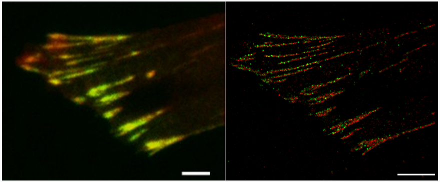
Approximating Singular Measures on the Torus with Moment Polynomials

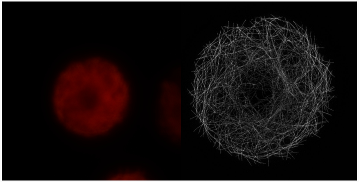
Paul Catala. Joint work with M. Hockmann, S. Kunis and M. Wageringel
University of Osnabrück.

Curves and Surfaces, 24.06.22

Motivation

Super-resolution. Estimate a signal from a few coarse linear measurements





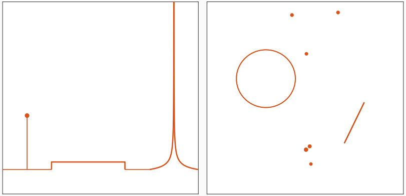
- Ubiquitous problem in imaging and data science (low-pass filtering)
  - Fluorescence microscopy
  - X-ray crystallography
  - Astronomical imaging
  - Mixture estimation
- Signals of interest are often structured: pointwise sources, curves, surfaces...

[0] images from the cell image library (http://cellimagelibrary.org/)

Data model

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

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

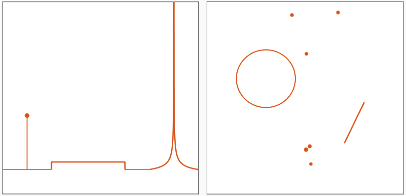


Singular measures μ

Data model

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

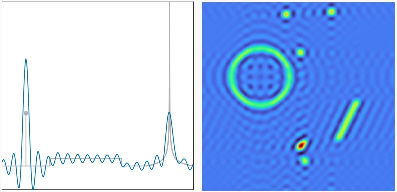
$$\mu \in \mathcal{M}(\mathbb{T}^d)$$



Singular measures μ

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

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

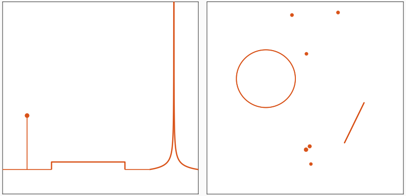


Fourier partial sum $S_n\mu$ $(n = 20)$

Data model

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

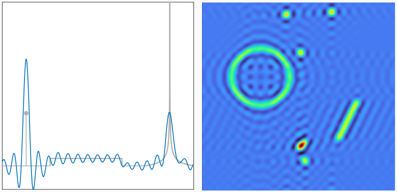
$$\mu \in \mathcal{M}(\mathbb{T}^d)$$



Singular measures μ

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

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

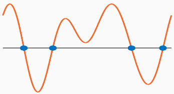


Fourier partial sum $S_n\mu$ $(n = 20)$

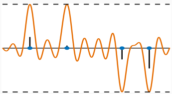
How can we recover μ from $\{\hat{\mu}(k)\}$, $k \in \{-n, \dots, n\}^d$?

Previous works

- For discrete measures → "interpolation"
  - Prony's method [R. de Prony, 1795], ESPRIT [Roy and Kailath, 1989], MUSIC [Schmidt, 1986], matrix pencils [Hua and Sarkar, 1989].
  - Off-the-grid optimization [Candès and Fernandez-Granda, 2014]



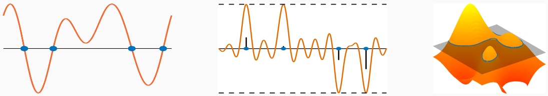
Prony (discrete)



Dual polynomial

Previous works

- For discrete measures → "interpolation"
  - Prony's method [R. de Prony, 1795], ESPRIT [Roy and Kailath, 1989], MUSIC [Schmidt, 1986], matrix pencils [Hua and Sarkar, 1989].
  - Off-the-grid optimization [Candès and Fernandez-Granda, 2014]

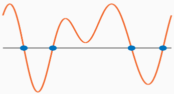


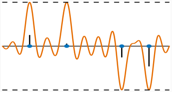
Prony (discrete) | Dual polynomial | Prony (curve)

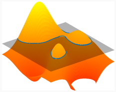
- For general measures,
  - FRI approaches [Pan, Blu, and Dragotti, 2014] (still interpolation)

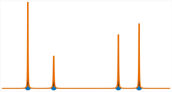
Previous works

- For discrete measures → "interpolation"
  - Prony's method [R. de Prony, 1795], ESPRIT [Roy and Kailath, 1989], MUSIC [Schmidt, 1986], matrix pencils [Hua and Sarkar, 1989].
  - Off-the-grid optimization [Candès and Fernandez-Granda, 2014]


Prony (discrete)


Dual polynomial

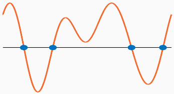

Prony (curve)

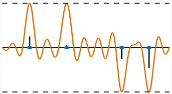

Approximation

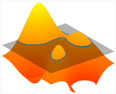
- For general measures,
  - FRI approaches [Pan, Blu, and Dragotti, 2014] (still interpolation)
  - Polynomial approximations [Mhaskar, 2019]
  - Christoffel approximations (rational) [Pauwels, Putinar, and Lasserre, 2020]

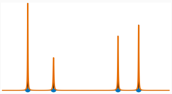
Previous works

- For discrete measures → "interpolation"
  - Prony's method [R. de Prony, 1795], ESPRIT [Roy and Kailath, 1989], MUSIC [Schmidt, 1986], matrix pencils [Hua and Sarkar, 1989].
  - Off-the-grid optimization [Candès and Fernandez-Granda, 2014]


Prony (discrete)


Dual polynomial


Prony (curve)


Approximation

- For general measures,
  - FRI approaches [Pan, Blu, and Dragotti, 2014] (still interpolation)
  - Polynomial approximations [Mhaskar, 2019]
  - Christoffel approximations (rational) [Pauwels, Putinar, and Lasserre, 2020]

- In this work:
  - easily computable polynomial approximations, with sharp rates in $\mathcal{W}_1$ metric (similarities with [Mhaskar, 2019], use of different distance between measures)
  - easily computable polynomial interpolant for algebraic varieties

Overview

Preliminaries

Moment Matrix

Definition (Moment matrix). Given $\{\hat{\mu}(k)\}$, $k \in \{-n, \ldots, n\}^d$, we define the moment matrix

$$T_n \stackrel{\text{def.}}{=} \left[\hat{\mu}(k-l)\right]_{k,l \in \{0,\ldots,n\}^d}.$$

- central in parametric approaches (Prony, ESPRIT, MUSIC, ...)
- important in off-the-grid optimization (Lasserre's hierarchies) [Castro et al., 2017]

Moment Matrix

Definition (Moment matrix). Given $\{\hat{\mu}(k)\}$, $k \in \{-n, \dots, n\}^d$, we define the moment matrix

$$T_n \stackrel{\text{def.}}{=} \left[\hat{\mu}(k-l)\right]_{k,l \in \{0,\dots,n\}^d}.$$

- central in parametric approaches (Prony, ESPRIT, MUSIC, ...)
- important in off-the-grid optimization (Lasserre's hierarchies) [Castro et al., 2017]

One important difference between the discrete and non-discrete cases

- If $\mu = \sum_{j=1}^{r} \lambda_j \delta_{x_j}$, T_n admits the Vandermonde decomposition

$$T_n = A \Lambda A^*$$

where $A = \left[e^{-2i\pi\langle k, x_j\rangle}\right]_{k \in \{0,\dots,n\}^d, j \in [\![1,r]\!]}$ and $\Lambda = \mathrm{Diag}(\lambda)$.

Moment Matrix

Definition (Moment matrix). Given $\{\hat{\mu}(k)\}$, $k \in \{-n, \ldots, n\}^d$, we define the moment matrix

$$T_n \overset{\text{def.}}{=} \left[\hat{\mu}(k-l)\right]_{k,l \in \{0,\ldots,n\}^d}.$$

- central in parametric approaches (Prony, ESPRIT, MUSIC, ...)
- important in off-the-grid optimization (Lasserre's hierarchies) [Castro et al., 2017]

One important difference between the discrete and non-discrete cases

- If $\mu = \sum_{j=1}^r \lambda_j \delta_{x_j}$, T_n admits the Vandermonde decomposition

$$T_n = A \Lambda A^*$$

where $A = \left[e^{-2i\pi\langle k, x_j\rangle}\right]_{k \in \{0,\ldots,n\}^d, j \in [\![1,r]\!]}$ and $\Lambda = \mathrm{Diag}(\lambda)$.

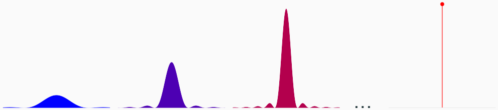
- No such decomposition in general $\to$ rank-revealing SVD provides useful tools

$$T_n = \sum_{j=1}^r \sigma_j u_j^{(n)} v_j^{(n)*}$$

Wasserstein-1 distance

- We need a distance between measures
- Examples include f-divergences, MMD, and Wasserstein distances
- Wasserstein distances metrize the weak* topology (on compact sets) [Santambrogio, 2015], *i.e.*

$$\mu_n \rightharpoonup \mu \iff \mathcal{W}_p(\mu_n, \mu) \to 0$$



- Wasserstein-1 further admits the dual formulation

$$\mathcal{W}_1(\mu, \nu) = \sup_{f \in \mathscr{C}(\mathbb{T}^d), \mathrm{Lip}(f) \leqslant 1} \int f \mathrm{d}(\mu - \nu)$$

→ requires no positivity, only $\mu(\mathbb{T}^d) = \nu(\mathbb{T}^d)$
→ $\mathrm{Lip}(f) \leqslant 1$ means $|f(x) - f(y)| \leqslant \min_{k \in \mathbb{Z}^d} \|x - y + k\|_1, \forall x, y$

Polynomial Approximations

Best Polynomial Approximation

- Assume μ is of finite total variation, $\|\mu\|_{TV} = 1$
- We make no further assumptions to provide a worst-case error bound

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

$$\sup_{\mu \in \mathcal{M}} \min_{\deg(p) \leqslant n} \mathcal{W}_1(p, \mu) \geqslant \frac{1}{4(n+1)}.$$

Idea of the proof

- Best approximation in the worst-case:

$$\begin{aligned}\sup_\mu \min_p \mathcal{W}_1(p,\mu) &\geqslant \min_p \mathcal{W}_1(p,\delta_0)\\ &= \min_p \sup_{\mathrm{Lip}(f)\leqslant 1} \|f - \check{p} * f\|_\infty \qquad (\check{p}(x) = p(-x))\\ &\geqslant \sup_{\mathrm{Lip}(f)\leqslant 1} \min_p \|f - p\|_\infty\end{aligned}$$

$\rightarrow$ worst-case error for best polynomial approximation of Lipschitz functions

- Generalization of a univariate argument of [Fisher, 1977] to the multivariate case

Sharpness (Lower Bound)

- For this worst-case bound, sharpness is revealed in the univariate case
- Make use of the Bernoulli spline

$$\mathcal{B}_1 : t \in \mathbb{T} \mapsto \sum_{k=1}^{\infty} \frac{\sin 2\pi k t}{\pi k} = \frac{1}{2} - t$$

Lemma. For $\mu, \nu \in \mathcal{M}(\mathbb{T})$, we have

$$\mathcal{W}_1(\mu, \nu) = \int_{\mathbb{T}} |\mathcal{B}_1 * \mu(t) - \mathcal{B}_1 * \nu(t)| \mathrm{d}t$$

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

- If $\mu = \delta_0$, then $\mathcal{W}_1(p^*, \delta_0) = \frac{1}{4}(n+1)^{-1}$, matching our lower bound
- If μ is absolutely continuous, leads to unicity of the best approximation

Fejér approximation

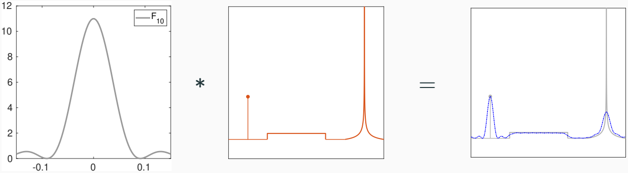
- The Fejér kernel F_n is defined by

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

- Consider the polynomial $p_n \overset{\text{def.}}{=} F_n * \mu$
  - Alternatively,

$$p_n(x) = (n+1)^{-d} \sum \hat{\mu}(k-l) e^{-2i\pi(k-l)x} = (n+1)^{-d} e(x)^* T_n e(x)$$

  - Computed using Fast Fourier Transforms



Fejér approximation

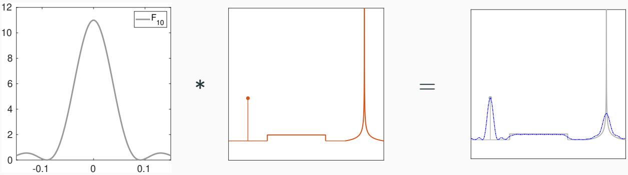
- The Fejér kernel F_n is defined by

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

- Consider the polynomial $p_n \overset{\text{def.}}{=} F_n * \mu$
  - Alternatively,

$$p_n(x) = (n+1)^{-d} \sum \hat{\mu}(k-l) e^{-2i\pi(k-l)x} = (n+1)^{-d} e(x)^* T_n e(x)$$

  - Computed using Fast Fourier Transforms



Theorem (Weak* convergence). We have that $p_n \rightharpoonup \mu$. More precisely,

$$\mathcal{W}_1(p_n, \mu) \leqslant \frac{d}{\pi^2} \frac{\log(n+1)+3}{n}$$

Sharpness (Upper 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

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

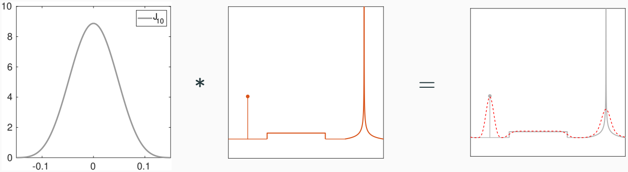
- For instance $\mathrm{d}\mu/\mathrm{d}x = 1 + \cos(2\pi x) := w(x)$ yields $\mathcal{W}_1(p_n, w) \geqslant (4\pi)^{-1}(n+1)^{-1}$
- However, $\mathcal{W}_1(p_n, \delta_0) \geqslant \frac{d}{\pi^2}\left(\frac{\log(n+2)}{n+1} + \frac{1}{n+3}\right)$

Jackson approximation

- The Jackson kernel J_n is defined by

$$J_{2m}(x) \overset{\text{def.}}{=} \frac{3}{m(2m^2+1)} \prod_{i=1}^{d} \frac{\sin^4((m+1)\pi x_i)}{\sin^4(\pi x_i)}$$

- Consider the polynomial $q_n \overset{\text{def.}}{=} J_n * \mu$
  - Computed with Fast Fourier Transforms

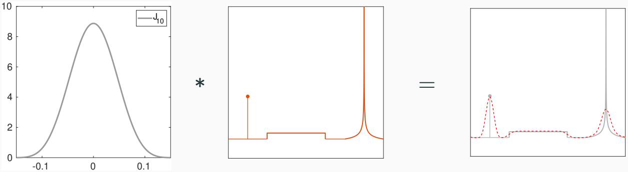


Jackson approximation

- The Jackson kernel J_n is defined by

$$J_{2m}(x) \overset{\text{def.}}{=} \frac{3}{m(2m^2+1)} \prod_{i=1}^{d} \frac{\sin^4((m+1)\pi x_i)}{\sin^4(\pi x_i)}$$

- Consider the polynomial $q_n \overset{\text{def.}}{=} J_n * \mu$
  - Computed with Fast Fourier Transforms



Theorem. (Weak* convergence) We have that $q_n \rightharpoonup \mu$. More precisely,

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

Polynomial Interpolation

Interpolating Polynomial

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

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

$\rightarrow$ unweighted counterpart of $p_n = F_n * \mu$. Note that $0 \leqslant p_{1,n} \leqslant 1$.

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

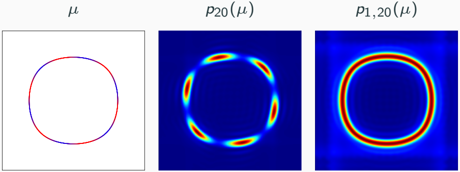
$$p_{1,n}(x) = \frac{1}{(n+1)^d} \sum_{j=1}^{r} |u_j^{(n)}(x)|^2$$

$\rightarrow$ unweighted counterpart of $p_n = F_n * \mu$. Note that $0 \leqslant p_{1,n} \leqslant 1$.

- Let $V \overset{\text{def.}}{=} \overline{\operatorname{Supp}\mu}^Z$ be the smallest algebraic set containing $\operatorname{Supp}\mu$
Let $\mathcal{V}(\operatorname{Ker} T_n)$ be the set of common roots of all polynomials in $\operatorname{Ker} T_n$.

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

$\rightarrow$ $\mathcal{V}(\operatorname{Ker} T_n) = V$ always holds for sufficiently large n if μ is discrete [Kunis et al., 2016],[Sauer, 2017] or nonnegative [Wageringel, 2022]
$\rightarrow$ generalizes [Ongie and Jacob, 2016] to varieties of arbitrary dimension



Pointwise convergence

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

Theorem. Let $y \in \mathbb{T}^d \setminus V$, and let g be a polynomial of max-degree m such that $g(y) \neq 0$ and g vanishes on $\operatorname{Supp}\mu$. 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). We have

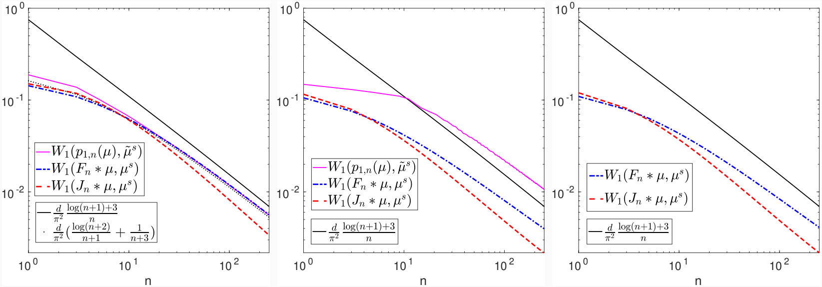
$$\frac{p_{1,n}}{\|p_{1,n}\|_{L^1}} \rightharpoonup \frac{1}{r} \sum_{j=1}^{r} \delta_{x_j}$$

Numerical Illustrations

Numerical Illustrations

- We consider three synthetic examples
  - discrete, $r = 15$ points, λ random — moments analytical
  - algebraic curve, $r = 3000$ points, λ uniform — numerical integration
  - circle, $r = 3000$ points, λ uniform — analytical

- We compute the semidiscrete optimal transport between the discretized approximation μ^r and the density p_n



Conclusion

Conclusion

Summary.

- New insights on Wasserstein-1 approximation of measures
- Computationally efficient polynomial approximations
- Pointwise convergence towards the characteristic function of the support

Outlook.

- Extension to the noisy regime
- Connection with Christoffel functions

Preprint available: arXiv.2203.10531

Thank you for your attention!

References

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

Castro, Y. de et al. (2017). “Exact solutions to Super Resolution on semi-algebraic domains in higher dimensions”. In: *IEEE Trans. Inform. Theory* 63.1, pp. 621–630.

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

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

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

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

Ongie, G. and M. Jacob (2016). “Off-the-grid recovery of piecewise constant images from few Fourier samples”. In: *SIAM J. Imaging Sci.* 9.3, pp. 1004–1041.

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

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

R. de Prony, G. (1795). “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, pp. 24–76.

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

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

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

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

Wageringel, M. (2022). "Truncated moment problems on positive-dimensional algebraic varieties".