# Trigonometric Approximations of the Sparse Super-Resolution Problem in Wasserstein Distances

Paul Catala, University of Osnabrück.
Joint work with M. Hockmann, S. Kunis and M. Wageringel

GAMM 2023, Dresden, 30.05.23 - 2.06.23

# Sparse Super-Resolution

- **Problem:** Recover a signal from a few coarse linear measurements

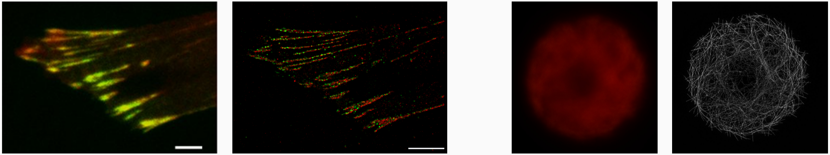
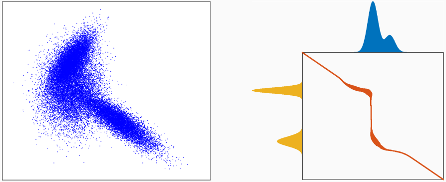


**Image Processing:** Fluorescence microcospy (source - `www.cellimagelibrary.org`), astronomical imaging, . . .



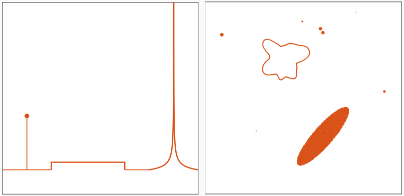
**Machine Learning:** Mixture estimation, optimal transport, . . .

- **Signals of interest are often structured:** pointwise sources, curves, graphs of functions...

# Data Model

- **Radon measures**
  $d \in \mathbb{N} \setminus \{0\}$, $\mathbb{T} \overset{\text{def.}}{=} \mathbb{R}/\mathbb{Z}$ (Torus),

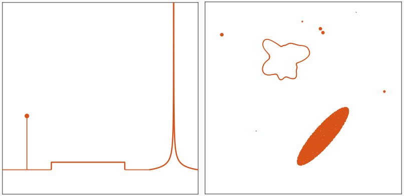
$$\mu \in \mathcal{M}(\mathbb{T}^d)$$



Singular measures $\mu$

- **Radon measures**
  $d \in \mathbb{N} \setminus \{0\}$, $\mathbb{T} \overset{\text{def.}}{=} \mathbb{R}/\mathbb{Z}$ (Torus),

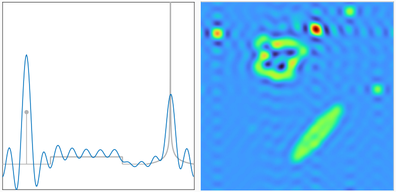
$$\mu \in \mathcal{M}(\mathbb{T}^d)$$



Singular measures $\mu$

- **Trigonometric moments**
  $k \in \Omega \subset \mathbb{Z}^d$, here $\mathbb{Z}_n^d = \{-n, \ldots, n\}^d$

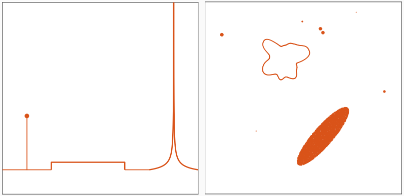
$$\hat{\mu}(k) \overset{\text{def.}}{=} \int_{\mathbb{T}^d} e^{-2\imath\pi \langle k, x \rangle} \mathrm{d}\mu(x)$$



Fourier partial sum $S_n\mu$ ($n = 13$)

- **Radon measures**
  $d \in \mathbb{N} \setminus \{0\}$, $\mathbb{T} \overset{\text{def.}}{=} \mathbb{R}/\mathbb{Z}$ (Torus),

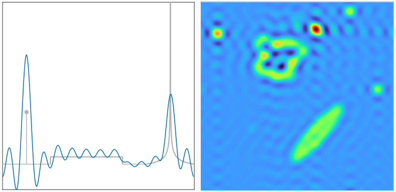
$$\mu \in \mathcal{M}(\mathbb{T}^d)$$



Singular measures $\mu$

- **Trigonometric moments**
  $k \in \Omega \subset \mathbb{Z}^d$, here $\mathbb{Z}_n^d = \{-n, \dots, n\}^d$

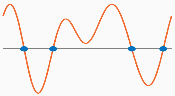
$$\hat{\mu}(k) \overset{\text{def.}}{=} \int_{\mathbb{T}^d} e^{-2\imath\pi\langle k, x\rangle} \mathrm{d}\mu(x)$$

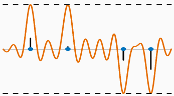


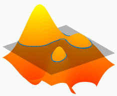
Fourier partial sum $S_n\mu$ $(n = 13)$

?

# Recovery Approaches

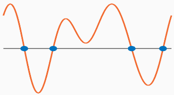

Prony


BLASSO (Dual)

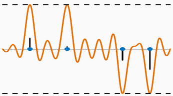

FRI

- Discrete support: exact recovery
  - Prony's method [R. de Prony, 1795], and subspace methods: ESPRIT [Roy and Kailath, 1989], MUSIC [Schmidt, 1986], matrix pencils [Hua and Sarkar, 1989], …
  - Off-the-grid optimization [Candès and Fernandez-Granda, 2014]
- Structured support: exact recovery
  - Finite Rate of Innovation [Pan, Blu, and Dragotti, 2014]
  - Super-resolution of lines [Polisano et al., 2017]

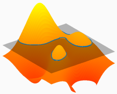
# Recovery Approaches



Prony



BLASSO (Dual)

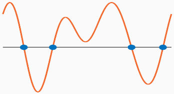


FRI

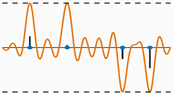
Christoffel

- Discrete support: exact recovery
  - Prony's method [R. de Prony, 1795], and subspace methods: ESPRIT [Roy and Kailath, 1989], MUSIC [Schmidt, 1986], matrix pencils [Hua and Sarkar, 1989], ...
  - Off-the-grid optimization [Candès and Fernandez-Granda, 2014]
- Structured support: exact recovery
  - Finite Rate of Innovation [Pan, Blu, and Dragotti, 2014]
  - Super-resolution of lines [Polisano et al., 2017]
- General case: asymptotic guarantees
  - Polynomial approximations [Mhaskar, 2019]
  - Christoffel approximations (rational) [Pauwels, Putinar, and Lasserre, 2020]

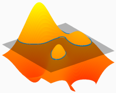
# Recovery Approaches



Prony



BLASSO (Dual)



FRI

Christoffel

- Discrete support: exact recovery
  - Prony's method [R. de Prony, 1795], and subspace methods: ESPRIT [Roy and Kailath, 1989], MUSIC [Schmidt, 1986], matrix pencils [Hua and Sarkar, 1989], ...
  - Off-the-grid optimization [Candès and Fernandez-Granda, 2014]
- Structured support: exact recovery
  - Finite Rate of Innovation [Pan, Blu, and Dragotti, 2014]
  - Super-resolution of lines [Polisano et al., 2017]
- General case: asymptotic guarantees
  - Polynomial approximations [Mhaskar, 2019]
  - Christoffel approximations (rational) [Pauwels, Putinar, and Lasserre, 2020]
- In this talk: polynomial approximations of

| $\mu$ | via convolution | generic support | rates in $p$-$\mathcal{W}$asserstein |
|---|---|---|---|
| $i_{\mathrm{Supp}\,\mu}$ | via SVD | algebraic support | rates in discrete case |

# Overview

# Moment Matrix

For $\mu \in \mathcal{M}(\mathbb{T}^d)$ and $n \in \mathbb{N}$, we define the moment matrix of $\mu$ of order n by

$$T_n \overset{\text{def.}}{=} \left(\hat{\mu}(k-l)\right)_{k,l \in \mathbb{N}_n^d}$$

where $\mathbb{N}_n^d \overset{\text{def.}}{=} \left\{k \in \mathbb{N}^d \;;\; \|k\|_\infty \leqslant n\right\}$.

### Remark

- $T_n \in \mathbb{C}^{N \times N}$ with $N \overset{\text{def.}}{=} (n+1)^d$.
- $T_n$ is multi-level Toeplitz: $T_{k+s,l} = T_{k,l-s}$, for all $k, s, l \in \mathbb{Z}^d$
- for instance with $d = 1$

$$T_n = \begin{bmatrix} \hat{\mu}(0) & \hat{\mu}(-1) & \hat{\mu}(-2) & \dots \\ \hat{\mu}(1) & \hat{\mu}(0) & \hat{\mu}(1) & \dots \\ \hat{\mu}(2) & \hat{\mu}(1) & \hat{\mu}(0) & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

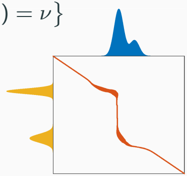
# Wasserstein distances

- **Distances between measures:** f-divergences, discrepancies, Wasserstein distances, ...

$$\mathcal{W}_p^p(\mu,\nu) = \inf \left\{ \int d(x,y)^p \mathrm{d}\pi(x,y) \; ; \; \pi \in \Pi(\mu,\nu) \right\}$$

[Kantorovich, 1942]

  - set of couplings: $\Pi(\mu,\nu) \stackrel{\text{def.}}{=} \{\pi \in \mathcal{M}_+(\mathbb{T}^d \times \mathbb{T}^d) \; ; \; \pi(\cdot, \mathbb{T}^d) = \mu, \; \pi(\mathbb{T}^d, \cdot) = \nu\}$
  - $d$ distance on $\mathbb{T}^d$: we use $d(x,y) = \|x-y\|_{p,\mathbb{T}} \stackrel{\text{def.}}{=} \min_{k\in\mathbb{Z}^d} \|x-y+k\|_p$.

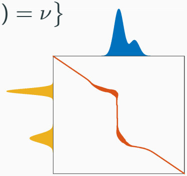


- **Distances between measures:** f-divergences, discrepancies, Wasserstein distances, ...

$$\mathcal{W}_p^p(\mu,\nu) = \inf\left\{\int d(x,y)^p \mathrm{d}\pi(x,y)\ ;\ \pi \in \Pi(\mu,\nu)\right\}$$

[Kantorovich, 1942]

  - set of couplings: $\Pi(\mu,\nu) \stackrel{\text{def.}}{=} \{\pi \in \mathcal{M}_+(\mathbb{T}^d \times \mathbb{T}^d)\ ;\ \pi(\cdot,\mathbb{T}^d) = \mu,\ \pi(\mathbb{T}^d,\cdot) = \nu\}$
  - $d$ distance on $\mathbb{T}^d$: we use $d(x,y) = \|x-y\|_{p,\mathbb{T}} \stackrel{\text{def.}}{=} \min_{k\in\mathbb{Z}^d} \|x-y+k\|_p$.



- **Dual problem:**

$$\mathcal{W}_p(\mu,\nu) = \sup\left\{\int \varphi \mathrm{d}\mu + \int \psi \mathrm{d}\nu\ ;\ \varphi(x)+\psi(y) \leqslant \|x-y\|_{p,\mathbb{T}}^p\right\}$$

  - for $\mathcal{W}_1$, $\psi_\star = -\varphi_\star$ and the constraint reads $|\varphi(x)-\varphi(y)| \leqslant \|x-y\|_{1,\mathbb{T}},\ \forall x,y \in \mathbb{T}^d$

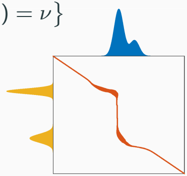
# Wasserstein distances

- **Distances between measures:** f-divergences, discrepancies, Wasserstein distances, ...

$$\mathcal{W}_p^p(\mu,\nu) = \inf \left\{ \int d(x,y)^p \mathrm{d}\pi(x,y) \; ; \; \pi \in \Pi(\mu,\nu) \right\}$$

[Kantorovich, 1942]

  - set of couplings: $\Pi(\mu,\nu) \overset{\text{def.}}{=} \{\pi \in \mathcal{M}_+(\mathbb{T}^d \times \mathbb{T}^d) \; ; \; \pi(\cdot, \mathbb{T}^d) = \mu, \; \pi(\mathbb{T}^d, \cdot) = \nu\}$
  - $d$ distance on $\mathbb{T}^d$: we use $d(x,y) = \|x-y\|_{p,\mathbb{T}} \overset{\text{def.}}{=} \min_{k\in\mathbb{Z}^d} \|x-y+k\|_p$.



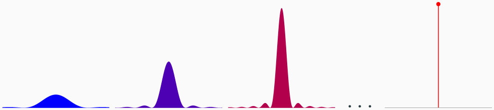
- **Dual problem:**

$$\mathcal{W}_p(\mu,\nu) = \sup \left\{ \int \varphi \mathrm{d}\mu + \int \psi \mathrm{d}\nu \; ; \; \varphi(x) + \psi(y) \leqslant \|x-y\|_{p,\mathbb{T}}^p \right\}$$

  - for $\mathcal{W}_1$, $\psi_\star = -\varphi_\star$ and the constraint reads $|\varphi(x) - \varphi(y)| \leqslant \|x-y\|_{1,\mathbb{T}}, \; \forall x,y \in \mathbb{T}^d$

- $\mathcal{W}_p$ metrizes the weak* topology (on compact sets)

$$\left( \forall \varphi \in \mathscr{C}(\mathbb{T}^d), \; \int \varphi \mathrm{d}\mu_n \to \int \varphi \mathrm{d}\mu \right) \iff \mathcal{W}_p(\mu_n, \mu) \to 0$$



# Polynomial Approximants

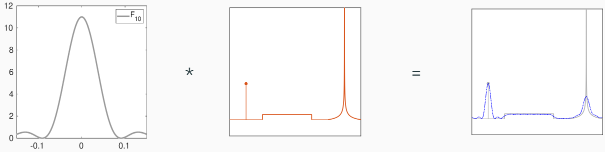
# Fejér approximation

- The Fejér kernel $F_n$ is defined by

$$F_n(x) \overset{\text{def.}}{=} \frac{1}{N} \prod_{i=1}^{d} \frac{\sin^2\left((n+1)\pi x_i\right)}{\sin^2\left(\pi x_i\right)}$$

- For arbitrary $\mu$, consider the polynomial

$$p_n \overset{\text{def.}}{=} F_n * \mu, \quad i.e. \quad p_n(x) = \int F_n(y-x)\mathrm{d}\mu(y)$$



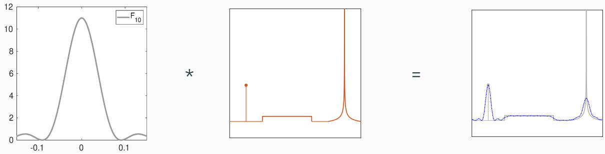
# Fejér approximation

- The Fejér kernel $F_n$ is defined by

$$F_n(x) \overset{\text{def.}}{=} \frac{1}{N} \prod_{i=1}^{d} \frac{\sin^2\left((n+1)\pi x_i\right)}{\sin^2\left(\pi x_i\right)}$$

- For arbitrary $\mu$, consider the polynomial

$$p_n \overset{\text{def.}}{=} F_n * \mu, \quad i.e. \quad p_n(x) = \int F_n(y-x)\mathrm{d}\mu(y)$$




- $p_n$ has a simple expression in terms of the moment matrix:

$$p_n(x) = N^{-1} v_n(x)^* T_n v_n(x), \quad \text{where} \quad v_n(x) = \left(e^{2\imath\pi\langle k, x\rangle}\right)_{k \in \mathbb{N}_n^d}$$

- $p_n$ can be computed using Fast Fourier Transforms:

$$p_n\left(\frac{j}{M}\right) = N^{-1} \sum_{k \in \mathbb{Z}_n^d} w(k)\hat{\mu}(k) e^{2\imath\pi\langle \frac{k}{M}, j\rangle}, \quad \text{where} \quad w(k) = \widehat{F_n}(k) = \prod_{i=1}^{d}\left(1 - \frac{|k_i|}{n+1}\right)$$

# Convergence (Fejér)

**Theorem (Weak-* convergence).** Assuming (only) that $\mu$ has finite total variation, we have that $p_n \rightharpoonup \mu$. More precisely,

$$\begin{cases} \mathcal{W}_1(p_n, \mu) \leqslant \dfrac{d}{\pi^2} \dfrac{\log(n+1) + 3}{n} \\ \mathcal{W}_p(p_n, \mu) \leqslant \left(\dfrac{2d}{p-1}\right)^{1/p} \dfrac{1}{(n+1)^{1/p}}, \quad p > 1 \end{cases}$$

These bounds are tight in the worst case:

$$\begin{cases} \dfrac{d}{\pi^2}\left(\dfrac{\log(n+2)}{n+1} + \dfrac{1}{n+3}\right) \leqslant \sup\limits_{\mu \in \mathcal{M}(\mathbb{T}^d)} \mathcal{W}_1(p_n, \mu) \\ \left(\dfrac{d}{2\pi^2(p-1)}\right)^{1/p} \dfrac{1}{4n^{1/p}} \leqslant \sup\limits_{\mu \in \mathcal{M}(\mathbb{T}^d)} \mathcal{W}_p(p_n, \mu), \quad p > 1 \end{cases}$$

*First step of the proof:*

Use the dual formulation of $\mathcal{W}_p$ to derive the relation

$$\mathcal{W}_p(F_n * \mu, \mu)^p \leqslant \int F_n(x) \|x\|_p^p \mathrm{d}x.$$

# Saturation

- Further assumptions on $\mu$ do not improve so much this bound.

**Theorem (Saturation).** For every measure $\mu \in \mathcal{M}(\mathbb{T}^d)$ not being the Lebesgue measure, there exists a constant $c$ such that

$$\mathcal{W}_1(p_n, \mu) \geqslant \frac{c}{n+1}$$

- For instance $\mathrm{d}\mu = (1 + \cos(2\pi x))\mathrm{d}x =: w(x)\mathrm{d}x$ yields

$$\mathcal{W}_1(p_n, w) \geqslant \frac{1}{4\pi(n+1)}$$

# Powers of Fejér approximations

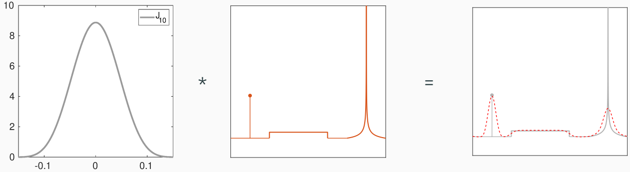
- We define the higher localized kernels

$$r = \left\lceil \frac{p}{2} + 1 \right\rceil, \quad m = \left\lfloor \frac{n}{r} \right\rfloor, \quad K_{n,p}(x) \overset{\text{def.}}{=} C_{m,d} \prod_{i=1}^{d} \frac{\sin^{2r}((m+1)\pi x_i)}{\sin^{2r}(\pi x_i)},$$

$\rightarrow$ For instance, with $p = 1$, $K_{n,1}(x) = \frac{3}{m(2m^2+1)} \frac{\sin^4((m+1)\pi x)}{\sin^4(\pi x)}$ is the Jackson kernel

- Consider the polynomial

$$q_{n,p} \overset{\text{def.}}{=} K_{n,p} * \mu, \quad \textit{i.e.} \quad q_{n,p}(x) = \int K_{n,p}(x-y)\mathrm{d}\mu(y)$$




- $q_{n,p}$ is of degree at most $n$
- $q_{n,p}$ can be computed with Fast Fourier Transforms

$$q_{n,p}\left(\frac{j}{M}\right) = C_n^{-1} \sum_{k} \widehat{K_{n,p}}(k)\hat{\mu}(k) e^{2\imath\pi \langle \frac{k}{M}, j \rangle}$$

# Convergence

**Theorem. (Weak* convergence)** Assuming that $\mu$ has finite total variation, we have that $q_{n,p} \rightharpoonup \mu$. More precisely, there exists $C_p$ independent of n such that

$$\mathcal{W}_p(q_{n,p}, \mu) \leqslant \frac{C_p}{n}.$$

This rate is sharp in the worst-case

$$\frac{d^{(1-p)/p}}{4(n+1)} \leqslant \sup_{\mu \in \mathcal{M}(\mathbb{T}^d)} \mathcal{W}_p(q_{n,p}, \mu)$$

- In particular, with the Jackson kernel, we have for instance

$$\mathcal{W}_1(K_{n,1} * \mu, \mu) \leqslant \frac{3d}{2(n+2)}$$

**Theorem (Worst-case bound).** For every $d, n \in \mathbb{N}$, for every $\mu \in \mathcal{M}(\mathbb{T}^d)$ with finite total variation, there exists a polynomial of best approximation in the Wasserstein-1 distance. Moreover

$$\sup_{\mu \in \mathcal{M}(\mathbb{T}^d)} \min_{\deg(p) \leqslant n} \mathcal{W}_p(p, \mu) \geqslant \frac{d^{(1-p)/p}}{4(n+1)}.$$

*Idea of proof*

- Dual $\mathcal{W}_1 \longleftrightarrow$ worst-case error for best polynomial approximation of Lipschitz functions
- Extend to $\mathcal{W}_p$ using $\mathcal{W}_p(\mu, \nu) \geqslant d^{(1-p)/p} \mathcal{W}_1(\mu, \nu)$ from Jensen's and Hölder's inequality

# Sharpness for $\mathcal{W}_1$

- For this worst-case bound, sharpness is revealed in the univariate case

**Theorem.** With $x \in \mathbb{T}$ we have $\mathcal{W}_1(p_\star, \delta_x) = \frac{1}{4}(n+1)^{-1}$.

- For this worst-case bound, sharpness is revealed in the univariate case

**Theorem.** With $x \in \mathbb{T}$ we have $\mathcal{W}_1(p_\star, \delta_x) = \frac{1}{4}(n+1)^{-1}$.

- Proof involves the relation

$$\mathcal{W}_1(\mu, \nu) = \|\mathcal{B}_1 * \mu - \mathcal{B}_1 * \nu\|_{L^1}, \quad \text{where} \quad \mathcal{B}_1 : t \in \mathbb{T} \mapsto \frac{1}{2} - t$$

(Periodic analog of the cumulative distribution formulation of $\mathcal{W}_1$ on $\mathbb{R}$)

- Transfer (by deconvolution) results on unicity of best $L^1$-approximation to unicity of our best polynomial approximation in some cases (*e.g.* $\mu$ a.c., or $\mu = \delta_x$)

# Sharpness for $\mathcal{W}_1$

- For this worst-case bound, sharpness is revealed in the univariate case

**Theorem.** With $x \in \mathbb{T}$ we have $\mathcal{W}_1(p_\star, \delta_x) = \frac{1}{4}(n+1)^{-1}$.

  - Proof involves the relation

$$\mathcal{W}_1(\mu, \nu) = \|\mathcal{B}_1 * \mu - \mathcal{B}_1 * \nu\|_{L^1}, \quad \text{where} \quad \mathcal{B}_1 : t \in \mathbb{T} \mapsto \frac{1}{2} - t$$

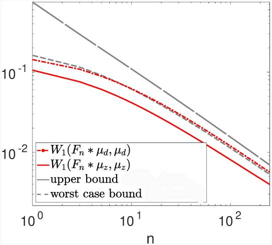
  (Periodic analog of the cumulative distribution formulation of $\mathcal{W}_1$ on $\mathbb{R}$)

  - Transfer (by deconvolution) results on unicity of best $L^1$-approximation to unicity of our best polynomial approximation in some cases (*e.g.* $\mu$ a.c., or $\mu = \delta_x$)

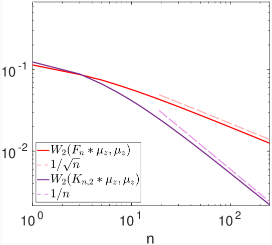
- De La Vallée Poussin approximation [Mhaskar, 2019] also achieves optimal rate but dependency with $d$ is worse in the constant

# Numerical Verification

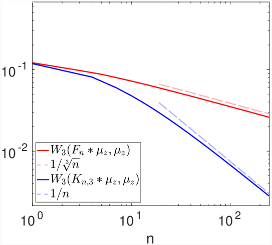
- Test on two measures:
  - a discrete measure $\mu_d$, $s = 15$
  - a (discretized) algebraic curve $\mu_z$, $s = 3000$
- Semidiscrete algorithm to compute the Wasserstein distance between polynomial density and singular measure.




(m) $\mathcal{W}_1$-rates for Fejér ($\mu_d$, $\mu_z$)




(n) $W_2$-rates for Fejér and $K_{n,1}$ ($\mu_z$)




(o) $W_3$-rates for Fejér and $K_{n,2}$ ($\mu_z$)

# Polynomial Interpolants

# Interpolating Polynomial

- The singular value decomposition: $T_n = \sum_{j=1}^{r} \sigma_j u_j^{(n)} v_j^{(n)*}$ allows to define

$$p_{1,n}(x) = \frac{1}{N} \sum_{j=1}^{r} |u_j^{(n)}(x)|^2$$

$\rightarrow$ unweighted counterpart of $p_n = N^{-1} e(x)^* T_n e(x) = N^{-1} \sum \sigma_j u_j^{(n)}(x) v_j^{(n)}(x)^*$.
We have $0 \leqslant p_{1,n} \leqslant 1$.

# Interpolating Polynomial

- The singular value decomposition: $T_n = \sum_{j=1}^r \sigma_j u_j^{(n)} v_j^{(n)*}$ allows to define

$$p_{1,n}(x) = \frac{1}{N} \sum_{j=1}^{r} |u_j^{(n)}(x)|^2$$

$\rightarrow$ unweighted counterpart of $p_n = N^{-1} e(x)^* T_n e(x) = N^{-1} \sum \sigma_j u_j^{(n)}(x) v_j^{(n)}(x)^*$.
We have $0 \leqslant p_{1,n} \leqslant 1$.

- Assume that $Z \overset{\text{def.}}{=} \operatorname{Supp} \mu$ is an algebraic variety (*i.e.* defined by polynomial equations)
Let $\mathcal{V}(\operatorname{Ker} T_n)$ be the set of common roots of all polynomials in $\operatorname{Ker} T_n$.

**Theorem (Interpolation).** If $\mathcal{V}(\operatorname{Ker} T_n) = Z$, then $p_{1,n}(x) = 1$ iff $x \in V$.

- The singular value decomposition: $T_n = \sum_{j=1}^{r} \sigma_j u_j^{(n)} v_j^{(n)*}$ allows to define

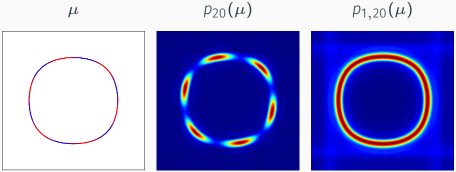
$$p_{1,n}(x) = \frac{1}{N} \sum_{j=1}^{r} |u_j^{(n)}(x)|^2$$

→ unweighted counterpart of $p_n = N^{-1} e(x)^* T_n e(x) = N^{-1} \sum \sigma_j u_j^{(n)}(x) v_j^{(n)}(x)^*$.
We have $0 \leqslant p_{1,n} \leqslant 1$.

- Assume that $Z \overset{\text{def.}}{=} \operatorname{Supp} \mu$ is an algebraic variety (*i.e.* defined by polynomial equations)
Let $\mathcal{V}(\operatorname{Ker} T_n)$ be the set of common roots of all polynomials in $\operatorname{Ker} T_n$.

**Theorem (Interpolation).** If $\mathcal{V}(\operatorname{Ker} T_n) = Z$, then $p_{1,n}(x) = 1$ iff $x \in V$.

→ $\mathcal{V}(\operatorname{Ker} T_n) = Z$ always holds for sufficiently large $n$ if $\mu$ is discrete [Kunis et al., 2016],[Sauer, 2017] or nonnegative [Wageringel, 2022]



# Pointwise convergence

- We assume that $Z \neq \mathbb{T}^d$

**Theorem.** Let $y \in \mathbb{T}^d \setminus Z$, and let $g$ be a polynomial of max-degree $m$ such that $g(y) \neq 0$ and $g$ vanishes on $Z$. Then, for all $n \geqslant m$,

$$p_{1,n+m}(y) \leqslant \frac{\|g\|_{L^2}^2}{|g(y)|} \frac{m(4m+2)^d}{n+1} + \frac{dm}{n+m+1}$$

- In combination with the interpolation property, this proves pointwise convergence to the characteristic function of the support, with rate $O(n^{-1})$.

# The Discrete Case

- If $\mu = \sum_{j=1}^{r} \lambda_j \delta_{x_j}$, stronger results are derived with the help of the Vandermonde decomposition of $T_n$

**Theorem (Pointwise convergence).** Let $x \neq x_j$ for all $j$. If $n+1 > \frac{4d}{\min_{j \neq l} \|x_j - x_l\|_\infty}$, then

$$p_{1,n}(x) \leqslant \frac{1}{3(n+1)^2} \frac{\lambda_{\max}}{\lambda_{\min}} \sum \frac{1}{\|x - x_j\|_\infty^2}$$

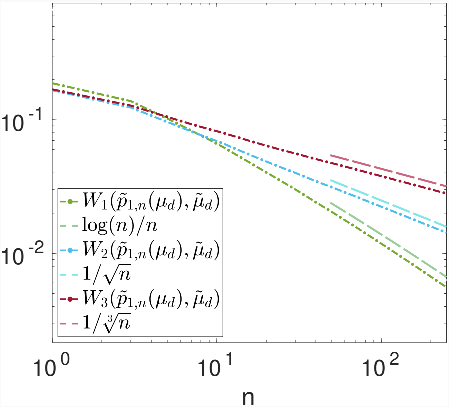
**Theorem (Weak* convergence).** Let $\tilde{p}_{1,n} \overset{\text{def.}}{=} p_{1,n}/\|p_{1,n}\|_{L^1}$. We have

$$\tilde{p}_{1,n} \rightharpoonup \tilde{\mu} \overset{\text{def.}}{=} \frac{1}{r} \sum_{j=1}^{r} \delta_{x_j}.$$

More specifically

$$\begin{cases} \mathcal{W}_1(\tilde{p}_{1,n}, \tilde{\mu}) = O(\frac{\log n}{n}) \\ \mathcal{W}_p(\tilde{p}_{1,n}, \tilde{\mu}) = O(n^{-1/p}) \end{cases}$$

# Numerical Verification




$\mathcal{W}_{\{1,2,3\}}$-rates for $\tilde{p}_{1,n}$

# Conclusion

# Conclusion

**Summary.**

- New insights on Wasserstein approximation of measures/support
- Computationally efficient polynomial approximations

**Outlook.**

- Extension to the noisy regime
- Extension to rational approximations (ongoing)
- Practical usage: approximation of Wasserstein distances $\mathcal{W}(\mu, \nu)$ by $\mathcal{W}(p_n, q_n)$

Preprint available: arXiv.2203.10531

Thank you for your attention!

# References

E.J. Candès and C. Fernandez-Granda. “Towards a Mathematical Theory of Super-Resolution”. In: *Comm. Pure Appl. Math.* 67.6 (2014), pp. 906–956.

Stephen D. Fisher. “Best Approximation by Polynomials”. In: *J. Approx. Th.* 21.1 (1977), pp. 43–59.

Y. Hua and T.K. Sarkar. “Generalized Pencil-of-Function Method for Extracting Poles of an EM System from its Transient Response”. In: *IEEE Trans. Antennas Propagation* 37.2 (1989).

L. Kantorovich. “On the Transfer of Masses (in Russian)”. In: *Doklady Akademii Nauk* 37.2 (1942), pp. 227–229.

S. Kunis et al. “A Multivariate Generalization of Prony’s Method”. In: *Linear Algebra Appl.* 490 (2016), pp. 31–47.

H. N. Mhaskar. “Super-Resolution Meets Machine Learning: Approximation of Measures”. In: *J. Fourier Anal. Appl.* 25.6 (2019), pp. 3104–3122.

H. Pan, T. Blu, and P.L. Dragotti. “Sampling Curves With Finite Rate of Innovation”. In: *IEEE Trans. Signal Process.* 62.2 (2014), pp. 458–471.

K. Polisano et al. “A convex approach to super-resolution and regularization of lines in images”. In: (2017). URL: `https://hal.archives-ouvertes.fr/hal-01599010`.

E. Pauwels, M. Putinar, and J.B. Lasserre. “Data Analysis From Empirical Moments and the Christoffel Function”. In: *F. Comp. Math.* 21 (2020), pp. 243–273.

G. R. de Prony. “Essai Expérimental et Analytique: Sur les Lois de la Dilatabilité des Fluides Élastiques et sur celles de la Force Expansive de la Vapeur de l’Eau et de la Vapeur de l’Alkool, à différentes températures”. In: *Journal de l’École Polytechnique Floréal et Plairial* 1.cahier 22 (1795), pp. 24–76.

R. Roy and T. Kailath. “ESPRIT-Estimation of Signal Parameters via Rotational Invariance Techniques”. In: *IEEE Trans. Acoustics Speech Signal Process.* 37.7 (1989), pp. 984–995.

F. Santambrogio. “Optimal transport for Applied Mathematicians”. In: *Birkäuser, NY* 55.58-63 (2015), p. 94.

T. Sauer. “Prony’s Method in Several Variables”. In: *Numer. Math.* 136 (2017), pp. 411–438.

R. Schmidt. “Multiple Emitter Location and Signal Parameter Estimation”. In: *IEEE Trans. Antennas Propagation* 34.3 (1986), pp. 276–280.

M. Wageringel. “Truncated moment problems on positive-dimensional algebraic varieties”. 2022.