OPEN ACCESS
IOP Publishing
Inverse Problems
Inverse Problems **41** (2025) 045008 (35pp)
https://doi.org/10.1088/1361-6420/adc0b5

Convergence analysis of the discretization of continuous-domain inverse problems

Vincent Guillemet*, **Julien Fageot and Michael Unser**

Biomedical Imaging Group, EPFL, Lausanne, Vaud CH-1015, Switzerland

E-mail: vincent.guillemet@epfl.ch, julien.fageot@epfl.ch and michael.unser@epfl.ch

Received 18 July 2024; revised 12 December 2024
Accepted for publication 14 March 2025
Published 25 March 2025

Abstract

We study continuous-domain linear inverse problems that involve a general data-fidelity term and a regularisation term. We consider a regularisation that is formed by the sparsity-promoting total-variation norm, pre-composed with a differential operator that specifies some underlying dictionary of atoms. It has been previously shown that such problems have sparse spline solutions with adaptive knots. These knots are part of the parameterization of the solution and their estimation is itself a difficult non-convex problem. To alleviate this difficulty, we rely on an exact discretization of the optimization problem, where the spline knots are chosen on a dense regular grid. We then follow a multiresolution strategy to refine this grid. In this work, we investigate the convergence of the discretization to the original continuous-domain problem when the grid goes from coarse to fine. We provide an in-depth study of this convergence, concluding that its strength depends on the regularity of the Green's function of the differential operator. We show that uniform convergence holds in very general settings. We carry a numerical analysis to illustrate our theoretical results.

Keywords: sparsity, convergence of optimization problems, rate of convergence, splines, total variation norm

---

* Author to whom any correspondence should be addressed.

1. Introduction

Inverse problems (IPs) play a central role in biomedical imaging [3, 33]. When the problem is underdetermined or the data noisy, its efficient resolution is still a major challenge. In this paper, we focus our attention on the recovery of an *unknown continuous* function $f:\mathbb{R}^D\to\mathbb{R}$ from a *finite-dimensional* measurement vector $\mathbf{y}=(y_m)_{m=1}^M\in\mathbb{R}^M$. The data $\mathbf{y}$ have been queried by the application of a *known* (linear) measurement operator $\boldsymbol{\nu}=(\nu_1,\ldots,\nu_M)$ on a function f with

$$y_m=\langle f,\nu_m\rangle+\epsilon_m=\int_{\mathbb{R}^D}\nu_m(x)f(x)\,\mathrm{d}x+\epsilon_m,\quad m=1,\ldots,M, \tag{1}$$

where ϵ_m is a noise, usually assumed to be small and independent of $\langle f,\nu_m\rangle$. For the estimation of $f:\mathbb{R}^D\to\mathbb{R}$, there exist two philosophies, depending on whether the IP is formulated in finite or infinite dimensions.

1.1. Discrete-domain IP

In this setting, f is assumed to be an element of a finite-dimensional vector space, so that $f=\sum_{k=1}^K a_k\phi_k$, where $\{\phi_k\}_{k=1}^K$ is a basis for the vector space in which f lives, and $\mathbf{a}=(a_k)_{k=1}^K\in\mathbb{R}^K$ is the vector of coefficients parameterizing f. The estimation of f from $\mathbf{y}$ then reduces to the estimation of $\mathbf{a}$. This *discrete-domain* inverse problem is typically solved, for $p=1,2$, by the resolution of

$$\operatorname*{argmin}_{\mathbf{a}\in\mathbb{R}^K}\left(\|\mathbf{y}-\mathbf{A}\mathbf{a}\|_2^2+\lambda\|\mathbf{L}\mathbf{a}\|_p^p\right),\quad\text{with}\quad\mathbf{A}_{m,k}=\langle\phi_k,\nu_m\rangle=\int_{\mathbb{R}^D}\nu_m(x)\,\phi_k(x)\,\mathrm{d}x, \tag{2}$$

where $p=2$ corresponds to the Tikhonov ℓ^2-regularisation [40] and $p=1$ corresponds to the sparsity-promoting regularisation typical of compressed sensing [21, 28]. Here, promoting $\|\mathbf{y}-\mathbf{A}\mathbf{a}\|_2^2$ plays the role of the data-fidelity term. The term $\lambda\|\mathbf{L}\mathbf{a}\|_p^p$ in (2) corresponds to the sparsity regularisation whose strength is modulated by λ. There, the (regularizing) matrix $\mathbf{L}$ acts as a change of basis for the coefficients $\mathbf{a}$ and is typically chosen as the *discretization* of a classical differential operator (gradient, for example [37]). Although broadly used, this philosophy has limitations. For instance, one can regret the arbitrariness of the choice of the basis functions ϕ_k that implicitly control the resolution of the reconstruction. Another issue is that ϕ_k may interfere with the regularisation. Recent advances at the intersection of variational splines and inverse problems [42] aim at overcoming some of the limitations of discrete-domain IP.

1.2. Continuous-domain IP

In infinite dimensions, which is the focus of this paper, f is assumed to live in a continuum (infinite-dimensional vector space), thus leading to a *continuous-domain* inverse problem. In this paper, we fix $D=1$, and consider a formulation that is the continuous analog of (2) for $p=1$ and that fits within the general framework of [43]. Specifically, we aim at solving:

$$\operatorname*{argmin}_{f\in\mathcal{M}_N(\mathbb{R})}\left(\|\mathbf{y}-\langle f,\boldsymbol{\nu}\rangle\|_2^2+\lambda\left\|\mathrm{D}^N f\right\|_{\mathcal{M}}\right),\quad N\geqslant 2, \tag{3}$$

where $\mathcal{M}_N(\mathbb{R})$ is the space of functions whose Nth weak derivative is a bounded measure, as formally defined in (11). The total variation norm $\|\cdot\|_{\mathcal{M}}$ from measure theory is the continuous counterpart of $\|\cdot\|_1$. The derivative D^N acts as a regularizing operator and is the continuous counterpart of $\mathbf{L}$, while N is chosen greater than 1 for technical reasons. The measurement operator $\boldsymbol{\nu}:\mathcal{M}_N(\mathbb{R})\to\mathbb{R}^M$ is required to be (weak*)continuous on $\mathcal{M}_N(\mathbb{R})$. Problem (3) is sensibly the same as Problem (2), the search space being extended and the regularizing operator acting in the continuous world. The formulation (3) has several advantages

- The formulation is explicit and D^N is a differential operator whose behavior is well understood.
- Like for discrete-domain IPs, the solution of continuous-domain IPs still involves basis functions. However, the bases of a continuous-domain IP are not arbitrary anymore but imposed by the choice of the continuous-domain regularisation.
- There is no need for a choice of discretization steps.

The continuous-domain formulation (3) has the remarkable property that, under suitable conditions, sparse solutions exist and are parameterized by M atoms, themselves being related to D^N [43]. Specifically, the extreme points of the solution set of (3) take the form

$$f^\star(x)=\sum_{k=1}^{M-N}a_k\frac{(x-x_k)_+^{N-1}}{(N-1)!}+\sum_{n=1}^{N}b_n\frac{x^{n-1}}{(n-1)!},\tag{4}$$

where $f^\star$ is a spline of degree $(N-1)$. The atoms $\frac{(x-x_k)_+^{N-1}}{(N-1)!}$ are shifted replicates of $\frac{(x)_+^{N-1}}{N!}$, the causal Green's function of D^N. The a_k are the amplitudes of the atoms and x_k the spline knots, which are unknown *a priori*. In order to find a solution to the IP, one therefore has to find the best $(\mathbf{a},\mathbf{x},\mathbf{b})$ for $\mathbf{a}=(a_k)_{k=1}^{M-N}, \mathbf{x}=(x_k)_{k=1}^{M-N}, \mathbf{b}=(b_n)_{n=1}^{N}$. To do so, we inject (4) inside (3). This yields the new finite-dimensional optimization problem

$$\underset{(\mathbf{a},\mathbf{x},\mathbf{b})\in\mathbb{R}^{M-N}\times\mathbb{R}^{M-N}\times\mathbb{R}^N}{\operatorname{argmin}}\left(\left\|\mathbf{y}-\tilde{\mathbf{A}}(\mathbf{x})\mathbf{a}-\mathbf{B}\mathbf{b}\right\|_2^2+\lambda\|\mathbf{a}\|_1\right),\quad\text{with}\quad\tilde{\mathbf{A}}(\mathbf{x})\in\mathbb{R}^{M\times(M-N)},\quad \mathbf{B}\in\mathbb{R}^{M\times N},\tag{5}$$

with $\tilde{\mathbf{A}}_{m,k}(\mathbf{x})=\left\langle\frac{(\cdot-x_k)_+^{N-1}}{(N-1)!},\nu_m\right\rangle$ and $\mathbf{B}_{m,n}=\left\langle\frac{(\cdot)^{n-1}}{(n-1)!},\nu_m\right\rangle$. It is the *exact discretization* of (3), in the sense that it perfectly represents the true continuous-domain problem in (3). In contrast to the *a priori* discretization (2), we view (5) as an *a posteriori* exact *discretization*. Unfortunately, this problem is non-convex due to the dependence of $\tilde{\mathbf{A}}(\mathbf{x})$ on the unknown knots x_k.

1.3. Solving of continuous-domain IP by gridding

Our solution to (5) does not rely on non-convex optimization. Instead, we assume that the knots x_k of (4) lie on the fixed grid $\mathbb{X}_h=\{x_k^h\}_{k=1}^{K_h}$, with stepsize h. Problem (5) then yields

$$(\mathbf{a}_h^\star,\mathbf{b}_h^\star)\in\underset{(\mathbf{a}_h,\mathbf{b}_h)\in\mathbb{R}^{K_h}\times\mathbb{R}^N}{\operatorname{argmin}}\left(\left\|\mathbf{y}-\tilde{\mathbf{A}}_h\mathbf{a}_h-\mathbf{B}\mathbf{b}_h\right\|_2^2+\lambda\|\mathbf{a}_h\|_1\right),\tag{6}$$

with $[\tilde{\mathbf{A}}_h]_{m,k} = \left\langle \frac{(\cdot - x_k^h)_+^{N-1}}{(N-1)!}, \nu_m \right\rangle$ and the optimization problem is now convex. For $\mathcal{M}_N(\mathbb{R}, \mathbb{X}_h)$, the space of splines (4) whose knots x_k lie on the grid $\mathbb{X}_h$, (6) is the *exact discretization* of

$$f_h^\star \in \underset{f \in \mathcal{M}_N(\mathbb{R}, \mathbb{X}_h)}{\arg\min} \left(\| \mathbf{y} - \langle f, \boldsymbol{\nu} \rangle \|_2^2 + \lambda \left\| \mathrm{D}^N f \right\|_{\mathcal{M}} \right), \quad \text{with} \quad \mathcal{M}_N(\mathbb{R}, \mathbb{X}_h) \subset \mathcal{M}_N(\mathbb{R}). \tag{7}$$

Nevertheless, since the knots x_k are now forced to live on the grid $\mathbb{X}_h$, (6) is still only an approximation of (3), which can be expected to get better as the grid size h gets smaller. Simply said, the goal of this paper is to show convergence of the grid-based discretization (6) to (3). We claim that the *a posteriori* discretization (6) is more advantageous then the standard *a priori* discretization (2).

- Like in the continuous problem (3), the choice of basis functions (atoms) in the gridded solution (6) is optimal as they parameterize the solution set (3).
- A sequence of solutions $(\mathbf{a}_h^\star, \mathbf{b}_h^\star)$ can be efficiently calculated using a multiresolution algorithm [19].
- The problem in (6) is not computationally harder to solve than Problem (2).

In practice we perform an iterative estimate of solutions $f_h^\star$ and decrease h until $f_h^\star$ in (7) is sufficiently close to an $f^\star$ in (3), up to a given tolerance. Therefore, for this grid-based discretization to be meaningful and practical, one must know with confidence that $\lim_{h \to 0} f_h^\star = f^\star$. This said, we still need to state the sense in which the convergence happens, and which region of $\mathbb{R}$ should the grids $\mathbb{X}_\ell$ be discretizing. Observe that, since the solution set of (3) may contain more than one element, and since the norm convergence $\lim_{h \to 0} \|f_h^\star - f^\star\|_{\mathcal{M}_N(\mathbb{R})} = 0$ cannot be achieved, the choice of an adequate notion of convergence is in itself a nontrivial issue.

1.4. Contributions

In this paper, we study the convergence $\lim_{h \to 0} f_h^\star = f^\star$, while keeping in mind that we cannot count on the uniqueness of solutions. For the following contributions, the sequence of grids $\mathbb{X}_h$ neither need to be embedded into one another, nor need their knots x_k to be uniformly spaced.

- **Section 2.2-Localization.** In theorem 2, we present an upgraded version of the representer theorem of [43], where the knots in (4) must fall in the closed convex hull of the generalized support of the measurement functional $\boldsymbol{\nu}$. This shows that, in practice, one can use a grid $\mathbb{X}_h$ that discretizes only the support of $\boldsymbol{\nu}$, without loss of information.
- **Section 3.2-Consistency (Generalized interpolation).** For a sequence of grids $\mathbb{X}_h$ discretizing the support of $\boldsymbol{\nu}$, we show in theorem 3 that $f_h^\star$ is asymptotically consistent with the solutions $f^\star$ of (2), in the sense that they both yield the same loss $\lim_{h \to 0} \mathcal{J}(f_h^\star) = \mathcal{J}(f^\star)$, the same measurements $\lim_{h \to 0} \langle f_h^\star, \boldsymbol{\nu} \rangle = \langle f^\star, \boldsymbol{\nu} \rangle$, and the same regularisation $\lim_{h \to 0} \left\| \mathrm{D}^N f_h^\star \right\|_{\mathcal{M}} = \left\| \mathrm{D}^N f^\star \right\|_{\mathcal{M}}$. In theorem 4 we reveal that the order of convergence for the loss is $\mathcal{O}(h)$ while the order of convergence for the measurements and the regularization is $\mathcal{O}\left(h^{\frac{1}{2}}\right)$.
- **Section 3.3 & 3.4-Convergence of the discretization.** In theorem 5, we establish how the convergence of $f_h^\star$ to $f^\star$ happens. We show that, for any sequence of minimizers $(f_h^\star)_{h>0}$ and any weak* convergent subsequence $(f_{h_\ell}^\star)_{\ell=1}^\infty$ with $\lim_{\ell \to 0} h_\ell = 0$, the limit $f^\star$ must be part

of the solution set (3) and is such that $\lim_{\ell\to\infty}\left\|f_{h_\ell}^{\star,(d)}-f^{\star,(d)}\right\|_{\infty,[\nu^-,\nu^+]}=0$, where $[\nu^-,\nu^+]$ is the support of $\boldsymbol{\nu}$ and d is the order of the derivative with $d\leqslant(N-2)$. In addition, if the solution is unique with $\mathcal{V}=\{f^\star\}$, then the whole sequence $f_h^\star$ is locally uniformly convergent to $f^\star$. We reveal in Corollary 1 that this convergence holds even when the calculation of $f_h^\star$ is inexact.

- **Section 4.** We investigate the convergence rate of $f_h^\star$ to $f^\star$ with Monte-Carlo simulations.

1.5. *Related work*

1.5.1. Continuous-domain inverse problems. There is a rich literature on the resolution of continuous-domain inverse problems for signals defined over the Euclidean space $\mathbb{R}^d$[9, 43], the sphere $\mathbb{S}^d$ [39], the torus $\mathbb{T}^d$[25], or that are elements of some abstract Banach spaces [6, 7, 27, 42]. The topological structure of the solution set is often studied for general regularisations. By contrast, in this paper, we focus more on the numerical resolution of these problems.

1.5.2. Convergence of grid-based discretization. In the context of continuous-domain TV-regularized problems, grid-based methods have been studied by several authors, such as [17, 19, 22–24], who often prove their convergence results in the weak* sense. This convergence is well suited to the space of Radon measures and the problem of Dirac-mass recovery, but it is not informative enough in our setting because it does not exploit the smoothness of splines. Nevertheless, several results that dependant on the regularity of the minimizer and the (proved or implicit) unicity of the solution $f^\star$, have been derived to illustrate the true nature of this convergence.

1.5.3. Convergence of grid-based discretization: finite-dimensional data. The authors of [19], working on the same problem as us ($N\geqslant 2$), have shown convergence of the loss functional (sum of the data-fidelity and the regularizer) $\lim_{h\to 0}\mathcal{J}(f_h^\star)=\mathcal{J}(f^\star)$, but on uniform grids. In addition, for the reconstruction of periodic functions measured by samples of the Fourier transform, it has been show that the solution $f^\star$ is unique and that $\lim_{h\to 0}\|f_h^\star-f^\star\|_\infty=0$ holds [17]. These two results are particular cases of theorems 3 and 5.

1.5.4. Convergence of grid-based discretization: infinite-dimensional data. If one assumes to have access to infinite-dimensional data $\mathbf{y}$, then the solution $f^\star$ can be shown to be unique [1]. This is an easier setting than the one we consider in this paper. When no smoothness is imposed on the solution ($N=0$), and when the measurement is the solution of an elliptic PDE, the authors of [35] built on [10, 15] to establish the convergence rates $(\mathcal{J}(f_h^\star)-\mathcal{J}(f^\star))=\mathcal{O}\left(h^{4-D}|\ln(h)|^\gamma\right)$ and $\|\langle f_h^\star-f^\star,\nu\rangle\|_{\mathcal{L}_2}=\mathcal{O}\left(h^{2-\frac{D}{2}}|\ln(h)|^{\frac{\gamma}{2}}\right)$. There, D represents the dimension of the ambient space ($D=2,3$) and γ is an appropriate constant. We comment in section 3.2 on why this rate is better than ours. On $N=1$, the authors of [1, 2] extended the prior work of Chambolle and collaborators [11, 12], and of Lucier and Wang [44] and obtained the convergence rate $\|f_h^\star-f^\star\|_{\mathcal{L}_2}=\mathcal{O}\left(h^{\frac{1}{2}}\right)$. There, the authors worked on the TV-regularized interpolation problem.

1.5.5. Convergence of adaptive grid-based discretization: conditional and over-parametrized gradient descents. It is often desirable to iteratively update, or modify the knots of a grid $\mathbb{X}_\ell$. These modifications are informed, in the sense that the new knots allow for a refinement of the solution $f^\star$. Two approaches exist.

First, the conditional gradient-descent or Frank–Wolfe (FW) algorithm [30] iteratively adds knots to the grid $\mathbb{X}_\ell$, giving a new $\mathbb{X}_{\ell+1}$, with the associated optimal solutions being $f_\ell^\star$ and $f_{\ell+1}^\star$. Modern treatments of this algorithm leverage the Fenchel duality theory, so that the new knots are chosen in order to saturate the primal-dual gap. Working on $N=0$, the authors of [5] showed that $(\mathcal{J}(f_\ell^\star)-\mathcal{J}(f^\star))\leqslant\frac{C}{2+\ell}$. Also working on $N=0$, the authors of [8, 26, 36], under additional regularity and non-degeneracy conditions, in particular, that a unique (sparse) solution $f^\star$ exists, established that the linear convergence $(\mathcal{J}(f_\ell^\star)-\mathcal{J}(f^\star))\leqslant C\zeta^\ell$ holds for several versions of the FW algorithm. The authors of [20] also showed that the sliding FW algorithm terminates in a finite number of iterations.

Second, as is being done in ReLU neural network (NN) optimization, the weights and the knots of the spline can be optimized at the same time with a non-convex gradient descent. This is particularly relevant in our setting because of the link between D^2 regularization and ReLU NN [34]. The author of [13] showed that the over-parametrized gradient descent achieves the linear convergence $(\mathcal{J}(f_\ell^\star)-\mathcal{J}(f^\star))\leqslant C\zeta^\ell$. This method can be seen as an adaptive grid-based discretization where the knots are updated through gradient-descent.

Our approach is fundamentally different in that, similarly to the FEM setting [35], we perform an agnostic refinement of the grid. This incurs some penalty in our convergence rates, which are not linear. Yet, these rates are, up to a certain extent, comparable to the ones found in [15, 35].

2. Representer theorem with support localization

We are interested in general optimization problems of the form

$$\underset{f\in\mathcal{M}_N(\mathbb{R})}{\operatorname{argmin}}\,\mathcal{J}(f),\tag{8}$$

with the *loss functional*

$$\mathcal{J}(f)=E(\mathbf{y},\langle f,\boldsymbol{\nu}\rangle)+\psi\left(\left\|\mathrm{D}^N f\right\|_{\mathcal{M}}\right),\tag{9}$$

where

- the vector $\mathbf{y}\in\mathbb{R}^M$ is the vector of available measurements of fixed dimension M;
- the parameter $N\geqslant 2$ encodes the regularity of the minimizer (see proposition 1);
- the operator $\boldsymbol{\nu}:\mathcal{M}_N(\mathbb{R})\to\mathbb{R}^M$ is the forward measurement operator that models the data-acquisition process and is represented by a matrix in finite dimensions;
- the functional $E(\mathbf{y},\langle f,\boldsymbol{\nu}\rangle)$ is the data-fidelity functional that measures the discrepancy between the data $\mathbf{y}$ and the simulated measurement $\langle f,\boldsymbol{\nu}\rangle$; it is usually tuned to the type of underlying measurement noise. E has been illustrated in the introduction by the classical $\|\mathbf{y}-\langle f,\boldsymbol{\nu}\rangle\|_2^2$;
- the regularizer $\left\|\mathrm{D}^N\cdot\right\|_{\mathcal{M}}$ is a semi-norm for the search space $\mathcal{M}_N(\mathbb{R})$, defined in section 2.1;
- the gradation map $\psi:\mathbb{R}^+\to\mathbb{R}^+$ is a continuous increasing function, the standard choice being $\psi(t)=\lambda t$; other configurations may be used to enforce a more or less progressive gradation of regularisation;
- the search space $\mathcal{M}_N(\mathbb{R})$ is a fundamental element of our work and is specified in section 2.1.

2.1. Functional setting

Let $\mathcal{M}(\mathbb{R})$ be the space of Radon measures. By the Riesz–Markov–Kakutani theorem [38], it is the functional dual of the space $\mathcal{C}_0(\mathbb{R})$ of continuous vanishing functions endowed with the sup norm $\|\cdot\|_\infty$. The total-variation norm on $\mathcal{M}(\mathbb{R})$, for which it forms a Banach space [38][Chapter 6], is given by

$$\|\cdot\|_{\mathcal{M}} := \sup_{\phi \in \mathcal{C}_0(\mathbb{R}), \phi \neq 0} \frac{|\langle \cdot, \phi \rangle|}{\|\phi\|_\infty}. \tag{10}$$

The $\|\cdot\|_{\mathcal{M}}$ norm can be seen as a generalization of the $\|\cdot\|_1$ norm. Indeed, if $f \in \mathcal{L}_1(\mathbb{R})$, then $\|f\|_1 = \|f\|_{\mathcal{M}}$. Let $\mathcal{S}'(\mathbb{R})$ be the space of tempered distributions. The search space $\mathcal{M}_N(\mathbb{R})$ is defined as

$$\mathcal{M}_N(\mathbb{R}) = \left\{ f \in \mathcal{S}'(\mathbb{R}) : \quad \left\| \mathrm{D}^N f \right\|_{\mathcal{M}} < \infty \right\}. \tag{11}$$

Moreover, equipped with the correct norm, it is a Banach space (appendix, proposition 6). This search space has been studied for general spline-admissible operators in [43]. From now on, $\mathcal{M}_N(\mathbb{R})$ is equipped with the weak* topology. General facts on this topology are recalled in A.1.1 and the specific weak* topology equipped on $\mathcal{M}_N(\mathbb{R})$ is described in (appendix, theorem 6)

The main theorem of section 2 states that the extreme points of the solution set of Problem (8) are D^N-splines whose knot points (definition 1) are inside the support of $\boldsymbol{\nu}$ (definition 3).

Definition 1 (D^N-spline). A generalized function $f \in \mathcal{S}'(\mathbb{R})$ is said to be a D^N-spline if

$$\mathrm{D}^N f = \sum_{k=1}^{K} a_k \delta_{x_k} \Leftrightarrow f = \sum_{k=1}^{K} a_k \frac{(\cdot - x_k)_+^{N-1}}{(N-1)!} + p \quad \text{with} \quad p \in \mathcal{N}_N, \tag{12}$$

where $\mathcal{N}_N = \left\{ \frac{(\cdot)^{n-1}}{(n-1)!} \right\}_{n=1}^{N}$ is the nullspace of D^N, K is the number of knots, $x_k \in \mathbb{R}$ are the distinct knots of the spline, $a_k \in \mathbb{R} \setminus \{0\}$, and δ_{x_k} is a Dirac-mass centered in x_k.

Then, while the notion of support supp of a function is classic, we recall the notion of support for generalised functions in definition 2, from Chapter 24 of [41]. This shall allow us to establish in theorem 2 where the solutions of (8) are localised.

Definition 2. Let $f \in \mathcal{S}'(\mathbb{R})$ be a generalized function.

(1) The distribution f is said to vanish on an open set $S \subset \mathbb{R}$ if

$$\forall \psi \in \mathcal{S}(\mathbb{R}) \quad \text{s.t.} \quad \mathrm{supp}(\psi) \subset S : \quad \langle f, \psi \rangle = 0, \tag{13}$$

where $\mathcal{S}(\mathbb{R})$ is the Schwartz space of smooth and rapidly decaying functions. Here, $\mathrm{supp}(\psi) = \overline{\psi^{-1}(\{0\}^c)}$ is closed.

(2) The complement in $\mathbb{R}$ of the largest open set O in which f vanishes is called the *support* of f and is denoted by $\mathrm{supp}(f)$.

This definition of support is exactly the one expected. In particular, when f is continuous, it coincides with the usual notion of support. If $f = \delta_{x_0}$, then $\mathrm{supp}(f) = \{x_0\}$. Likewise, if $f = \sum_{k=1}^{K} a_k \delta_{x_k}$, then $\mathrm{supp}(f) = \bigcup_{k=1}^{K} \{x_k\}$.

Definition 3 (Compact support).

(1) We denote by $\mathrm{Csupp}(f)$ the closed convex hull of $\mathrm{supp}(f)$.
(2) Let $\boldsymbol{\nu}:\mathcal{M}_N\to\mathbb{R}^M$ be the measurement operator. We say that $\boldsymbol{\nu}=(\nu_1,\ldots,\nu_M)$ is *compactly supported* if every set $\mathrm{Csupp}(\nu_m)$ is compact. Further, we write $\mathrm{Csupp}(\nu_m)=[\nu_m^-,\nu_m^+]$ with $\nu_m^-,\nu_m^+\in\mathbb{R}$ such that $\nu_m^-\leqslant\nu_m^+$ and, by extension, $\mathrm{Csupp}(\boldsymbol{\nu})=[\nu^-,\nu^+]$, where $\nu^-=\min\limits_{m\in[1,M]}\nu_m^-$ and $\nu^+=\max\limits_{m\in[1,M]}\nu_m^+$.

Observe that $\mathrm{Csupp}(f)$ is the convexification of $\mathrm{supp}(f)$. For example, if $\mathrm{supp}(f)=[0,1]\cup[2,3]$, then $\mathrm{Csupp}(f)=[0,3]$. Likewise, if $f=\sum_{k=1}^K\delta_{x_k}$, then $\mathrm{Csupp}(f)=\left[\min\limits_{1\leqslant k\leqslant K}x_k,\ \max\limits_{1\leqslant k\leqslant K}x_k\right]$.

2.2. Representer theorem

We first state a known result that gives the generic spline form of the extreme points of the solution set of Problem (8).

Theorem 1 [42, theorem 2] and [29, theorem 4].
We consider the following setting.

(1) *The derivative* D^N *is the derivative of order* $N\geqslant 2$ *with nullspace* $\mathcal{N}_N$.
(2) *The measurement operator* $\boldsymbol{\nu}=(\nu_1,\ldots,\nu_m):\mathcal{M}_N(\mathbb{R})\to\mathbb{R}^M$ *is a* weak* *continuous linear measurement operator with nullspace* $\mathcal{N}_{\boldsymbol{\nu}}$.
(3) *We assume that* $\mathcal{N}_{\boldsymbol{\nu}}\cap\mathcal{N}_N=0$. *Equivalently, we assume that* $\boldsymbol{\nu}:\mathcal{N}_N\to\mathbb{R}^M$ *is injective.*
(4) *The loss functional* $E:\mathbb{R}^M\times\mathbb{R}^M\to\mathbb{R}^+\cup\{\infty\}$ *is proper, coercive, lower semi-continuous, and convex in its second argument.*
(5) *The gradation map* $\psi:\mathbb{R}^+\to\mathbb{R}^+$ *is continuous, increasing, and convex.*

Then, the solution set

$$\mathcal{V}=\underset{f\in\mathcal{M}_N(\mathbb{R})}{\mathrm{argmin}}\,\mathcal{J}(f)\tag{14}$$

is the nonempty, weak* *compact,* weak* *closed convex hull of its extreme points, which all are* D^N*-splines with at most* $(M-N)$ *knots.*

Let $\boldsymbol{\nu}$ be compactly supported with $\mathrm{Csupp}(\nu)=[\nu^-,\nu^+]$. We are now in the position to state and prove the central result of section 2; namely, that the minimisation of $\mathcal{J}$ over the space $\mathcal{M}_N(\mathbb{R})$, whose spline elements have knots x_k in $\mathbb{R}$, is equivalent to the minimisation of $\mathcal{J}$ over the 'localized' space $\mathcal{M}_N([\nu^-,\nu^+])$, whose spline members have knots x_k that are restricted to $[\nu^-,\nu^+]$. The space $\mathcal{M}_N([\nu^-,\nu^+])$ is defined as

$$\mathcal{M}_N\left([\nu^-,\nu^+]\right)=\left\{f\in\mathcal{S}'(\mathbb{R}):\quad \mathrm{D}^Nf\in\mathcal{M}\left([\nu^-,\nu^+]\right)\right\}.\tag{15}$$

This space, unlike $\mathcal{M}_N(\mathbb{R})$, enjoys a simple integral representation. This idea is formalized in proposition 1, where the regularity of functions in $\mathcal{M}_N([\nu^-,\nu^+])$ is also specified. The proof is given in A.2.

Proposition 1. *A generalized function f belongs to the space* $\mathcal{M}_N([\nu^-,\nu^+])$ *if and only if it is of the form*

$$f = \int_{\nu^-}^{\nu^+} \frac{(\cdot - y)_+^{N-1}}{(N-1)!} \mathrm{d}m(y) + \sum_{n=1}^{N} b_n \frac{(\cdot)^{n-1}}{(n-1)!} : \quad m \in \mathcal{M}\left(\left[\nu^-,\nu^+\right]\right), b_n \in \mathbb{R}. \tag{16}$$

Moreover, one has the proper inclusion

$$\mathcal{M}_N\left(\left[\nu^-,\nu^+\right]\right) \subset \mathcal{C}^{N-2}(\mathbb{R}), \tag{17}$$

where $\mathcal{C}^{N-2}(\mathbb{R})$ *is the space of* $(N-2)$ *times continuously differentiable functions.*

Next arrives the main result, that leverages proposition 1. Because the proof reveals important techniques used in this paper, it is given here, after the theorem statement.

Theorem 2. *In addition to the assumptions of theorem 1, let* $\boldsymbol{\nu}$ *be compactly supported with* $\mathrm{Csupp}(\boldsymbol{\nu}) = [\nu^-,\nu^+]$. *Then, it holds that*

$$\mathcal{V} = \operatorname*{argmin}_{f\in\mathcal{M}_N(\mathbb{R})} \mathcal{J}(f) = \operatorname*{argmin}_{f\in\mathcal{M}_N([\nu^-,\nu^+])} \mathcal{J}(f), \tag{18}$$

and every extreme point of the solution set $\mathcal{V}$ *is of the form*

$$e^\star = \sum_{k=1}^{M-N} a_k \frac{(\cdot - x_k)_+^{N-1}}{(N-1)!} + \sum_{n=1}^{N} b_n \frac{(\cdot)^{n-1}}{(n-1)!}, \qquad x_k \in]\nu^-,\nu^+[. \tag{19}$$

Proof of theorem2. We first show that all the extreme points $e^\star$ of $\mathcal{V}$ are of the form (19). Theorem 1 states that $e^\star$ is such that

$$e^\star(\cdot) = \sum_{k=1}^{K} a_k \frac{(\cdot - x_k)_+^{N-1}}{(N-1)!} + \sum_{n=1}^{N} b_n \frac{(\cdot)^{n-1}}{(n-1)!}, \quad K+N \leqslant M. \tag{20}$$

It follows for $\nu_{m'}$, the m'th component of $\boldsymbol{\nu}$, that

$$\langle e^\star, \nu_{m'}\rangle = \sum_{k=1}^{K} a_k \left\langle \frac{(\cdot - x_k)_+^{N-1}}{(N-1)!}, \nu_{m'} \right\rangle + \sum_{n=1}^{N} b_n \left\langle \frac{(\cdot)^{n-1}}{(n-1)!}, \nu_{m'} \right\rangle \tag{21}$$

$$= \underbrace{\sum_{k\in[1,K]:x_k\leqslant\nu^-} a_k \left\langle \frac{(\cdot - x_k)_+^{N-1}}{(N-1)!}, \nu_{m'} \right\rangle}_{(A)} + \underbrace{\sum_{\substack{k\in[1,K]:\\ \nu^+>x_k>\nu^-}} a_k \left\langle \frac{(\cdot - x_k)_+^{N-1}}{(N-1)!}, \nu_{m'} \right\rangle}_{(B)}$$

$$+ \underbrace{\sum_{n=1}^{N} b_n \left\langle \frac{(\cdot)^{n-1}}{(n-1)!}, \nu_{m'} \right\rangle}_{(C)}, \tag{22}$$

where we used the fact that $\sum_{k\in[1,K]:x_k\geqslant\nu^+} a_k \left\langle \frac{(\cdot-x_k)_+^{N-1}}{(N-1)!}, \nu_{m'} \right\rangle = 0$, since $(\cdot - x_k)_+ = 0$ over the integration domain (the support of $\nu_{m'}$). Thus, a knot $x_k \geqslant \nu^+$ has no impact on the calculation of $\langle e^\star, \nu_{m'}\rangle$ and necessarily increases the regularisation cost $\left\|\mathrm{D}^N e^\star\right\|_{\mathcal{M}}$. Hence, if $e^\star$

is a minimizer then the support of $m = \sum_{k=1}^{K} a_k \delta_{x_k}$ must be included in $]-\infty, \nu^+[$. Then, we show that (A) in (22) can be moved into (B) (nullspace term):

$$(A) = \sum_{k\in[1,K]:x_k\leqslant\nu^-} a_k \left\langle \frac{1}{(N-1)!} \sum_{n=0}^{N-1} \binom{N-1}{n} (-x_k)^{N-1-n} (\cdot)^n, \nu_{m'} \right\rangle \tag{23}$$

$$= \sum_{n=0}^{N-1} \sum_{k\in[1,K]:x_k\leqslant\nu^-} \frac{a_k}{(N-n-1)!} (-x_k)^{N-n-1} \left\langle \frac{(\cdot)^n}{(n)!}, \nu_{m'} \right\rangle \tag{24}$$

$$= \sum_{n=1}^{N} \tilde{b}_n \left\langle \frac{(\cdot)^{n-1}}{(n-1)!}, \nu_{m'} \right\rangle. \tag{25}$$

In (23) we used the fact that $(x-y)_+ = (x-y)$ over the integration domain. Set $\tilde{b}_n = \sum_{k\in[1,K]:x_k\leqslant\nu^-} \frac{a_k}{(N-n)!} (-x_k)^{N-n}$. Injecting (25) into (22) yields

$$\langle e^\star, \nu_{m'} \rangle = \sum_{k\in[1,K]:\nu^+>x_k>\nu^-} a_k \left\langle \frac{(\cdot - x_k)_+^{N-1}}{(N-1)!}, \nu_{m'} \right\rangle + \sum_{n=1}^{N} \left(b_n + \tilde{b}_n\right) \left\langle \frac{(\cdot)^{n-1}}{(n-1)!}, \nu_{m'} \right\rangle. \tag{26}$$

Let $\tilde{e}$ be the new candidate solution

$$\tilde{e} = \sum_{k\in[1,K]:\nu^+>x_k>\nu^-} a_k \frac{(\cdot - x_k)_+^{N-1}}{(N-1)!} + \sum_{n=1}^{N} \left(b_n + \tilde{b}_n\right) \frac{(\cdot)^{n-1}}{(n-1)!}, \tag{27}$$

such that

(1) $\forall m \in [1, M] \quad \langle e^\star, \nu_{m'} \rangle = \langle \tilde{e}, \nu_{m'} \rangle$;
(2) $\left\| \mathrm{D}^N e^\star \right\|_{\mathcal{M}} = \| m \|_{\mathcal{M}} \geqslant \left\| m \mathbb{1}_{]\nu^-,\nu^+[} \right\|_{\mathcal{M}} = \left\| \mathrm{D}^N \tilde{e} \right\|_{\mathcal{M}}$.

Then, $\mathcal{J}(e^\star) \geqslant \mathcal{J}(\tilde{e})$ with equality iff $m(\cdot)\mathbb{1}_{]\nu^-,\nu^+[}(\cdot) = m(\cdot)$. Thus, for $e^\star$ to be a minimizer it is necessary that $m(\cdot)\mathbb{1}_{]\nu^-,\nu^+[}(\cdot) = m(\cdot)$, what implies that $x_k \in]\nu^-, \nu^+[$.

We then prove that any solution $f^\star \in \mathcal{V}$ can be written as

$$f^\star(\cdot) = \int_{\nu^-}^{\nu^+} \frac{(\cdot - y)_+^{N-1}}{(N-1)!} \mathrm{d}m(y) + \sum_{n=1}^{N} b_n \frac{(\cdot)^{n-1}}{(n-1)!}, \tag{28}$$

with $b_n \in \mathbb{R}$ and $m \in \mathcal{M}([\nu^-, \nu^+])$. Indeed, this directly implies that $\mathcal{V} \subset \mathcal{M}_N([\nu^-, \nu^+])$ and therefore that $\mathcal{V} = \underset{f\in\mathcal{M}_N(\mathbb{R})}{\operatorname{argmin}} \mathcal{J}(f) = \underset{f\in\mathcal{M}_N([\nu^-,\nu^+])}{\operatorname{argmin}} \mathcal{J}(f)$. What concludes the proof. We know that $\mathcal{V}$ is the weak* closed convex hull of extreme points of the form (19). Thus, $f^\star$ is the weak* limit of a sequence $f_\ell = \frac{(\cdot)_+^{N-1}}{(N-1)!} * m_\ell + p_\ell$, where $(p_\ell)_{\ell=1}^{\infty} \subset \mathcal{N}_N$ and $(m_\ell)_{\ell=1}^{\infty} \subset \mathcal{M}([\nu^-, \nu^+])$. This implies that $\mathrm{D}^N f_\ell = m_\ell$ is weak* convergent, in $\mathcal{M}(\mathbb{R})$, to $\mathrm{D}^N f^\star$, which in turn implies that $\mathrm{D}^N f^\star \in \mathcal{M}([\nu^-, \nu^+])$. This shows, by definition of $\mathcal{M}_N([\nu^-, \nu^+])$, that $f^\star \in \mathcal{M}_N([\nu^-, \nu^+])$. ☐

Theorem 2 has important implications for our analysis and the development of numerical solvers. On the theoretical hand, (18) informs us that the optimisation of the functional $\mathcal{J}$ on a continuous-domain vector space $\mathcal{M}_N(\mathbb{R})$ is equivalent to the optimisation on the continuous-compact-domain vector space $\mathcal{M}_N([\nu^-, \nu^+])$, which is simpler to study. This will have a great influence on the convergence studies in section 3 and, combined with proposition 1, it will imply that all minimizers are $(N-2)$ times continuously differentiable. On the practical hand, (19) informs us that the grid $\mathbb{X}_\ell$ only needs to discretize the interval $]\nu^-,\nu^+[$, and not the entire $\mathbb{R}$, which is time-saving for the practitioner.

The compact support of $\boldsymbol{\nu}$ is the cornerstone of our proof. Indeed, we use the fact that, even though the signal is nonzero outside $[\nu^-,\nu^+]$, the signal contribution from knot points (x_k) outside $[\nu^-,\nu^+]$ can be moved in a costless fashion into the nullspace of D^N. As explained in the section 1, the direct estimation of a solution of the form (19) is hard because the solution depends on the knots x_k in a non-convex fashion. To bypass this difficulty, we shall discretize the problem on a grid [19] and further study the convergence as the grid gets finer.

3. Convergence of the grid-based discretization of the continuous-domain problem

3.1. Discretization scheme

We consider a sequence of grids $\mathbb{X}_\ell = \{x_k^\ell\}_{k=1}^{K_\ell}$, on which we exactly discretize the IP (8). The index ℓ encodes the degree of refinement of the grid and the stepsize h is calculated *a posteriori* using the nodal width (definition 4). Among the parameters that determine a solution to (8), we only impose the knots $x_k^\ell \in \mathbb{X}_\ell$. The space of splines with knots in $\mathbb{X}_\ell$ is defined as

$$\mathcal{M}_N(\mathbb{R},\mathbb{X}_\ell) = \left\{ \sum_{k=1}^{K_\ell} a_k \frac{(\cdot - x_k^\ell)_+^{N-1}}{(N-1)!}, \quad a_k \in \mathbb{R} \right\} \oplus \mathcal{N}_N. \tag{29}$$

The solution to the new optimization problem becomes

$$\mathcal{V}_\ell = \underset{f\in\mathcal{M}_N(\mathbb{R},\mathbb{X}_\ell)}{\operatorname{argmin}} \mathcal{J}(f), \tag{30}$$

with exact discretization

$$\underset{(\mathbf{a}_\ell,\mathbf{b}_\ell)\in\mathbb{R}^{K_\ell}\times\mathbb{R}^N}{\operatorname{argmin}} \left(\left\| \mathbf{y} - \tilde{\mathbf{A}}_\ell \mathbf{a}_\ell - \mathbf{B}_\ell \mathbf{b}_\ell \right\|_2^2 + \lambda \|\mathbf{a}_\ell\|_1 \right), \tag{31}$$

$$\text{and with} \quad \left[\tilde{\mathbf{A}}_\ell\right]_{m,k} = \left\langle \frac{(\cdot - x_k^\ell)_+^{N-1}}{(N-1)!}, \nu_m \right\rangle \quad \text{and} \quad [\mathbf{B}_\ell]_{m,n} = \left\langle \frac{(\cdot)^{n-1}}{(n-1)!}, \nu_m \right\rangle. \tag{32}$$

It is to be solved numerically. The main topic of this paper is whether or not $\mathcal{V}_\ell$ is a good approximation of $\mathcal{V}$, and in which sense. In section 3.2 we establish several consistency results and in section 3.3 we state local uniform convergence.

In order to show all these convergences, we build on the notion of nodal width. It intuitively measures how far a grid $\mathbb{X}$ is to the compact set K.

Definition 4 (Nodal width). Let $\mathbb{X}_\ell = \{x_k^\ell\}_{k=1}^{K_\ell}$ be a sequence of grids. For a compact set K, we define the *nodal width* as $\Theta(K,\mathbb{X}_\ell) = \max_{y\in K}\min_{x\in\mathbb{X}_\ell} ||x-y||_2$.

Our primary assumption about the grids is that the nodal width calculated on the support of the measurement operator $\boldsymbol{\nu}$ goes to 0. The grids are free otherwise, in the sense that we neither require them to be embedded into one another, nor that their nodes be uniformly spaced.

3.2. Consistency of the grid-based discretization

As first step toward the understanding of the convergence from $f_\ell^\star$ to $f^\star$, theorem 3 presents two results: consistency in the simulated measurement $\nu(f_\ell^\star)$; and consistency in the regularisation $\left\|\mathrm{D}^N f_\ell^\star\right\|_{\mathcal{M}}$. The proof of theorem 3 is in A.3.

Theorem 3. *Let all the assumptions of theorem 2 be fulfilled. Suppose in addition that, for a fixed measurement* $\mathbf{y} \in \mathbb{R}^M$*,*

(1) *the nodal width vanishes, with* $\lim\limits_{\ell\to\infty} \Theta([\nu^-,\nu^+],\mathbb{X}_\ell) = 0$;
(2) *the data-fidelity* $E(\cdot,\cdot)$ *is strictly convex in its second variable;*
(3) *the data-fidelity is lower-bounded, so that* $\inf\limits_{\mathbf{z}\in\mathbb{R}^M} E(\mathbf{y},\mathbf{z}) = A > -\infty$;
(4) *the set* $\mathrm{dom}(E(\mathbf{y},\cdot)) = \left\{\mathbf{z}\in\mathbb{R}^M : \quad E(\mathbf{y},\mathbf{z}) < \infty\right\}$ *is open.*

Then, for any sequence $(f_\ell^\star)_{\ell=1}^\infty$ *of solutions with* $f_\ell^\star \in \mathcal{V}_\ell$ *and for any solution* $f^\star \in \mathcal{V}$*, the following convergences hold:*

(i) *over the regularisation,* $\lim\limits_{\ell\to\infty} \left\|\mathrm{D}^N f_\ell^\star\right\|_{\mathcal{M}} = \left\|\mathrm{D}^N f^\star\right\|_{\mathcal{M}}$;
(ii) *over the measurement,* $\lim\limits_{\ell\to\infty} \langle f_\ell^\star, \boldsymbol{\nu}\rangle = \langle f^\star, \boldsymbol{\nu}\rangle$;
(iii) *over the data-fidelity,* $\lim\limits_{\ell\to\infty} E(\mathbf{y}, \langle f_\ell^\star, \boldsymbol{\nu}\rangle) = E(\mathbf{y}, \langle f^\star, \boldsymbol{\nu}\rangle)$.

Theorem 3 implies the convergence $\lim\limits_{\ell\to\infty} \mathcal{J}(f_\ell^\star) = \mathcal{J}(f^\star)$. On the theoretical hand, it also shows that our grid-based discretization is coherent with respect to the measurements. This reveals that the minimisation of $\mathcal{J}$ on a (fine enough) grid $\mathbb{X}_\ell$ yields an estimate $\langle f_\ell^\star, \boldsymbol{\nu}\rangle$ that is close to the optimal value $\langle f^\star, \boldsymbol{\nu}\rangle$ and to the data $\mathbf{y}$ as well, by means of the data-fidelity E. Moreover, it implies that the values of $\langle f^\star, \boldsymbol{\nu}\rangle$ and $\left\|\mathrm{D}^N f^\star\right\|_{\mathcal{M}}$ are unique, even when $f^\star$ is not. On the practical hand, the convergences established in theorem 3, as well as the uniqueness of the limits which are implied, show that the relative error $\frac{\left\|\langle f_{\ell+1}^\star - f_\ell^\star, \boldsymbol{\nu}\rangle\right\|_2}{\left\|\langle f_\ell^\star, \boldsymbol{\nu}\rangle\right\|_2}$ can be used reliably to monitor the convergence over ℓ.

Remark (On theorem 3).

- Theorem 2 allows us to only require in theorem 3 that $\lim\limits_{\ell\to\infty} \Theta([\nu^-,\nu^+],\mathbb{X}_\ell) = 0$ (as opposed to $\lim\limits_{\ell\to\infty} \Theta(\mathbb{R},\mathbb{X}_\ell) = 0$), because of the density of knots in $[\nu^-,\nu^+]$.
- The assumption that $E(\mathbf{y},\cdot)$ is strictly convex is not necessary and can be lifted. If done so, theorem 3 holds up to a subsequence: there exists a subsequence $(f_{\ell_k}^\star)_{k=1}^\infty$ with $\ell_{k+1} > \ell_k$ and a function $f^{\star\star} \in \mathcal{V}$ such that $\lim\limits_{k\to\infty} \left\|\mathrm{D}^N f_{\ell_k}^\star\right\|_{\mathcal{M}} = \left\|\mathrm{D}^N f^{\star\star}\right\|_{\mathcal{M}}$ and $\lim\limits_{k\to\infty} \langle f_{\ell_k}^\star, \boldsymbol{\nu}\rangle = \langle f^{\star\star}, \boldsymbol{\nu}\rangle$. In practice, strict convexity is not a restrictive assumption. For example, the squared ℓ_2 norm $\|\mathbf{y} - \cdot\|_2^2$ is strictly convex. More generally all p-norms are strictly convex for $1 < p < \infty$. Likewise, the Kullback–Leibler divergence $\mathrm{KL}(\mathbf{y}|\mathbf{q}) = \sum_{i=1}^M y_m \log\left(\frac{y_m}{q_m}\right)$ is strictly convex in $\mathbf{q}$ if all y_m are nonzero.
- Theorem 3 is a refinement of the convergence result [19][theorem 3] that shows that $\lim\limits_{\ell\to\infty} \mathcal{J}(f_\ell^\star) = \mathcal{J}(f^\star)$ for uniform grids. Indeed, our theorem shows that the convergence actually holds for both the data and the regularisation, as well as for nonuniform grids.

- The functional $E(\mathbf{y}, \langle \cdot, \boldsymbol{\nu} \rangle) : \mathcal{M}_N(\mathbb{R}) \to \mathbb{R}$ is not strictly convex, even though $E(\mathbf{y}, \cdot) : \mathbb{R}^M \to \mathbb{R}$ is strictly convex and $\langle \cdot, \boldsymbol{\nu} \rangle : \mathcal{N}_N \to \mathbb{R}$ is injective. Consequently, the solution set $\mathcal{V}$ is not necessarily a singleton.

In theorem 4, whose proof is in A.3, we characterize the speed of the convergences in 3, under additional regularity conditions. For its statement we notate the mesh size $h_\ell = \Theta([\nu^-, \nu^+], \mathbb{X}_\ell)$.

Theorem 4. *Let all the assumptions of theorem 3 be verified. In addition, let us consider the following scenarios:*

(1) *the data fidelity* $E(\mathbf{y}, \cdot)$ *is continuously differentiable on* $\mathrm{dom}(E(\mathbf{y}, \cdot))$*;*
(2) *for* $1 \leqslant m \leqslant M$, ν_m *is a combination of an* $\mathcal{L}_\infty(\mathbb{R})$ *function and a sum of Dirac masses;*
(3) *the data fidelity* $E(\mathbf{y}, \cdot)$ *is strongly convex on any compact subset of* $\mathrm{dom}(E(\mathbf{y}, \cdot))$*;*
(4) *the measurement operator* $\boldsymbol{\nu} : \mathcal{M}_N(\mathbb{R}) \to \mathbb{R}^M$ *is surjective.*

If 1. & 2. hold, then for any sequence $(f_\ell^\star)_{\ell=1}^\infty$ *of solutions with* $f_\ell^\star \in \mathcal{V}_\ell$ *and for any solution* $f^\star \in \mathcal{V}$*, the following convergence hold:*

(i) *over the loss functional,* $|\mathcal{J}(f^\star) - \mathcal{J}(f_\ell^\star)| = \mathcal{O}(h_\ell)$ *as* $\ell \to \infty$.

If 1. to 4. hold, then the following convergences hold:

(ii) *over the measurement,* $\|\langle f^\star - f_\ell^\star, \boldsymbol{\nu} \rangle\|_2 = \mathcal{O}\left(h_\ell^{\frac{1}{2}}\right)$ *as* $\ell \to \infty$;
(iii) *over the data fidelity,* $|E(\mathbf{y}, \langle f^\star, \boldsymbol{\nu} \rangle) - E(\mathbf{y}, \langle f_\ell^\star, \boldsymbol{\nu} \rangle)| = \mathcal{O}\left(h_\ell^{\frac{1}{2}}\right)$ *as* $\ell \to \infty$;
(iv) *over the regularization,* $\left| \|\mathrm{D}^N f^\star\|_{\mathcal{M}} - \|\mathrm{D}^N f_\ell^\star\|_{\mathcal{M}} \right| = \mathcal{O}\left(h_\ell^{\frac{1}{2}}\right)$ *as* $\ell \to \infty$.

While theorem 4 is a first step in obtaining convergence rates, these estimates are probably not optimal. Observe that the regularity $(N-2)$ of the solution $f^\star$ does not appear in the given orders, as could have been expected from the approximation-theory results (see for example [14]), or from wavelet-based approximation (see for example [32]). In addition, these orders are also slower than those established in the sparse elliptic PDE framework [35]. There, the authors do not directly formulate the optimization on the discretized search space. Instead, they discretize the measurement operators and the associated PDE. Here, our proof relies on the construction of a good approximation $\hat{f}_\ell \in \mathcal{M}_N(\mathbb{R}, \mathbb{X}_\ell)$ of $f^\star$. The difficulty lies in the building of an $\hat{f}_\ell$ such that the convergence of $\left\|\mathrm{D}^N \hat{f}_\ell\right\|_{\mathcal{M}}$ to $\|\mathrm{D}^N f^\star\|_N$ reflects the regularity of $f^\star$ while still being a good approximation of $f^\star$. The construction of more favourable approximations and the possible refinement of the orders in theorem 4 are reserved for a future work.

Theorems 3 and 4 provide consistency on the measurements side but are not informative on the convergence of the sequence of functions $f_\ell^\star$ itself, which is the topic of section 3.3.

3.3. Sub-sequential local uniform convergence of the grid-based discretization

In this section, we upgrade theorem 3 by proving that, up to a subsequence, $f_\ell^\star$ is locally uniformly convergent to $f^\star$, with the same holding true for its derivatives up to order $(N-2)$. In addition, we show that the coefficients of the polynomials that encode the tails of $f_\ell^\star$ are converging to the coefficients of the polynomials that encode the tails of $f^\star$. This, as well as local uniform convergence, is formalized in the next theorem, whose proof is in A.4.

Theorem 5. *Let all the hypothesis of theorem 3 be verified. Then, for any sequence* $(f^\star_\ell)_{\ell=1}^\infty$ *of solutions with* $f^\star_\ell \in \mathcal{V}_\ell$*, for any subsequence* $(f^\star_{\ell_k})_{k=1}^\infty$ *with* $\ell_{k+1} > \ell_k$ *that is* weak* *convergent to a limit* $f^\star \in \mathcal{M}_N(\mathbb{R})$*, it holds that*

(1) *the limit* $f^\star$ *is a solution of the continuous-domain problem, with* $f^\star \in \mathcal{V}$*;*

(2) *on* $]-\infty,\nu^-]$ *the functions* $f^\star_{\ell_k}, f^\star$ *are fully described by elements of the nullspace* $\mathcal{N}_N$*, whose coefficients are convergent*

$$\forall x \in]-\infty,\nu^-], \quad f^\star_{\ell_k}(x) = \sum_{n=1}^{N} b_{\ell_k,n}\frac{x^{n-1}}{(n-1)!}, f^\star(x) = \sum_{n=1}^{N} b_n \frac{x^{n-1}}{(n-1)!},$$
$$\forall 1 \leqslant n \leqslant N, \quad \lim_{k\to\infty} b_{\ell_k,n} = b_n; \tag{33}$$

(3) *on* $[\nu^-,\nu^+]$*, uniform convergence holds*

$$\forall d \leqslant N-2, \quad \lim_{k\to\infty} \left\| f^{\star,(d)}_{\ell_k} - f^{\star,(d)} \right\|_{\infty,[\nu^-,\nu^+]} = 0; \tag{34}$$

(4) *on* $[\nu^+,\infty[$ *the functions* $f^\star_{\ell_k}, f^\star$ *are fully described by elements of the nullspace* $\mathcal{N}_N$*, whose coefficients are convergent (beware, these coefficients differ from those on* $]-\infty,\nu^-]$*, because of the causal innovations in* $[\nu^-,\nu^+]$*)*

$$\forall x \in [-\nu^+,\infty[, \quad f^\star_{\ell_k}(x) = \sum_{n=1}^{N} \tilde{b}_{\ell_k,n}\frac{x^{n-1}}{(n-1)!}, f^\star(x) = \sum_{n=1}^{N} \tilde{b}_n \frac{x^{n-1}}{(n-1)!},$$
$$\forall 1 \leqslant n \leqslant N, \quad \lim_{\ell_k\to\infty} \tilde{b}_{\ell_k,n} = \tilde{b}_n. \tag{35}$$

In addition, the following holds:

(5) *there exists a subsequence* $(f^\star_{\ell_k})_{k=1}^\infty$ *with* $\ell_{k+1} > \ell_k$*, that is* weak* *convergent to a limit* $f^\star \in \mathcal{M}_N(\mathbb{R})$*;*

(6) *in the case where the solution* $\mathcal{V}$ *is unique, the convergences above (Items 2–4) hold for the sequence* $(f^\star_\ell)_{\ell=1}^\infty$ *itself.*

Theorem 5 is a surprisingly strong result as it states that, even though there are finitely many measurements $\mathbf{y}$, convergence is observed, up to a subsequence, on the whole real line. It indicates that the error that one makes when imposing the x_k to live on $\mathbb{X}_\ell$ vanishes as $\ell \to \infty$. This tells us that our grid-based approximation is coherent, in the sense that it asymptotically recovers a solution $f^\star$ of the continuous-domain problem. The theorem also describes how strong this convergence is, depending on N. Observe that it not only reveals 'local' or pointwise convergence of the iterates $f^\star_\ell$ but also a convergence of the derivatives and of the tails, which can be of high importance to the practitioner.

In the case where the solution $\mathcal{V}$ of the continuous-domain problem is unique, there is no need for a subsequence and the described convergences hold for the sequence $(f^\star_\ell)_{\ell=1}^\infty$ itself. However, situations where the solution is unique are in general not well understood. For the particular case of periodic functions measured by samples of the Fourier transform, it has been shown that the solution $f^\star$ in (3) is unique and that the solutions $f^\star_\ell$ in (30) converge, in the uniform norm, to $f^\star$ (see [18]).

In the case where the solution is non-unique, theorem 5 states that there exists a subsequence of $(f_\ell^\star)_{\ell=1}^\infty$ for which the convergences (33),(34), and (35) hold.

3.4. Local uniform convergence of the grid-based discretization

While theorem 5 is a strong theoretical result, it still lacks practicality: i) one usually only solves a problem up to a certain precision and finds $\tilde{f}_\ell \notin \mathcal{V}_\ell$; and ii) the handling of a subsequence is not something that is easy to do in practice. A remedy to i) is to assume that $\tilde{f}_\ell$ is sufficiently close to some $f_\ell^\star \in \mathcal{V}_\ell$. In contrast to the convergences in theorem 3, the proof of uniform convergence without resort to a subsequence is difficult. Therefore, ii) is still an open problem. We henceforth assume that $\mathcal{V} = \{f^\star\}$. Then $(\tilde{f}_\ell)_{\ell=1}^\infty$ must be convergent to $f^\star$.

Consider a sequence $(\mathbb{X}_\ell)_{\ell=1}^\infty$ of grids that are embedded one into another, with $\mathbb{X}_\ell \subset \mathbb{X}_{\ell+1}$. Without loss of generality, we assume that $\mathbb{X}_\ell = (x_k)_{k=1}^{K_\ell}$. (Observe that we do not require the nodes to be uniformly spaced.) We now consider the grid-based discretization (30), together with its exact discretization (31) on the grid $\mathbb{X}_\ell$. Suppose that the discretized solutions to (31) are obtained through an algorithm that produces an approximate solution $(\tilde{\mathbf{a}}_\ell, \tilde{\mathbf{b}}_\ell) \in \mathbb{R}^{K_\ell} \times \mathbb{R}^M$ with $\tilde{f}_\ell(\cdot) = \sum_{k=1}^{K_\ell} \tilde{a}_{\ell,k} \frac{(\cdot - x_k)_+^{N-1}}{(N-1)!} + \sum_{n=1}^{N} \tilde{b}_{\ell,n} \frac{(\cdot)^{n-1}}{(n-1)!}$. Since $K_\ell < K_{\ell+1} < \infty$, without loss of generality we may assume that $(\tilde{\mathbf{a}}_\ell, \tilde{\mathbf{b}}_\ell) \in \ell_1(\mathbb{N}) \times \mathbb{R}^M$, where $\ell_1(\mathbb{N})$ is the space of sequences whose entries have a finite norm $\|\cdot\|_1$, and $\mathbb{N} = \{1, 2, \ldots\}$

Observe that, with this representation, the mass of the coefficients $\tilde{\mathbf{a}}_\ell$ may get shifted toward larger and larger indices. Nevertheless, this is not an issue because we do not consider the limit of $\tilde{\mathbf{a}}_\ell$ directly. We now state corollary 1, whose proof is in A.5.

Corollary 1. *Let* $((\tilde{\mathbf{a}}_\ell, \tilde{\mathbf{b}}_\ell))_{\ell=1}^\infty \in \left(\ell_1(\mathbb{N}) \times \mathbb{R}^M\right)^{\mathbb{N}}$ *be a sequence of parameters of the approximate solutions. Suppose that*

(i) *the solution is unique, with* $\mathcal{V} = \{f^\star\}$*;*
(ii) *the coefficients* $((\tilde{\mathbf{a}}_\ell, \tilde{\mathbf{b}}_\ell))_{\ell=1}^\infty$ *are such that there exists a sequence of solutions* $(\mathbf{a}_\ell^\star, \mathbf{b}_\ell^\star)$ *of (31) such that*
  $\lim_{\ell \to \infty} \left\|(\mathbf{a}_\ell^\star, \mathbf{b}_\ell^\star) - (\tilde{\mathbf{a}}_\ell, \tilde{\mathbf{b}}_\ell)\right\|_1 = 0$;
(iii) *the grids are well-localized:* $\forall \ell \in \mathbb{N}, \mathbb{X}_\ell \subset [\nu^-, \nu^+]$.

Then, (33),(34), and (35) hold with $\tilde{f}_\ell$ *replacing* $f_{\ell_k}^\star$.

For a numerical resolution of the IP, corollary 1 shows that the acquired sequence $(\tilde{f}_\ell)_{\ell=1}^\infty$ of approximate solutions converges uniformly to a limit solution $f^\star \in \mathcal{V}$, provided that $\mathcal{V} = \{f^\star\}$. The assumption that there exists a sequence of solutions $(\mathbf{a}_\ell^\star, \mathbf{b}_\ell^\star)$ of (31) such that $\lim_{\ell \to \infty} \left\|(\mathbf{a}_\ell^\star, \mathbf{b}_\ell^\star) - (\tilde{\mathbf{a}}_\ell, \tilde{\mathbf{b}}_\ell)\right\|_1 = 0$, simply states that each discretized problem (31) has to be solved with increasing precision, as is often done in practice. Now, unicity is often assumed in order to derive strong convergence results [8, 26, 36] but is difficult to verify in practice. One possibility to mitigate the issues related to non-unicity is to take the solution $\tilde{f}_\ell$ as the initialization of the search for $f_{\ell+1}^\star$ [19]. Doing so, or taking a constant initialization, has allowed us to escape the issues related to non-unicity in section 4.

The speed of convergence of $\tilde{f}_\ell$ to $f^\star$ is investigated in section 4.

4. Monte-Carlo simulations

We perform numerical simulations to illustrate the results of theorem 5 and to investigate the rate of convergence (to be defined) of $f_\ell^\star$ to $f^\star$. We consider the a continuous-domain optimization problem that results in

$$f^\star \in \underset{f\in\mathcal{M}_N(\mathbb{R})}{\operatorname{argmin}} \left(\|\mathbf{y} - \langle f, \boldsymbol{\nu}\rangle\|_2^2 + \lambda \left\|\mathrm{D}^N f\right\|_{\mathcal{M}} \right) \tag{36}$$

and in its discretized version

$$f_\ell^\star \in \underset{f\in\mathcal{M}_N(\mathbb{R},\mathbb{X}_\ell)}{\operatorname{argmin}} \left(\|\mathbf{y} - \langle f, \boldsymbol{\nu}\rangle\|_2^2 + \lambda \left\|\mathrm{D}^N f\right\|_{\mathcal{M}} \right), \tag{37}$$

which we identify into the *equivalent* problem

$$\underset{(\mathbf{a}_\ell,\mathbf{b}_\ell)\in\mathbb{R}^{K_\ell}\times\mathbb{R}^N}{\operatorname{argmin}} \left(\left\|\mathbf{y} - \tilde{\mathbf{A}}_\ell \mathbf{a}_\ell - \mathbf{B}_\ell \mathbf{b}_\ell\right\|_2^2 + \lambda \|\mathbf{a}_\ell\|_1 \right), \tag{38}$$

$$\text{with} \quad \left[\tilde{\mathbf{A}}_\ell\right]_{m,k} = \left\langle \frac{(\cdot - x_k^\ell)_+^{N-1}}{(N-1)!}, \nu_m \right\rangle \quad \text{and} \quad [\mathbf{B}_\ell]_{m,n} = \left\langle \frac{(\cdot)^{n-1}}{(n-1)!}, \nu_m \right\rangle. \tag{39}$$

We solve (38) numerically with the following specificities.

- The solution in 37 is transformed into an equivalent formulation, based on a B-spline representation that which spans the same functional spaces as the Green's function $(\cdot)_+^{N-1}$ [19]. The B-spline-related matrices $\mathbf{A}^\ell$ enjoy a good numerical conditioning. The new optimization problem is solved using a primal-dual splitting algorithm [16], which is stopped at the relative tolerance $\mathrm{tol} = 10^{-5}$.
- The regularizing operator takes the form D^N. We provide simulations for $N = 2$ and $N = 3$.
- The measurement operator $\boldsymbol{\nu} : \mathcal{M}_N(\mathbb{R}) \to \mathbb{R}^M$ is the concatenation of Fourier measurements on the compact set $[0,6]$, so that $\langle f, \nu_m\rangle = \int_0^6 f(x)\mathrm{e}^{-\mathrm{i}\omega_m x}\mathrm{d}x$ at some locations ω_m uniformly sampled in the frequency range $[0,40]$. It follows that $\boldsymbol{\nu}$ is compactly supported with $\mathrm{Csupp}(\boldsymbol{\nu}) = [0,6]$. We chose $M = 70$ and verified experimentally that the unknown function can be correctly reconstructed with this number of measurements.
- We take $\mathbf{y} \in \mathbb{R}^M$ to be the noiseless measurements of a function f_0, which is a linear combination of shifted linear B-splines. The amplitudes and shifts are chosen uniformly random in $[-1,5]$ and $[-1,1]$, in accordance with the strategy of Monte-Carlo simulations.
- The spline knots x_k^ℓ are chosen uniformly spaced in the interval $[-1,7]$, with the parameter n representing the number of nodes. We chose the augmented $[-1,7]$ to include B-splines whose support intersect the targeted domain $[0,6]$.
- Being in a noiseless setting, we choose a small regularisation parameter $\lambda = 10^{-7}$.

The task is to reconstruct the unknown function f_0 through the sequence of estimates $f_\ell^\star$ of (37). Theorem 5 informs us that, up to a subsequence, we ought to observe that

$$\forall 0 \leqslant d \leqslant (N-2), \qquad \lim_{\ell\to\infty} \left\| f_\ell^{\star,(d)} - f^{\star,(d)} \right\|_{2,[0,6]} = 0, \tag{40}$$

where we use the $\mathcal{L}^2$-norm instead of the ∞-norm for practical purpose. We shall study numerically at which rate of convergence R this happens. The rate R cannot, in practice, be computed

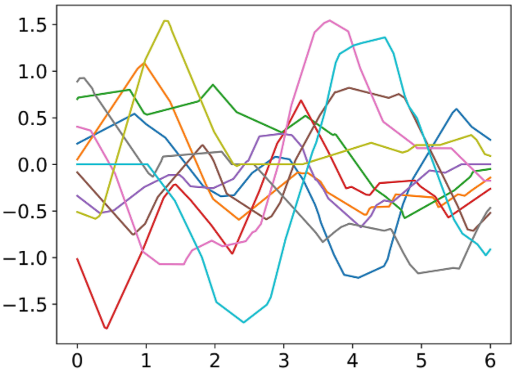

Figure 1. Ten realisations of the Monte-Carlo randomised data.

exactly. Thus, we rely on an *approximate rate of convergence* $\hat{R}$, calculated as follows. We first notice that

$$\exists C > 0 : \quad \lim_{\ell\to\infty} \frac{\|f_\ell^\star - f^\star\|_{2,[0,6]}}{\ell^{-R}} = C \Rightarrow \lim_{\ell\to\infty} \frac{\log_2\left(\|f_\ell^\star - f^\star\|_{2,[0,6]}\right)}{\log_2(\ell)} = -R, \tag{41}$$

which implies that $\log_2(\|f_\ell^\star - f^\star\|_{2,[0,6]})$ behaves approximately linearly in $\log_2(\ell)$. We then replace $f^\star$ by $f_{1024}^\star$ as the finest solution, assuming that $f^\star$ is very close to $f_{1024}^\star$ (theorem 5). Finally, the approximate $\hat{R}$ is calculated by a linear regression in the log–log plot scale according to (41).

To reduce the dependance on specific data, we run the algorithm for 100 random choices of f_0. Ten of them are shown in figure 1, while a single realisation has been considered in 2, where we display the error $\log_2(\|f_\ell^\star - f^\star\|_{2,[0,6]})$. Figure 2 illustrates that convergence happens, approximately linearly in the log scale, and that the error curve is close to the average of all error curves.

We formalize the idea of closeness in table 1, where $\hat{R}$ is a random variable that depends on the realisation of f_0, and where we provide the mean (approximate) rate of convergence $\mu(\hat{R})$, the standard deviation $\sigma(\hat{R})$ of the rate of convergence, as well as its minimum and maximum values min($\hat{R}$) and max($\hat{R}$), respectively. The rate of convergence is slightly better for smoother splines. Finally, we observe a good concentration (small standard deviation) of the rate of convergence around its mean. This indicate that the convergence of the algorithm is stable with respect to the unknown function f_0.

To investigate the rate of convergence of the error $\|f_\ell^\star - f^\star\|_{2,[0,6]}$, one can observe that this error can only be worse (larger) than $\min_{f\in\mathcal{M}_N(\mathbb{R},\mathbb{X}_\ell)} \|f - f^\star\|_{2,[0.6]}$. Interestingly, the rate of convergence of the latter quantity can be calculated. This is done in proposition 2, which is left proof-less as being an extension of [32][theorem 9.14].

Proposition 2. *For all* $f^\star \in \mathcal{M}_N([0,6])$ *and for all* $0 \leqslant d \leqslant N-1$

$$\min_{f\in\mathcal{M}_N(\mathbb{R},\mathbb{X}_\ell)} \left\| f^{(d)} - f^{\star,(d)} \right\|_{2,[0,6]} = \mathcal{O}\left(\|f^\star\|_{\mathcal{M}_N} 2^{-\frac{(N-d)}{2}\ell} \right). \tag{42}$$

Proposition 2 reveals that the rate of convergence of $\min_{f\in\mathcal{M}_N(\mathbb{R},\mathbb{X}_\ell)} \|f - f^\star\|_{2,[0.6]}$ is never smaller than $\frac{N-d}{2}$. For the three examples taken in the table 1, proposition 2 suggests the rates

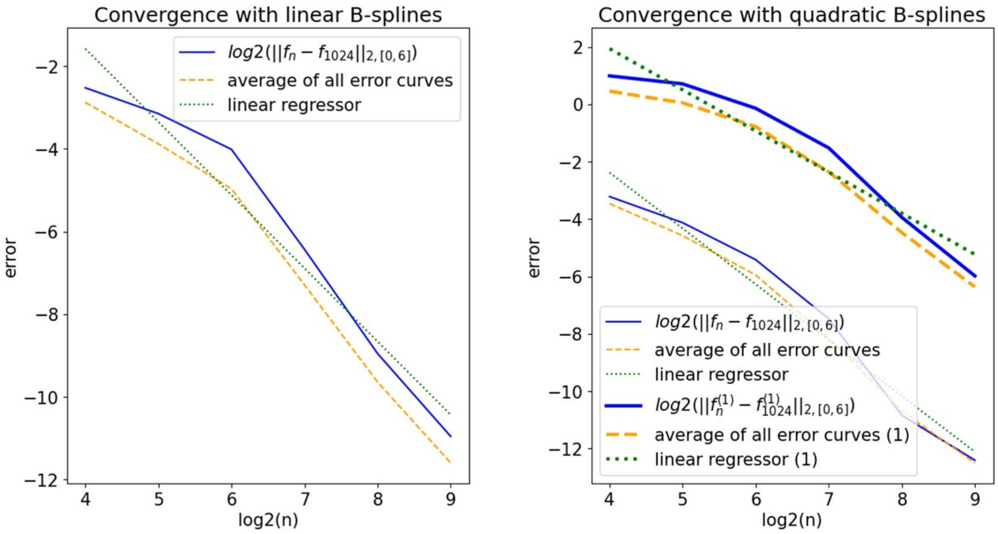

Figure 2. Convergence of the algorithm for one realisation of f_0. Convergence with B-spline of Order 2 means that the reconstruction has been done with B-splines of order 2 (equivalently $N = 2$). The linear regressor corresponds to the linear regression on the error curve. The average of all error curves corresponds to the mean error curve on all 100 simulations. Finally, linear regressor (1) and average of all curves (1) mean the same as before but applied on the error curves of the derivatives.

Table 1. Spline2-Derivative0 refers to the approximation error of the derivatives of order 0 and to the reconstruction made with linear B-splines (order 2). Likewise, Spline3 refers to quadratic B-splines (order 3).

	$\mu(\hat{R})$	$\sigma(\hat{R})$	$\min(\hat{R})$	$\max(\hat{R})$
Spline2-Derivative0	1.81	0.06	1.61	1.99
Spline3-Derivative0	1.88	0.21	1.41	2.36
Spline3-Derivative1	1.41	0.11	1.08	1.64

$\{1, 1.5, 1\}$. We observed an average of $\{1.81, 1.88, 1.41\}$, such that the rates are always higher, in these examples, than the suggested (lowest) ones. To conclude, theorem 5 predicted that local uniform convergence hold, without predicting the rate of convergence, latter suggested by proposition 2. Our numerical experiments lead to the following conclusions.

- The rate of convergence (and, thus, the convergence) is slightly better for smoother splines (higher N).
- The rate of convergence is stable with respect to the the unknown f_0 (small standard variation).

Data availability statement

The data cannot be made publicly available upon publication because they are not available in a format that is sufficiently accessible or reusable by other researchers. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgment

Vincent Guillemet was supported in part by the European Research Council (ERC Project FunLearn) under Grant 101020573, in part by the Swiss National Science Foundation (SNSF) under grant 200020_219356.

Appendix. Proofs

A.1. Mathematical preliminaries

A.1.1. On the notion of continuity in banach spaces. We recall the notions of continuity, weak* continuity and weak* topology on Banach spaces.

Definition 5. Let $\mathcal{B}$ be a Banach space and $\mathcal{B}'$ be its functional dual.

- A sequence $\{f_n\}_{n=1}^{\infty} \subset \mathcal{B}'$ of functionals is said to be norm or strongly convergent to $f \in \mathcal{B}'$ if $\lim\limits_{n\to\infty} \|f_n - f\|_{\mathcal{B}'} = 0$, where the dual norm $\|\cdot\|_{\mathcal{B}'}$ is defined as $\|f\|_{\mathcal{B}'} = \sup\limits_{\phi\in\mathcal{B}, \|\phi\|_{\mathcal{B}}=1} |\langle f, \phi\rangle|$.
- A sequence $\{f_n\}_{n=1}^{\infty} \subset \mathcal{B}'$ of functionals is said to be weak* convergent to $f \in \mathcal{B}'$ if $\forall \phi \in \mathcal{B}$: $\lim\limits_{n\to\infty} \langle f_n, \phi\rangle = \langle f, \phi\rangle$.
- The weak* topology on $\mathcal{B}'$ is the coarsest topology that makes the mappings $\phi \mapsto \langle f, \phi\rangle$ continuous for all $\phi \in \mathcal{B}$.

A classical result from functional analysis states that norm convergence implies weak* convergence. The converse is false.

A.1.2. On the structure of the space $\mathcal{M}_N(\mathbb{R})$ and its predual. The function space $\mathcal{M}_N(\mathbb{R})$ is defined as $\mathcal{M}_N(\mathbb{R}) = \{f \in \mathcal{S}'(\mathbb{R}) : \quad \|\mathrm{D}^N f\|_{\mathcal{M}} < \infty\}$, and the nullspace span $\left\{\frac{(\cdot)^{n-1}}{(n-1)!}\right\}_{n=1}^{N}$ of its semi norm $\|\mathrm{D}^N \cdot\|_{\mathcal{M}}$ is denoted by $\mathcal{N}_N$. From now on, we use the notation $\forall n \in [1, N]$, $p_n = \frac{(\cdot)^{n-1}}{(n-1)!}$, $\rho_N(\cdot) = \frac{(\cdot)_+^{N-1}}{(N-1)!}$. The goal of this section is to recall that there exists a space $\mathcal{X}$, to be defined, such that $\mathcal{X}' = \mathcal{M}_N(\mathbb{R})$, and that some derivatives of Diracs belong to $\mathcal{X}$.

We are going to 'invert' the operator D^N to study the topological structure of $\mathcal{M}_N(\mathbb{R})$. To this end, we introduce a biorthogonal sequence from which we will derive a right-inverse.

Definition 6. Let $\mathbf{p} = (p_n)_{n=1}^N$ be as previously defined, and $\boldsymbol{\phi} = (\phi_n)_{n=1}^N$ be a complementary set of functions. Then, the system $(\mathbf{p}, \boldsymbol{\phi})$ is said to be D^N- admissible if

(1) the basis functions are biorthogonal, so that $\forall m, n \in [1, N] \quad \langle p_n, \boldsymbol{\phi}_m\rangle = \delta_{m,n}$;
(2) all functions ϕ_n are test functions, so that $\forall 1 \leqslant n \leqslant N, \phi_n \in \mathcal{S}(\mathbb{R})$;
(3) all functions $\boldsymbol{\phi}_n$ are compactly supported and causal, so that $\exists \phi^+ \in \mathbb{R}^+$ such that

$$\forall 1 \leqslant n \leqslant N, \quad \mathrm{supp}\,(\phi_n) \subset \left[0, \phi^+\right]. \tag{A.1}$$

Item 2 of definition 6 assures that $\langle p_n, \phi_m\rangle$ is well defined. Let $(\mathrm{D}^N)_{\phi}^{-1}$ be defined as

$$\left(\mathrm{D}^N\right)_{\phi}^{-1} : \begin{cases} \mathcal{M}(\mathbb{R}) \to \left(\mathrm{D}^N\right)_{\phi}^{-1} \mathcal{M}(\mathbb{R}) \\ m \mapsto \int_{\mathbb{R}} g_{\phi}(\cdot, y)\,\mathrm{d}m(y), \end{cases} \qquad \text{with} \quad g_{\phi}(x,y) = \rho_N(x-y) - \sum_{n=1}^{N} p_n(x)\left(\rho_N^{\star} * \phi_n\right)(y), \tag{A.2}$$

where $\rho_N^{\star} = \frac{(-\cdot)_+^{N-1}}{(N-1)!}$ is the Green's function of $\mathrm{D}^{N,\star}$, the adjoint of D^N.

This rest of this section is organised as follows: theorem 6 states that $(\mathrm{D}^N)_{\phi}^{-1}$ is well-defined and exhibits $\mathcal{M}_N(\mathbb{R})$ as the direct sum of simpler spaces. Theorem 7 states that $\mathcal{C}_N(\mathbb{R})$ is a predual of $\mathcal{M}_N(\mathbb{R})$, which shows that $\mathcal{M}_N(\mathbb{R})$ can be equipped with a weak* topology. Theorem 7 also reveals (Item 4) a sufficient, easy-to-verify condition for a generalized function ρ to belong in $\mathcal{C}_N(\mathbb{R})$. Corollary 2 takes advantage of this condition to state that some derivatives of Diracs belong to $\mathcal{C}_N(\mathbb{R})$.

Theorem 6 ([43](theorem 4&5)). *Let $(\mathbf{p}, \phi)$ be a D^N-admissible system. Then, the following holds true.*

(1) *The space $(\mathrm{D}^N)_{\phi}^{-1}\mathcal{M}(\mathbb{R}) = \{\int_{\mathbb{R}} g_{\phi}(\cdot,y)\mathrm{d}m(y), \quad m \in \mathcal{M}(\mathbb{R})\}$ is a Banach space for the norm $\|\mathrm{D}^N \cdot\|_{\mathcal{M}}$. Moreover,*

$$\forall m \in \mathcal{M}(\mathbb{R}), \mathrm{D}^N \int_{\mathbb{R}} g_{\phi}(\cdot, y)\,\mathrm{d}m(y) = m, \tag{A.3}$$

$$\forall n \in [1,N], \forall m \in \mathcal{M}(\mathbb{R}), \left\langle \left(\mathrm{D}^N\right)_{\phi}^{-1} m, \phi_n \right\rangle = 0. \tag{A.4}$$

(2) *The space $\mathcal{M}_N(\mathbb{R})$ is a Banach space for the norm $\|f\|_{\mathcal{M}_N} = \|\mathrm{D}^N f\|_{\mathcal{M}} + \sqrt{\sum_{n=1}^{N}\langle f, \phi_n\rangle^2}$. Moreover, its admits the direct sum decomposition $\mathcal{M}_N(\mathbb{R}) = (\mathrm{D}^N)_{\phi}^{-1}\mathcal{M}(\mathbb{R}) \oplus \mathcal{N}_N$.*

Theorem 7. *Let $(\mathbf{p}, \phi)$ be a D^N-admissible system. Then, the function space*

$$\mathcal{C}_N(\mathbb{R}) = \left\{ \mathrm{D}^{N,\star} v + \sum_{n=1}^{N} a_n \phi_n : v \in \mathcal{C}_0(\mathbb{R}), \mathbf{a} = (a_n)_{n=1}^{N} \in \mathbb{R}^N \right\} \tag{A.5}$$

has the following properties:

(1) *it has the direct-sum representation $\mathcal{C}_N(\mathbb{R}) = \mathrm{D}^{N,\star}\mathcal{C}_0(\mathbb{R}) \oplus \mathrm{span}\{\phi_n\}_{n=1}^{N}$;*
(2) *it is a Banach space for the norm $\|g\|_{\mathcal{C}_N} = \max(\|\mathbf{v}\|_{\infty}, \|\mathbf{a}\|_2), \forall g \in \mathcal{C}_N(\mathbb{R})$;*
(3) *it is the predual of $\mathcal{M}_N(\mathbb{R})$, so that $\mathcal{C}_N(\mathbb{R})' = \mathcal{M}_N(\mathbb{R})$, where $\mathcal{M}_N(\mathbb{R})$ is equipped with the norm $\|\mathrm{D}^N \cdot\|_{\mathcal{M}} + \left\|(\langle \cdot, \phi_n\rangle)_{n=1}^{N}\right\|_2$;*
(4) *if $\rho \in \mathcal{S}'(\mathbb{R})$ is compactly supported and is such that the distribution*

$$\int_{\mathbb{R}} g_{\phi}(x,y)\,\rho(x)\,\mathrm{d}x = \left(\rho_N^{\star} * \rho\right)(y) - \sum_{n=1}^{N} \langle \rho, p_n\rangle \left(\rho_N^{\star} * \phi_n\right)(y) \tag{A.6}$$

is in $\mathcal{C}_0(\mathbb{R})$, then $\rho \in \mathcal{C}_N(\mathbb{R})$.

Proof of theorem 7. Items 1, 2, and 3 of theorem 7 are simple restatements of [43][theorem 6]. Thus, we only prove Item 4. We claim that

$$\mathcal{C}_N(\mathbb{R}) \ni \tilde{\rho}(\cdot) = \mathrm{D}^{N,\star} \int_{\mathbb{R}} g_{\boldsymbol{\phi}}(x,\cdot)\,\rho(x)\,\mathrm{d}x + \sum_{n=1}^{N} \phi_n(\cdot)\,\langle \rho, p_n \rangle \tag{A.7}$$

is such that $\tilde{\rho} = \rho$, which would prove the claim. We calculate that

$$\begin{aligned}
\tilde{\rho} &= \mathrm{D}^{N,\star}\left(\rho_N^\star * \rho - \sum_{n=1}^{N} \langle \rho, p_n \rangle \rho_N^\star * \phi_n\right) + \sum_{n=1}^{N} \phi_n \langle \rho, p_n \rangle && \text{(A.8)}\\
&= \mathrm{D}^{N,\star} \rho_N^\star * \rho - \sum_{n=1}^{N} \langle \rho, p_n \rangle \mathrm{D}^{N,\star} \rho_N^\star * \phi_n + \sum_{n=1}^{N} \phi_n \langle \rho, p_n \rangle && \text{(A.9)}\\
&= \delta * \rho - \sum_{n=1}^{N} \langle \rho, p_n \rangle \delta * \phi_n + \sum_{n=1}^{N} \phi_n \langle \rho, p_n \rangle = \rho. && \text{(A.10)}
\end{aligned}$$

In (A.9) we used the associativity of convolution, which holds as soon as the distributions are one-sided. For example, causal. ☐

Corollary 2. *Let* $(\mathbf{p}, \boldsymbol{\phi})$ *be a* D^N*-admissible system. Then,*

$$\forall 0 \leqslant d \leqslant N-2, \forall x' \in \mathbb{R}: \quad \delta_{x'}^{(d)} \in \mathcal{C}_N(\mathbb{R}). \tag{A.11}$$

Proof. The result follows directly from Item 4 of theorem 7 and the fact, to be shown, that $g_{\boldsymbol{\phi}}(x',\cdot)$ is compactly supported and in $\mathcal{C}^{N-2}(\mathbb{R})$. We first show that $\forall x' \in \mathbb{R}$, $\mathrm{supp}(g_{\boldsymbol{\phi}}(x',\cdot)) \subset [\min(0,x'), \max(\phi^+, x')]$. Indeed, $\mathrm{supp}(\rho_N(\cdot - y)) = [y, \infty[$ and $\mathrm{supp}(\phi_n) \subset [0, \phi^+]$. Therefore if $y > \max(x', \phi^+)$ then $\rho_N(x' - y) = 0$ and $\langle \rho_N(\cdot - y), \phi_n \rangle = 0$ because the supports are disjoint. Hence, $g_{\boldsymbol{\phi}}(x', y) = 0$ if $y > \max(x', \phi^+)$. When $y < \min(0, x')$, we compute that

$$\begin{aligned}
g_{\boldsymbol{\phi}}(x',y) &= \rho_N(x'-y) - \sum_{n=1}^{N} p_n(x')\,\langle \rho_N(\cdot - y), \phi_n \rangle && \text{(A.12)}\\
&= \rho_N(x'-y) - \sum_{n=1}^{N} \frac{x'n-1}{(N-1)!} \left\langle u(\cdot - y)\,\frac{(\cdot - y)^{N-1}}{(N-1)!}, \phi_n \right\rangle && \text{(A.13)}\\
&= \rho_N(x'-y) - \sum_{n=1}^{N} \frac{x'n-1}{(N-1)!} \left\langle \sum_{k=0}^{N-1} \frac{(\cdot)^k\,(-y)^{N-1-k}}{k!\,(N-1-k)!}, \phi_n \right\rangle && \text{(A.14)}\\
&= \rho_N(x'-y) - \sum_{n=1}^{N} \frac{x'n-1}{(N-1)!}\,\frac{(-y)^{N-1-n}}{(N-1-n)!} = \rho_N(x'-y) - \frac{(x'-y)^{N-1}}{(N-1)!} = 0. && \text{(A.15)}
\end{aligned}$$

This shows that $g_{\boldsymbol{\phi}}(x',\cdot)$ is compactly supported in $[\min(0,x'), \max(x', \phi^+)]$. Furthermore, $\forall x' \in \mathbb{R}$, $g_{\boldsymbol{\phi}}(x',\cdot) \in \mathcal{C}^{N-2}(\mathbb{R})$. Indeed $\rho_N(x' - \cdot) \in \mathcal{C}^{N-2}(\mathbb{R})$ by definition and $\rho_N^\star * \phi_n \in \mathcal{C}^{N-2}(\mathbb{R})$ because the convolution between a $\mathcal{C}^\infty(\mathbb{R})$ function and a $\mathcal{C}^{N-2}(\mathbb{R})$ function is always in $\mathcal{C}^\infty(\mathbb{R})$. ☐

A.2. Proofs of section 2.2

Proof of proposition1. Theorem 6 reveals that

$$\mathcal{M}_N(\mathbb{R}) = \left(\mathrm{D}^N\right)_\phi^{-1} \mathcal{M}(\mathbb{R}) \oplus \mathcal{N}_N. \tag{A.16}$$

Using the fact that $(\mathrm{D}^N)_\phi^{-1}$ is represented by the kernel

$$g_\phi(x,y) = \frac{(x-y)_+^{N-1}}{(N-1)!} - \sum_{n=1}^N \left\langle \frac{(\cdot - y)_+^{N-1}}{(N-1)!}, \phi_n \right\rangle \frac{(x)^{n-1}}{(n-1)!}, \tag{A.17}$$

we conclude that all functions $f \in \mathcal{M}_N([\nu^-,\nu^+])$ are of the form, for $m \in \mathcal{M}([\nu^-,\nu^+])$ and $\mathbf{b} = (b_n)_{n=1}^N \in \mathbb{R}^N$,

$$f(x) = \int_{\nu^-}^{\nu^+} g_\phi(x,y)\,\mathrm{d}m(y) + \sum_{n=1}^N b_n \frac{x^{n-1}}{(n-1)!} \tag{A.18}$$

$$= \int_{\nu^-}^{\nu^+} \left(\frac{(x-y)_+^{N-1}}{(N-1)!} - \sum_{n=1}^N \left\langle \frac{(\cdot - y)_+^{N-1}}{(N-1)!}, \phi_n \right\rangle \frac{(x)^{n-1}}{(n-1)!} \right) \mathrm{d}m(y) + \sum_{n=1}^N b_n \frac{x^{n-1}}{(n-1)!} \tag{A.19}$$

$$= \int_{\nu^-}^{\nu^+} \frac{(x-y)_+^{N-1}}{(N-1)!} \mathrm{d}m(y) + \sum_{n=1}^N \tilde{b}_n \frac{x^{n-1}}{(n-1)!}, \tag{A.20}$$

$$\text{for} \quad \tilde{b}_n = b_n - \int_{\nu^-}^{\nu^+} \left\langle \frac{(\cdot - y)_+^{N-1}}{(N-1)!}, \phi_n \right\rangle \mathrm{d}m(y). \tag{A.21}$$

We used the fact that two integrals $\int_{\nu^-}^{\nu^+} \frac{(x-y)_+^{N-1}}{(N-1)!}\mathrm{d}m(y)$ and $\int_{\nu^-}^{\nu^+} \left\langle \frac{(\cdot-y)_+^{N-1}}{(N-1)!}, \phi_n \right\rangle \mathrm{d}m(y)$ happen to be well-defined, which does not arrive when $]\nu^-,\nu^+[=]-\infty,\infty[$. Based on that, we also used the linearity of the integral with respect to the x variable. Further, the inclusion

$$\mathcal{M}_N\left(\left[\nu^-,\nu^+\right]\right) \subset \mathcal{C}^{N-2}(\mathbb{R}) \tag{A.22}$$

comes directly from the application of the Lebesgue dominated-convergence theorem to the integral representation (A.21), to find that

$$\mathrm{D}^{N-2} f = \mathrm{D}^{N-2} \left(\frac{\int_{\nu^-}^{\nu^+} (\cdot - y)_+^{N-1}}{(N-1)!} \mathrm{d}m(y) + \sum_{n=1}^N \frac{\tilde{b}_n}{(n-1)!} (\cdot)^{n-1} \right) \tag{A.23}$$

$$= \int_{\nu^-}^{\nu^+} \mathrm{D}^{N-2} \frac{(\cdot - y)_+^{N-1}}{(N-1)!} \mathrm{d}m(y) + \tilde{b}_N(\cdot) + \tilde{b}_{N-1} \tag{A.24}$$

$$= \int_{\nu^-}^{\nu^+} (\cdot - y)_+ \,\mathrm{d}m(y) + \tilde{b}_N(\cdot) + \tilde{b}_{N-1}, \tag{A.25}$$

which is a continuous function. This concludes the proof. □

A.3. Proofs of section 3.2

Proof of theorem3. Due to its length, the proof is split into three parts. Prior to the start, we recall that a lower-semicontinuous convex functional is continuous on the interior of the set where it takes finite values. Then, due to Assumption 4, we conclude that $E(\mathbf{y},\cdot)$ is continuous on this set, which we name $\mathcal{U}$. The weak* continuity of the mapping $\boldsymbol{\nu}:\mathcal{M}_N(\mathbb{R})\to\mathbb{R}^M$ implies that that $\mathcal{U}'=\boldsymbol{\nu}^{-1}(\mathcal{U})$ is an open set in the weak* topology. Finally we remark two important points.

- The set $\mathcal{U}'$ is exactly the set where $E(\mathbf{y},\boldsymbol{\nu}(\boldsymbol{\cdot}))$ takes finite values.
- The data-fidelity $E(\mathbf{y},\boldsymbol{\nu}(\cdot)):\mathcal{U}'\to\mathbb{R}^M$ is weak* continuous.

Using the previous remark, we assume throughout this proof that $E(\mathbf{y},\boldsymbol{\nu}(\cdot))$ is weak* continuous.

Step 1: here, we prove that

$$\lim_{\ell\to\infty}\mathcal{J}\left(f_\ell^\star\right)=\mathcal{J}\left(f^\star\right).\tag{A.26}$$

Let $f^\star\in\mathcal{V}$ be an extreme point and $(f_\ell^\star)_{\ell=1}^\infty$ be a sequence of solutions such that $f_\ell^\star\in\mathcal{V}_\ell$. Theorem 2 states that

$$f^\star(\cdot)=\sum_{m=1}^{M-N}a_m\frac{(\cdot-x_m)_+^{N-1}}{(N-1)!}+\sum_{n=1}^{N}b_n\frac{(\cdot)^{n-1}}{(n-1)!},\quad x_m\in\left[\nu^-,\nu^+\right].\tag{A.27}$$

Define

$$\hat{x}_m^\ell=\underset{z\in\mathbb{X}_\ell}{\mathrm{argmin}}|x_m-z|,\quad \widehat{f_\ell}=\sum_{m=1}^{M-N}a_m\frac{\left(\cdot-\hat{x}_m^\ell\right)_+^{N-1}}{(N-1)!}+\sum_{n=1}^{N}\frac{b_n(\cdot)^{n-1}}{(n-1)!}\in\mathcal{M}_N(\mathbb{R},\mathbb{X}_\ell)\tag{A.28}$$

and observe that Assumption 1 implies that

$$\lim_{\ell\to\infty}|\hat{x}_m^\ell-x_m|=0.\tag{A.29}$$

We first want to show that $\hat{f}_\ell$ is weak* convergent to $f^\star$. Recall that $\mathcal{C}_N(\mathbb{R})=\mathrm{D}^{N,\star}\mathcal{C}_0(\mathbb{R})\oplus\mathrm{span}\{\phi_n\}_{n=1}^N$ (proposition 7). Since the term $\sum_{n=1}^N\frac{b_n(\cdot)^{n-1}}{(n-1)!}$ is independant of ℓ, it is sufficient to show that

(i) for all $\phi\in\mathcal{C}_0(\mathbb{R})$: $\lim\limits_{\ell\to\infty}\left\langle\frac{(\cdot-\hat{x}_m^\ell)_+^{N-1}}{(N-1)!},\mathrm{D}^{N,\star}\phi\right\rangle=\left\langle\frac{(\cdot-x_m)_+^{N-1}}{(N-1)!},\mathrm{D}^{N,\star}\phi\right\rangle$;

(ii) for all $\sum_{n=1}^N c_n\phi_n\in\mathrm{span}\{\phi_n\}_{n=1}^N$: $\lim\limits_{\ell\to\infty}\left\langle\frac{(\cdot-\hat{x}_m^\ell)_+^{N-1}}{(N-1)!},\sum_{n=1}^N c_n\phi_n\right\rangle=\left\langle\frac{(\cdot-x_m)_+^{N-1}}{(N-1)!},\sum_{n=1}^N c_n\phi_n\right\rangle$.

For Item i, using the continuity of ϕ and the fact that $\frac{(\cdot)_+^{N-1}}{N-1!}$ is a Green's function of D^N, we find that

$$\lim_{\ell\to\infty}\left\langle \frac{\left(\cdot-\hat{x}_m^\ell\right)_+^{N-1}}{(N-1)!},\mathrm{D}^{N,\star}\phi\right\rangle = \lim_{\ell\to\infty}\left\langle \mathrm{D}^N\frac{\left(\cdot-\hat{x}_m^\ell\right)_+^{N-1}}{(N-1)!},\phi\right\rangle = \lim_{\ell\to\infty}\left\langle \delta_{\hat{x}_m^\ell}(N-1)!,\phi\right\rangle \tag{A.30}$$

$$= \langle (N-1)!\delta_{x_m},\phi\rangle = \left\langle \frac{(\cdot-x_m)_+^{N-1}}{(N-1)!},\mathrm{D}^{N,\star}\phi\right\rangle. \tag{A.31}$$

For Item ii, we find that

$$\lim_{\ell\to\infty}\left\langle \frac{\left(\cdot-\hat{x}_m^\ell\right)_+^{N-1}}{(N-1)!},\sum_{n=1}^N c_n\phi_n\right\rangle = \lim_{\ell\to\infty}\sum_{n=1}^N c_n\left\langle \frac{\left(\cdot-\hat{x}_m^\ell\right)_+^{N-1}}{(N-1)!},\phi_n\right\rangle \tag{A.32}$$

$$= \left\langle \frac{(\cdot-x_m)_+^{N-1}}{(N-1)!},\sum_{n=1}^N c_n\phi_n\right\rangle, \tag{A.33}$$

where to pass to the limit we used the fact that ϕ_ℓ is a compactly supported function in $\mathcal{S}(\mathbb{R})$ and the fact that $\frac{(\cdot-x_m)_+^{N-1}}{(N-1)!}\in\mathcal{C}^{N-2}(\mathbb{R})$. It follows that $f^\star$ is indeed the weak* limit of $\hat{f}_\ell$. Moreover, $\mathcal{J}$ is weak* continuous because $E(\mathbf{y},\boldsymbol{\nu}(\cdot))$ and $\left\|\mathrm{D}^N\cdot\right\|_{\mathcal{M}}$ are weak* continuous and ψ is continuous. In consequence,

$$\mathcal{J}(f^\star)\leqslant\lim_{\ell\to\infty}\mathcal{J}(f_\ell^\star)\leqslant\lim_{\ell\to\infty}\mathcal{J}\left(\widehat{f}_\ell\right)=\mathcal{J}(f^\star)\Rightarrow\lim_{\ell\to\infty}\mathcal{J}(f_\ell^\star)=\mathcal{J}(f^\star). \tag{A.34}$$

Step 2: here, we prove that, for any sequence of solutions $(f_\ell^\star)_{\ell=1}^\infty$, there exists a subsequence $(f_{\ell_k}^\star)_{k=1}^\infty$ of $(f_\ell^\star)_{\ell=1}^\infty$ that is weak* convergent to some minimizer $\tilde{f}\in\mathcal{V}$. As starting point, we use the direct-sum decomposition of theorem 6, to produce

$$f_\ell^\star = g_\ell^\star + h_\ell^\star,\quad\text{with}\quad g_\ell^\star=\left(\mathrm{D}^N\right)_{\boldsymbol{\phi}}^{-1}m_\ell\quad\text{and}\quad h_\ell^\star\in\mathcal{N}_N. \tag{A.35}$$

We know by Step 1 that $\lim\limits_{\ell\to\infty}\mathcal{J}(f_\ell^\star)$ is a convergent sequence. In consequence, for $C=\max\limits_{\ell\in\mathbb{N}}|\mathcal{J}(f_\ell^\star)|<\infty$ and A such that $\inf\limits_{\mathbf{z}\in\mathbb{R}^M}E(\mathbf{y},\mathbf{z})=A>(-\infty)$, it follows that, $\forall\ell\in\mathbb{N}$,

$$E(y,\langle f_\ell^\star,\boldsymbol{\nu}\rangle)+\psi\left(\left\|\mathrm{D}^N f_\ell^\star\right\|_{\mathcal{M}}\right)\leqslant C\Rightarrow\psi\left(\left\|\mathrm{D}^N f_\ell^\star\right\|_{\mathcal{M}}\right)\leqslant C-E(y,\langle f_\ell^\star,\boldsymbol{\nu}\rangle) \tag{A.36}$$

$$\Rightarrow\psi\left(\left\|\mathrm{D}^N f_\ell^\star\right\|_{\mathcal{M}}\right)\leqslant C-A \tag{A.37}$$

$$\Rightarrow\left\|\mathrm{D}^N f_\ell^\star\right\|_{\mathcal{M}}=\left\|\mathrm{D}^N g_\ell^\star\right\|_{\mathcal{M}}\leqslant\psi^{-1}(C-A). \tag{A.38}$$

Thus, the sequence $g_\ell^\star$ lives in the ball of radius $\psi^{-1}(C-A)$, in the space $(\mathrm{D}^N)_{\boldsymbol{\phi}}^{-1}\mathcal{M}(\mathbb{R})$, equipped with the norm $\left\|\mathrm{D}^N\cdot\right\|_{\mathcal{M}}$. Moreover, $\forall\ell\in\mathbb{N}$,

$$E(y,\langle f_\ell^\star,\boldsymbol{\nu}\rangle)+\psi\left(\left\|\mathrm{D}^N f_\ell^\star\right\|_{\mathcal{M}}\right)\leqslant C\Rightarrow E(y,\langle g_\ell^\star,\boldsymbol{\nu}\rangle+\langle h_\ell^\star,\boldsymbol{\nu}\rangle)\leqslant C. \tag{A.39}$$

We then prove that $\sup\limits_{\ell\in\mathbb{N}}\|\langle h_\ell^\star,\boldsymbol{\phi}\rangle\|_2<\infty$. Suppose by contradiction that $\sup\limits_{\ell\in\mathbb{N}}\|\langle h_\ell^\star,\boldsymbol{\phi}\rangle\|_2=\infty$. Then, we find a subsequence $(h_{\ell_k}^\star)_{k=1}^\infty$ such that

$$\lim_{k\to\infty}\left\|\langle h^{\star}_{\ell_k},\boldsymbol{\phi}\rangle\right\|_2=\infty\Rightarrow\lim_{k\to\infty}\left\|\langle h^{\star}_{\ell_k},\boldsymbol{\nu}\rangle\right\|_2=\infty \tag{A.40}$$

$$\Rightarrow\lim_{k\to\infty}\left\|\langle f^{\star}_{\ell_k},\boldsymbol{\nu}\rangle\right\|_2\geqslant\lim_{k\to\infty}\left\|\langle h^{\star}_{\ell_k},\boldsymbol{\nu}\rangle\right\|_2-\sup_{k\in\mathbb{N}}\left\|\langle g^{\star}_{\ell_k},\boldsymbol{\nu}\rangle\right\|_2=\infty \tag{A.41}$$

$$\Rightarrow\lim_{k\to\infty}E\left(\mathbf{y},\langle f^{\star}_{\ell_k},\boldsymbol{\nu}\rangle\right)=\infty, \tag{A.42}$$

where (A.40) holds because the assumption $\mathcal{N}_{\boldsymbol{\nu}}\cap\mathcal{N}_N=\emptyset$ implies that $\boldsymbol{\nu}:\mathcal{N}_N\to\mathbb{R}^M$ is an isomorphism onto its range. Hence, $\|\langle\cdot,\boldsymbol{\phi}\rangle\|_2$ is a norm on $\mathcal{N}_N$. Thus, if $\left\|\langle h^{\star}_{\ell_k},\boldsymbol{\phi}\rangle\right\|_2$ diverges, then so does $\left\|\langle h^{\star}_{\ell_k},\boldsymbol{\nu}\rangle\right\|_2$. Implication (A.41) holds because of the triangular inequality and the fact that $\sup\limits_{k\in\mathbb{N}}\left\|\langle g^{\star}_{\ell_k},\boldsymbol{\nu}\rangle\right\|_2<\infty$, which is a consequence of the boundedness of $\{g^{\star}_{\ell}\}_{\ell=1}^{\infty}$ and the weak* continuity of $\boldsymbol{\nu}$. Finally, A.42 holds because of the coercivity of $E(\mathbf{y},\cdot)$. Observe that (A.42) is in contradiction with (A.39). In consequence, we find that $B=\sup\limits_{\ell\in\mathbb{N}}\|\langle h^{\star}_{\ell},\boldsymbol{\phi}\rangle\|_2<\infty$. Thus,

$$\|f^{\star}_{\ell}\|_{\mathcal{M}_N}=\left\|\mathrm{D}^N g^{\star}_{\ell}\right\|_{\mathcal{M}}+\|\langle h^{\star}_{\ell},\boldsymbol{\phi}\rangle\|_2\leqslant\psi^{-1}(C-A)+B \tag{A.43}$$

and the sequence $(f^{\star}_{\ell})_{\ell=1}^{\infty}$ lives in a ball (in $\mathcal{M}_N(\mathbb{R})$) of radius $\psi^{-1}(C-A)+B$. According to the Banach–Alaoglu theorem, this ball is weak* compact and we may find a subsequence of $(f^{\star}_{\ell_k})_{k=1}^{\infty}$ such that $\lim\limits_{k\to\infty}f^{\star}_{\ell_k}=\tilde{f}$ in the weak* sense. Moreover, we get that

$$\mathcal{J}(f^{\star})=\lim_{k\to\infty}\mathcal{J}\left(f^{\star}_{\ell_k}\right)=\mathcal{J}\left(\tilde{f}\right), \tag{A.44}$$

where the first equality follows from Step 1 and the second equality follows from the weak* convergence of $f^{\star}_{\ell_k}$ to $\tilde{f}$ and the weak* continuity of $\mathcal{J}$. Thus, we have found that $(f^{\star}_{\ell})_{\ell=1}^{\infty}$ has a subsequence $(f^{\star}_{\ell_k})_{k=1}^{\infty}$ that is convergent to a solution of the continuous problem $\tilde{f}\in\mathcal{V}$. Using the weak* continuity of $E(\mathbf{y},\boldsymbol{\nu}(\cdot))$ and $\left\|\mathrm{D}^N\cdot\right\|_{\mathcal{M}}$, we also find that

$$\lim_{k\to\infty}\left\|\mathrm{D}^N f^{\star}_{\ell_k}\right\|_{\mathcal{M}}=\left\|\mathrm{D}^N\tilde{f}\right\|_{\mathcal{M}},\qquad\lim_{k\to\infty}\langle f^{\star}_{\ell_k},\boldsymbol{\nu}\rangle=\langle\tilde{f},\boldsymbol{\nu}\rangle. \tag{A.45}$$

<u>**Step 3:**</u> we upgrade the convergences established in Step 2 to the whole sequence $(f^{\star}_{\ell})_{\ell=1}^{\infty}$, taking into account that $(\langle f^{\star}_{\ell},\boldsymbol{\nu}\rangle)_{\ell=1}^{\infty}$ is a bounded sequence. Suppose by contradiction that it is not convergent. Then, we may find two subsequences $\left(\langle f^{\star}_{\sigma(\ell)},\boldsymbol{\nu}\rangle\right)_{\ell=1}^{\infty}$ and $\left(\langle f^{\star}_{\sigma'(\ell)},\boldsymbol{\nu}\rangle\right)_{\ell=1}^{\infty}$ with different limits $\mathbf{a}$ and $\mathbf{a}'$, where σ and σ' are increasing mappings from $\mathbb{N}$ to $\mathbb{N}$. Note that, by Step 1,

$$\lim_{\ell\to\infty}\mathcal{J}\left(f^{\star}_{\sigma(\ell)}\right)=\lim_{\ell\to\infty}\mathcal{J}\left(f^{\star}_{\ell}\right)=\mathcal{J}\left(\tilde{f}\right),\quad\text{and}\quad\lim_{\ell\to\infty}\mathcal{J}\left(f^{\star}_{\sigma'(\ell)}\right)=\lim_{\ell\to\infty}\mathcal{J}\left(f^{\star}_{\ell}\right)=\mathcal{J}\left(\tilde{f}\right). \tag{A.46}$$

Moreover, by convexity of $\mathcal{J}$, the relation

$$\lim_{\ell\to\infty}\mathcal{J}\left(\frac{1}{2}f^{\star}_{\sigma(\ell)}+\frac{1}{2}f^{\star}_{\sigma'(\ell)}\right)\leqslant\lim_{\ell\to\infty}\left(\frac{1}{2}\mathcal{J}\left(f^{\star}_{\sigma(\ell)}\right)+\frac{1}{2}\mathcal{J}\left(f^{\star}_{\sigma'(\ell)}\right)\right)=\mathcal{J}\left(\tilde{f}\right) \tag{A.47}$$

implies, by minimality of $\tilde{f}$, that

$$\lim_{\ell\to\infty}\mathcal{J}\left(\frac{1}{2}f^{\star}_{\sigma(\ell)}+\frac{1}{2}f^{\star}_{\sigma'(\ell)}\right)=\mathcal{J}\left(\tilde{f}\right). \tag{A.48}$$

We conclude with

$$\begin{aligned}
\mathcal{J}\left(\tilde{f}\right) &= \lim_{\ell\to\infty} \mathcal{J}\left(\frac{1}{2}f^{\star}_{\sigma(\ell)} + \frac{1}{2}f^{\star}_{\sigma'(\ell)}\right) && \text{(A.49)}\\
&= \lim_{\ell\to\infty} E\left(\mathbf{y}, \frac{1}{2}\langle f^{\star}_{\sigma(\ell)}, \boldsymbol{\nu}\rangle + \frac{1}{2}\langle f^{\star}_{\sigma'(\ell)}, \boldsymbol{\nu}\rangle\right) + \psi\left(\left\|\mathrm{D}^N\left(\frac{1}{2}f^{\star}_{\sigma(\ell)} + \frac{1}{2}f^{\star}_{\sigma'(\ell)}\right)\right\|_{\mathcal{M}}\right) && \text{(A.50)}\\
&= E\left(\mathbf{y}, \frac{1}{2}\mathbf{a} + \frac{1}{2}\mathbf{a}'\right) + \lim_{\ell\to\infty} \psi\left(\left\|\mathrm{D}^N\left(\frac{1}{2}f^{\star}_{\sigma(\ell)} + \frac{1}{2}f^{\star}_{\sigma'(\ell)}\right)\right\|_{\mathcal{M}}\right) && \text{(A.51)}\\
&< \frac{1}{2}\left(E\left(\mathbf{y},\mathbf{a}\right) + \lim_{\ell\to\infty} \psi\left(\left\|\mathrm{D}^N f^{\star}_{\sigma(\ell)}\right\|_{\mathcal{M}}\right)\right) + \frac{1}{2}\left(E\left(\mathbf{y},\mathbf{a}'\right) + \lim_{\ell\to\infty} \psi\left(\left\|\mathrm{D}^N f^{\star}_{\sigma'(\ell)}\right\|_{\mathcal{M}}\right)\right) && \text{(A.52)}\\
&= \frac{1}{2}\lim_{\ell\to\infty} \mathcal{J}\left(f^{\star}_{\sigma(\ell)}\right) + \frac{1}{2}\lim_{\ell\to\infty} \mathcal{J}\left(f^{\star}_{\sigma'(\ell)}\right) = \mathcal{J}\left(\tilde{f}\right), && \text{(A.53)}
\end{aligned}$$

where we have used the strict convexity of E and the convexity of the regularisation term. This is a contradiction. It follows that $(\langle f^{\star}_{\ell}, \boldsymbol{\nu}\rangle)_{\ell=1}^{\infty}$ is a convergent sequence. Since, from Step 2, $\left(f^{\star}_{\ell_k}\right)_{k=1}^{\infty}$ is weak* convergent to $\tilde{f}$, it must be that $(\langle f^{\star}_{\ell}, \boldsymbol{\nu}\rangle)_{\ell=1}^{\infty}$ is convergent to $\langle \tilde{f}, \boldsymbol{\nu}\rangle$. In addition, an argument similar to Derivation (A.53) also shows that the value $\langle f^{\star}, \boldsymbol{\nu}\rangle$ does not depend on $f^{\star} \in \mathcal{V}$. Indeed, if such a dependence would exist, then one could consider the convex combination of the values and reach a contradiction using the strict convexity of E. We conclude that

$$\begin{aligned}
\mathcal{J}\left(\tilde{f}\right) &= \lim_{\ell\to\infty} \left(E\left(\mathbf{y}, \langle f^{\star}_{\ell}, \boldsymbol{\nu}\rangle\right) + \psi\left(\left\|\mathrm{D}^N f^{\star}_{\ell}\right\|_{\mathcal{M}}\right)\right) = E\left(\mathbf{y}, \langle \tilde{f}, \boldsymbol{\nu}\rangle\right) + \lim_{\ell\to\infty} \psi\left(\left\|\mathrm{D}^N f^{\star}_{\ell}\right\|_{\mathcal{M}}\right) && \text{(A.54)}\\
&\Rightarrow \lim_{\ell\to\infty} \psi\left(\left\|\mathrm{D}^N f^{\star}_{\ell}\right\|_{\mathcal{M}}\right) = \psi\left(\left\|\mathrm{D}^N \tilde{f}\right\|_{\mathcal{M}}\right) && \text{(A.55)}\\
&\Rightarrow \lim_{\ell\to\infty} \psi^{-1}\left(\psi\left(\left\|\mathrm{D}^N f^{\star}_{\ell}\right\|_{\mathcal{M}}\right)\right) = \psi^{-1}\left(\psi\left(\left\|\mathrm{D}^N \tilde{f}\right\|_{\mathcal{M}}\right)\right) && \text{(A.56)}\\
&\Rightarrow \lim_{\ell\to\infty} \left\|\mathrm{D}^N f^{\star}_{\ell}\right\|_{\mathcal{M}} = \left\|\mathrm{D}^N \tilde{f}\right\|_{\mathcal{M}}. && \text{(A.57)}
\end{aligned}$$

Finally, using the fact that $E(\mathbf{y}, \cdot)$ is continuous on the set where its takes finite values, and using the convergence $\lim\limits_{\ell\to\infty} \langle f^{\star}_{\ell}, \boldsymbol{\nu}\rangle = \langle f^{\star}, \boldsymbol{\nu}\rangle$, we find that

$$\lim_{\ell\to\infty} E\left(\mathbf{y}, \langle f^{\star}_{\ell}, \boldsymbol{\nu}\rangle\right) = E\left(\mathbf{y}, \langle f^{\star}, \boldsymbol{\nu}\rangle\right). \tag{A.58}$$

This concludes the proof. ☐

Proof of theorem 4. **Proof of Item i.** Without loss of generality, we assume that $f^{\star}$ is of the form

$$f^{\star}(\cdot) = \sum_{m=1}^{M-N} a_m \frac{(\cdot - x_m)_{+}^{N-1}}{(N-1)!} + \sum_{n=1}^{N} b_n \frac{(\cdot)^{n-1}}{(n-1)!}, \quad x_m \in \left[\nu^{-}, \nu^{+}\right], \quad \mathbf{a} = (a_m)_{m=1}^{M} \in \mathbb{R}^M. \tag{A.59}$$

Then, consider the same approximant $\hat{f}_{\ell}$ of $f^{\star}_{\ell}$ as in Step 1 of the proof of theorem 3. We recall that $\hat{f}_{\ell}$ is weak*-convergent to $f^{\star}$. In addition, if the grid $\mathbb{X}_{\ell}$ is sufficiently fine, then the spline

$\hat{f}_\ell$ has as many knots as $f^\star_\ell$ and $\left\|\mathrm{D}^N \hat{f}_\ell\right\|_{\mathcal{M}} = \left\|\mathrm{D}^N f^\star_\ell\right\|_{\mathcal{M}}$. We find, using a Taylor expansion of $E(\mathbf{y},\cdot)$ around $\langle f^\star, \boldsymbol{\nu}\rangle \in \mathbb{R}^M$, that

$$|\mathcal{J}(f^\star) - \mathcal{J}(f^\star_\ell)| \leqslant \mathcal{J}(f^\star_\ell) - \mathcal{J}(f^\star) \leqslant \mathcal{J}\left(\hat{f}_\ell\right) - \mathcal{J}(f^\star) = E\left(\mathbf{y}, \langle \hat{f}_\ell, \boldsymbol{\nu}\rangle\right) - E(\mathbf{y}, \langle f^\star, \boldsymbol{\nu}\rangle) \tag{A.60}$$

$$= \boldsymbol{\nabla} E(\mathbf{y}, \langle f^\star, \boldsymbol{\nu}\rangle)^\mathsf{T} \left(\langle \hat{f}_\ell - f^\star, \boldsymbol{\nu}\rangle\right) + \mathcal{O}\left(\left\|\langle \hat{f}_\ell - f^\star, \boldsymbol{\nu}\rangle\right\|_2\right) \tag{A.61}$$

$$\leqslant \|\boldsymbol{\nabla} E(\mathbf{y}, \langle f^\star, \boldsymbol{\nu}\rangle)\|_2 \left\|\langle \hat{f}_\ell - f^\star, \boldsymbol{\nu}\rangle\right\|_2 + \mathcal{O}\left(\left\|\langle \hat{f}_\ell - f^\star, \boldsymbol{\nu}\rangle\right\|_2\right) \tag{A.62}$$

$$= \mathcal{O}\left(\left\|\langle \hat{f}_\ell - f^\star, \boldsymbol{\nu}\rangle\right\|_2\right). \tag{A.63}$$

From the Newton binomial theorem and from the fact that $\sup\limits_{1\leqslant m\leqslant M} |x_m - \hat{x}^\ell_m| \leqslant h_\ell$, when $\nu_{m'} \in \mathcal{L}_\infty(\mathbb{R})$ and for $\nu_\pm = \max(|\nu^-|, |\nu^+|)$, we upperbound the error $|\langle \hat{f}_\ell - f^\star, \nu_{m'}\rangle|$ as

$$\leqslant \sum_{m=1}^{M-N} \frac{|a_m| \, \|\nu_{m'}\|_\infty}{(N-1)!} \int_{\min(x_m, \hat{x}^\ell_m)}^{\nu^+} \left|\left((x - x_m)_+^{N-1} - (x - \hat{x}^\ell_m)_+^{N-1}\right)\right| \mathrm{d}x \tag{A.64}$$

$$\leqslant \sum_{m=1}^{M-N} \frac{|a_m| \, \|\nu_{m'}\|_\infty}{(N-1)!} \int_{\min(x_m, \hat{x}^\ell_m)}^{\max\left(x_m, \hat{x}^\ell_m\right)} \left|\left(x - \min\left(x_m, \hat{x}^\ell_m\right)\right)^{N-1}\right| \mathrm{d}x$$

$$+ \sum_{m=1}^{M-N} \frac{|a_m| \, \|\nu_{m'}\|_\infty}{(N-1)!} \int_{\max(x_m, \hat{x}^\ell_m)}^{\nu^+} \left|\left((x - x_m)^{N-1} - (x - \hat{x}^\ell_m)^{N-1}\right)\right| \mathrm{d}x \tag{A.65}$$

$$\leqslant \frac{\|\mathbf{a}\|_1 \, \|\nu_{m'}\|_\infty}{(N-1)!} h_\ell^{N-1} + \sum_{m=1}^{M-N}\sum_{n=1}^{N-1} \frac{|a_m| \, \|\nu_{m'}\|_\infty}{(N-1-n)!n!} \int_{\max(x_m, \hat{x}^\ell_m)}^{\nu^+} |x|^{N-1-n} |x_m^n - \hat{x}_m^{\ell,n}| \mathrm{d}x \tag{A.66}$$

$$\leqslant C_1 h_\ell^{N-1} + \sum_{m=1}^{M-N}\sum_{n=1}^{N-1} \frac{|a_m| \, \|\nu_{m'}\|_\infty \, \nu_\pm^{N-1-n} \left(\nu^+ - \nu^-\right)}{(N-1-n)!n!} |x_m - \hat{x}^\ell_m| \sum_{n'=0}^{n-1} |x_m^{n-n'-1} \hat{x}_m^{\ell,n'}| \tag{A.67}$$

$$\leqslant C_1 h_\ell^{N-1} + \sum_{m=1}^{M-N}\sum_{n=1}^{N-1} \frac{|a_m| \, \|\nu_{m'}\|_\infty \, \nu_\pm^{N-1-n} \left(\nu^+ - \nu^-\right)}{(N-1-n)!n!} h_\ell n \nu_\pm^{n-1} \tag{A.68}$$

$$\leqslant C_1 h_\ell^{N-1} + C_2 h_\ell, \tag{A.69}$$

where $C_1 = \frac{\|\mathbf{a}\|_1 \|\nu_{m'}\|_\infty}{(N-1)!}$ and $C_2 = \sum_{m=1}^{M-N} \sum_{n=1}^{N-1} \frac{|a_m| \|\nu_{m'}\|_\infty \nu_\pm^{N-2}(\nu^+ - \nu^-)}{(N-1-n)!(n-1)!}$. Likewise, if $\nu_{m'}$ is the combination of an $\mathcal{L}_\infty(\mathbb{R})$ function and a sum of Dirac, one can follow a similar derivation to find that there exists a constant C such that

$$|\langle \hat{f}_\ell - f^\star, \nu_{m'}\rangle| \leqslant C h_\ell \quad \Rightarrow \quad |\mathcal{J}(f^\star) - \mathcal{J}(f^\star_\ell)| = \mathcal{O}(h_\ell), \tag{A.70}$$

where we used (A.63).

Proof of Item ii. We take advantage of the Fenchel duality theory. Denote by $E^\star_{\mathbf{y}}(\cdot)$ and by $\psi^\star_{\mathcal{M}_N}$ the Fenchel conjugate (see [4][p.134]) of $E(\mathbf{y},\cdot)$ and $\psi(\left\|\mathrm{D}^N \cdot\right\|_{\mathcal{M}})$. In addition, define the predual operator $\boldsymbol{\nu}^\star$ as

$$\boldsymbol{\nu}^\star : \begin{cases} \mathbb{R}^M \to \mathcal{C}_N(\mathbb{R}) \\ (z_m)_{m=1}^M \mapsto \sum_{m=1}^M z_m \nu_m, \end{cases} \tag{A.71}$$

which is such that

$$\forall \mathbf{z}\in\mathbb{R}^M, \forall f\in\mathcal{M}_N(\mathbb{R}): \quad \langle\langle f,\boldsymbol{\nu}\rangle,\mathbf{z}\rangle = \langle f,\boldsymbol{\nu}^\star(\mathbf{z})\rangle. \tag{A.72}$$

Then, it follows from [4][theorem 4.4.3] that the strong duality holds, with

$$\inf_{f\in\mathcal{M}_N(\mathbb{R})}\left\{E(\mathbf{y},\langle f,\boldsymbol{\nu}\rangle)+\psi\left(\left\|\mathrm{D}^N f\right\|_{\mathcal{M}}\right)\right\} = \sup_{\mathbf{z}\in\mathbb{R}^M}\left\{-\psi^\star_{\mathcal{M}_N}(\boldsymbol{\nu}^\star(\mathbf{z}))-E^\star_{\mathbf{y}}(-\mathbf{z})\right\}. \tag{A.73}$$

On the LHS (RHS, respectively) of (A.73) is the primal (dual, respectively) problem. The requirement for this duality to hold is that [4] [equation 4.3.2] holds, which translates here into

$$\operatorname{dom}(E(\mathbf{y},\cdot))\cap\left\langle\operatorname{dom}\left(\psi\left(\left\|\mathrm{D}^N\cdot\right\|_{\mathcal{M}}\right)\right),\boldsymbol{\nu}\right\rangle\neq\emptyset \Leftrightarrow \operatorname{dom}(E(\mathbf{y},\cdot))\cap\langle\mathcal{M}_N(\mathbb{R}),\boldsymbol{\nu}\rangle\neq\emptyset \tag{A.74}$$

$$\Leftrightarrow \operatorname{dom}(E(\mathbf{y},\cdot))\cap\mathbb{R}^M\neq\emptyset, \tag{A.75}$$

where we used Assumption 4 that $\boldsymbol{\nu}$ is surjective and the fact that $\operatorname{dom}(E(\mathbf{y},\cdot))$ is nonempty because $E(\mathbf{y},\cdot)$ is proper. Then, a primal-dual pair (of solutions) $(f^\star,\mathbf{z}^\star)$ verifies the first-order optimality condition

$$\boldsymbol{\nu}^\star(\mathbf{z}^\star)\in\partial\psi\left(\left\|\mathrm{D}^N\cdot\right\|_M\right)|_{f^\star}, \quad -\mathbf{z}^\star\in\partial E(\mathbf{y},\cdot)|_{\langle f^\star,\boldsymbol{\nu}\rangle}. \tag{A.76}$$

Since we assumed that $E(\mathbf{y},\cdot)$ is continuously differentiable, (A.76) is equivalent to

$$\boldsymbol{\nu}^\star(\mathbf{z}^\star)\in\partial\psi\left(\left\|\mathrm{D}^N\cdot\right\|_M\right)|_{f^\star}, \quad \mathbf{z}^\star=-\boldsymbol{\nabla}E(\mathbf{y},\langle f^\star,\boldsymbol{\nu}\rangle). \tag{A.77}$$

Furthermore, it follows from the LHS inclusion of (A.77) that (by definition of the subdifferential)

$$\forall f\in\mathcal{M}_N(\mathbb{R}): \quad \langle f-f^\star,\boldsymbol{\nu}^\star(\mathbf{z}^\star)\rangle+\psi\left(\left\|\mathrm{D}^N f^\star\right\|\right)\leqslant\psi\left(\left\|\mathrm{D}^N f\right\|\right) \tag{A.78}$$

and, by replacing f by $f^\star_\ell$, $\mathbf{z}^\star$ by $-\boldsymbol{\nabla}E(\mathbf{y},\langle f^\star,\boldsymbol{\nu}\rangle)$, we find that

$$\langle\langle f^\star_\ell-f^\star,\boldsymbol{\nu}\rangle,-\boldsymbol{\nabla}E(\mathbf{y},\langle f^\star,\boldsymbol{\nu}\rangle)\rangle+\psi\left(\left\|\mathrm{D}^N f^\star\right\|\right)\leqslant\psi\left(\left\|\mathrm{D}^N f^\star_\ell\right\|\right) \tag{A.79}$$

$$\Rightarrow 0\leqslant\psi\left(\left\|\mathrm{D}^N f^\star_\ell\right\|\right)-\psi\left(\left\|\mathrm{D}^N f^\star\right\|\right)+\langle\langle f^\star_\ell-f^\star,\boldsymbol{\nu}\rangle,\boldsymbol{\nabla}E(\mathbf{y},\langle f^\star,\boldsymbol{\nu}\rangle)\rangle. \tag{A.80}$$

The derivation from (A.73) to (A.80) can be replicated on the Banach space $\mathcal{M}_N(\mathbb{R},\mathbb{X}_\ell)$ and the local solution $f^\star_\ell$ with the approximation $\hat{f}_\ell$ in order to establish that

$$0\leqslant\psi\left(\left\|\mathrm{D}^N\hat{f}_\ell\right\|\right)-\psi\left(\left\|\mathrm{D}^N f^\star_\ell\right\|\right)+\langle\langle\hat{f}_\ell-f^\star_\ell,\boldsymbol{\nu}\rangle,\boldsymbol{\nabla}E(\mathbf{y},\langle f^\star_\ell,\boldsymbol{\nu}\rangle)\rangle. \tag{A.81}$$

It follows, by developing the terms in (A.81), that

$$\begin{aligned}0\leqslant{}&\psi\left(\left\|\mathrm{D}^N\hat{f}_\ell\right\|\right)-\psi\left(\left\|\mathrm{D}^N f^\star_\ell\right\|\right)+\langle\langle f^\star-f^\star_\ell,\boldsymbol{\nu}\rangle,\boldsymbol{\nabla}E(\mathbf{y},\langle f^\star_\ell,\boldsymbol{\nu}\rangle)\rangle\\&+\langle\langle\hat{f}_\ell-f^\star,\boldsymbol{\nu}\rangle,\boldsymbol{\nabla}E(\mathbf{y},\langle f^\star_\ell,\boldsymbol{\nu}\rangle)\rangle.\end{aligned} \tag{A.82}$$

Recall that $\left\|\mathrm{D}^N f^\star\right\|_{\mathcal{M}} = \left\|\mathrm{D}^N \hat{f}_\ell\right\|$. It follows, by adding (A.82) to (A.80), that

$$0 \leqslant \langle\langle f_\ell^\star - f^\star, \boldsymbol{\nu}\rangle, \boldsymbol{\nabla} E(\mathbf{y}, \langle f^\star, \boldsymbol{\nu}\rangle - \boldsymbol{\nabla} E(\mathbf{y}, \langle f_\ell^\star, \boldsymbol{\nu}\rangle))\rangle + \langle\langle \hat{f}_\ell - f^\star, \boldsymbol{\nu}\rangle, \boldsymbol{\nabla} E(\mathbf{y}, \langle f_\ell^\star, \boldsymbol{\nu}\rangle)\rangle \tag{A.83}$$

$$\Rightarrow \langle\langle f_\ell^\star - f^\star, \boldsymbol{\nu}\rangle, \boldsymbol{\nabla} E(\mathbf{y}, \langle f_\ell^\star, \boldsymbol{\nu}\rangle) - \boldsymbol{\nabla} E(\mathbf{y}, \langle f^\star, \boldsymbol{\nu}\rangle)\rangle \leqslant \langle\langle \hat{f}_\ell - f^\star, \boldsymbol{\nu}\rangle, \boldsymbol{\nabla} E(\mathbf{y}, \langle f_\ell^\star, \boldsymbol{\nu}\rangle)\rangle \tag{A.84}$$

$$\Rightarrow \gamma \left\|\langle f_\ell^\star - f^\star, \boldsymbol{\nu}\rangle\right\|_2^2 \leqslant \langle\langle \hat{f}_\ell - f^\star, \boldsymbol{\nu}\rangle, \boldsymbol{\nabla} E(\mathbf{y}, \langle f_\ell^\star, \boldsymbol{\nu}\rangle)\rangle \leqslant \left\|\langle \hat{f}_\ell - f^\star, \boldsymbol{\nu}\rangle\right\|_2 \left\|\boldsymbol{\nabla} E(\mathbf{y}, \langle f_\ell^\star, \boldsymbol{\nu}\rangle)\right\|_2 \tag{A.85}$$

$$\Rightarrow \gamma \left\|\langle f_\ell^\star - f^\star, \boldsymbol{\nu}\rangle\right\|_2^2 \leqslant C C' h_\ell, \tag{A.86}$$

where we used in (A.85) that $E(\mathbf{y}, \cdot)$ is strongly convex on a closed ball, centred on $\langle f^\star, \boldsymbol{\nu}\rangle$, of an appropriate radius. We note γ the strong convexity constant. In (A.86), we defined $\sup_{\ell \in \mathbb{N}} \left\|\boldsymbol{\nabla} E(\mathbf{y}, \langle f_\ell^\star, \boldsymbol{\nu}\rangle)\right\|_2 = C' < \infty$ and C such that $\left\|\langle \hat{f}_\ell - f^\star, \boldsymbol{\nu}\rangle\right\|_2 \leqslant C h_\ell$, whose existence has been established in the proof of Item i. This concludes the proof of Item ii.

Proof of Items iii and iv. Concerning Item iii, we use a Taylor expansion of $E(\mathbf{y}, \cdot)$ around $\langle f^\star, \boldsymbol{\nu}\rangle$ and around $\langle f_\ell^\star, \boldsymbol{\nu}\rangle$ to find that

$$E(\mathbf{y}, \langle f_\ell^\star, \boldsymbol{\nu}\rangle) - E(\mathbf{y}, \langle f^\star, \boldsymbol{\nu}\rangle) \leqslant \boldsymbol{\nabla} E(\mathbf{y}, \langle f^\star, \boldsymbol{\nu}\rangle)^\mathsf{T} (\langle f_\ell^\star - f^\star, \boldsymbol{\nu}\rangle) + \mathcal{O}(\left\|\langle f_\ell^\star - f^\star, \boldsymbol{\nu}\rangle\right\|_2) \tag{A.87}$$

$$E(\mathbf{y}, \langle f^\star, \boldsymbol{\nu}\rangle) - E(\mathbf{y}, \langle f_\ell^\star, \boldsymbol{\nu}\rangle) \leqslant \boldsymbol{\nabla} E(\mathbf{y}, \langle f_\ell^\star, \boldsymbol{\nu}\rangle)^\mathsf{T} (\langle f^\star - f_\ell^\star, \boldsymbol{\nu}\rangle) + \mathcal{O}(\left\|\langle f_\ell^\star - f^\star, \boldsymbol{\nu}\rangle\right\|_2) \tag{A.88}$$

$$\Rightarrow \left|E(\mathbf{y}, \langle f_\ell^\star, \boldsymbol{\nu}\rangle) - E(\mathbf{y}, \langle f^\star, \boldsymbol{\nu}\rangle)\right| \leqslant \mathcal{O}\left(h_\ell^{\frac{1}{2}}\right), \tag{A.89}$$

where we used Item ii to derive (A.89). Item iv follows from the combination of Item i and Item iii. This concludes the proof. □

A.4. Proofs of section 3.3

In this appendix, we aim at showing theorem 5. We first provide two preliminary results. Lemma 1 is given without proof as it is a classical result.

Lemma 1. *Let $f_1, \ldots, f_N$ be a free (independent) family of functions on an interval $[a,b]$. Then, there exist points $\{x_n\}_{n=1}^N \subset [a,b]$ such that the matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ defined as $\mathbf{A}_{\tilde{n},n} = f_n(x_{\tilde{n}})$ is invertible.*

Proposition 3. *Let $f_1, \ldots, f_N$ be a free (independent) family of functions on an interval $[a,b]$. Suppose that there exists a sequence of functions $g_\ell = \sum_{n=1}^N b_{n,\ell} f_n$ that are pointwise convergent on $[a,b]$ to some function g. Then, the sequence $\{b_{n,\ell}\}_{\ell=1}^\infty$ is such that $\lim_{\ell \to \infty} b_{n,\ell} = b_n$ for some limits $\mathrm{b_n}$ and*

$$g = \sum_{n=1}^N b_n f_n. \tag{A.90}$$

Proof. Using proposition 1, we find points $x_1, \ldots, x_N \in [a,b]$ such that the matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ defined as $[\mathbf{A}]_{\tilde{n},n} = f_n(x_{\tilde{n}})$ is invertible. Define $\mathbf{b}_\ell \in \mathbb{R}^N$ as $[\mathbf{b}_\ell]_n = b_{n,\ell}$ and observe that

$[\mathbf{A}\mathbf{b}_\ell]_{\tilde{n}} = \sum_{n=1}^{N} b_{n,\ell} f_n(x_{\tilde{n}}) = g_\ell(x_{\tilde{n}})$. The pointwise convergence of g_ℓ to g, applied on those points $x_{\tilde{n}}$ translates into

$$\lim_{\ell\to\infty} \mathbf{A}\mathbf{b}_\ell = (g(x_1),\ldots,g(x_N)) \Rightarrow \lim_{\ell\to\infty} \mathbf{b}_\ell = \mathbf{A}^{-1}(g(x_1),\ldots,g(x_N)). \quad \text{(A.91)}$$

Therefore, the sequence $b_{n,\ell}$ is convergent in ℓ, so we may define $b_n = \lim_{\ell\to\infty} b_{n,\ell}$ and $\tilde{g} = \sum_{n=1}^{N} b_n f_n$. We conclude that, for all $x \in [a,b]$,

$$g(x) = \lim_{\ell\to\infty} g_\ell(x) = \lim_{\ell\to\infty} \sum_{n=1}^{N} b_{n,\ell} f_n(x) = \sum_{n=1}^{N} f_n(x) \lim_{\ell\to\infty} b_{n,\ell} = \sum_{n=1}^{N} f_n(x) b_n = \tilde{g}(x). \quad \text{(A.92)}$$

☐

Then, we state a version of the Ascoli–Arzela theorem as theorem 8.

Definition 7. (1) A sequence $(f_\ell)_{\ell=1}^{\infty} \subset \mathcal{C}([a,b])$ is said to be *equicontinuous* if

$$\forall x \in [a,b], \forall \epsilon > 0, \exists \delta > 0: \quad (\forall x' \in [a,b]: |x - x'| < \delta, \forall \ell \in \mathbb{N}) \Rightarrow |f_\ell(x') - f_\ell(x)| < \epsilon. \quad \text{(A.93)}$$

(2) A sequence $(f_\ell)_{\ell=1}^{\infty} \subset \mathcal{C}([a,b])$ is said to be *uniformly Lipschitz* if

$$\exists L > 0, \forall x, y \in \mathbb{R}, \forall \ell \in \mathbb{N}: \quad |f_\ell(x) - f_\ell(y)| \leqslant L|x - y|. \quad \text{(A.94)}$$

Observe that uniform Lipschitzness implies equicontinuity.

Theorem 8 ([31] (theorem 7, Chapter 7)). *Let $F = \{f_\ell\}_{\ell=1}^{\infty}$ be a sequence of continuous functions on the interval $[a,b]$. If the family F is equicontinuous, then pointwise convergence is equivalent to uniform convergence.*

Proof of theorem 5. We have proved in Step 2 of the proof of theorem 3 that $(f_\ell^\star)_{\ell=1}^{\infty}$ is a bounded sequence in $\mathcal{M}_N(\mathbb{R})$ and, therefore, has a subsequence $(f_{\ell_k}^\star)_{k=1}^{\infty}$ that is weak* convergent to some solution $f^\star \in \mathcal{V}$ (previously written $\tilde{f}$).

If the solution $f^\star$ is unique, then $\mathcal{V} = \{f^\star\}$ and the whole sequence $(f_\ell^\star)_{\ell=1}^{\infty}$ must be weak* convergent to $f^\star$.

If the solution is not unique, we continue to work with the subsequence $(f_{\ell_k}^\star)_{k=1}^{\infty}$. Nevertheless, to simplify the notation, we write $(f_\ell^\star)_{\ell=1}^{\infty}$ instead of $(f_{\ell_k}^\star)_{k=1}^{\infty}$.

Proof of Item 1: this is proved in the last part of Step 2 of the proof of theorem 3.

Proof of Item 2: by theorem 2, we know that $f_\ell^\star, f^\star$ admit the decomposition

$$f_\ell^\star(\cdot) = \int_{\nu^-}^{\nu^+} \frac{(\cdot - y)_+^{N-1}}{(N-1)!} \mathrm{d}m_\ell(y) + \sum_{n=1}^{N} b_{\ell,n} \frac{(\cdot)^{n-1}}{(n-1)!}, \quad \text{(A.95)}$$

$$f^\star(\cdot) = \int_{\nu^-}^{\nu^+} \frac{(\cdot - y)_+^{N-1}}{(N-1)!} \mathrm{d}m(y) + \sum_{n=1}^{N} b_n \frac{(\cdot)^{n-1}}{(n-1)!}, \quad \text{(A.96)}$$

with m_ℓ and m being composed of at most $(M-N)$ Diracs. Since $(z)_+^{N-1} = 0$ for $z \leqslant 0$, it holds that

$$\forall x \in]-\infty, \nu^-], \quad f_\ell(x) = \sum_{n=1}^{N} b_{\ell,n} \frac{x^{n-1}}{(n-1)!}, \quad f^\star(x) = \sum_{n=1}^{N} b_n \frac{x^{n-1}}{(n-1)!}. \quad \text{(A.97)}$$

Corollary 2 informs us that $\delta_x \in \mathcal{C}_N(\mathbb{R})$. Hence,

$$\lim_{\ell\to\infty} \langle f_\ell, \delta_x\rangle = \langle f, \delta_x\rangle \Leftrightarrow \lim_{\ell\to\infty} f_\ell(x) = f(x), \tag{A.98}$$

so that f_ℓ is pointwise convergent to f on $\mathbb{R}$. We conclude, using proposition 3, that $\forall n \in [1, N]$, $\lim_{\ell\to\infty} b_{\ell,n} = b_n$.

Proof of Item 3: for $d \leqslant (N-2)$ we define $F_d = \left(f_\ell^{\star,(d)}\right)_{\ell=1}^{\infty} \subset \mathcal{C}([\nu^-, \nu^+])$. By Corollary 2, and the argumentation of Step 1, we know that the sequence F_d is pointwise convergent, on $[\nu^-, \nu^+]$, to the limit $f^{\star,(d)}$. We want to apply theorem 8 to the sequence F_d in order to conclude with uniform convergence on $[\nu^-, \nu^+]$. To do so, we have to verify that the sequence F_d is equicontinuous by proving the stronger result that, for $0 \leqslant d \leqslant (N-2)$, F_d is uniformly Lipschitz. Using the assumption that $\boldsymbol{\nu}$ is compactly supported (definition 2) with $\mathrm{Csupp}(\boldsymbol{\nu}) = [\nu^-, \nu^+]$, we apply theorem 2 to find that

$$f_\ell^{\star,(d)}(x) = \sum_{m=1}^{M-N} a_{\ell,m} \frac{(x - x_{\ell,m})_+^{N-1-d}}{(N-1-d)!} + \sum_{n=d+1}^{N} b_{\ell,n-d} \frac{x^{n-1-d}}{(N-1-d)!}, \tag{A.99}$$

where the knots $x_{\ell,m}$ verify that $x_{\ell,m} \in [\nu^-, \nu^+]$. Observe that, using proposition 1,

$$\|f\|_{\mathcal{M}_N} = \|m\|_{\mathcal{M}} + \left\|(b_n)_{n=1}^N\right\|_2, \quad \text{for} \quad f = \frac{(\cdot)_+^{N-1}}{(N-1)!} * m + \sum_{n=1}^{N} \frac{b_n (\cdot)^{n-1}}{(n-1)!} \tag{A.100}$$

is a norm on $\mathcal{M}_N([\nu^-, \nu^+])$. The following inequalities hold $\forall x, y \in [\nu^-, \nu^+]$

$$\begin{aligned}
\left|f_\ell^{\star,(d)}(x) - f_\ell^{\star,(d)}(y)\right| &\leqslant \sum_{m=1}^{M-N} |a_{\ell,m}| \left| \frac{(x - x_{\ell,m})_+^{N-1-d}}{(N-1-d)!} - \frac{(y - x_{\ell,m})_+^{N-1-d}}{(N-1-d)!} \right| \\
&\quad + \sum_{n=d+1}^{N} |b_{\ell,n-d}| \left| \frac{x^{n-1-d} - y^{n-1-d}}{(n-1-d)!} \right| && \text{(A.101)} \\
&\leqslant |x-y| \left(L^d \sum_{m=1}^{M-N} |a_{\ell,m}| + \max_{d+1\leqslant n\leqslant N} L_n^d \sum_{n=1}^{N} |b_{\ell,n}| \right) && \text{(A.102)} \\
&\leqslant |x-y| \left(L^d \left\|m_\ell^\star\right\|_{\mathcal{M}} + \max_{d+1\leqslant n\leqslant N} L_n^d \left\|(b_{\ell,n})_{n=1}^N\right\|_1 \right) && \text{(A.103)} \\
&\leqslant |x-y| \left(L^d \left\|m_\ell^\star\right\|_{\mathcal{M}} + \sqrt{N} \max_{d+1\leqslant n\leqslant N} L_n^d \left\|(b_{\ell,n})_{n=1}^N\right\|_2 \right) && \text{(A.104)} \\
&\leqslant |x-y| \left(\max\left(L^d, \sqrt{N} \max_{d+1\leqslant n\leqslant N} L_n^d \right) \left\|f_\ell^\star\right\|_{\mathcal{M}_N} \right), && \text{(A.105)}
\end{aligned}$$

where $m_\ell^\star = \sum_{m=1}^{M-N} a_{\ell,m}\delta_{x_{\ell,m}}$, L^d is the Lipschitz constant of $\frac{(\cdot)_+^{N-1-d}}{(N-1-d)!}$ over the interval $[\nu^- - \nu^+, \nu^+ - \nu^-]$, and L_n^d is the Lipschitz constant of $\frac{(\cdot)^{n-1-d}}{(n-1-d)!}$ over the interval $[\nu^-, \nu^+]$. It follows that the sequence F_d is uniformly Lipschitz with Lipschitz constant $\max\left(L^d, \sqrt{N} \max\limits_{d+1\leqslant n\leqslant N} L_n^d\right) \sup\limits_{\ell\in\mathbb{N}} \|f_\ell^\star\|_{\mathcal{M}_N}$, hence, equicontinuous, and finally, uniformly convergent.

Proof of Item 4: observe that Item 4 is very similar to Item 1, with the difference that the coefficients $b_{n,\ell}$ need to be adapted to the causal innovations in $[\nu^-, \nu^+]$. For $x \in [\nu^+, \infty[$, we find that

$$f_\ell^\star(x) = \sum_{m=1}^{M-N} a_{\ell,m} \frac{(x - x_{\ell,m})^{N-1}}{(N-1)!} + \sum_{n=1}^{N} b_{\ell,n} \frac{x^{n-1}}{(n-1)!} \tag{A.106}$$

$$= \sum_{m=1}^{M-N} \sum_{n=0}^{N-1} \frac{a_{\ell,m}}{(N-1)!} \binom{N-1}{n} (-x_{\ell,m})^{N-1-n} x^n + \sum_{n=1}^{N} b_{\ell,n} \frac{x^{n-1}}{(n-1)!} \tag{A.107}$$

$$= \sum_{n=1}^{N} \left(\sum_{m=1}^{M-N} \frac{a_{\ell,m}}{(N-n)!} (-x_{\ell,m})^{N-n} \right) \frac{x^{n-1}}{(n-1)!} + \sum_{n=1}^{N} b_{\ell,n} x^{n-1} \tag{A.108}$$

$$= \sum_{n=1}^{N} \tilde{b}_{\ell,n} x^{n-1}, \quad \text{for} \quad \tilde{b}_{\ell,n} = \left(\sum_{m=1}^{M-N} \frac{a_{\ell,m}}{(N-n)!} (-x_{\ell,m})^{N-n} \right) + b_{\ell,n}. \tag{A.109}$$

We proceed to the same decomposition with $f^\star$ and conclude the proof with the same technique as in the proof of Item 2.

Proof of Item 5: this is a consequence of Step 2 of the proof of theorem 3.

Proof of Item 6: let $f^\star$ be the unique solution, so that $\mathcal{V} = \{f^\star\}$. Moreover, let $\{f_\ell^\star\}_{\ell=1}^\infty$ be a sequence of solutions with $f_\ell^\star \in \mathcal{V}_\ell$. We know from the proof of theorem 3 that any weak* convergent subsequence of $\{f_\ell^\star\}_{\ell=1}^\infty$ must be convergent to a solution in $\mathcal{V}$. By unicity of the solution, the limit must be $f^\star$. To conclude, we use the fact that in the weak* topology, if all the convergent subsequences converge to the same limit, then the sequence itself must be convergent to this limit. ☐

A.5. Proofs of section 3.4

Proof of Corollary 1. By assumption, there exists a sequence of solutions $(\mathbf{a}_\ell^\star, \mathbf{b}_\ell^\star)$ of (31) such that $\lim\limits_{\ell\to\infty} \left\|(\mathbf{a}_\ell^\star, \mathbf{b}_\ell^\star) - (\tilde{\mathbf{a}}_\ell, \tilde{\mathbf{b}}_\ell)\right\|_1 = 0$. Define the sequence of solutions $f_\ell^\star \in \mathcal{V}_\ell$ by $f_\ell^\star(\cdot) = \sum_{k=1}^{K_\ell} a_{\ell,k}^\star \frac{(\cdot - x_k)_+^{N-1}}{(N-1)!} + \sum_{n=1}^{N} b_{\ell,n}^\star \frac{(\cdot)^{n-1}}{(n-1)!}$. First, the following inequalities hold

$$\left\|\mathrm{D}^d\left(f_\ell^\star - \tilde{f}_\ell\right)\right\|_{\infty,[\nu^-,\nu^+]}$$

$$= \left\|\mathrm{D}^d\left(\sum_{k=1}^{\infty} (a_{\ell,k}^\star - \tilde{a}_{\ell,k}) \frac{(\cdot - x_k)_+^{N-1}}{(N-1)!} + \sum_{n=1}^{N} \left(b_{\ell,n}^\star - \tilde{b}_{\ell,n}\right) \frac{(\cdot)^{n-1}}{(n-1)!}\right)\right\|_{\infty,[\nu^-,\nu^+]} \tag{A.110}$$

$$
\begin{aligned}
&\leqslant \left\| \sum_{k=1}^{\infty} \left(a^{\star}_{\ell,k} - \tilde{a}_{\ell,k}\right) \frac{(\cdot - x_k)_+^{N-1-d}}{(N-1-d)!} \right\|_{\infty,[\nu^-,\nu^+]} \\
&\quad + \left\| \sum_{n=d+1}^{N} \left(b^{\star}_{\ell,n-d} - \tilde{b}_{\ell,n-d}\right) \frac{(\cdot)^{n-1-d}}{(N-1-d)!} \right\|_{\infty,[\nu^-,\nu^+]} && \text{(A.111)} \\
&\leqslant \left\| \mathbf{a}^{\star}_{\ell} - \tilde{\mathbf{a}}_{\ell} \right\|_1 \max_{x,y\in[\nu^-,\nu^+]} \left| \frac{(x-y)_+^{N-1-d}}{(N-1-d)!} \right| + \left\| \mathbf{b}^{\star}_{\ell} - \tilde{\mathbf{b}}_{\ell} \right\|_1 \max_{x\in[\nu^-,\nu^+],0\leqslant n\leqslant N-1-d} \left| \frac{(x)^n}{n!} \right| && \text{(A.112)} \\
&\leqslant \left\| \left(\mathbf{a}^{\star}_{\ell} - \tilde{\mathbf{a}}_{\ell}, \mathbf{b}^{\star}_{\ell} - \tilde{\mathbf{b}}_{\ell} \right) \right\|_1 C, && \text{(A.113)}
\end{aligned}
$$

with $C = \max\left(\max_{x,y\in[\nu^-,\nu^+]} \left| \frac{(x-y)_+^{N-1-d}}{(N-1-d)!} \right|, \max_{x\in[\nu^-,\nu^+],1\leqslant n\leqslant N-1-d} \left| \frac{x^n}{n!} \right| \right)$. We know from theorem 5 that $(f_{\ell}^{\star,(d)})_{\ell=1}^{\infty}$ is such that

$$
\lim_{\ell\to\infty} \left\| \mathrm{D}^d \left(f^{\star}_{\ell} - f^{\star}\right) \right\|_{\infty,[\nu^-,\nu^+]} = 0. \tag{A.114}
$$

It follows that

$$
\begin{aligned}
\lim_{\ell\to\infty} \left\| \mathrm{D}^d \left(\tilde{f}_{\ell} - f^{\star}\right) \right\|_{\infty,[\nu^-,\nu^+]} &\leqslant \lim_{\ell\to\infty} \left\| \mathrm{D}^d \left(\tilde{f}_{\ell} - f^{\star}_{\ell}\right) \right\|_{\infty,[\nu^-,\nu^+]} + \lim_{\ell\to\infty} \left\| \mathrm{D}^d \left(f^{\star}_{\ell} - f^{\star}\right) \right\|_{\infty,[\nu^-,\nu^+]} && \text{(A.115)} \\
&= \lim_{\ell\to\infty} \left\| \mathrm{D}^d \left(\tilde{f}_{\ell} - f^{\star}_{\ell}\right) \right\|_{\infty,[\nu^-,\nu^+]} + 0 && \text{(A.116)} \\
&\leqslant C \lim_{\ell\to\infty} \left\| \left(\mathbf{a}^{\star}_{\ell}, \mathbf{b}^{\star}_{\ell}\right) - \left(\tilde{\mathbf{a}}_{\ell}, \tilde{\mathbf{b}}_{\ell}\right) \right\|_1 = 0. && \text{(A.117)}
\end{aligned}
$$

This shows that (34) holds with the substitution of $f^{\star}_{\ell_k}$ for $\tilde{f}_{\ell}$. Concerning (33), it follows from $\mathbb{X}_{\ell} \subset [\nu^-,\nu^+]$ that $\mathrm{supp}(\mathrm{D}^N \tilde{f}_{\ell}) \subset [\nu^-,\nu^+]$. In addition, $\mathrm{supp}(\mathrm{D}^N f^{\star}) \subset [\nu^-,\nu^+]$. Therefore, $\tilde{f}_{\ell} = \sum_{n=1}^{N} [\tilde{\mathbf{b}}_{\ell}]_n (\cdot)^{n-1}$ over $]-\infty,\nu^-]$, where $\tilde{\mathbf{b}}_{\ell}$ contains the coefficients associated to the nullspace part of $\mathcal{M}_N(\mathbb{R})$, used in the statement of Corollary 1. The same way, there exists $\mathbf{b} \in \mathbb{R}^N$ such that $f^{\star} = \sum_{n=1}^{N} [\mathbf{b}]_n (\cdot)^{n-1}$ over $]-\infty,\nu^-]$. It follows that

$$
\begin{aligned}
\lim_{\ell\to\infty} \left\| \mathbf{b} - \tilde{\mathbf{b}}_{\ell} \right\|_1 &\leqslant \lim_{\ell\to\infty} \| \mathbf{b} - \mathbf{b}^{\star}_{\ell} \|_1 + \lim_{\ell\to\infty} \left\| \mathbf{b}^{\star}_{\ell} - \tilde{\mathbf{b}}_{\ell} \right\|_1 && \text{(A.118)} \\
&\leqslant \lim_{\ell\to\infty} \| \mathbf{b} - \mathbf{b}^{\star}_{\ell} \|_1 + \lim_{\ell\to\infty} C \left\| \left(\mathbf{a}^{\star}_{\ell}, \mathbf{b}^{\star}_{\ell}\right) - \left(\tilde{\mathbf{a}}_{\ell}, \tilde{\mathbf{b}}_{\ell}\right) \right\|_1 = 0, && \text{(A.119)}
\end{aligned}
$$

where $\lim_{\ell\to\infty} \| \mathbf{b} - \mathbf{b}^{\star}_{\ell} \|_1 = 0$ because of (33). This shows that (33) holds with the substitution of $f^{\star}_{\ell_k}$ for $\tilde{f}_{\ell}$. The proof of the convergence of the coefficients on $[\nu^+,\infty[$ is similar, with the addition of the same trick as in Step 4 of the proof of theorem 5. ☐

ORCID iD

Vincent Guillemet https://orcid.org/0009-0008-0154-7389

References

[1] Bartels S 2012 Total variation minimization with finite elements: convergence and iterative solution *SIAM J. Numer. Anal.* **50** 1162–80

[2] Bartels S, Nochetto R and Salgado A 2015 A total variation diminishing interpolation operator and applications *Math. Comput.* **84** 2569–87

[3] Bertero M, Boccacci P and De Mol C 2021 *Introduction to Inverse Problems in Imaging* (CRC Press)

[4] Borwein J M and Zhu Q J 2005 Variational techniques in convex analysis *Techniques of Variational Analysis* pp 111–63

[5] Boyd N, Schiebinger G and Recht B 2017 The alternating descent conditional gradient method for sparse inverse problems *SIAM J. Optim.* **27** 616–39

[6] Boyer C, Chambolle A, Castro Y D, Duval V, De Gournay F and Weiss P 2019 On representer theorems and convex regularization *SIAM J. Optim.* **29** 1260–81

[7] Bredies K and Carioni M 2020 Sparsity of solutions for variational inverse problems with finite-dimensional data *Calculus Variations PDE* **59** 14

[8] Bredies K, Carioni M, Fanzon S and Walter D 2024 Asymptotic linear convergence of fully-corrective generalized conditional gradient methods *Math. Program.* **205** 135–202

[9] Campos J, Aziznejad S and Unser M 2022 Learning of continuous and piecewise-linear functions with Hessian total-variation regularization *IEEE Open J. Signal Process.* **3** 36–48

[10] Casas E, Clason C and Kunisch K 2012 Approximation of elliptic control problems in measure spaces with sparse solutions *SIAM J. Control Optim.* **50** 1735–52

[11] Chambolle A 2004 An algorithm for total variation minimization and applications *J. Math. Imaging Vis.* **20** 89–97

[12] Chambolle A 2005 Total variation minimization and a class of binary MRF models *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition* (Springer) pp 136–52

[13] Chizat L 2022 Sparse optimization on measures with over-parameterized gradient descent *Math. Program.* **194** 487–532

[14] Ciarlet P G and Raviart P-A 1972 General lagrange and hermite interpolation in R^n with applications to finite element methods *Arch. Ration. Mech. Anal.* **46** 177–99

[15] Clason C and Kunisch K 2011 A duality-based approach to elliptic control problems in non-reflexive banach spaces *ESAIM: Control Optimisation Calculus Variations 17* **1** 243–66

[16] Condat L 2013 A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms *J. Optim. Theory Appl.* **158** 460–79

[17] Debarre T, Denoyelle Q and Fageot J 2022 On the uniqueness of solutions for the basis pursuit in the continuum *Inverse Problems* **38** 125005

[18] Debarre T, Denoyelle Q and Fageot J 2023 Tv-based spline reconstruction with fourier measurements: uniqueness and convergence of grid-based methods *J. Comput. Appl. Math.* **422** 114937

[19] Debarre T, Fageot J, Gupta H and Unser M 2019 B-spline-based exact discretization of continuous-domain inverse problems with generalized tv regularization *IEEE Trans. Inf. Theory* **65** 4457–70

[20] Denoyelle Q, Duval V, Peyré G and Soubies E 2019 The sliding Frank-Wolfe algorithm and its application to super-resolution microscopy *Inverse Problems* **36** 014001

[21] Donoho D L 2006 Compressed sensing *IEEE Trans. Inf. Theory* **52** 1289–306

[22] Duval V and Peyré G 2014 Exact support recovery for sparse spikes deconvolution *Found. Comput. Math.* **15** 1315–55

[23] Duval V and Peyré G 2017 Sparse regularization on thin grids I: the lasso *Inverse Problems* **33** 055008

[24] Duval V and Peyré G 2017 Sparse spikes super-resolution on thin grids II: the continuous basis pursuit *Inverse Problems* **33** 095008

[25] Fageot J and Simeoni M 2020 Tv-based reconstruction of periodic functions *Inverse Problems* **36** 115015

[26] Flinth A, de Gournay F and Weiss P 2023 Grid is good: adaptive refinement algorithms for off-the-grid total variation minimization (arXiv: 2301.07555)

[27] Flinth A and Weiss P 2019 Exact solutions of infinite dimensional total-variation regularized problems *Inf. Inference J. IMA* **8** 407–43

[28] Foucart S, Rauhut H, Foucart S and Rauhut H 2013 *An Invitation to Compressive Sensing* (Springer)

[29] Gupta H, Fageot J and Unser M 2018 Continuous-domain solutions of linear inverse problems with Tikhonov versus generalized tv regularization *IEEE Trans. Signal Process.* **66** 4670–84

[30] Jaggi M 2013 Revisiting Frank-Wolfe: projection-free sparse convex optimization *Int. Conference on Machine Learning* (PMLR) pp 427–35

[31] Kelley J L 2017 *General Topology* (Courier Dover Publications)

[32] Mallat S 1999 *A Wavelet Tour of Signal Processing* (Elsevier)

[33] McCann M T and Unser M *et al* 2019 Biomedical image reconstruction: from the foundations to deep neural networks *Found. Trends Signal Process.* **13** 283–359

[34] Parhi R and Nowak R D 2021 Banach space representer theorems for neural networks and ridge splines *J. Mach. Learn. Res.* **22** 1–40

[35] Pieper K and Vexler B 2013 A priori error analysis for discretization of sparse elliptic optimal control problems in measure space *SIAM J. Control Optim.* **51** 2788–808

[36] Pieper K and Walter D 2021 Linear convergence of accelerated conditional gradient algorithms in spaces of measures *ESAIM: Control Optim. Calculus Variations* **27** 38

[37] Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica* D **60** 259–68

[38] Rudin W 1987 *Real and Complex Analysis* 3rd edn (McGraw-Hill)

[39] Simeoni M M J A Functional inverse problems on spheres: theory, algorithms and applications *Technical report* EPFL 2020

[40] Tikhonov A N 1963 Solution of incorrectly formulated problems and the regularization method *Sov. Dok.* **4** 1035–8

[41] Treves F 2016 *Topological Vector Spaces, Distributions and Kernels Pure and Applied Mathematics* vol 25 (Elsevier)

[42] Unser M 2021 A unifying representer theorem for inverse problems and machine learning *Found. Comput. Math.* **21** 941–60

[43] Unser M, Fageot J and Ward J P 2017 Splines are universal solutions of linear inverse problems with generalized tv regularization *SIAM Rev.* **59** 769–93

[44] Wang J and Lucier B J 2011 Error bounds for finite-difference methods for Rudin–Osher–Fatemi image smoothing *SIAM J. Numer. Anal.* **49** 845–68