

# Non-uniform spline recovery from small degree polynomial approximation

Yohann De Castro, Guillaume Mijoule

*Département de Mathématiques (CNRS UMR 8628), Bâtiment 425, Faculté des Sciences d'Orsay,  
Université Paris-Sud 11, F-91405 Orsay Cedex, France.*

---

## Abstract

We investigate the sparse spikes deconvolution problem onto spaces of algebraic polynomials. Our framework encompasses the measure reconstruction problem from a combination of noiseless and noisy moment measurements. We study a TV-norm regularization procedure to localize the support and estimate the weights of a target discrete measure in this frame. Furthermore, we derive quantitative bounds on the support recovery and the amplitudes errors under a Chebyshev-type minimal separation condition on its support. Incidentally, we study the localization of the knots of non-uniform splines when a Gaussian perturbation of their inner-products with a known polynomial basis is observed (i.e. a small degree polynomial approximation is known) and the boundary conditions are known. We prove that the knots can be recovered in a grid-free manner using semidefinite programming.

*Keywords:* LASSO, Super-resolution, Non-uniform splines, Algebraic polynomials,  $\ell_1$ -minimization.

---

## 1. Introduction

### 1.1. Non-uniform spline recovery

Our framework involves the recovery of non-uniform splines, i.e. a smooth polynomial function that is piecewise-defined on subintervals of different lengths.

---

*Email addresses:* [yohann.decastro@math.u-psud.fr](mailto:yohann.decastro@math.u-psud.fr) (Yohann De Castro),  
[guillaume.mijoule@math.u-psud.fr](mailto:guillaume.mijoule@math.u-psud.fr) (Guillaume Mijoule)  
*URL:* [www.math.u-psud.fr/~decastro](http://www.math.u-psud.fr/~decastro) (Yohann De Castro)

More precisely, we investigate a grid-free procedure to estimate a non-uniform spline from a polynomial approximation of small degree. Our estimation procedure can be used as a post-processing technique in various fields such as data assimilation [16], shape optimization [15] or spectral methods in PDE's [14].

For instance, one gets a polynomial approximation of the solution of a PDE when using spectral methods such as the Galerkin method. In this setting, one seeks a weak solution of a PDE using bounded degree polynomials as test functions. Then, the Lax-Milgram theorem grants the existence of a unique weak solution  $\mathbf{f}$  for which a polynomial approximation  $P$  can be computed. Moreover, Céa's lemma shows that the Galerkin approximation  $P$  is comparable to the best polynomial approximation  $\mathbf{p}(\mathbf{f})$  of the weak solution  $\mathbf{f}$ . This situation can be depicted by Assumption 1. Hence, if one knows the weak solution  $\mathbf{f}$  is a non-uniform spline then our (post-processing) procedure can provide a grid-free estimate  $\hat{\mathbf{f}}$  from the Galerkin approximation  $P$ . Moreover, Theorem 2 shows that the recovered spline has large discontinuities near the large discontinuities of the target spline  $\mathbf{f}$ . Hence, the location of the large enough discontinuities of the weak solution  $\mathbf{f}$  can be quantitatively and in a grid-free manner estimated from the Galerkin approximation using our algorithm.

As an example, Figure 1 illustrates how our procedure improves a polynomial approximation of a non-uniform spline. Observe that discontinuities of splines make them difficult to approximate by polynomials. Consider an approximation (thin black line) of the spline (thick dashed gray line). It seems rather difficult to localize the discontinuities of the spline from the knowledge of this polynomial approximation and the boundary conditions. Nevertheless, our procedure produces a non-uniform spline (thick black line) whose large discontinuities are close to the knots of the target spline.

The method we propose is as follows. Following an idea of [3], we aim at reconstructing a spline of degree  $d$  by recovering its  $d + 1$  distributional derivative, using tools of the super-resolution theory [10, 8, 9]. More precisely, consider an univariate spline  $\mathbf{f}$  of degree  $d$ , defined on  $[-1, 1]$ . The  $d + 1$  distributional derivative of  $\mathbf{f}$ , denoted  $\mathbf{f}^{(d+1)}$ , is a discrete signed measure whose support are the knots of

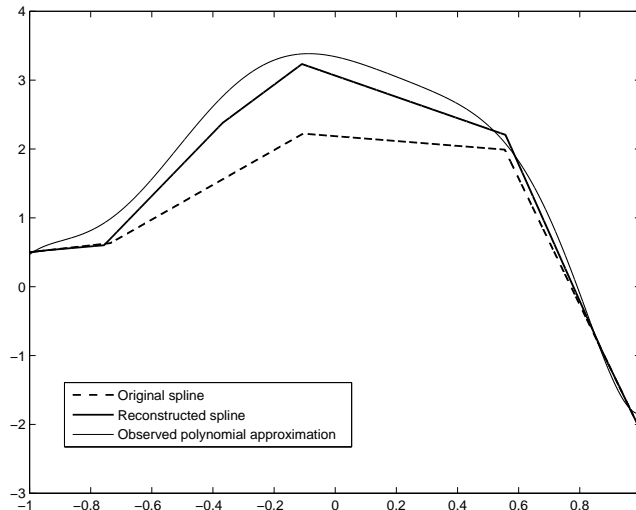


Figure 1: Estimated spline (thick black line) of a non-uniform spline (thick dashed gray line) and its knots from a polynomial approximation (thin black line).

the spline. Using an integration by parts, one can show that the first  $m + 1$  polynomial moments of  $\mathbf{f}^{(d+1)}$  can be expressed as a linear combination of the first  $m - d$  moments of  $\mathbf{f}$  and its  $2(d + 1)$  boundary conditions; moreover, the first  $d + 1$  moments of  $\mathbf{f}^{(d+1)}$  only depend on the boundary conditions (details are in Lemma 5). As a consequence, observing  $m - d$  noisy moments and the (noiseless) boundary conditions of  $\mathbf{f}$  is equivalent to observing  $d + 1$  noiseless and  $m - d$  noisy moments of  $\mathbf{f}^{(d+1)}$ . This observation is the motivation of the theoretical work of this paper.

### 1.2. Sparse spikes deconvolution onto spaces of algebraic polynomials

In this paper, we extend some recent results in spike deconvolution to the frame of algebraic polynomials. Beyond the theoretical interest, we focus on this model in order to bring tools and quantitative guarantees from the super-resolution theory to the companion problem of the recovery of knots of non-uniform splines [3]. At first glance, this setting can be depicted as a deconvolution problem where one wants to recover the location of the support of a discrete measure from the observation of its convolution with an algebraic polynomial of given degree  $m$ . More precisely,

we aim at recovering a discrete measure from the knowledge of the true  $(d + 1)$  first moments and a noisy version of the  $(m - d)$  next ones.

### 1.3. Previous works

The super-resolution problem has been intensively investigated in the last years. In [5, 9] the authors give an exact recovery condition for the noiseless problem in a general setting. In the Fourier frame, this analysis was greatly refined in [8] which shows that the exact recovery condition is satisfied for all measure satisfying a “minimum separation condition”. The recovery from noisy samplings was investigated in [7] which characterizes the reconstruction error as the resolution increases. The first results on quantitative localization was brought by the authors of [1] who give bounds on the support detection error in a general frame. This analysis was derived in terms of the amplitude of the target measure in [13]. In the Fourier frame, the optimal rates in prediction error have been investigated in [18]. Lastly, the behavior and the stability of  $TV$ -norm regularization in the space of measures has been investigated in [12] when observing small noise errors.

The spline recovery problem in the noiseless case has been studied in [3] where the authors assume that one knows the orthogonal projection  $\mathbf{p}(\mathbf{f})$  of the non-uniform spline  $\mathbf{f}$ . Our frame extends their point of view to the noisy case where one observes a polynomial approximation  $P$ . To the best of our knowledge, there is no result on a quantitative localization of the knots of non-uniform splines from noisy measurements.

## 2. General model and notation

Let  $[-1, 1]$  be equipped with the distance:

$$\forall u, v \in [-1, 1], \quad d(u, v) = |\arccos u - \arccos v|.$$

Let  $\mathbf{x}$  be a signed measure on  $[-1, 1]$  with finite support of unknown size  $s$ . In particular,  $\mathbf{x}$  admits a polar decomposition:

$$\mathbf{x} = \sum_{k=1}^s a_k \delta_{t_k}, \tag{1}$$

where  $a_k \in \mathbb{R} \setminus \{0\}$ ,  $t_k \in [-1, 1]$ , and  $\delta_t$  denotes the Dirac measure at point  $t$ . Let  $m$  be a positive integer and  $\mathcal{F} = \{\varphi_0, \varphi_1, \dots, \varphi_m\}$  be such that  $\varphi_0 = 1$  and for  $k = 1, \dots, m$ ,

$$\varphi_k = \sqrt{2} T_k,$$

where  $T_k(t) = \cos(k \arccos(t))$  is the  $k$ -th Chebyshev polynomial of the first kind. Observe that the family  $\mathcal{F}$  is an orthonormal family with respect to the probability measure  $\mathbb{M}(dt) = (1/\pi)(1-t^2)^{-1/2} \mathcal{L}(dt)$  on  $[-1, 1]$ , where  $\mathcal{L}$  denotes the Lebesgue measure. Define the  $k$ -th generalized moment of a signed measure  $\mu$  on  $[-1, 1]$  as:

$$c_k(\mu) = \int_{[-1,1]} \varphi_k d\mu,$$

for  $k = 0, 1, \dots, m$ . Assume that we observe  $c_k(\mathbf{x})$  for  $0 \leq k \leq d$  and a noisy version of  $c_k(\mathbf{x})$  for  $d+1 \leq k \leq m$ , where possibly  $d = -1$ . Define  $y_k = c_k(\mathbf{x}) + \varepsilon_k$  such as  $\varepsilon_k = 0$  for  $0 \leq k \leq d$  and  $\varepsilon_k$  are i.i.d.  $\mathcal{N}(0, \sigma^2)$  for  $d+1 \leq k \leq m$ . This can be written as:

$$\mathbf{y} = \mathbf{c}(\mathbf{x}) + \mathbf{e}, \tag{2}$$

where  $\mathbf{c}(\mathbf{x}) = (c_k(\mathbf{x}))_{k=0}^m$  and  $\mathbf{e} = (0, \mathbf{n})$  with  $\mathbf{n} \sim \mathcal{N}(0, \sigma^2 \text{Id}_{m-d})$ . Note we know the first true moments up to the order  $d$  and a noisy version of them up to the order  $m$ . Moreover, the degree  $d$  is allowed to be  $-1$ .

### 2.1. An $L_1$ -minimization procedure

Our analysis follows recent proposals on  $\ell_1$ -minimization [5, 1, 18, 12]. Denote by  $\mathcal{M}$  the set of all finite signed measures on  $[-1, 1]$  endowed with the total variation norm  $\|\cdot\|_{TV}$ , which is isometrically isomorphic to the dual  $\mathcal{C}([-1, 1])^*$  of continuous function endowed with the supremum norm. We recall that for all  $\mu \in \mathcal{M}$ ,

$$\|\mu\|_{TV} = \sup_{\mathcal{P}} \sum_{E \in \mathcal{P}} |\mu(E)|,$$

where the supremum is taken over all partitions  $\mathcal{P}$  of  $[-1, 1]$  into a finite number of disjoint measurable subsets. Consider a modified version of the convex program BLASSO [1] given by:

$$\hat{\mathbf{x}} \in \arg \min_{\mu \in \mathcal{C}_d(\mathbf{x})} \frac{1}{2} \|\mathbf{c}(\mu) - \mathbf{y}\|_2^2 + \lambda \|\mu\|_{TV}, \tag{3}$$

where  $\mathbf{C}_d(\mathbf{x}) := \{\mu \in \mathcal{M}; \forall k = 0, \dots, d, c_k(\mu) = c_k(\mathbf{x})\}$  and  $\lambda > 0$  is a tuning parameter. Questions immediately arise:

- How close is the recovered spike measure from the target  $\mathbf{x}$ ?
- How accurate is the localization of (3) in terms of the noise and the amplitude of the recovered/original spike?

To the best of our knowledge, this paper is the first to quantitatively address these questions in the frame of algebraic polynomials.

## 2.2. Contribution

**Definition 1** (Minimum separation). *Let  $\mathbf{T} \subset [-1, 1]$ . We define  $\Delta(\mathbf{T})$ , the minimum separation of  $\mathbf{T}$ , by*

$$\Delta(\mathbf{T}) = \min_{(t, t') \in \mathbf{T}^2; t \neq t'} \min \{d(t, t'), \pi - d(t, t')\},$$

that is the minimum modulus between two points of  $\arccos(\mathbf{T}) + \pi\mathbb{Z}$ .

Let  $\epsilon(\mathbf{T})$  denote the distance from  $\mathbf{T} \setminus \{-1, 1\}$  to the edges of  $[-1, 1]$ :

$$\epsilon(\mathbf{T}) = \min \{\min(d(t, 1), d(t, -1)); t \in \mathbf{T} \setminus \{-1, 1\}\}.$$

**Theorem 1.** *Assume  $m \geq 128$ . Let  $\eta > 0$  and set:*

$$\lambda_0 := 2\sigma[2(1 + \eta)(m - d)\log(5(m + d + 1))]^{1/2},$$

then with probability greater than  $1 - [\frac{1}{5(m+d)}]^\eta$  the following holds. If  $\lambda \geq \lambda_0$  and

$$\min\{\Delta(\mathbf{T}), 2\epsilon(\mathbf{T})\} \geq \frac{5\pi}{m}, \quad (4)$$

then there exists a solution  $\hat{\mathbf{x}}$  to (3) with finite support  $\hat{\mathbf{x}} = \sum_{k=1}^{\hat{s}} \hat{a}_k \delta_{\hat{t}_k}$  satisfying:

(i) *Global control:*

$$\sum_{k=1}^{\hat{s}} |\hat{a}_k| \min \left\{ m^2 \min_{t \in \mathbf{T}} d(t, \hat{t}_k)^2; c_0^2 \right\} \leq c_1 \lambda,$$

(ii) *Local control:*

$$\forall i = 1, \dots, s, \quad \left| a_i - \sum_{\hat{t}_k \in \text{Supp}(\hat{\mathbf{x}}) \mid d(t_i, \hat{t}_k) \leq \frac{c_0}{m}} \hat{a}_k \right| \leq c_2 \lambda,$$

(iii) *Large spike localization:*

$$\forall i = 1, \dots, s, \text{ s.t. } |a_i| > c_2 \lambda, \exists \hat{t} \in \text{Supp}(\hat{\mathbf{x}}) \text{ s.t. } d(t_i, \hat{t}) \leq \left[ \frac{c_1 \lambda}{|a_i| - c_2 \lambda} \right]^{1/2} \frac{1}{m},$$

where  $c_0 = 1.0361$ ,  $c_1 = 235.85$ , and  $c_2 = 220.72$ .

In the proof of the theorem, we will need the two following lemmas, which capitalize on the recent papers [8] and build an explicit dual certificate in the frame of algebraic polynomials. More precisely, we explicitly bound from above the dual certificates by a quadratic function near the support points, as done in [8].

**Lemma 1.** *Assume (4) holds. Then for all  $t_j \in \mathbf{T}$ , there exists a polynomial  $q_{t_j}$  of degree  $m$  such that:*

1.  $q_{t_j}(t_j) = 1$ ,
2.  $\forall t_l \in \mathbf{T} \setminus \{t_j\}, \quad q_{t_j}(t_l) = 0$ ,
3. if  $d(t, t_j) \leq c_0/m$  then:

$$1 - C_2 m^2 d(t, t_j)^2 \leq q_{t_j}(t) \leq 1 - C_1 m^2 d(t, t_j)^2,$$

4. if  $d(t, t_l) \leq c_0/m$  and  $t_l \in \mathbf{T} \setminus \{t_j\}$  then:

$$C_1 m^2 d(t, t_j)^2 \leq q_{t_j}(t) \leq C_2 m^2 d(t, t_j)^2,$$

5. if  $d(t, t_l) > c_0/m$  for all  $t_l \in \mathbf{T}$  then:

$$c_0^2 C_1 \leq q_{t_j}(t) \leq 1 - c_0^2 C_1,$$

where  $c_0 = 2\pi \cdot 0.1649$ ,  $C_1 = 0.00424$ , and  $C_2 = 0.25$ .

*Proof.* By symmetrizing the support, we can use existing results for real trigonometric polynomials. Let  $X = \frac{1}{2\pi} (\arccos(\mathbf{T}) \cup [-\arccos(\mathbf{T})]) + \frac{1}{2}$ . Note that  $X \subset [0, 1]$ . It is easy to check that (4) implies:

$$\min_{(x, x') \in X; x \neq x'} |x - x'| \geq 2.5/m \quad (5)$$

Thus, according to Proposition 2.1 and Lemma 2.5 of [8], for all  $x_j \in X$ , there exists a real trigonometric polynomial of degree  $m$ ,  $\tilde{q}_{x_j} : x \mapsto \sum_{k=-m}^m c_k e^{2i\pi kx}$ , such that:

- $\tilde{q}_{x_j}(x_j) = \tilde{q}_{x_j}(-x_j) = 1$ ,
- $|\tilde{q}_{x_j}(x)| < 1, x \in [0, 1] \setminus X$ ,
- $\tilde{q}_{x_j}(x_l) = -1, x_l \in X \setminus \{x_j, -x_j\}$ ,
- $\forall (x, x_l) \in [0, 1] \times X, |x - x_l| \leq 0.1649/m$ ,

$$|\tilde{q}_{x_j}(x)| \leq 1 - 0.3353m^2(x - x_l)^2,$$

- $\forall x \in [0, 1], \forall x_l \in X, |x - x_l| > 0.1649/m$ ,

$$|\tilde{q}_{x_j}(x)| \leq 1 - 0.3353 \cdot 0.1649^2.$$

We stress that the polynomial  $\tilde{q}_{x_j}$  as constructed in Lemma 2.2 of [8] is even. We detail the argument here. Let  $K$  stand for the square of the Fejér kernel, defined by

$$K(t) = \left[ \frac{\sin((m/2 + 1)\pi t)}{(m/2 + 1)\sin(\pi t)} \right]^4.$$

Then, in the proof of Lemma 2.5 in [8], it is shown that there exists a unique polynomial of the form

$$q(t) = \sum_{x_i \in X} \alpha_i K(t - x_i) + \beta_i K'(t - x_i) \quad (6)$$

satisfying

$$\begin{aligned} q(x_j) &= q(-x_j) = 1, \\ q(x_l) &= -1, \quad \forall x_l \in X \setminus \{x_j, -x_j\}, \\ q'(x_l) &= 0, \quad \forall x_l \in X, \end{aligned} \quad (7)$$



where  $\alpha_i$  and  $\beta_i$  are complex numbers. Using the symmetry of  $K$ , the anti-symmetry of  $K'$  and the symmetry of  $X$ , we see that the polynomial  $\tilde{q} := t \mapsto q(-t)$  is also of the form (6). Using again the symmetry of  $X$ , we have that  $\tilde{q}$  satisfies (7). By unicity,  $\tilde{q} = q$ .

Thus, the trigonometric polynomial function  $p_{x_j} : x \in [-\pi, \pi] \mapsto \tilde{q}_{x_j} \left( \frac{1}{2\pi}x + \frac{1}{2} \right)$  is real and even, so we have the expansion:

$$p_{x_j}(x) = \sum_{k=0}^m a_k \cos(kx).$$

Moreover, since  $\sup_{x \in [0, 2\pi]} |p_{x_j}(x)| = 1$ , Bernstein's inequality [4] implies:

$$\sup_{x \in [0, 1]} |p_{x_j}''(x)| \leq m^2. \quad (8)$$

Let  $t_j \in \mathbf{T}$  and  $x_j = \arccos(t_j)$ . We define:

$$q_{t_j}(t) = \frac{1}{2} p_{x_j}(\arccos t) + \frac{1}{2} = \frac{1}{2} \sum_{k=0}^m a_k T_k(t) + \frac{1}{2},$$

where  $T_k$  is the  $k$ -th Chebyshev polynomial of the first kind. Lemma 1 is a direct consequence of the properties verified by  $\tilde{q}_{x_j}$  and (8).  $\square$

**Lemma 2.** *Assume (4) holds. Then for all  $(v_1, \dots, v_S)$  such that  $|v_j| = 1$ , there exists a polynomial  $q$  of degree  $m$  such that:*

1.  $\forall j \in [1, S], q(t_j) = v_j$ ,
2. if  $d(t, t_j) \leq c_0/m$  then:

$$1 - |q(t)| \geq 2C_1 m^2 d(t, t_j)^2,$$

3. if  $d(t, t_l) > 2\pi \cdot 0.1649/m$  for all  $t_l \in \mathbf{T}$  then:

$$1 - |q(t)| \geq 2c_0^2 C_1,$$

where  $c_0 = 2\pi \cdot 0.1649$  and  $C_1 = 0.00424$ .

*Proof.* Similarly as previous lemma, if  $X = \frac{1}{2\pi} (\arccos(\mathbf{T}) \cup [-\arccos(\mathbf{T})]) + \frac{1}{2}$ , then we can construct a trigonometric polynomial  $\tilde{q} : x \mapsto \sum_{k=-m}^m c_k e^{2i\pi kx}$ , such that:

- $\tilde{q}(x_j) = \tilde{q}(-x_j) = v_j, \forall j \in [1, S],$
- $|\tilde{q}(x)| < 1, \forall x \in [0, 1] \setminus X,$
- $\forall (x, x_l) \in [0, 1] \times X, |x - x_l| \leq 0.1649/m,$

$$|\tilde{q}(x)| \leq 1 - 0.3353 m^2 (x - x_l)^2,$$

- $\forall x \in [0, 1], \forall x_l \in X, |x - x_l| > 0.1649/m,$

$$|\tilde{q}(x)| \leq 1 - 0.3353 \cdot 0.1649^2.$$

Then  $p : x \in [-\pi, \pi] \mapsto \tilde{q}\left(\frac{1}{2\pi}x + \frac{1}{2}\right)$  is even, so we have the expansion  $p(x) = \sum_{k=0}^m a_k \cos(kx)$  where  $a_k \in \mathbb{R}$ . Putting

$$q : t \mapsto \sum_{k=0}^m a_k \cos(k \arccos t) = \sum_{k=0}^m a_k T_k(t),$$

we can show  $q$  verifies the needed properties.  $\square$

*Proof of Theorem 1.* We mention that the proof of (ii), which uses Lemma 1, follows the one in [13].

Assume that  $\lambda \geq \lambda_0$  where  $\lambda_0$  is described by the following lemma (the dependence in  $\eta$  has been omitted).

**Lemma 3.** Set  $\lambda_R := \sigma [8(m-d) \log(5(m+d+1))]^{1/2}$  and  $\lambda > \lambda_R$ , then:

$$\mathbb{P} \left( \left\| \sum_{k=0}^m \varepsilon_k \varphi_k \right\|_{\infty} > \lambda \right) \leq \exp \left[ -\frac{\lambda^2 - \lambda_R^2}{8\sigma^2(m-d)} \right].$$

In particular, for all  $\eta > 0$ , if

$$\lambda_0(\eta) := \sigma [8(1+\eta)(m-d) \log(5(m+d+1))]^{1/2},$$

then

$$\mathbb{P} \left( \left\| \sum_{k=0}^m \varepsilon_k \varphi_k \right\|_{\infty} > \lambda_0(\eta) \right) \leq \frac{1}{[5(m+d+1)]^\eta}. \quad (9)$$

A proof of Lemma 3 can be found in Appendix A. Observe that the condition of the following lemma is met.

**Lemma 4.** Let  $\hat{\mathbf{x}}$  be a solution to (3). Then the following holds:

$$\forall P \in \text{Span}(\mathcal{F}), \quad \left| \int_{-1}^1 P d(\hat{\mathbf{x}} - \mathbf{x}) \right| \leq (\lambda + \lambda_0) \|P\|_\infty, \quad (10)$$

where  $\lambda_0 \geq \left\| \sum_{k=0}^m \varepsilon_k \varphi_k \right\|_\infty$ .

A proof of Lemma 4 can be found in Appendix C. One can prove that there exists a solution  $\hat{\mathbf{x}}$  to (3) with finite support, see Lemma 10. Set:

$$\hat{\mathbf{x}} = \sum_{k=1}^s \hat{a}_k \delta_{\hat{i}_k}.$$

Set  $v_j = \bar{a}_j / |a_j|$  for  $j = 1, \dots, s$  and consider  $q = \sum_{k=0}^m \beta_k \varphi_k$  the algebraic polynomial described in Lemma 2. Set:

$$\mathcal{D} := \|\hat{\mathbf{x}}\|_{TV} - \|\mathbf{x}\|_{TV} - \int_{-1}^1 q d(\hat{\mathbf{x}} - \mathbf{x}).$$

Note that  $\mathcal{D} \geq 0$ . Since  $\mathbf{x}$  is feasible, it holds:

$$\frac{1}{2} \|\mathbf{c}(\hat{\mathbf{x}}) - \mathbf{y}\|_2^2 + \lambda \mathcal{D} + \lambda \int_{-1}^1 q d(\hat{\mathbf{x}} - \mathbf{x}) \leq \frac{1}{2} \|\mathbf{e}\|_2^2.$$

Hence, using the fact that for any  $\mu \in \mathcal{M}$ ,  $\langle \mathbf{c}(\mu), \beta \rangle = \int_{-1}^1 q d\mu$ ,

$$\frac{1}{2} \|\mathbf{c}(\hat{\mathbf{x}}) - \mathbf{y} + \lambda \beta\|_2^2 + \lambda \mathcal{D} \leq \frac{1}{2} \|\mathbf{e}\|_2^2 + \frac{1}{2} \|\lambda \beta\|_2^2 - \lambda \langle \mathbf{e}, \beta \rangle.$$

Eventually,

$$\mathcal{D} \leq \frac{\lambda}{2} \left\| \beta - \frac{\mathbf{e}}{\lambda} \right\|_2^2.$$

Using Lemma 3, we have with probability greater than  $1 - \frac{1}{[5(m+d+1)]^\eta}$ :

$$\left\| \sum_{k=0}^m (\beta_k - \varepsilon_k / \lambda) \varphi_k \right\|_\infty \leq 2,$$

so that:

$$\mathcal{D} \leq 2\lambda. \quad (11)$$

Moreover, using Lemma 2, note that:

$$\begin{aligned}
\mathcal{D} &= \|\hat{\mathbf{x}}\|_{TV} - \int_{-1}^1 q d\hat{\mathbf{x}}, \\
&\geq \sum_{k=1}^{\hat{s}} |\hat{a}_k| (1 - |q(\hat{t}_k)|), \\
&\geq \sum_{k=1}^{\hat{s}} |\hat{a}_k| \min\{2C_1 m^2 \min_{t \in \mathbf{T}} d(t, \hat{t}_k)^2; 2c_0^2 C_1\}, \tag{12}
\end{aligned}$$

where  $c_0 = 2\pi \cdot 0.1649$  and  $C_1 = 0.00424$  and the proof of (i) follows.

Now, let  $t_j \in \mathbf{T}$  and consider the polynomial  $q_{t_j}$  described in Lemma 1. Using (12) we get that:

$$\begin{aligned}
& \left| \sum_{\{k \mid d(t_j, \hat{t}_k) > \frac{c_0}{m}\}} \hat{a}_k q_{t_j}(\hat{t}_k) + \sum_{\{k \mid d(t_j, \hat{t}_k) \leq \frac{c_0}{m}\}} \hat{a}_k (q_{t_j}(\hat{t}_k) - 1) \right| \\
& \leq \sum_{\{k \mid d(t_j, \hat{t}_k) > \frac{c_0}{m}\}} |\hat{a}_k| |q_{t_j}(\hat{t}_k)| + \sum_{\{k \mid d(t_j, \hat{t}_k) \leq \frac{c_0}{m}\}} |\hat{a}_k| |q_{t_j} - 1|(\hat{t}_k), \\
& \leq \sum_{k=1}^{\hat{s}} |\hat{a}_k| \min\{C_2 m^2 \min_{t \in \mathbf{T}} d(t, \hat{t}_k)^2; 1 - c_0^2 C_1\}, \\
& \leq C' \times \sum_{k=1}^{\hat{s}} |\hat{a}_k| \min\{2C_1 m^2 \min_{t \in \mathbf{T}} d(t, \hat{t}_k)^2; 2c_0^2 C_1\}, \\
& \leq 2C' \lambda. \tag{13}
\end{aligned}$$

where  $C_2 = 0.25$  and  $C' = \max\{\frac{C_2}{2C_1}, \frac{1-c_0^2 C_1}{2c_0^2 C_1}\} = 109.36$ . Invoking (10), we deduce that for all  $i = 1, \dots, s$ ,

$$\begin{aligned}
|a_i - \hat{\mathbf{x}}(t_i + \mathcal{B}(c_0/m))| &\leq \left| \int q_{t_i} d\mathbf{x} - \int q_{t_i} d\hat{\mathbf{x}} \right| \\
& \quad + \sum_{\{k \mid d(t_i, \hat{t}_k) > \frac{c_0}{m}\}} \hat{a}_k q_{t_i}(\hat{t}_k) + \sum_{\{k \mid d(t_i, \hat{t}_k) \leq \frac{c_0}{m}\}} \hat{a}_k (q_{t_i}(\hat{t}_k) - 1), \\
& \leq 2(C' + 1)\lambda,
\end{aligned}$$

where  $t_i + \mathcal{B}(c_0/m) = \{t \mid d(t_i, t) \leq c_0/m\}$ , proving (ii). Finally, observe that (iii) is a consequence of the aforementioned inequalities.  $\square$

### 3. Non-uniform spline reconstruction

#### 3.1. Notations

In this section, we assume that  $d \geq 0$ . Observe that the frame investigated in this paper covers the recovery problem of a non-uniform spline of degree  $d$  from its projection onto  $\mathbb{R}_{m-d-1}[X]$ , the space of algebraic polynomials of degree at most  $m - d - 1$ . Indeed, consider an univariate spline  $\mathbf{f}$  of degree  $d$  over the knot sequence  $\mathbf{T} = \{-1, t_1, t_2, \dots, t_s, 1\}$ , that is a continuously differentiable function  $\mathbf{f}$  of order  $d - 1$  piecewise-defined by:

$$\mathbf{f} = \mathbb{1}_{[-1, t_1)} \mathbf{P}_0 + \mathbb{1}_{[t_1, t_2)} \mathbf{P}_1 + \dots + \mathbb{1}_{[t_{s-1}, t_s)} \mathbf{P}_{s-1} + \mathbb{1}_{[t_s, 1]} \mathbf{P}_s,$$

where  $\mathbf{P}_k$  belongs to  $\mathbb{R}_d[X]$ , and for all subset  $E \subseteq [-1, 1]$ ,  $\mathbb{1}_E(t)$  equals 1 if  $t$  belongs to  $E$  and 0 otherwise. Consider  $\mathbf{f}^{(d+1)}$ , the  $(d + 1)$ -th distributional derivative of  $\mathbf{f}$ . We have :

$$\mathbf{f}^{(d+1)} = \sum_{k=1}^s (\mathbf{P}_k^{(d)} - \mathbf{P}_{k-1}^{(d)}) \delta_{t_k},$$

where  $\mathbf{P}_k^{(d)} \in \mathbb{R}$  is the  $d$ -th derivative of  $\mathbf{P}_k$ .

The next lemma links the moments of the spline  $\mathbf{f}$  to the ones of the signed measure  $\mathbf{f}^{(d+1)}$ .

**Lemma 5.**

$$\mathbf{c}(\mathbf{f}^{(d+1)}) = \begin{bmatrix} 0 & W_1 \\ (-1)^{d+1} \text{Id}_{m-d} & W_2 \end{bmatrix} \begin{pmatrix} \mathbf{p}(\mathbf{f}) \\ \mathbf{b} \end{pmatrix}, \quad (14)$$

where:

- $\mathbf{p}(\mathbf{f}) = (\langle \mathbf{f}, \varphi_{d+1}^{(d+1)} \rangle, \langle \mathbf{f}, \varphi_{d+2}^{(d+1)} \rangle, \dots, \langle \mathbf{f}, \varphi_m^{(d+1)} \rangle)$ ,
- $\mathbf{b} = (\mathbf{P}_0(-1), \dots, \mathbf{P}_0^{(d-1)}(-1), \mathbf{P}_0^{(d)}(-1), \mathbf{P}_s(1), \dots, \mathbf{P}_s^{(d-1)}(1), \mathbf{P}_s^{(d)}(1))$ ,
- and  $W_1, W_2$  are known matrices, defined by relations (16), (17) and (18), whose entries belong to the set  $\{-1, 1, \sqrt{2}(-1)^m \mathbf{w}_{k,l}\}$ ;  $m \in \{0, 1\}$  and  $k, l \in \mathbb{N}$  where  $\mathbf{w}_{k,l}$  are constants defined in (15).

*Proof.* By induction, for  $k = 0, 1, \dots, m$ ,

$$c_k(\mathbf{f}^{(d+1)}) = \langle \mathbf{f}^{(d+1)}, \varphi_k \rangle = \sum_{l=0}^d (-1)^l [\mathbf{f}^{(d-l)} \varphi_k^{(l)}]_{-1}^1 + (-1)^{d+1} \langle \mathbf{f}, \varphi_k^{(d+1)} \rangle.$$

Moreover, it is known that for all integers  $k, l$ ,  $T_k^{(l)}(-1) = (-1)^{k+l} \mathbf{w}_{k,l}$  and  $T_k^{(l)}(1) = \mathbf{w}_{k,l}$  where:

$$\mathbf{w}_{k,l} := \mathbb{1}_{\{k \geq l\}} \prod_{j=0}^{l-1} \frac{k^2 - j^2}{2j+1}. \quad (15)$$

Therefore, for  $m \geq k > d$ ,

$$\begin{aligned} c_k(\mathbf{f}^{(d+1)}) &= \sqrt{2} \sum_{l=0}^d (-1)^l \mathbf{w}_{k,l} \mathbf{P}_s^{(d-l)}(1) \\ &\quad + (-1)^{k+1} \sqrt{2} \sum_{l=0}^d \mathbf{w}_{k,l} \mathbf{P}_0^{(d-l)}(-1) + (-1)^{d+1} \langle \mathbf{f}, \varphi_k^{(d+1)} \rangle, \end{aligned} \quad (16)$$

for  $d \geq k \geq 1$ ,

$$c_k(\mathbf{f}^{(d+1)}) = \sqrt{2} \sum_{l=0}^k (-1)^l \mathbf{w}_{k,l} \mathbf{P}_s^{(d-l)}(1) + (-1)^{k+1} \sqrt{2} \sum_{l=0}^k \mathbf{w}_{k,l} \mathbf{P}_0^{(d-l)}(-1), \quad (17)$$

and

$$c_0(\mathbf{f}^{(d+1)}) = \mathbf{P}_s^{(d)} - \mathbf{P}_0^{(d)}, \quad (18)$$

as claimed.  $\square$

*Remark.* The family  $\{\varphi_{d+1}^{(d+1)}, \varphi_{d+2}^{(d+1)}, \dots, \varphi_m^{(d+1)}\}$  is a basis of  $\mathbb{R}_{m-d-1}[X]$ , so  $\mathbf{p}(\mathbf{f})$  is entirely determined by any projection of  $\mathbf{f}$  onto  $\mathbb{R}_{m-d-1}[X]$ .

*Remark.* Observe that  $\mathbf{b}$  describes the boundary conditions on  $\mathbf{f}$ . Recall that, in our model, we assume that the experimenter knows these boundary conditions. Furthermore, Equation (14) shows that the noiseless moments appearing in (19) can be determined by the boundary conditions  $\mathbf{b}$ .

### 3.2. Observation of a random perturbation

**Assumption 1** (Approximate projection of non-uniform splines). *We say that a random polynomial  $P$  with values in  $\mathbb{R}_{m-d-1}[X]$  satisfies **Assumption 1** if*

$$\Theta(P) \sim \mathcal{N}(\mathbf{p}(\mathbf{f}), \sigma^2 \text{Id}_{m-d}),$$

where  $\Theta(P) := (\langle P, \varphi_{d+1}^{(d+1)} \rangle, \langle P, \varphi_{d+2}^{(d+1)} \rangle, \dots, \langle P, \varphi_m^{(d+1)} \rangle)$ .

*Remark.* Note that Assumption 1 asserts that the experimenter observes a Gaussian perturbation (with known covariance matrix) of the inner-products of the non-uniform spline  $\mathbf{f}$  with the polynomial basis  $\{\varphi_{d+1}^{(d+1)}, \varphi_{d+2}^{(d+1)}, \dots, \varphi_m^{(d+1)}\}$ . In particular, observe that  $\|\varphi_m^{(d+1)}\|_2^2 = \mathcal{O}[(\frac{m!}{(m-d-1)!})^2]$  so that the signal-to-noise ratio (SNR) is of the order of  $(\frac{m!}{\sigma(m-d-1)!})^2$ . In applications, the standard assumption is that the SNR depends only on the noise variance. To match this situation, one needs to consider a noise level  $\sigma := \sigma_0 \frac{m!}{(m-d-1)!}$  in order to get a SNR of the order of  $1/\sigma_0^2$ . For sake of readability, we do not pursue on this idea but the simulations of this paper are made accordingly.

*Remark.* Remark that the noisy moments appearing in (19) are a Gaussian perturbation of the moments described by (14).

### 3.3. Algorithm and main theorem

Let  $P$  be a random vector with values in  $\mathbb{R}_{m-d-1}[X]$ . Set:

$$\hat{\mathbf{x}} \in \arg \min_{\mu \in \mathbf{C}_d(\mathbf{f}^{(d+1)})} \frac{1}{2} \|\mathbf{c}(\mu) - \mathbf{y}\|_2^2 + \lambda \|\mu\|_{TV}. \quad (19)$$

Recall that  $\mathbf{C}_d(\mathbf{f}^{(d+1)}) := \{\mu \in \mathcal{M}; \quad \forall k = 0, \dots, d, c_k(\mu) = c_k(\mathbf{f}^{(d+1)})\}$ ,  $\lambda > 0$  is a tuning parameter and

$$\mathbf{y} := \begin{bmatrix} 0 & W_1 \\ (-1)^{d+1} \text{Id}_{m-d} & W_2 \end{bmatrix} \begin{pmatrix} \Theta(P) \\ \mathbf{b} \end{pmatrix}.$$

*Remark.* Note that if a discrete measure  $\hat{\mathbf{x}}$  enjoys

$$\forall k = 0, \dots, d, \quad c_k(\hat{\mathbf{x}}) = c_k(\mathbf{f}^{(d+1)}) \quad (20)$$

then one can explicitly construct the unique non-uniform spline  $\hat{\mathbf{f}}$  with  $(d+1)$ -th derivative  $\hat{\mathbf{x}}$  and boundary conditions  $\mathbf{b}$ . Indeed, observe that we can uniquely construct a non-uniform spline  $\tilde{\mathbf{f}}$  from the knowledge of the  $(d+1)$  boundary conditions at point  $-1$  and its  $(d+1)$ -th derivative. Moreover, Eq.'s (20), (17) and (18) show that  $\hat{\mathbf{f}}$  satisfies the  $(d+1)$  boundary conditions at point 1 and so the boundary conditions  $\mathbf{b}$ .

**Input:** Boundary conditions  $\mathbf{b}$ , a polynomial approximation  $P$ , an upper bound  $\sigma$  on the noise standard deviation and  $\alpha > 0$  a tuning parameter.

**Output:** A non-uniform spline  $\hat{\mathbf{f}}$ .

1. Set  $d = \text{Size}(\mathbf{b})/2 - 1$  and  $m = \deg(P) + d + 1$ ,
2. Compute  $\Theta(P) = (\langle P, \varphi_{d+1}^{(d+1)} \rangle, \langle P, \varphi_{d+2}^{(d+1)} \rangle, \dots, \langle P, \varphi_m^{(d+1)} \rangle)$ ,
3. Compute  $\mathbf{y} = \begin{bmatrix} 0 & W_1 \\ (-1)^{d+1} \text{Id}_{m-d} & W_2 \end{bmatrix} \begin{pmatrix} \Theta(P) \\ \mathbf{b} \end{pmatrix}$ ,  
where  $W_1$  and  $W_2$  are described in Lemma 5.
4. Set  $\lambda = 4\sigma[2(1 + \alpha)(m - d)\log(5(m + d + 1))]^{1/2}$ ,
5. Find a discrete solution  $\hat{\mathbf{x}} = \sum_{k=1}^{\hat{s}} \hat{a}_k \delta_{\hat{\tau}_k}$  to (19)  
using SDP programming, see Appendix D,
6. Find the unique spline  $\hat{\mathbf{f}}$  of order  $d - 1$  such that  $\hat{\mathbf{f}}^{(d+1)} = \hat{\mathbf{x}}$  and  
 $(\hat{\mathbf{f}}_0(-1), \dots, \hat{\mathbf{f}}_0^{(d-1)}(-1), \hat{\mathbf{f}}_0^{(d)}, \hat{\mathbf{f}}_s(1), \dots, \hat{\mathbf{f}}_s^{(d-1)}(1), \hat{\mathbf{f}}_s^{(d)}) = \mathbf{b}$ .

**Algorithm 1:** Non-uniform spline recovery algorithm



**Theorem 2.** Let  $m > d \geq 0$ . Let  $\mathbf{f}$  be a non-uniform spline of degree  $d$  that can be written as:

$$\mathbf{f} = \mathbb{1}_{[-1, t_1)} \mathbf{P}_0 + \mathbb{1}_{[t_1, t_2)} \mathbf{P}_1 + \dots + \mathbb{1}_{[t_{s-1}, t_s)} \mathbf{P}_{s-1} + \mathbb{1}_{[t_s, 1]} \mathbf{P}_s,$$

where  $\mathbf{P}_k \in \mathbb{R}_d[X]$  and  $\mathbf{T} = \{-1, t_1, t_2, \dots, t_s, 1\}$  enjoys:

$$\min\{\Delta(\mathbf{T}), 2\epsilon(\mathbf{T})\} \geq \frac{5\pi}{m}.$$

Set  $\mathbf{b} = (\mathbf{P}_0(-1), \dots, \mathbf{P}_0^{(d-1)}(-1), \mathbf{P}_0^{(d)}, \mathbf{P}_s(1), \dots, \mathbf{P}_s^{(d-1)}(1), \mathbf{P}_s^{(d)})$  and let  $P$  be such that **Assumption 1** holds. Let  $\alpha > 0$  then, with probability greater than  $1 - \left[\frac{1}{5(m+d)}\right]^\alpha$ , any output  $\hat{\mathbf{f}}$  of **Algorithm 1** enjoys:

1. *Global control:*

$$\sum_{k=1}^s |\hat{\mathbf{P}}_k^{(d)} - \hat{\mathbf{P}}_{k-1}^{(d)}| \min \left\{ m^2 \min_{t \in \mathbf{T}} d(t, \hat{t}_k)^2; c_0^2 \right\} \leq c_1 \lambda,$$

2. *Large discontinuity localization:*  $\forall i = 1, \dots, s$ , s.t.  $|\mathbf{P}_i^{(d)} - \mathbf{P}_{i-1}^{(d)}| > c_2 \lambda$ ,

$$\exists \hat{t} \in \{\hat{t}_1, \dots, \hat{t}_s\} \text{ s.t. } d(t_i, \hat{t}) \leq \left[ \frac{c_1 \lambda}{|\mathbf{P}_i^{(d)} - \mathbf{P}_{i-1}^{(d)}| - c_2 \lambda} \right]^{1/2} \frac{1}{m},$$

where  $c_0 = 1.0361$ ,  $c_1 = 235.85$ ,  $c_2 = 220.72$ ,  $\lambda = 4\sigma[2(1 + \alpha)(m - d) \log(5(m + d + 1))]^{1/2}$  and  $\hat{\mathbf{f}}$  is written as:

$$\hat{\mathbf{f}} = \mathbb{1}_{[-1, \hat{t}_1)} \hat{\mathbf{P}}_0 + \mathbb{1}_{[\hat{t}_1, \hat{t}_2)} \hat{\mathbf{P}}_1 + \dots + \mathbb{1}_{[\hat{t}_{s-1}, \hat{t}_s)} \hat{\mathbf{P}}_{s-1} + \mathbb{1}_{[\hat{t}_s, 1]} \hat{\mathbf{P}}_s,$$

with  $\hat{\mathbf{P}}_k \in \mathbb{R}_d[X]$ .

*Proof.* From (14) deduce that if  $P$  satisfies **Assumption 1** then:

$$\mathbf{y} := \begin{bmatrix} 0 & W_1 \\ (-1)^{d+1} \text{Id}_{m-d} & W_2 \end{bmatrix} \begin{pmatrix} \Theta(P) \\ \mathbf{b} \end{pmatrix} = \mathbf{c}(\mathbf{f}^{(d+1)}) + (-1)^{d+1} \begin{pmatrix} 0 \\ \mathbf{n} \end{pmatrix},$$

where  $W_1$  and  $W_2$  are described in **Lemma 5**. Observe the result follows from **Theorem 1**.  $\square$

## 4. Numerical experiments

The semidefinite formulation of our procedure follows from standard arguments in super-resolution theory, see Appendix B and Appendix D.

We have run several numerical experiments and we have observed the following behaviour. In most cases, our approach succeeds in localizing the knots of the original spline and the amplitudes of its discontinuities while some small discontinuities may appear in the reconstructed spline.

Observe that, as can be seen in the second example of Figure 2, a small error in the estimation of the amplitude of a discontinuity may have a large impact on the reconstructed spline. More precisely, the  $\ell_\infty$ -distance between the original and reconstructed splines can be large. However, large discontinuities are well estimated (as proven in Theorem 2) so that the overall profile of the original spline is well depicted by the reconstructed spline.

Finally 3 and 4 show on an example the behaviour of our algorithm when increasing the noise level  $\sigma$ , and with degrees  $d$  higher than 1.

### A. Rice method

Define the Gaussian process  $\{X_{m,d}(t), t \in [-1, 1]\}$  by:

$$\forall t \in [-1, 1], \quad X_{m,d}(t) = \xi_{d+1}\varphi_{d+1}(t) + \xi_{d+2}\varphi_{d+2}(t) + \dots + \xi_m\varphi_m(t),$$

where  $\xi_{d+1}, \dots, \xi_m$  are i.i.d. standard normal. Its covariance function is:

$$r(s, t) = \varphi_{d+1}(t)\varphi_{d+1}(s) + \varphi_{d+2}(t)\varphi_{d+2}(s) + \dots + \varphi_m(t)\varphi_m(s),$$

where the dependence in  $m$  and  $d$  has been omitted. Observe its maximal variance is attained at point 1 and is given by  $\sigma_{m,d}^2 = 2(m-d)$ , and its variance function is  $\sigma_{m,d}^2(t) = \varphi_{d+1}(t)^2 + \varphi_{d+2}(t)^2 + \dots + \varphi_m(t)^2$ .

**Lemma 6.** Let  $\mathcal{X} = \max_{t \in [-1, 1]} |X_{m,d}(t)|$ , then:

$$\forall u > \sqrt{2(m-d)}, \quad \mathbb{P}\{\mathcal{X} > u\} \leq 5(m+d+1) \exp\left[-\frac{u^2}{8(m-d)}\right].$$

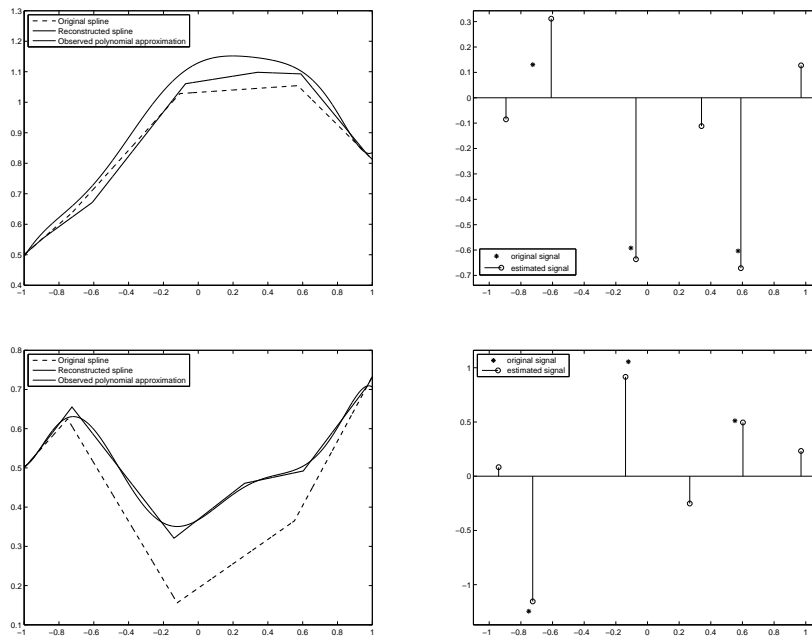


Figure 2: Left : estimated spline (thick black line) of a non-uniform spline  $f$  (thick dashed gray line) and its knots from a polynomial approximation (thin black line). Right :  $d + 1$ -derivative of the spline (stars: original spline; circles: reconstructed spline).

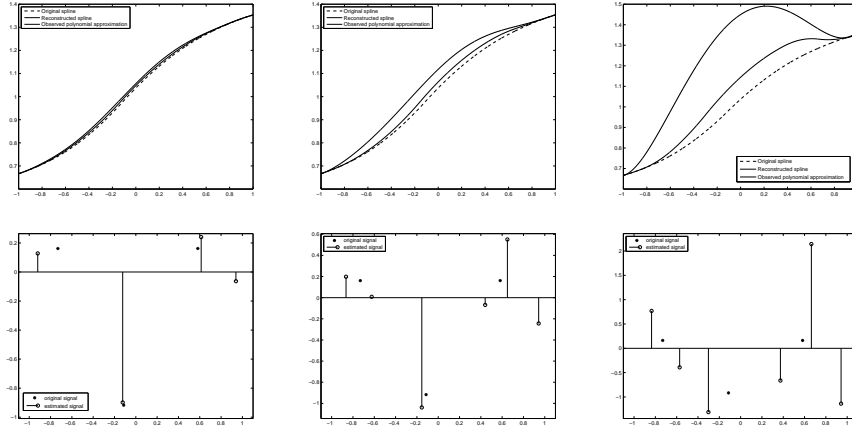


Figure 3: Top : estimated spline (thick black line) of a non-uniform spline  $f$  (thick dashed gray line) and its knots from a polynomial approximation (thin black line). Bottom : corresponding  $d + 1$ -derivative of the spline (stars: original spline; circles: reconstructed spline). Degree  $d = 2$ , number of observed noisy moments  $m - d = 8$ . Noise levels  $\sigma = \sigma_0 \frac{m!}{(m-d-1)!}$  with  $\sigma_0 \equiv 0.0005, 0.002, 0.01$  (from left to right).

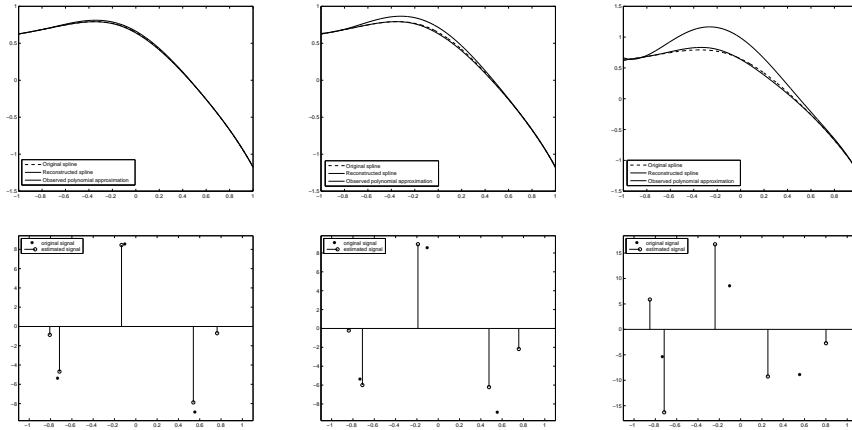


Figure 4: Top : estimated spline (thick black line) of a non-uniform spline  $f$  (thick dashed gray line) and its knots from a polynomial approximation (thin black line). Bottom : corresponding  $d + 1$ -derivative of the spline (stars: original spline; circles: reconstructed spline). Degree  $d = 3$ , number of observed noisy moments  $m - d = 7$ . Noise levels  $\sigma = \sigma_0 \frac{m!}{(m-d-1)!}$  with  $\sigma_0 \equiv 0.0005, 0.002, 0.01$  (from left to right).

*Proof.* By the change of variables  $t = \cos \theta$ , for all  $t \in [-1, 1]$ :

$$X_{m,d}(t) = X_{m,d}(\cos \theta) = \sqrt{2}\xi_{d+1} \cos((d+1)\theta) + \dots + \sqrt{2}\xi_m \cos(m\theta).$$

Set  $T(\theta) := X_{m,d}(t)$ . We recall that its variance function is given by:

$$\sigma_{m,d}^2(\theta) = 2 \cos^2((d+1)\theta) + \dots + 2 \cos^2(m\theta) = m - d + \frac{\mathbf{D}_m(2\theta) - \mathbf{D}_d(2\theta)}{2},$$

where  $\mathbf{D}_k$  denotes the Dirichlet kernel of order  $k$ . Observe that:

$$\forall \theta \in \mathbb{R}, \quad \sigma_{m,d}^2(\theta) \leq \sigma_{m,d}^2(0) = 2(m-d),$$

By the Rice method [2], for  $u > 0$ :

$$\begin{aligned} \mathbb{P}\{\mathcal{X} > u\} &\leq 2\mathbb{P}\left\{\max_{\theta \in [0, \pi]} T(\theta) > u\right\}, \\ &\leq 2\mathbb{P}\{T(0) > u\} + 2\mathbb{E}[U_u([0, \pi])], \\ &= 2 \left[ 1 - \Psi\left(\frac{u}{\sqrt{2(m-d)}}\right) \right] + 2 \int_0^\pi \mathbb{E}((T'(\theta))^+ | T(\theta) = u) \psi_{\sigma_{m,d}(\theta)}(u) d\theta \end{aligned}$$

where  $U_u$  is the number of crossings of the level  $u$ ,  $\Psi$  is the c.d.f. of the standard normal distribution, and  $\psi_\sigma$  is the density of the centered normal distribution with standard error  $\sigma$ . First, observe that for  $v > 0$ ,  $(1 - \Psi(v)) \leq \frac{1}{2} \exp(-v^2/2)$ . Hence,

$$1 - \Psi\left(\frac{u}{\sqrt{2(m-d)}}\right) \leq \frac{1}{2} \exp\left(-\frac{u^2}{4(m-d)}\right).$$

Moreover, regression formulas implies that:

$$\begin{aligned} \mathbb{E}(T'(\theta) | T(\theta) = u) &= \frac{r_{0,1}(\theta, \theta)}{r(\theta, \theta)} u, \\ \text{Var}(T'(\theta) | T(\theta) = u) &\leq \text{Var}(T'(\theta)) = r_{1,1}(\theta, \theta), \end{aligned}$$

where, for instance,  $r_{1,1}(v, \theta) = \frac{\partial^2 r(v, \theta)}{\partial v \partial \theta}$ . We recall that the covariance function is given by:

$$\begin{aligned} r(v, \theta) &= 2 \cos((d+1)v) \cos((d+1)\theta) + \dots + 2 \cos(mv) \cos(m\theta), \\ &= \frac{1}{2} [\mathbf{D}_m(v - \theta) + \mathbf{D}_m(v + \theta) - \mathbf{D}_d(v - \theta) - \mathbf{D}_d(v + \theta)]. \end{aligned}$$

Observe that:

$$r_{0,1}(\theta, \theta) = \frac{1}{2} [\mathbf{D}'_m(2\theta) - \mathbf{D}'_d(2\theta)] = - \sum_{k=d+1}^m k \sin(2k\theta),$$

$$r_{1,1}(\theta, \theta) = \sum_{k=d+1}^m k^2 (1 - \cos(2k\theta)).$$

On the other hand, if  $Z \sim \mathcal{N}(\mu, \sigma^2)$  then

$$\mathbb{E}(Z^+) = \mu \Psi\left(\frac{\mu}{\sigma}\right) + \sigma \psi\left(\frac{\mu}{\sigma}\right) \leq \mu^+ + \frac{\sigma}{\sqrt{2\pi}},$$

where  $\psi$  is the standard normal density. We get that:

$$\begin{aligned} & \int_0^\pi \mathbb{E}((T'(\theta))^+ | T(\theta) = u) \psi_{\sigma_{m,d}(\theta)}(u) dt \\ & \leq \int_0^\pi \frac{[\mathbf{D}'_m(2\theta) - \mathbf{D}'_d(2\theta)]^+}{2\sigma_{m,d}^2(\theta)} u \psi_{\sigma_{m,d}(\theta)}(u) d\theta \\ & + \frac{1}{\sqrt{2\pi}} \int_0^\pi \left[ \sum_{k=d+1}^m k^2 (1 - \cos(2k\theta)) \right]^{1/2} \psi_{\sigma_{m,d}(\theta)}(u) d\theta, \\ & = A + B. \end{aligned}$$

We use the following straightforward relations:

- $\forall 0 < \sigma_1 < \sigma_2 < u, \quad \psi_{\sigma_1}(u) \leq \psi_{\sigma_2}(u),$
- $\forall \theta, \quad [\mathbf{D}'_m(2\theta) - \mathbf{D}'_d(2\theta)]^+ \leq \sum_{k=d+1}^m k = \frac{(m+d+1)(m-d)}{2},$
- $\forall \theta \in [0, \pi],$

$$\frac{u}{2\sigma_{m,d}^2(\theta)} \psi_{\sigma_{m,d}(\theta)}(u) \leq \frac{1}{2\sqrt{2\pi}u^2} \frac{u^3}{\sigma_{m,d}^3(\theta)} e^{-\frac{u^2}{4\sigma_{m,d}^2(\theta)}} e^{-\frac{u^2}{4\sigma_{m,d}^2(\theta)}} \leq \frac{2}{3u^2} e^{-\frac{u^2}{8(m-d)}}.$$

Eventually, we get, for  $u > \sqrt{2(m-d)}$ :

$$A \leq \frac{\pi}{3} \frac{(m+d+1)(m-d)}{u^2} \exp\left(-\frac{u^2}{8(m-d)}\right),$$

$$B \leq \left[ \frac{\pi}{12} ((2m+1)(m+1)m - (2d+1)(d+1)d) \right]^{1/2} \psi_{\sqrt{2(m-d)}}(u).$$

and the result follows.  $\square$

As a corollary, we deduce Lemma 3.

## B. Fenchel dual and first order conditions

**Lemma 7.** *The program:*

$$\inf_{\mu \in \mathbf{C}_d(\mathbf{x})} \frac{1}{2} \|\mathbf{c}(\mu) - \mathbf{y}\|_2^2 + \lambda \|\mu\|_{TV}, \quad (\text{B.1})$$

has Fenchel dual program:

$$- \inf_{\|\sum_{k=0}^m \alpha_k \varphi_k\|_\infty \leq \lambda} \left\{ \langle \alpha, \mathbf{y} \rangle + \frac{1}{2} \sum_{k=d+1}^m \alpha_k^2 \right\}. \quad (\text{B.2})$$

Moreover, there is no duality gap.

*Proof.* The case  $d = -1$  has been treated in [1]. Assume that  $d \geq 0$ . Program (B.1) can be viewed as:

$$\inf_{\mu \in \mathcal{M}} h(\mathbf{c}(\mu)) + \psi_1(\mu) + \psi_2(\mu),$$

where  $h(c) := (1/2)\|c - \mathbf{y}\|_2^2$ ,  $\psi_1(\mu) := \lambda \|\mu\|_{TV}$  and  $\psi_2(\mu) := u_{\mathbf{C}_d(\mathbf{x})}(\mu)$ , with:

$$u_{\mathbf{C}_d(\mathbf{x})}(\mu) = \begin{cases} 0 & \text{if } \mu \in \mathbf{C}_d(\mathbf{x}), \\ \infty & \text{otherwise.} \end{cases}$$

Note the function  $h$  has Legendre conjugate:

$$\forall \alpha \in \mathbb{R}^{m+1}, \quad h^*(\alpha) = \langle \alpha, \mathbf{y} \rangle + \frac{1}{2} \|\alpha\|_2^2,$$

One can check that the function  $\psi_1$  has Legendre conjugate:

$$\forall f \in \mathcal{C}([-1, 1]), \quad \psi_1^*(f) = \sup_{\mu \in \mathcal{M}} \int f d\mu - \lambda \|\mu\|_{TV} = u_{\mathbf{B}_\infty(\lambda)}(f),$$

where:

$$u_{\mathbf{B}_\infty(\lambda)}(f) = \begin{cases} 0 & \text{if } \|f\|_\infty \leq \lambda, \\ \infty & \text{otherwise.} \end{cases}$$

Indeed, we have, for all  $\mu \in \mathcal{M}$ ,  $\int f d\mu - \lambda \|\mu\|_{TV} \leq (\|f\|_\infty - \lambda) \|\mu\|_{TV}$ , showing that the supremum over  $\mu$  is 0 if  $\|f\|_\infty \leq \lambda$ . If  $\|f\|_\infty > \lambda$ , define, for every  $A > 0$ ,  $\mu_A = A \operatorname{sg}(f(x_0)) \delta_{x_0}$  where  $x_0$  is such that  $\|f\|_\infty = |f(x_0)|$ . Then  $\int f d\mu_A - \lambda \|\mu_A\|_{TV} = A(\|f\|_\infty - \lambda)$  for every  $A > 0$ , which completes proving our claim.

Let us turn to the Legendre conjugate of  $\psi_2$ . We show that

$$\forall f \in \mathcal{C}([-1, 1]), \quad \psi_2^*(f) = \sup_{\mu \in \mathbf{C}_d(\mathbf{x})} \int f d\mu = \begin{cases} \sum_{k=0}^d a_k y_k & \text{if } f = \sum_{k=0}^d a_k \varphi_k, \\ \infty & \text{otherwise.} \end{cases}$$

Indeed, the result is obvious if  $f$  is of the form  $f = \sum_{k=0}^d a_k \varphi_k$ . In the other case, recall that  $\{\varphi_k\}_{k \geq 0}$  is a complete orthonormal family of  $L^2([-1, 1], \mathbb{M})$  where  $\mathbb{M}(dt) = (1/\pi)(1-t^2)^{-1/2} dt$  ( $dt$  denotes the Lebesgue measure). Thus, in this Hilbert space,  $f$  can be expanded as  $f = \sum_{k=0}^{\infty} a_k \varphi_k$  with  $a_p \neq 0$  for som  $p > d$ . Define the measure  $\mu_1(dt) = \varphi_p(t)\mathbb{M}(dt)$ . Observe that  $\int \varphi_k d\mu_1 = \delta_{kp}$  and  $\int f d\mu_1 = a_p$ . Let  $\mu_0 \in \mathbf{C}_d(\mathbf{x})$ , and  $\mu_A = \mu_0 + A\mu_1$  for every  $A \in \mathbb{R}$ . Then  $\mu_A \in \mathbf{C}_d(\mathbf{x})$  and  $\int f d\mu_A = \sum_{k=0}^d a_k y_k + Aa_p, \forall A \in \mathbb{R}$ . This proves our claim.

Let  $f \in \mathcal{C}([-1, 1])$ . The Legendre conjugate  $\psi^*$  of  $\psi := \psi_1 + \psi_2$  at  $f$  is given by:

$$\psi^*(f) = \inf_{f=f_1+f_2} \psi_1^*(f_1) + \psi_2^*(f_2) = \inf_{a_0, \dots, a_d \in \mathbb{R}} u_{\mathbb{B}_\infty(\lambda)}(f - \sum_{k=0}^d a_k \varphi_k) + \sum_{k=0}^d a_k y_k. \quad (\text{B.3})$$

Indeed, observe that the bi-conjugate of  $\psi_1$  (resp.  $\psi_2$ ) enjoys  $\psi_1^{**} = \psi_1$  (resp.  $\psi_2^{**} = \psi_2$ ) and it holds:

$$\begin{aligned} \psi^*(f) &= \sup_{\mu \in \mathcal{M}} \left\{ \int f d\mu - \psi_1(\mu) - \psi_2(\mu) \right\}, \\ &= \sup_{\mu \in \mathcal{M}} \left\{ \int f d\mu - \sup_{f_1} \left\{ \int f_1 d\mu - \psi_1^*(f_1) \right\} - \sup_{f_2} \left\{ \int f_2 d\mu - \psi_2^*(f_2) \right\} \right\}, \\ &= \inf_{f_1, f_2} \sup_{\mu \in \mathcal{M}} \left\{ \int (f - f_1 - f_2) d\mu + \psi_1^*(f_1) + \psi_2^*(f_2) \right\}, \\ &= \inf_{f=f_1+f_2} \psi_1^*(f_1) + \psi_2^*(f_2). \end{aligned}$$

Moreover, observe that the dual operator  $\mathbf{c}^*$  of  $\mathbf{c}$  is given by:

$$\forall \alpha \in \mathbb{R}^{m+1}, \quad \mathbf{c}^*(\alpha) = \sum_{k=0}^m \alpha_k \varphi_k.$$



Observe that the bi-conjugate of  $h$  enjoys  $h^{**} = h$ . Then, notice that:

$$\begin{aligned} \inf_{\mu \in \mathcal{M}} h(\mathbf{c}(\mu)) + \psi(\mu) &= \inf_{\mu \in \mathcal{M}} \sup_{\alpha \in \mathbb{R}^{m+1}} \{ \langle \alpha, \mathbf{c}(\mu) \rangle - h^*(\alpha) \} + \psi(\mu), \\ &= \sup_{\alpha \in \mathbb{R}^{m+1}} -h^*(\alpha) - \sup_{\mu \in \mathcal{M}} \{ \langle -\mathbf{c}^*(\alpha), \mu \rangle - \psi(\mu) \}, \\ &= - \inf_{\alpha \in \mathbb{R}^{m+1}} h^*(\alpha) + \psi^*(-\mathbf{c}^*(\alpha)). \end{aligned}$$

It follows that the program (B.1) has Fenchel dual:

$$- \inf_{\alpha \in \mathbb{R}^{m+1}} h^*(\alpha) + \psi^*(-\mathbf{c}^*(\alpha)) = - \inf_{\|\sum_{k=0}^m \alpha_k \varphi_k\|_\infty \leq \lambda} \left\{ \langle \alpha, \mathbf{y} \rangle + \frac{1}{2} \sum_{k=d+1}^m \alpha_k^2 \right\}.$$

Slater's condition shows that strong duality holds.  $\square$

**Lemma 8.** *The first order conditions read: There exists  $\hat{a}_0, \dots, \hat{a}_d \in \mathbb{R}$  such that*

$$\|\hat{P}\|_\infty \leq \lambda \quad \text{and} \quad \lambda \|\hat{\mathbf{x}}\|_{TV} = \int_{-1}^1 \hat{P} d(\hat{\mathbf{x}}), \quad (\text{B.4})$$

where:

$$\hat{P} = \sum_{k=0}^d \hat{a}_k \varphi_k + \sum_{k=d+1}^m (y_k - c_k(\hat{\mathbf{x}})) \varphi_k.$$

*Proof.* Let  $\mu \in \mathbf{C}_d(\mathbf{x})$  and  $\gamma \in (0, 1)$ . Set  $\nu = \hat{\mathbf{x}} + \gamma(\mu - \hat{\mathbf{x}})$  then, by convexity:

$$\|\mu\|_{TV} - \|\hat{\mathbf{x}}\|_{TV} \geq \frac{1}{\gamma} (\|\nu\|_{TV} - \|\hat{\mathbf{x}}\|_{TV}).$$

Observe that  $\nu \in \mathbf{C}_d(\mathbf{x})$ , by optimality:

$$\begin{aligned} \lambda (\|\nu\|_{TV} - \|\hat{\mathbf{x}}\|_{TV}) &\geq \frac{1}{2} (\|\mathbf{c}(\hat{\mathbf{x}}) - \mathbf{y}\|_2^2 - \|\mathbf{c}(\nu) - \mathbf{y}\|_2^2), \\ &= \gamma \langle \mathbf{y} - \mathbf{c}(\hat{\mathbf{x}}), \mathbf{c}(\mu) - \mathbf{c}(\hat{\mathbf{x}}) \rangle - \frac{\gamma^2}{2} \|\mathbf{c}(\mu) - \mathbf{c}(\hat{\mathbf{x}})\|_2^2. \end{aligned}$$

Letting  $\gamma$  go to 0, we deduce:

$$\forall \mu \in \mathbf{C}_d(\mathbf{x}), \quad \lambda (\|\mu\|_{TV} - \|\hat{\mathbf{x}}\|_{TV}) \geq \langle \mathbf{y} - \mathbf{c}(\hat{\mathbf{x}}), \mathbf{c}(\mu) - \mathbf{c}(\hat{\mathbf{x}}) \rangle. \quad (\text{B.5})$$

Conversely, if (B.5) holds then, for all  $\mu \in \mathbf{C}_d(\mathbf{x})$ :

$$\begin{aligned} \frac{1}{2} \|\mathbf{c}(\mu) - \mathbf{y}\|_2^2 + \lambda \|\mu\|_{TV} &\geq \frac{1}{2} \|\mathbf{c}(\hat{\mathbf{x}}) - \mathbf{y} + \mathbf{c}(\mu) - \mathbf{c}(\hat{\mathbf{x}})\|_2^2 \\ &\quad + \langle \mathbf{y} - \mathbf{c}(\hat{\mathbf{x}}), \mathbf{c}(\mu) - \mathbf{c}(\hat{\mathbf{x}}) \rangle + \lambda \|\hat{\mathbf{x}}\|_{TV}, \\ &= \frac{1}{2} \|\mathbf{c}(\hat{\mathbf{x}}) - \mathbf{y}\|_2^2 + \lambda \|\hat{\mathbf{x}}\|_{TV} + \frac{1}{2} \|\mathbf{c}(\mu) - \mathbf{c}(\hat{\mathbf{x}})\|_2^2, \\ &\geq \frac{1}{2} \|\mathbf{c}(\hat{\mathbf{x}}) - \mathbf{y}\|_2^2 + \lambda \|\hat{\mathbf{x}}\|_{TV}. \end{aligned}$$

Therefore, Eq. (B.5) is a necessary and sufficient condition for the measure  $\hat{\mathbf{x}}$  to be a solution to (3). In particular, it follows:

$$\lambda \|\hat{\mathbf{x}}\|_{TV} - \langle \mathbf{y} - \mathbf{c}(\hat{\mathbf{x}}), \mathbf{c}(\hat{\mathbf{x}}) \rangle \leq \inf_{\mu \in \mathcal{C}_d(\mathbf{x})} \{ \lambda \|\mu\|_{TV} - \langle \mathbf{y} - \mathbf{c}(\hat{\mathbf{x}}), \mathbf{c}(\mu) \rangle \} = -\psi^*(\hat{Q}),$$

where  $\psi^*$  is defined by (B.3) and  $\hat{Q} = \sum_{k=d+1}^m (y_k - c_k(\hat{\mathbf{x}}))\varphi_k$ . The optimality conditions can be deduced from (B.3).  $\square$

### C. Proof of Lemma 4

Let  $(a_k)_{k=0}^m$  be the coefficients of  $P$ , namely:

$$P = \sum_{k=0}^m a_k \varphi_k.$$

Since  $\mathcal{F}$  is an orthonormal family of  $L^2(\mathbb{M})$ , it holds

$$\begin{aligned} \int_{-1}^1 P d(\hat{\mathbf{x}} - \mathbf{x}) &= \sum_{k=0}^m a_k \int_{-1}^1 \varphi_k d(\hat{\mathbf{x}} - \mathbf{x}), \\ &= \sum_{k=d+1}^m a_k (c_k(\hat{\mathbf{x}}) - c_k(\mathbf{x})), \\ &= \int_{-1}^1 \left( \sum_{k=d+1}^m a_k \varphi_k \right) \left( \sum_{k=d+1}^m (c_k(\hat{\mathbf{x}}) - c_k(\mathbf{x})) \varphi_k \right) d\mathbb{M}, \\ &= \int_{-1}^1 \left( \sum_{k=d+1}^m a_k \varphi_k \right) \left( -\sum_{k=0}^d \hat{a}_k \varphi_k + \sum_{k=d+1}^m (c_k(\hat{\mathbf{x}}) - c_k(\mathbf{x})) \varphi_k \right) d\mathbb{M}, \\ &= \int_{-1}^1 \left( \sum_{k=d+1}^m a_k \varphi_k \right) \left( \sum_{k=0}^m \varepsilon_k \varphi_k - \hat{P} \right) d\mathbb{M}, \end{aligned}$$

where  $\hat{a}_0, \dots, \hat{a}_d \in \mathbb{R}$  and  $\hat{P}$  are given by Lemma B.4. By Cauchy-Schwarz inequality, it yields:

$$\begin{aligned} \int_{-1}^1 P d(\hat{\mathbf{x}} - \mathbf{x}) &\leq \left\| \sum_{k=d+1}^m a_k \varphi_k \right\|_2 \left\| \sum_{k=0}^m \varepsilon_k \varphi_k - \hat{P} \right\|_2, \\ &\leq \|P\|_2 \left\| \sum_{k=0}^m \varepsilon_k \varphi_k - \hat{P} \right\|_\infty, \\ &\leq \|P\|_\infty \left( \left\| \sum_{k=0}^m \varepsilon_k \varphi_k \right\|_\infty + \|\hat{P}\|_\infty \right), \end{aligned}$$

where  $\|\cdot\|_2$  stands for the norm associated to the Hilbert space  $L^2(\mathbb{M})$  for which  $\mathcal{F}$  is an orthonormal family. The result follows from (B.4).

## D. Background on Semi-Definite Programming in Super-Resolution

### *Zero-noise problem*

In the noiseless case, observe that  $\mathbf{n} = 0$ . Exact recovery from moment samples has been investigated in [1, 3] where one considers the program:

$$\mathbf{x}^0 \in \arg \min_{\mu \in \mathcal{M}} \|\mu\|_{TV} \quad \text{s.t.} \quad \int \Phi d\mu = \int \Phi dx, \quad (\text{D.1})$$

where  $\Phi = (\varphi_0, \dots, \varphi_m)$  is the Chebyshev moment curve. The optimality condition of (D.1) shows that the sub-gradient of the  $\ell_1$ -norm vanishes at any solution point  $\mathbf{x}^0$ . Therefore a sufficient condition for exact recovery is that  $\mathbf{x}$  satisfies the optimality condition. This is covered by the notion of “dual certificate” [9, 8] or equivalently the notion of “source condition” [6].

**Definition 2** (Dual certificate). *We say that a polynomial  $P = \sum_{k=0}^m \alpha_k \varphi_k$  is a dual certificate for the measure  $\mathbf{x}$  defined by (1) if and only if it satisfies the following properties:*

- *sign interpolation:  $\forall k \in \{1, \dots, S\}, P(\mathbf{t}_k) = a_k/|a_k|$ ,*
- *$\ell_\infty$ -constraint:  $\|P\|_\infty \leq 1$ .*

One can prove [9] that  $\mathbf{x}$  is a solution to (D.1) if and only if  $\mathbf{x}$  has a dual certificate.

### *Semi-noisy moment sample model*

In our model, we deal with an observation  $\mathbf{y}$  described by (2). In this case, the existence of a dual certificate is not sufficient to derive support localization, see for instance [1]. One needs to strengthen this notion using the Quadratic Isolation Condition [1].

**Definition 3** (Quadratic isolation condition). *A finite set  $\mathbf{T} = \{t_1, \dots, t_s\} \subset [-1, 1]$  satisfies the quadratic isolation condition with parameters  $C_a > 0$  and  $0 < C_b < 1$ ,*

denoted by  $\text{QIC}(C_a, C_b)$ , if and only if for all  $\{\theta_k\}_{k=1}^s \in \mathbb{R}^s$ , there exists  $P \in \text{Span}(\mathcal{F})$  such that for all  $k = 1, \dots, s$ ,  $P(\mathbf{t}_k) = \exp(-i\theta_k)$ , and

$$\forall x \in [-1, 1], \quad 1 - |P(x)| \geq \min_{t \in \mathbf{T}} \{C_a m^2 d(x, t)^2, C_b\}.$$

As showed by Lemma 2, if the support  $\mathbf{T}$  satisfy a minimal separation condition described in (4) then  $\mathbf{T}$  satisfies  $\text{QIC}(C_a, C_b)$  with constants  $C_a = 0.00848$  and  $C_b = 0.00879$ .

### Semi-definite programming

Observe that the Fenchel dual program of (3) is given by:

$$\tilde{\alpha} \in \arg \min_{\|\sum_{k=0}^m \alpha_k \varphi_k\|_\infty \leq \lambda} \left\{ \langle \alpha, y \rangle + \frac{1}{2} \sum_{k=d+1}^m \alpha_k^2 \right\}, \quad (\text{D.2})$$

and strong duality holds, see Lemma 7. Moreover, observe that the constraint  $\|\sum_{k=0}^m \alpha_k \varphi_k\|_\infty \leq \lambda$  can be re-cast as imposing that the algebraic polynomials:

$$P_1 := \lambda + \sum_{k=0}^m \alpha_k \varphi_k \geq 0 \quad \text{and} \quad P_2 := \lambda - \sum_{k=0}^m \alpha_k \varphi_k \geq 0. \quad (\text{D.3})$$

Considering the change of variables  $\theta = \arccos(t)$ , the aforementioned inequalities can be equivalently drawn for some trigonometric polynomials. Using Riesz-Fejér theorem, one can show that non-negative trigonometric polynomials are sums of squares polynomials (SOS). A standard result, see for instance [11], ensures that the convex set of sum of square polynomials (SOS) can be described as the intersection between the set of positive hermitian semi-definite (SDP) matrices and an affine constraint.

**Lemma 9.** *The constraint (D.3) can be re-casted into a semi-definite constraint.*

Hence, we can compute  $\tilde{\alpha}$  using a SDP program. Moreover, Fenchel's duality theorem shows that the dual polynomial:

$$\hat{P} = \frac{1}{\lambda} \sum_{k=0}^m \tilde{\alpha}_k \varphi_k,$$

is a sub-gradient of the  $TV$ -norm at point  $\hat{\mathbf{x}}$ . In particular, the support  $\hat{\mathbf{T}}$  of  $\hat{\mathbf{x}}$  is included in:

$$\{t \in [-1, 1], \quad |\hat{P}| = 1\}.$$

If  $\hat{P}$  is not constant, this level set has at most  $m + 1$  points and it defines the support of the solution. Hence, we can find the weights of  $\hat{\mathbf{x}}$  using a least-square-type estimator subject to the affine constraint given by the intersection between  $\mathbf{C}_d(\mathbf{x})$  and discrete measures with support included in  $\hat{\mathbf{T}}$ . In this case, the solution to (3) is unique and can be computed using the aforementioned SDP program. If  $\hat{P}$  is constant then there always exists a solution to (3) with finite support. Indeed, using the fact that there is no duality gap, one can check that the solution has non-negative (resp. non-positive) weights if  $\hat{P} = 1$  (resp.  $\hat{P} = -1$ ). Therefore, Carathéodory's theorem shows that there always exists a solution with finite support<sup>1</sup>. However, one can not use the dual program (D.2) to compute the solution to the primal program (3). We deduce the following lemma.

**Lemma 10.** *There always exists a solution to the primal problem (3) with a support of size at most  $m + 2$ . Moreover, if  $\hat{P}$  is not constant, the solution to (3) is unique, its support is included in the level set  $\{t \in [-1, 1], |\hat{P}| = 1\}$  and has size at most  $m + 1$ .*

- [1] Azais, J.-M., De Castro, Y., Gamboa, F., 2014. Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis*.
- [2] Azais, J.-M., Wschebor, M., 2008. A general expression for the distribution of the maximum of a Gaussian field and the approximation of the tail. *Stochastic Processes and their Applications* 118 (7), 1190–1218.
- [3] Bendory, T., Dekel, S., Feuer, A., 2014. Exact recovery of non-uniform splines from the projection onto spaces of algebraic polynomials. *Journal of Approximation Theory* 182, 7–17.

---

<sup>1</sup>The interested reader may find a valuable reference on the geometry of the cone of non-negative measures in [17].

- [4] Borwein, P., Erdélyi, T., Zhang, J., 1994. Chebyshev polynomials and Markov–Bernstein type inequalities for rational spaces. *Journal of the London Mathematical Society* 50 (3), 501–519.
- [5] Bredies, K., Pikkarainen, H. K., 2013. Inverse problems in spaces of measures. *ESAIM: Control, Optimisation and Calculus of Variations* 19 (01), 190–218.
- [6] Burger, M., Osher, S., 2004. Convergence rates of convex variational regularization. *Inverse problems* 20 (5), 1411.
- [7] Candès, E. J., Fernandez-Granda, C., 2013. Super-resolution from noisy data. *Journal of Fourier Analysis and Applications* 19 (6), 1229–1254.
- [8] Candès, E. J., Fernandez-Granda, C., 2014. Towards a Mathematical Theory of Super-resolution. *Communications on Pure and Applied Mathematics* 67 (6), 906–956.
- [9] De Castro, Y., Gamboa, F., 2012. Exact reconstruction using Beurling minimal extrapolation. *Journal of Mathematical Analysis and applications* 395 (1), 336–354.
- [10] Donoho, D., 1992. Superresolution via sparsity constraints. *SIAM Journal on Mathematical Analysis* 23 (5), 1309–1331.
- [11] Dumitrescu, B., 2007. *Positive trigonometric polynomials and signal processing applications*. Springer.
- [12] Duval, V., Peyré, G., 2013. Exact Support Recovery for Sparse Spikes Deconvolution. *arXiv preprint arXiv:1306.6909*.
- [13] Fernandez-Granda, C., 2013. Support detection in super-resolution. *arXiv preprint arXiv:1302.3921*.
- [14] Gottlieb, D., Orszag, S. A., 1977. *Numerical analysis of spectral methods*. Vol. 2. SIAM.

- [15] Henrot, A., Pierre, M., 2006. Variation et optimisation de formes: une analyse géométrique. Vol. 48. Springer.
- [16] Kalnay, E., 2003. Atmospheric modeling, data assimilation, and predictability. Cambridge university press.
- [17] Krein, M. G., Nudelman, A. A., 1977. The Markov moment problem and extremal problems, volume 50 of Translations of mathematical monographs. American Mathematical Society, Providence, Rhode Island.
- [18] Tang, G., Bhaskar, B. N., Recht, B., 2013. Near minimax line spectral estimation. In: Information Sciences and Systems (CISS), 2013 47th Annual Conference on. IEEE, pp. 1–6.