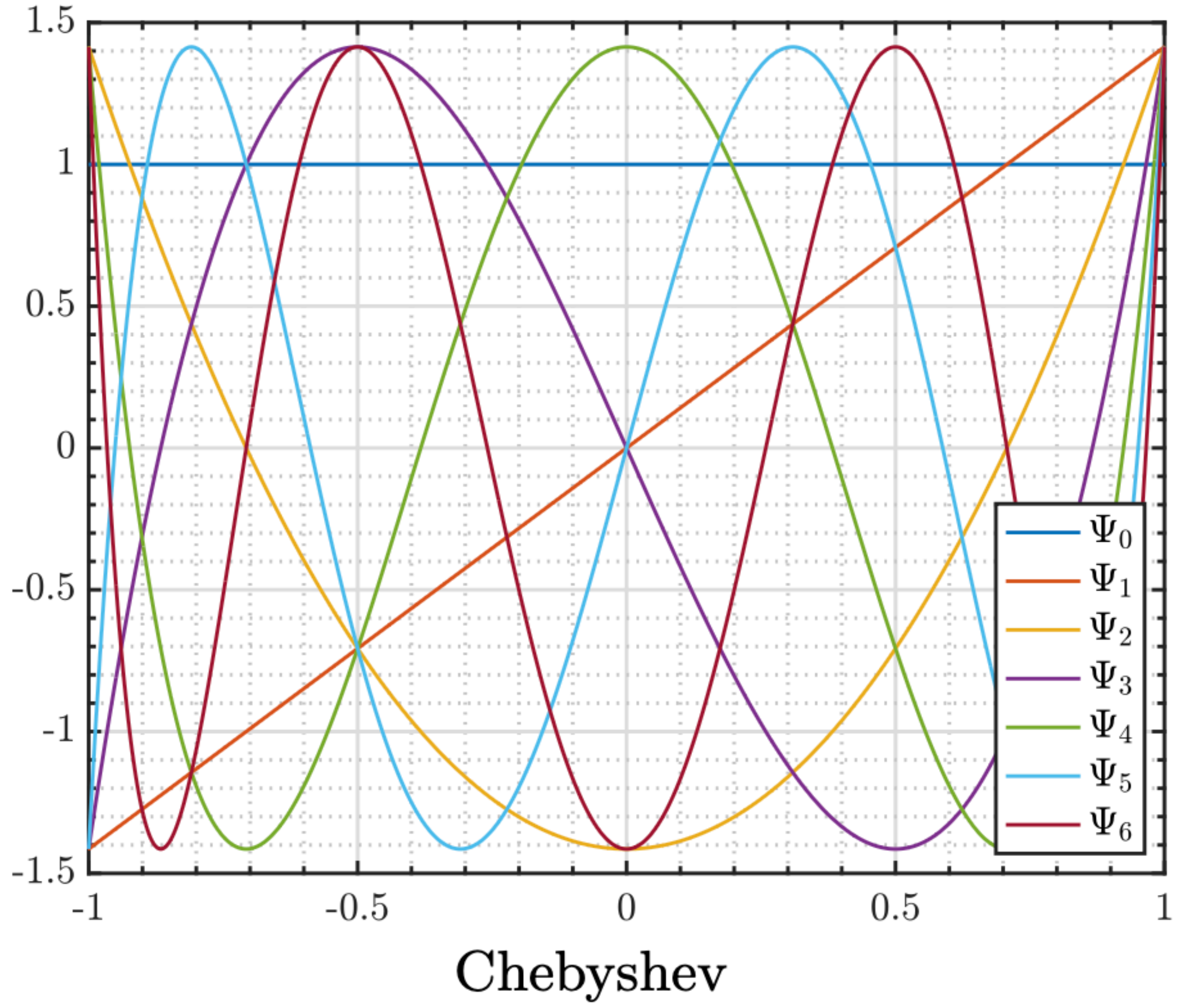
(NB: the nonlinearity comes from the sparsity requirement: $\|\hat{\mathbf{c}}\|_0 \leq s$)

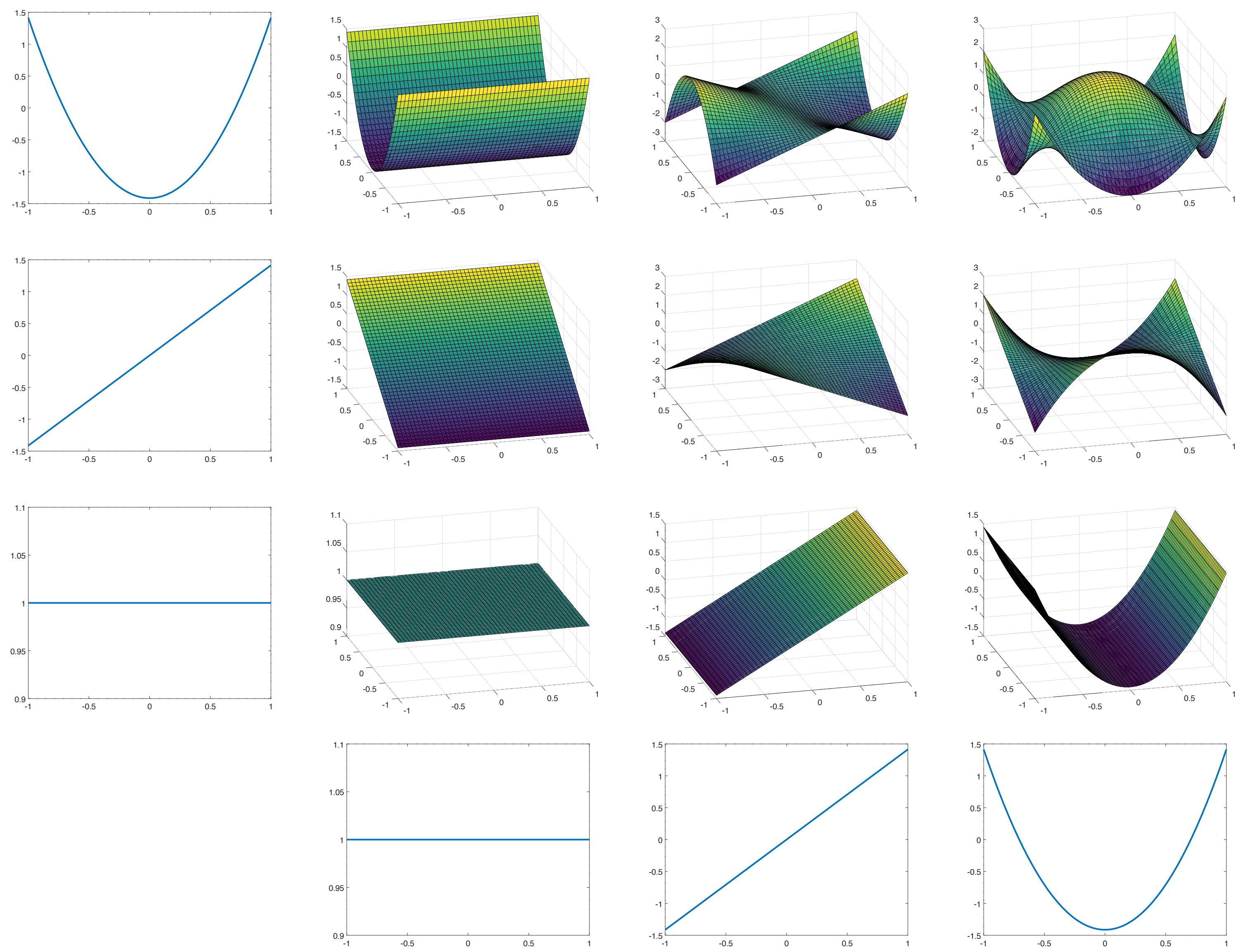
- Three main ingredients:
  - Polynomial basis
  - Multi-index sets
  - Approximation methods from pointwise data

Ingredient #1: the polynomial basis

- Orthogonal basis in L^2_ρ with respect to the sampling probability measure ρ
- Common options: **Legendre** polynomials, **Chebyshev** polynomials (and their associated tensor-product versions)



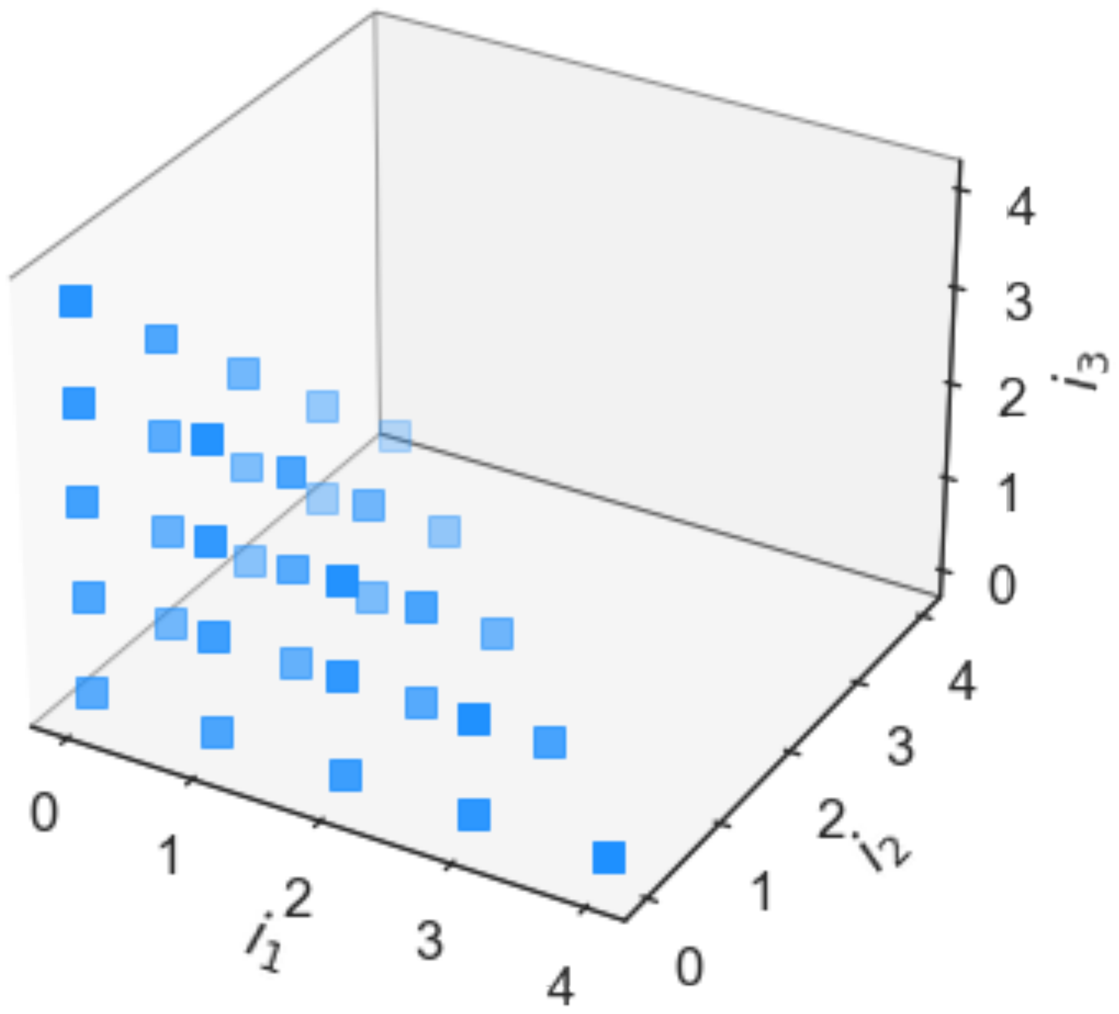


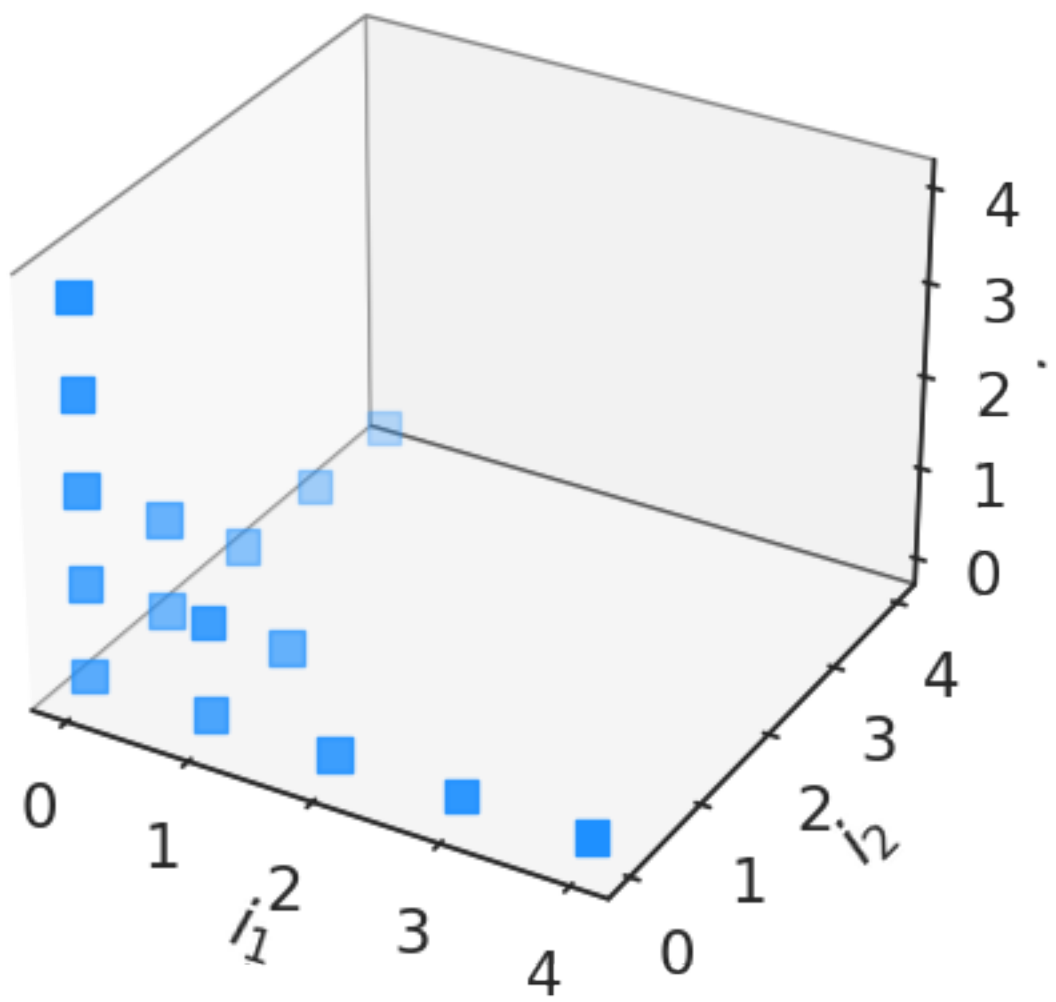


Ingredient #2: Multi-index sets

- Infinite search space $\mathbb{N}_0^d$ of polynomial coefficient indices ν - how to restrict it in a smart way?
- **Lower sets** are a good option for orthogonal polynomials
  - Intuitive idea: if some polynomial coefficient c_ν of a function f is large at some multi-index ν, we can expect that c_μ with $\mu \leq \nu$ will also be large
- Good news: the best s-term approximation error in lower sets decays as $\sigma_s(\mathbf{x})_p$
- Drawback: lower sets usually grows exponentially (or log-exponentially) with the dimension d of the function domain / parameter space

Ingredient #2: Multi-index sets





$$\Lambda_n^{\mathrm{TD}} := \left\{ \nu = (\nu_k)_{k=1}^{d} \in \mathbb{N}_0^d : \sum_{k=1}^{d} \nu_k \leq n \right\}$$

$$\Lambda_n^{\mathrm{HC}} := \left\{ \nu = (\nu_k)_{k=1}^{d} \in \mathbb{N}_0^d : \prod_{k=1}^{d} (\nu_k + 1) \leq n + 1 \right\}$$

Ingredient #3: Approximation methods

- **Least squares**. the target multi-index set S is known:

$$\min_{\mathbf{z}\in\mathbb{R}^s} \|\boldsymbol{\Psi}\mathbf{z} - \mathbf{b}\|_2^2$$

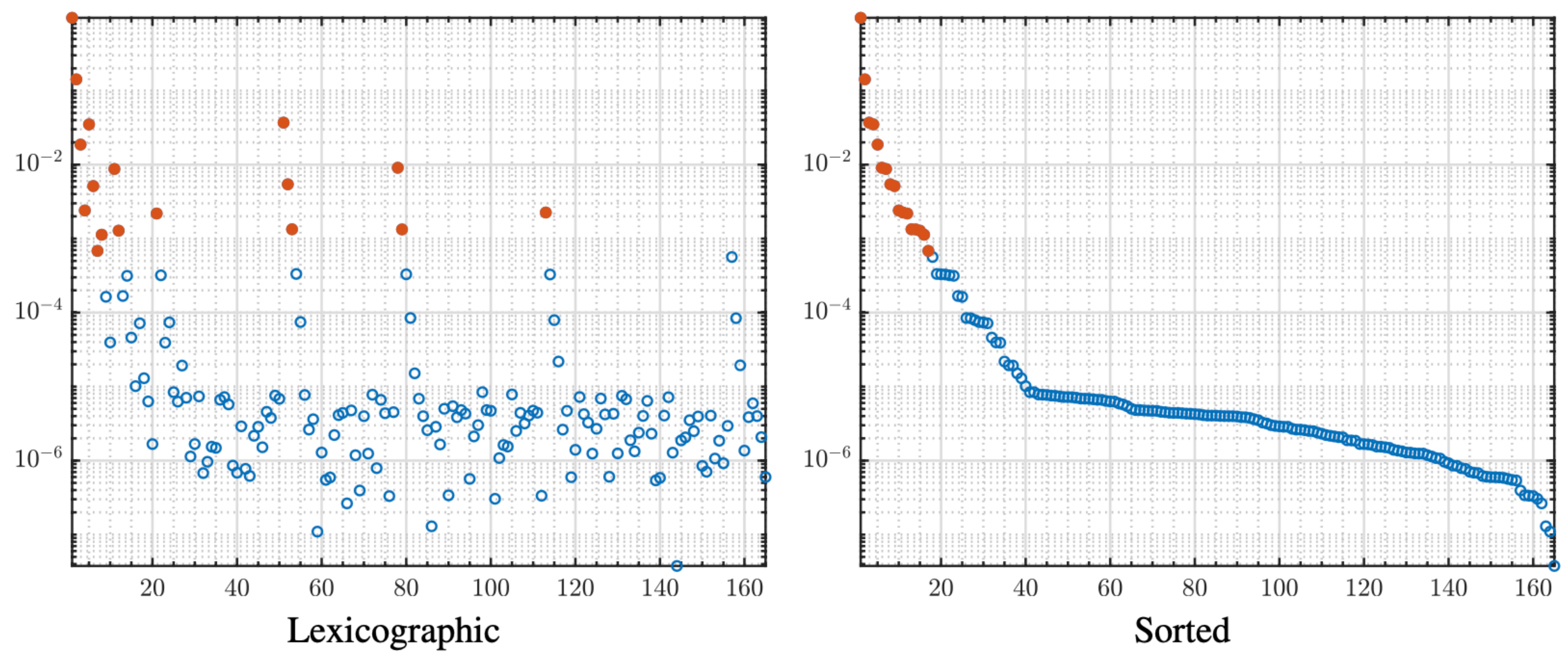
- **Compressed sensing.** The set S is not known a priori $\to$ we choose $\Lambda \subseteq \mathbb{N}_0^d$ that contains S.
  - **Square Root LASSO (SR-LASSO)** ℓ^1-minimization problem:

$$\min_{\mathbf{z}\in\mathbb{C}^N} \lambda\|\mathbf{z}\|_1 + \|\boldsymbol{\Psi}\mathbf{z} - \mathbf{b}\|_2$$

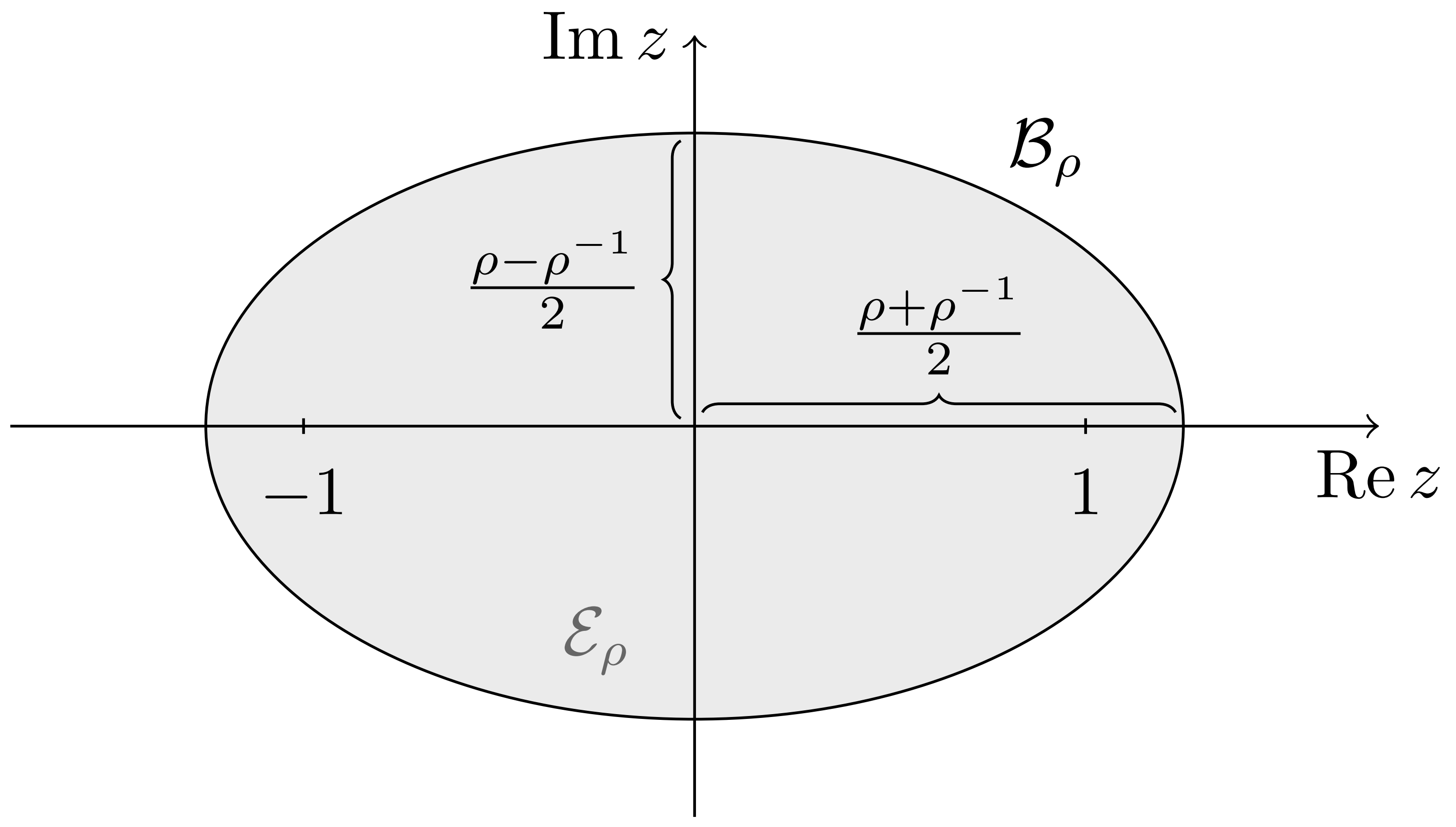
  - **Quadratically Constrained Basis Pursuit (QCBP)** ℓ^1-minimization problem:

$$\min_{\mathbf{z}\in\mathbb{C}^N} \|\mathbf{z}\|_1 \quad \text{subject to } \|\boldsymbol{\Psi}\mathbf{z} - \mathbf{b}\|_2 \leq \eta$$

Holomorphy meets sparsity



Coefficient decay of a holomorphic function ($d = 8$)



Bernstein's ellipse

If f is holomorphic in a **Bernstein's polyellipse** $\mathcal{E}_\rho := \mathcal{E}_{\rho_1} \times \cdots \times \mathcal{E}_{\rho_d} \subset \mathbb{C}^d$, then

$$\underbrace{\inf_{\|z\|_0 \le s} \left\| f - \sum_{\nu \in \mathbb{N}_0^d} z_\nu \Psi_\nu \right\|_{L^2}}_{\text{best } s\text{-term approx. error}} \lesssim \|f\|_{L^\infty(\mathcal{E}_\rho)} \cdot \exp\left(-\gamma \cdot s^{1/d}\right).$$

*This slide is courtesy of Simone Brugiapaglia

Holomorphy meets sparsity

Holomorphy is a reasonable assumption for many applications:

- Parametric PDEs (e.g. Harmonic oscillator, heat equation, parametric domains)

Example: parametric diffusion equation

$$\begin{cases} \nabla \cdot (a_\mu(x)\nabla u(x)) = g(x) & x \in \Omega \\ u(x) = h(x) & x \in \partial\Omega \end{cases}$$

- Parametric ODEs on graphs (focus of this talk)

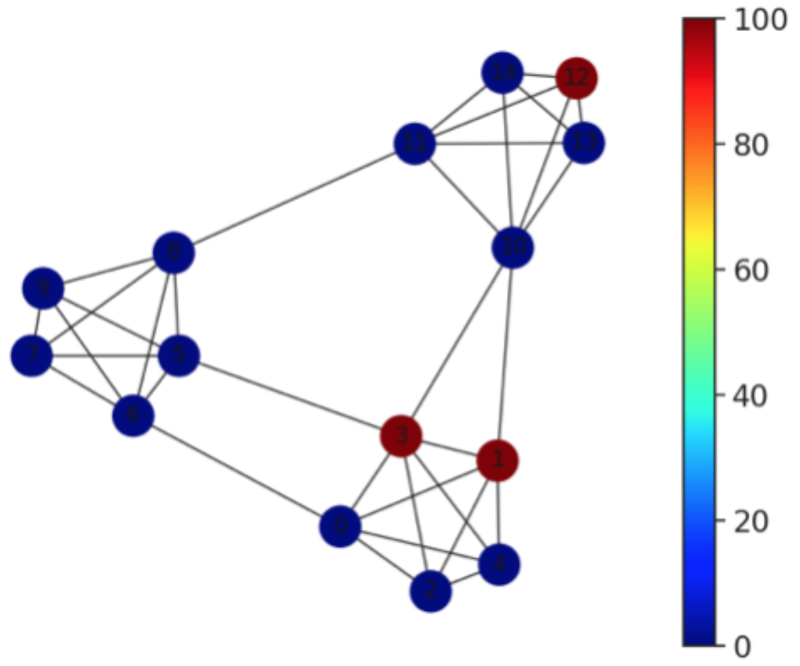
Sparse polynomials for diffusion on graphs

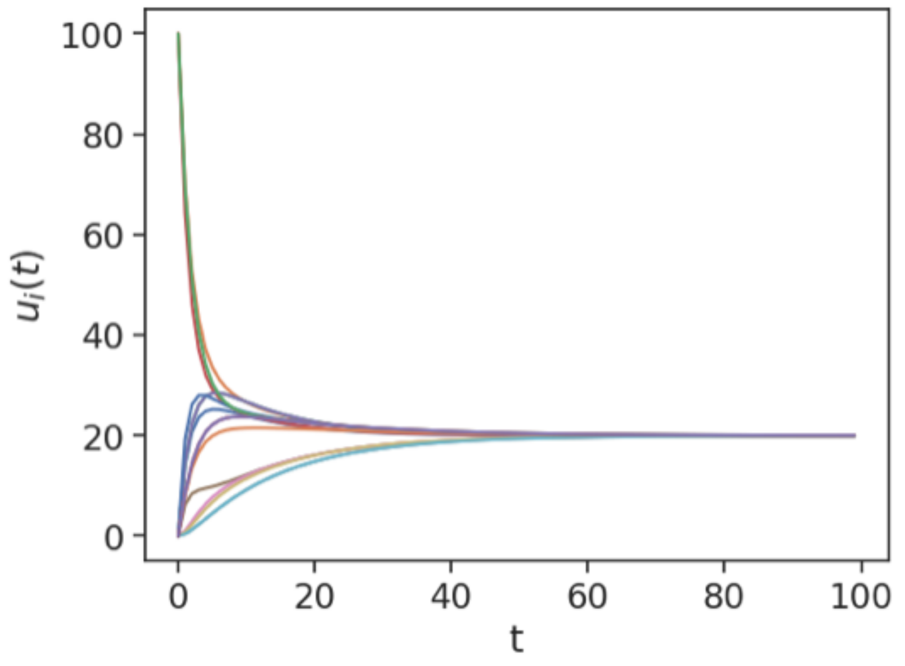
Diffusion on graphs

- Diffusion equation on graphs:

$$\dot{\mathbf{u}}(t) = \mathbf{M}(t)\mathbf{u}(t), t \in [0,T]$$

where $\mathbf{M}(t) = \mathbf{C}(t) \odot \mathbf{W}(t) - \mathbf{D}(\mathbf{C}(t), \mathbf{W}(t))$ with $\mathbf{C} : [0,T] \times \mathscr{U} \to \mathbb{C}^{|V|\times|V|}$ being the time-dependent diffusion coefficient matrix and $\mathbf{W}(t)$ the continuous time-dependent adjacency matrix.





Goal: find an approximation of the parameter-to-QoI map $f_v(\mathbf{z}) = \mathbf{u}_v(t, \mathbf{z})$ for a fixed node v at fixed timestep t.

Holomorphy assumption

Assumption 1

There exist a compact set $\mathscr{K} \subset \mathbb{C}^d$, with $\mathscr{U} \subset \mathscr{K}$, and a map $\tilde{\mathbf{C}} : [0,T] \times \mathscr{K} \to \mathbb{C}^{|V|\times|V|}$ such that $\tilde{\mathbf{C}}|_{[0,T]\times\mathscr{U}} = \mathbf{C}$, $\tilde{\mathbf{C}}$ is continuous in $[0,T] \times \mathscr{K}$ and $\tilde{\mathbf{C}}(t,\cdot)$ is holomorphic in $\mathring{\mathscr{K}}$ for every $t \in [0,T]$. Moreover, $\mathbf{C}(t)_{uv}$ is constant for every pair of nodes u, v belonging to the same community.

Example:

$$\mathbf{C}(t,\mathbf{y}) = \begin{pmatrix} \frac{y_{\sigma(1,1)}+1}{2} \cdot h_{\sigma(1,1)}(t)\mathbf{1} & \cdots & \frac{y_{\sigma(1,K)}+1}{2} \cdot h_{\sigma(1,K)}(t)\mathbf{1} \\ \vdots & \ddots & \vdots \\ \frac{y_{\sigma(K,1)}+1}{2} \cdot h_{\sigma(K,1)}(t)\mathbf{1} & \cdots & \frac{y_{\sigma(K,K)}+1}{2} \cdot h_{\sigma(K,K)}(t)\mathbf{1} \end{pmatrix}$$

Main results

Theorem

Let $T > 0$ and let $\mathbf{C}$ satisfy the holomorphy assumption for some compact $\mathscr{K} \subseteq \mathbb{C}^d$ such that $\mathscr{U} \subset \mathscr{K}$. Then the map $\mathbf{f} : \mathscr{U} = [-1,1]^d \to \mathbb{C}^{|V|}$ defined by $\mathbf{f}(\mathbf{y}) = \mathbf{u}(T, \mathbf{y})$ satisfying the diffusion equation on graphs admits a holomorphic extension $\tilde{\mathbf{f}}$ to $\mathring{\mathscr{K}}$. Moreover, for any compact subset $\mathscr{H} \subset \mathring{\mathscr{K}}$, we have

$$\|\tilde{\mathbf{f}}\|_{L^\infty(\mathscr{H}):=\sup_{\mathbf{z}\in\mathscr{H}} \|\tilde{\mathbf{f}}(\mathbf{z})\|_2} \leq B(\mathbf{u}_0, T, \tilde{\mathbf{M}}, \mathscr{H}),$$

where $B(\mathbf{u}_0, T, \tilde{\mathbf{M}}, \mathscr{H}) = \|\mathbf{u}_0\|_2 \cdot \exp\left(2 \cdot \int_0^T \sup_{\mathbf{z}\in\mathscr{H}} \|\tilde{\mathbf{M}}(t, \mathbf{z})\|_{2\to 2}\, dt\right) < \infty.$

Proof sketch:

- Apply Volterra's Lemma after checking that the hypotheses hold
- The bound is proved by means of Gronwall's lemma.

Main results

Corollary 1 (informal)

Under the hypothesis of the Theorem, we have that $\sigma_s(\mathbf{c})_q \leq \dfrac{B(u_0, T, \mathbf{M}) \cdot C(p, \rho, d)}{(s+1)^{\frac{1}{p}-\frac{1}{q}}}$.

Corollary 2 (informal)

Under the hypothesis of the Theorem, for $0 < p < 1$, the minimizer $\hat{f}$ of the LS objective satisfies with high probability

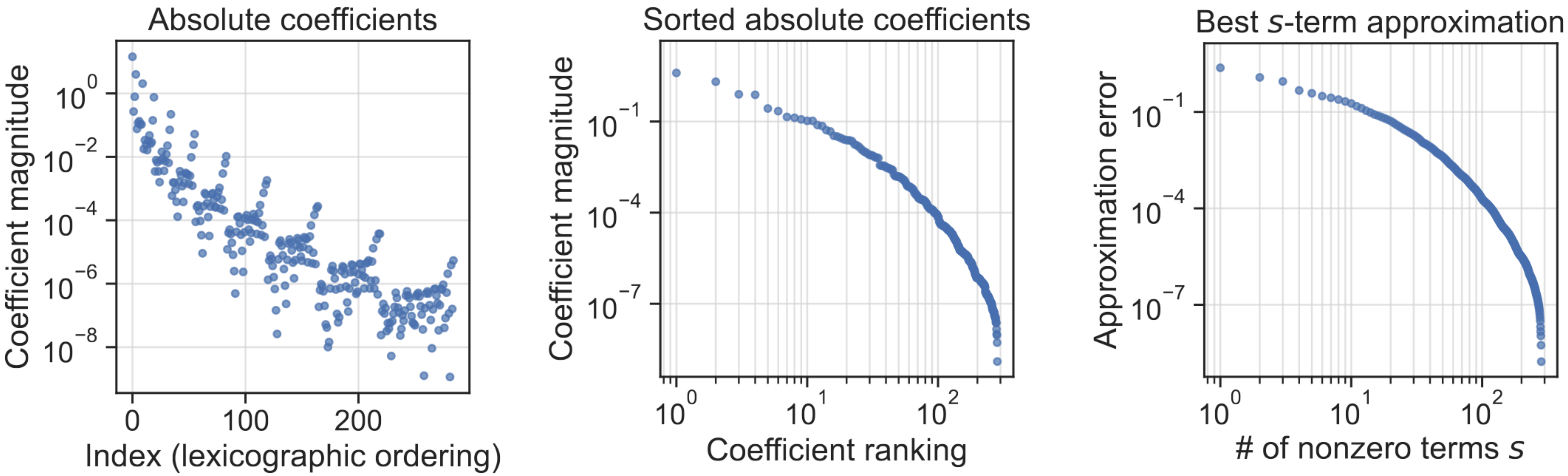
$$\|f_v - \hat{f}_v\|_{L^2_\varrho(\mathcal{U})} \leq C(\rho, p) \cdot (m/\log(m/\epsilon))^{\frac{1}{2}-\frac{1}{p}}.$$

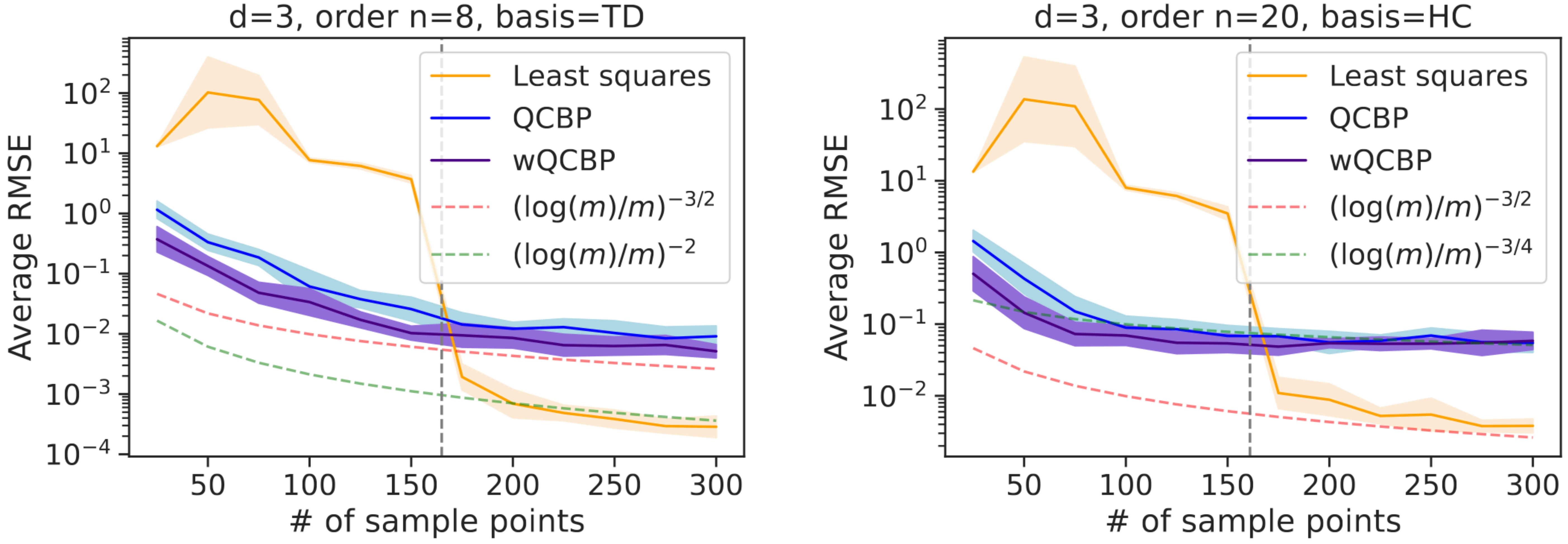
Corollary 3 (informal)

Under the hypothesis of the Theorem, for $0 < p < 1$, the minimizer $\hat{f}$ of the SR-LASSO objective satisfies with high probability

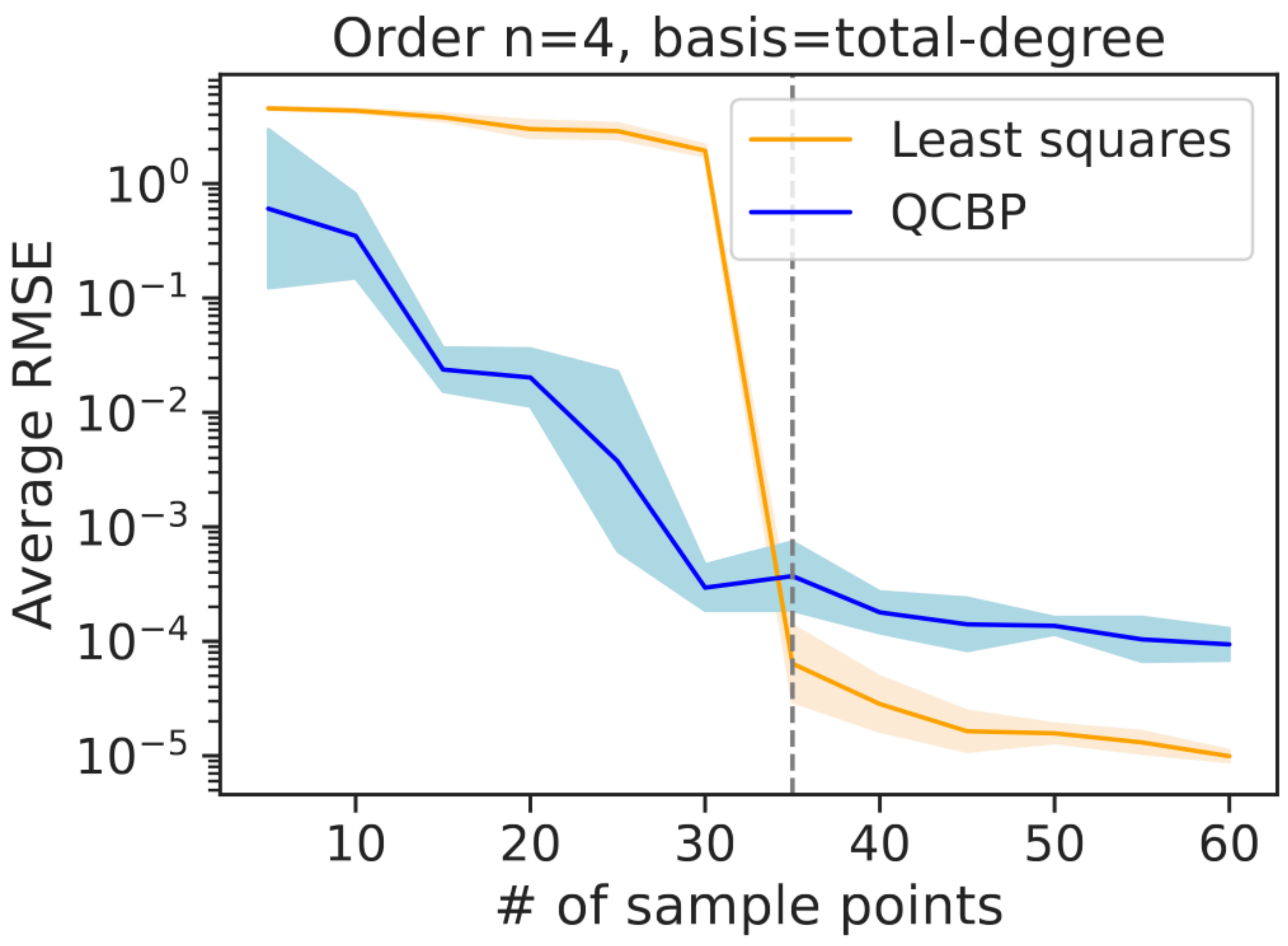
$$\|f_v - \hat{f}_v\|_{L^2_\varrho(\mathcal{U})} \leq \kappa_1 \cdot B(u_0, T, \rho, \mathbf{M}) \cdot \left(C_1 + C_2\right) \cdot \tilde{m}^{\frac{1}{2}-\frac{1}{p}}$$

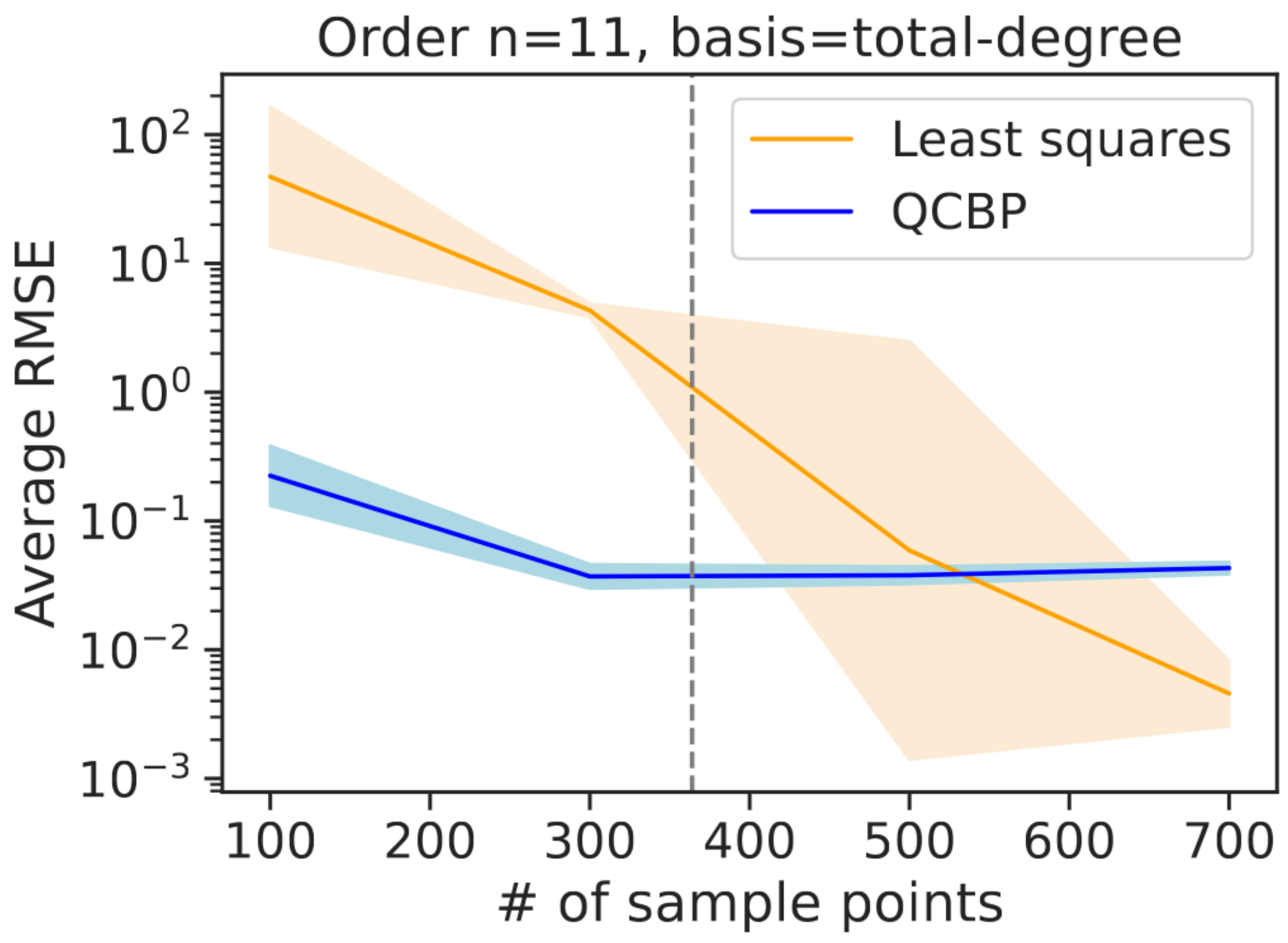
Numerical results - time-independent SBM





Numerical results - real-life datasets





Sparse polynomials on models for complex networks
(work in progress)

The Kuramoto-Sakaguchi model

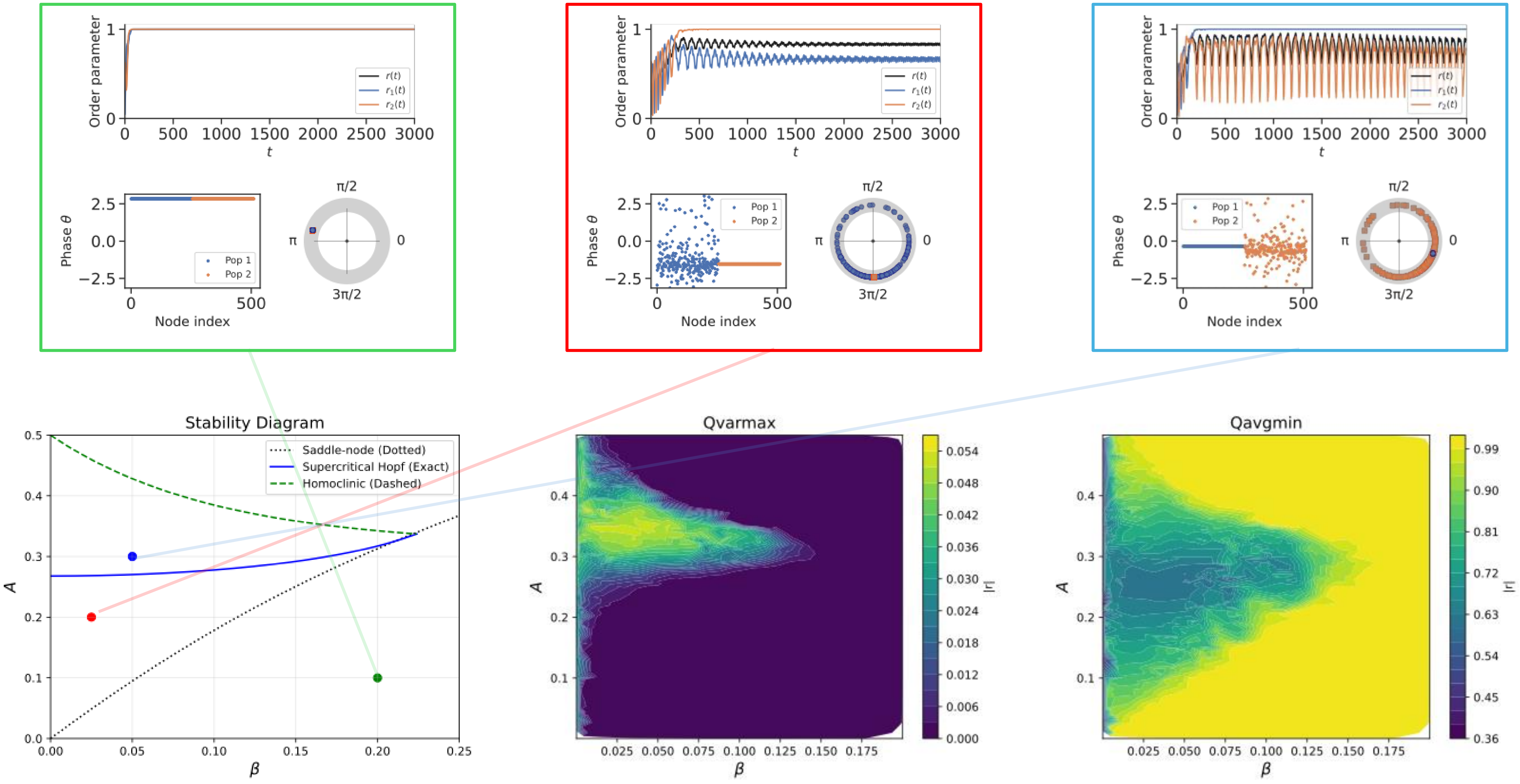
Model that describes the synchronization behavior of two population of oscillators:

$$\frac{d\theta_i^\sigma}{dt} = \omega + \sum_{\sigma'=1}^{2} \frac{K_{\sigma\sigma'}}{N_{\sigma'}} \sum_{j=1}^{N_{\sigma'}} \sin(\theta_j^{\sigma'} - \theta_i^\sigma - \alpha)$$

- Assumptions: $K_{11} = K_{22} = \mu > 0$, $K_{12} = K_{21} = \nu > 0$, with $\mu + \nu = 1$, $A = \mu - \nu$ and $\beta = \pi/2 - \alpha$
- Synchronization quantified by the **order parameter:** $Re^{i\psi(t)} = \frac{1}{N} \sum_{j=1}^{N} e^{i\theta_j(t)}$
- Abrams et al. (2008), Panaggio et al. (2016): characterization of **chimera states** in the $A - \beta$ plane
- Goal: detect the chimera states with sparse polynomials (computing QoIs: temporal variance and temporal mean of the community-wise order parameter)

Abrams, D. M., Mirollo, R., Strogatz, S. H., & Wiley, D. A. (2008). Solvable model for chimera states of coupled oscillators. *Physical review letters*, *101*(8), 084103.
Panaggio, M. J., Abrams, D. M., Ashwin, P., & Laing, C. R. (2016). Chimera states in networks of phase oscillators: the case of two small populations. *Physical Review E*, *93*(1), 012218.

The Kuramoto-Sakaguchi model



SIS epidemic dynamics on networks with parameterized degree heterogeneity

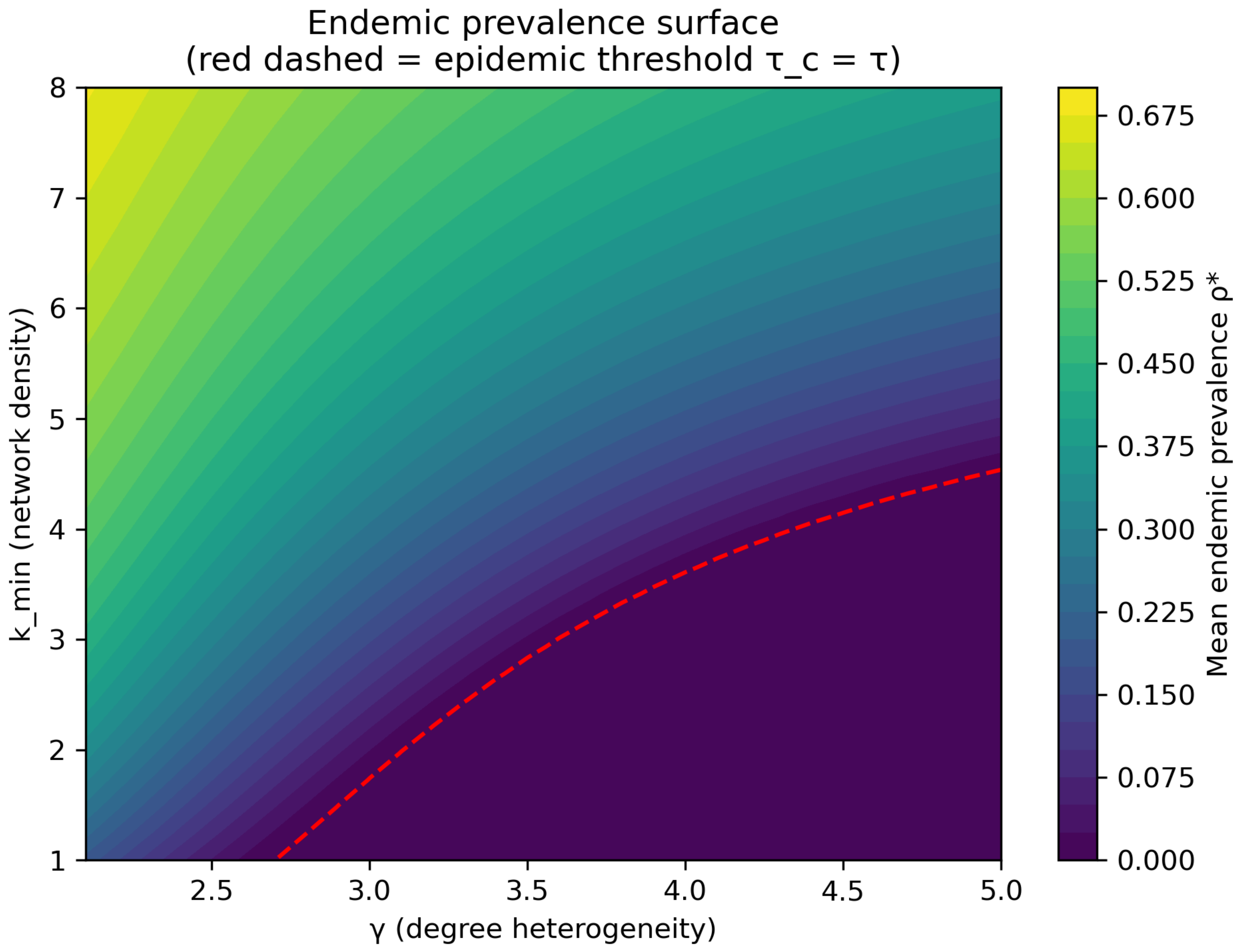
N-intertwined mean-field SIS model on weighted networks [Van Mieghem, 2008]

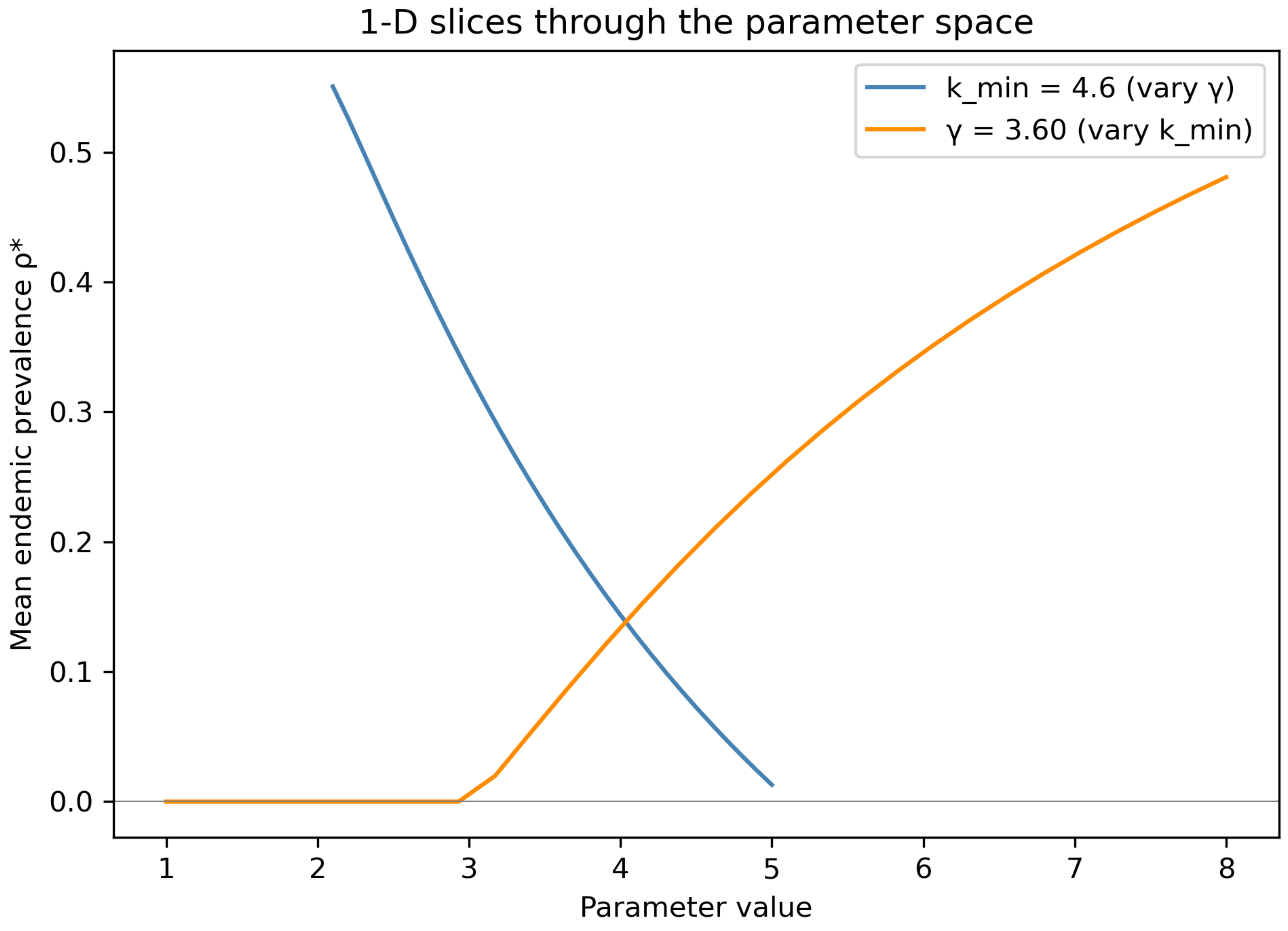
$$\dot{x}_i = -\delta x_i + \beta (1 - x_i) \sum_{j=1}^{N} P_{ij}(\gamma, k_{\min})\, x_j, \qquad i = 1, \ldots, N,$$

- $x_i(t) \in [0,1]$: infection probability at node i; $\beta > 0$: transmission rate per contact; $\delta > 0$ recovery rate.
- If $\tau = \beta/\delta < \tau_c$: disease-free equilibrium $\mathbf{x} = \mathbf{0}$ (globally stable); $\tau > \tau_c$: unique endemic equilibrium $\mathbf{x}^* > \mathbf{0}$ (globally asymptotically stable)
- $\mathbf{P}(\gamma, k_{\min})$ encodes the network structure of the *Chung-Lu model*: $P_{ij}(\mathbf{w}) = \min\left(\frac{w_i w_j}{\sum_{l=1}^{N} w_l}, 1 \right)$ with $w_i(\gamma, k_{\min}) = k_{\min} \cdot \left(\frac{N}{i} \right)^{1/(\gamma-1)}$

SIS epidemic dynamics on networks with parameterized degree heterogeneity

Goal: recover the endemic prevalence trend from $(\gamma, k_{\min})$





Main takeaways

- Sparse polynomials: powerful and theoretically sound tool for function approximation from samples
- Holomorphy: necessary but reasonable assumption → fits well in ODEs on graphs
- Theoretical guarantees on convergence rates

Perspectives:

- Study of models on complex networks with high-dimensional parameterization
- Sparse polynomials can help in theoretical analysis of deep learning models (through the so-called ***practical existence theorems***) → convergence rates for GNNs?

Main takeaway (for me)

Thank you for the attention!

References:

- Adcock, Brugiapaglia, Webster, ***SIAM,*** 2022:
  Sparse Polynomial Approximation of High-Dimensional Functions
- GAD, Ajavon, Brugiapaglia, ***Springer J. Sci. Comp.***, 2026 (in press):
  Surrogate models for Diffusions on graphs via sparse polynomials

Preprint

Backup slides: Volterra integral Lemma

Lemma 1 (Holomorphic dependence; Volterra integral equations) *Let $T > 0$, $n, d \in \mathbb{N}$, $\mathcal{K} \subset \mathbb{C}^d$ be a compact set, and $\boldsymbol{g} : [0,T] \times \mathcal{K} \to \mathbb{C}^n$ and $\boldsymbol{h} : [0,T]^2 \times \mathbb{C}^n \times \mathcal{K} \to \mathbb{C}^n$ be continuous functions. Moreover, suppose that there exists a constant $L > 0$ such that $\boldsymbol{h}$ satisfies the Lipschitz condition*

$$\|\boldsymbol{h}(t,s,\boldsymbol{u},\boldsymbol{z}) - \boldsymbol{h}(t,s,\boldsymbol{v},\boldsymbol{z})\| \leq L\|\boldsymbol{u}-\boldsymbol{v}\|, \quad \forall t,s \in [0,T], \quad \forall \boldsymbol{u},\boldsymbol{v} \in \mathbb{C}^n, \forall \boldsymbol{z} \in \mathcal{K}, \tag{20}$$

for some norm $\|\cdot\|$ over $\mathbb{C}^n$. Then the Volterra integral equation

$$\boldsymbol{u}(t,\boldsymbol{z}) = \boldsymbol{g}(t,\boldsymbol{z}) + \int_0^t \boldsymbol{h}(t,s,\boldsymbol{u}(s,\boldsymbol{z}),\boldsymbol{z})ds, \quad \forall t \in [0,T],$$

admits a unique solution $t \mapsto \boldsymbol{u}(t,\boldsymbol{z})$ for every $\boldsymbol{z} \in \mathcal{K}$ and the mapping $(t,\boldsymbol{z}) \mapsto \boldsymbol{u}(t,\boldsymbol{z})$ is continuous in $[0,T] \times \mathcal{K}$. In addition, let $\mathring{\mathcal{K}}$ be the interior of $\mathcal{K}$ and assume that $\boldsymbol{z} \mapsto \boldsymbol{g}(t,\boldsymbol{z})$ is holomorphic in $\mathring{\mathcal{K}}$ for any fixed $t \in [0,T]$ and $(\boldsymbol{v},\boldsymbol{z}) \mapsto \boldsymbol{h}(t,s,\boldsymbol{v},\boldsymbol{z})$ is holomorphic in $\mathbb{C}^n \times \mathring{\mathcal{K}}$ for any fixed $(t,s) \in [0,T]^2$. Then $\boldsymbol{z} \mapsto \boldsymbol{u}(t,\boldsymbol{z})$ is holomorphic in $\mathring{\mathcal{K}}$ for any fixed $t \in [0,T]$.