MULTI-DIMENSIONAL SPARSE SUPER-RESOLUTION

CLARICE POON,* GABRIEL PEYRÉ†

Abstract. This paper studies sparse super-resolution in arbitrary dimensions. More precisely, it develops a theoretical analysis of support recovery for the so-called BLASSO method, which is an off-the-grid generalisation of ℓ^1 regularization (also known as the LASSO). While super-resolution is of paramount importance in overcoming the limitations of many imaging devices, its theoretical analysis is still lacking beyond the 1-dimensional (1-D) case. The reason is that in the 2-dimensional (2-D) case and beyond, the relative position of the spikes enters the picture, and different geometrical configurations lead to different stability properties. Our first main contribution is a connection, in the limit where the spikes cluster at a given point, between solutions of the dual of the BLASSO problem and the least interpolant space for Hermite polynomial interpolation. This interpolation space, introduced by De Boor, can be computed by Gaussian elimination, and lead to an algorithmic description of limiting solutions to the dual problem. With this construction at hand, our second main contribution is a detailed analysis of the support stability and super-resolution effect in the case of a pair of spikes. This includes in particular a sharp analysis of how the signal-to-noise ratio should scale with respect to the separation distance between the spikes. Lastly, numerical simulations on different classes of kernels show the applicability of this theory and highlight the richness of super-resolution in 2-D.

1. Introduction. Sparse super-resolution is a fundamental problem of imaging sciences, where one seeks to recover the positions and amplitudes of pointwise sources (so-called "spikes") from linear measurements against *smooth* functions. A typical example is deconvolution, where the measurements correspond to a low-pass filtering (or equivalently the observation of low-frequency Fourier coefficients), and is at the heart of fluorescent microscopy techniques such as PALM and STORM [5, 52]. More generally, the measurements need not to be translation-invariant, and this is for instance the case in MEG/EEG [3] where the location of pointwise activity sources is crucial [34].

The fundamental question in all these fields is to understand the super-resolution limit (often called the "Rayleigh limit") of some computational method. This corresponds, for a given signal-to-noise ratio, to the minimum allowable separation limit between spikes so that their locations can be estimated. Certifying whether (or not) this limit goes to zero as the noise level drops, and at which speed, is a difficult problem, which until now, has mostly been addressed in 1-D. These questions are even more involved in higher-dimensions, because the geometry of the spikes configurations is much richer, and in contrary to the 1-D setting, these geometric configurations (e.g. whether 3 spikes are aligned or not) are expected to impact the super-resolution ability. It is the purpose of this paper to shed some light on higher-dimensional super-resolution phenomena.

1.1. BLASSO and Super-resolution. In this work, we focus on inverse problems for arbitrary dimension d. To this end, we consider the underlying domain to be $\mathcal{X} = \mathbb{R}^d$ or $\mathcal{X} = \mathbb{R}^d/\mathbb{Z}^d$ the d-dimensional torus. The sparse recovery methods we consider are framed as optimization problems on the Banach space $\mathcal{M}(\mathcal{X})$ of bounded Radon measures on $\mathcal{X}$. The unknown data to recover is thus a measure $m_0 \in \mathcal{M}(\mathcal{X})$, and one has access to linear measurements y in some Hilbert space $\mathcal{H}$, with

$$y = \Phi m_0 + w \in \mathcal{H} \tag{1.1}$$

*DAMTP, UNIVERSITY OF CAMBRIDGE, C.M.H.S.POON@MATHS.CAM.AC.UK

†DMA, ECOLE NORMALE SUPÉRIEURE, 45 RUE D'ULM, F-75230 PARIS CEDEX 05, FRANCE, GABRIEL.PEYRE@ENS.FR

where $w \in \mathcal{H}$ models the acquisition noise, and $\Phi : \mathcal{M}(\mathcal{X}) \to \mathcal{H}$ is an operator of the form

$$\Phi : m \in \mathcal{M}(\mathcal{X}) \mapsto \int_{\mathcal{X}} \varphi(x)\mathrm{d}m(x),$$

defined through its kernel $\varphi : x \in \mathcal{X} \mapsto \varphi(x) \in \mathcal{H}$. This kernel is assumed to be smooth, and it is indeed this smoothness that makes the operator "coherent" and ill-posed, thus requiring some sort of regularization technique to invert the forward imaging model (1.1).

A typical example is deconvolution, which corresponds to the translation invariant problem, where $\varphi(x) = \psi(x-\cdot)$ for some $\psi \in L^2(\mathcal{X})$ and $\mathcal{H} = L^2(\mathcal{X})$. On $\mathcal{X} = \mathbb{R}^d/\mathbb{Z}^d$, when ψ has a finitely supported Fourier spectrum $(\hat{\psi}(\omega))_{\omega\in\Omega}$, with $|\Omega| < +\infty$, up to a rescaling of the measurement, this is equivalent to consider directly a sampling of the Fourier frequencies, i.e. $\varphi(x) = (e^{2\mathrm{i}\pi\langle x,\,\omega\rangle})_{\omega\in\Omega}$ and $\mathcal{H} = \mathbb{C}^{|\Omega|}$. Non-translation invariant operators are also ubiquitous in imaging, for instance non-stationary blurs or indirect observation on the boundary of the domain as for instance in EEG/MEG. Denoting $S \subset \mathcal{X}$ the domain of interest (for instance the boundary of the skull for brain imaging), these techniques can be modelled as a sub-sampled convolution, i.e. $\varphi(x) = (\psi(x-z))_{z\in\partial S}$, where typically ψ is some kernel associated to a stationary electric or magnetic field. Another example of non-convolution problems routinely encountered in imaging is the Laplace transform $\varphi(x) = e^{-\langle x,\,\cdot\rangle}$. Lastly, let us mention the problem of mixture estimation, which can also be framed as a multi-dimensional sparse recovery problem [35].

This work focusses on the recovery of a superposition of point sources, which are measures of the form $m = \sum_{i=1}^N a_i\delta_{z_i}$, where $a\delta_z$ is a Dirac (or a "spike") at position $z \in \mathcal{X}$ and with amplitude $a \in \mathbb{R}$. Following several recent works (reviewed in Section 1.2 below), we regularize using the total variation norm $|m|(\mathcal{X})$ of $m \in \mathcal{M}(\mathcal{X})$, which is defined as

$$|m|(\mathcal{X}) \stackrel{\text{def.}}{=} \sup\left\{\int_{\mathcal{X}} \eta(x)\mathrm{d}m(x)\ ;\ \eta \in \mathscr{C}(\mathcal{X}),\ \|\eta\|_{L^\infty} \leqslant 1\right\}.$$

This corresponds to the generalization of the ℓ^1 and L^1 norms of vectors and functions, since one has $|\sum_i a_i\delta_{z_i}|(\mathcal{X}) = \|a\|_{\ell^1}$ and if m has a density $f = \frac{\mathrm{d}m}{\mathrm{d}x}$ with respect to the Lebesgue measure $\mathrm{d}x$, then $|m|(\mathcal{X}) = \|f\|_{L^1(\mathrm{d}x)}$.

We thus consider a least-squares total-variation regularization

$$\underset{m\in\mathcal{M}(\mathcal{X})}{\operatorname{argmin}}\ |m|(\mathcal{X}) + \frac{1}{2\lambda}\|\Phi m - y\|_{\mathcal{H}}^2, \tag{$\mathcal{P}_\lambda(y)$}$$

where $\lambda > 0$ is the regularization parameter, which should be adapted to the noise level $\|w\|$ (theoretical results such as ours advocate a linear scaling between λ and $\|w\|_{\mathcal{H}}$). In the case of noiseless measurements y (i.e. $w = 0$), the limit of ($\mathcal{P}_\lambda(y)$) as $\lambda \to 0$ is the constrained problem

$$\underset{m\in\mathcal{M}(\mathcal{X})}{\operatorname{argmin}}\ |m|(\mathcal{X}) \text{ subject to } \Phi m = y. \tag{$\mathcal{P}_0(y)$}$$

The purpose of this paper is to study the structure (and in particular the support) of the solutions to ($\mathcal{P}_\lambda(y)$) when $\|w\|$ and λ are small, and the positions of the spikes $(z_i)_i$ of the sought after measure m_0 cluster around a fixed point. The key question which we address is as follows: *if* $y = \Phi m_0 + w$ *where* $m_0 = \sum_{i=1}^N a_i \delta_{tz_i}$ *is the true measure and* $w \in \mathcal{H}$ *models some additive noise, how should* $\|w\|$ *and* λ *scale relative to the separation distance* t *such that the solution of* ($\mathcal{P}_\lambda(y)$) *recovers precisely* N *spikes, whose amplitudes and positions are stable with respect to* w *and* λ*?*

1.2. Previous Works.

On-the-grid LASSO. There is a long history on the use of convex methods, and in particular the ℓ^1 norm, to recover sparse signals on a grid. This was pioneered by geophysicists [16, 42, 54] and then became mainstream in signal processing [15] and for model selection in statistics [63]. General theoretical approaches (only constraining the sparsity of the signal), such as those used to study compressed sensing [28, 14], are however ineffective for super-resolution, where the linear measurement operator is highly coherent. The theory of super-resolution requires to impose constraints on the minimum separation distance between the spikes in place of the total number of those spikes.

Off-the-grid BLASSO, minimum-separation condition. In order to avoid introducing a-priori a fixed grid, which might be detrimental both theoretically and computationally, it makes sense to consider the "off-the-grid" setting introduced in the previous section 1.1, as exposed by several authors, including [23, 10, 13]. The ground breaking work of Candès and Fernandez-Granda [13], for the low-pass filter with cut-off frequency f_c, proved that under a $\mathcal{O}(1/f_c)$ minimum separation condition between the spikes, exact recovery is achieved. This initial result has been extended to include noise stability [12, 2]. Further refinements have been proposed, for instance statistical bounds [6] and exact support recovery condition [31] for more general (not necessarily translation invariant) measurements. This line of theoretical study works in arbitrary dimensions, but this does not corresponds to a super-resolution regime ($O(1/f_c)$ being often considered as the natural Rayleigh limit). For spikes with arbitrary signs, it is easy to construct counter examples below the Rayleigh limit where the BLASSO does not work even when there is no noise (this is to be contrasted with Prony-type approaches, see below).

BLASSO, positive spikes in 1-D. Going below the Rayleigh limit requires an analysis of the signal-to-noise scaling, i.e. at which speed the signal-to-noise ratio (SNR) should drop to zero as a function of the spikes separation distance. This scaling is typically polynomial, and the exponent depends on the number of spikes that cluster around a given location. The study of this scaling was initiated by [27], for combinatorial search methods (thus not numerically tractable). With such non-convex methods, the optimal scaling is obtained with no constraint on the sign pattern of the spikes, see also [25] for a refined analysis.

As noticed above, BLASSO ($\mathcal{P}_\lambda(y)$) in general cannot go below the Rayleigh limit for measures with arbitrary signs. For positive spikes, BLASSO achieves super-resolution for some classes of measurement operators Φ. This was initially studied for the LASSO problem (on a grid) [29, 33], and for the BLASSO, this holds for low-pass Fourier measurements and polynomial moments [23]. This exact recovery of positive spikes can also cope with sub-sampling [57, 30], but this is restricted to the 1-D setting.

Stability to noise for the LASSO (on discrete grid) is studied in [47], with a signal-to-noise scaling $\|w\| = O(t^{2N})$ where N is the number of spikes clustered in a radius of t. This result holds in 1-D and 2-D, see also [4]. In the 1-D BLASSO case, a signal-to-noise scaling $\|w\| = O(t^{2N-1})$ actually leads to exact support recovery under a non-degeneracy condition (which is true for the Gaussian convolution kernel) [26].

Our work proposes extensions of this last result in arbitrary dimensions.

BLASSO, positive spikes in multiple-dimensions. Only few theoretical works have studied sparse super-resolution in more than one dimension. Let us single out the work of [47], which proves that 2-D super resolution on a discrete grid has a similar signal-to-noise scaling with spikes separation distance t as in 1-D. This however does not provide an understanding of how super-resolution and support recovery operate in multiple dimensions, which in turn requires to work off-the-grid, using the BLASSO, as we do in this paper. Let us also note the work of [60] which studies scaling of statistical decision-theoretic bounds for pairs of spikes. This is inline with our study in Section 3, which shows that the BLASSO achieves the same signal-to-noise scaling.

Prony type methods. While this paper is dedicated to convex ℓ^1-type methods, there is a large body of methods and analysis that use non-convex or non-variational approaches. These methods are often generalizations of the initial idea of Prony [50] which encodes the spikes positions as the zeros of some polynomial, whose coefficients are derived from the measurements, see [61] and the review paper [40]. Let us for instance cite MUSIC [58], matrix pencil [36], ESPRIT [51], finite rate of innovation [7], Cadzow's denoising [11, 19].

It is not the purpose of this paper to advocate for or against the use of convex methods, and there are pros-and-cons both in term of both practical performances and theoretical understanding. An important advantage of Prony-based approaches is that, in the noiseless setting, $w = 0$, they achieve exact recovery without any condition on the sign of the spikes (whereas BLASSO requires either a minimum separation distance or positivity). The theoretical analysis of these approaches in the presence of noise is however more intricate, and only partial results exist. Furthermore, they rely on root finding computations, which might be unstable when detecting clusters of spikes. Stability issues also appear in non-root finding methods (such as the Frank-Wolfe algorithm we use in the numerical section). Developing stabilized solvers might be possible, but this is beyond the scope of this paper. Cramer-Rao statistical bounds can be derived [18] and non-asymptotic bounds have been proposed under minimum-separation conditions [43, 46]. A second difficulty with Prony based approaches is that they are non-trivial to extend to higher dimensions, and there is no general agreement on a canonical formulation even in 2-D. We refer for instance to [48, 41, 56, 53, 17, 1] and the references therein for several such extensions. Furthermore, in the noiseless setting, in sharp contrast with the 1-D setting, the number of recovered spikes does not scale linearly with the number of measurements [39].

Numerical Methods for the BLASSO. While it is not the purpose of this paper to develop numerical methods, let us sketch some pointers to solvers for the BLASSO problem. The BLASSO is computationally challenging because it is an infinite dimensional optimization problem. The most straightforward approach is to approximate the problem on a grid, which then becomes a finite dimensional LASSO. This however leads to quantization artifacts, typically doubling the number of observed spikes in 1-D [32].

In the case of a finite number f_c of Fourier frequencies, the dual optimization problem is finite dimensional. In 1-D, it can be solved by lifting to a semi-definite program in $O(f_c^2)$ variables [13]. In 2-D and on more general semi-algebraic domains, this lifting is more involved numerically, and requires the use of a Lasserre hierarchy [24]. For general measurements, [10] proposes to use the Frank-Wolfe algorithm, which operates by adding in a greedy-manner new spikes, and is a convex counterpart of the celebrated matching pursuit algorithm [45] over a continuous dictionary [37].

It has a slow convergence rate, which is improved by interleaving non-convex optimization steps, also used in [9]. This algorithm is in practice efficient and leads to state of the art result in 2-D and 3-D resolution, for instance with application to single-molecule imaging [9].

1.3. Dual Certificates. In this section we give the necessary background to state our contributions, which is the standard machinery of so-called "dual certificates" routinely used to analyze solutions of sparsity-regularized variational problems such as ($\mathcal{P}_\lambda(y)$).

Certificates. A positive discrete measure

$$m_0 = \sum_{i=1}^{N} a_i \delta_{z_i} \in \mathcal{M}(\mathcal{X}),$$

with $a_i \geqslant 0$ for all $i = 1, \dots, N$, is solution to ($\mathcal{P}_0(y)$) if and only if $\Phi m_0 = y$ and the set of Lagrange multipliers of the constraint, often referred to in the literature as "dual certificates"

$$\mathcal{D}(Z) = \{\eta \in \mathrm{Im}(\Phi^*) \; ; \; \|\eta\|_\infty \leqslant 1, \forall i, \eta(z_i) = 1\}$$

is non-empty, where we denote by $Z = (z_i)_{i=1}^N \in \mathcal{X}^N$ the spikes locations. Here we denoted $\Phi^* : p \in \mathcal{H} \mapsto \eta \overset{\text{def.}}{=} \Phi^* p \in \mathscr{C}(\mathcal{X})$ where $\eta(x) \overset{\text{def.}}{=} \langle \varphi(x), p \rangle_{\mathcal{H}}$. Proving that $\mathcal{D}(Z) \neq \emptyset$ is thus a Lagrange interpolation problem using continuous interpolating functions in $\mathrm{Im}(\Phi^*)$, and with the additional constraint that η should be bounded by 1.

To perform sensitivity analysis of the solutions of ($\mathcal{P}_\lambda(y)$) with respect to small perturbations (w, λ), one needs to ensure that some certificate $\eta \in \mathcal{D}(Z)$ is "non-degenerate". In the following, we say that a smooth function η is non-degenerate for the positions Z if

$$\forall\, x \notin Z, \quad \eta(x) < 1, \quad \text{and} \quad \forall\, i = 1, \dots, N, \; \eta(z_i) = 1 \quad \text{and} \quad \nabla^2 \eta(z_i) \prec 0. \quad (\mathrm{ND}(Z))$$

Here, we denote by $\nabla^2 \eta(x) \in \mathbb{R}^{d \times d}$ the Hessian matrix of η at $x \in \mathcal{X}$, and $A \prec 0$ means that the matrix $A \in \mathbb{R}^{d \times d}$ is negative definite. The condition ($\mathrm{ND}(Z)$) is a strengthening of the condition of being a dual certificate, and is reminiscent of the non-degeneracy condition on the relative interior of the sub-differential which is standard in sensitivity analysis in finite dimensions [8] (recall that we are here dealing with infinite-dimensional optimization problems).

As initially shown in [31], support stability of the solution to ($\mathcal{P}_\lambda(y)$) with small (w, λ) is governed by a specific dual certificate of minimum norm

$$\eta_{0,Z} \overset{\text{def.}}{=} \Phi^* p_{0,Z} \quad \text{where} \quad p_{0,Z} \overset{\text{def.}}{=} \underset{p \in \mathcal{H}}{\mathrm{argmin}} \; \{\|p\|_{\mathcal{H}} \; ; \; \Phi^* p \in \mathcal{D}(Z)\}.$$

Ensuring support stability requires $\eta_{0,Z}$ to be "non degenerate" in the sense that it should satisfy condition ($\mathrm{ND}(Z)$) above.

More precisely, if $\eta_{0,Z}$ is non degenerate (i.e. satisfies ($\mathrm{ND}(Z)$)), then when using $(\|w\| / \lambda, \lambda) = O(1)$, [31] shows that the solution of ($\mathcal{P}_\lambda(y)$) is unique and composed of N spikes, whose positions and amplitudes converge smoothly toward (a, Z) as $\lambda \to 0$. Note that even if m_0 is solution to ($\mathcal{P}_0(y)$), it might be the case that $\eta_{0,Z}$ is degenerate, meaning that for small noise, the support of m_0 is unstable. This is for instance illustrated in Section 4.2 (see in particular Figure 4.1).

Vanishing derivatives pre-certificate. Direct analysis of this $\eta_{0,Z}$, and in particular proving that it satisfies condition (ND(Z)), is difficult, mainly because of the non-linear constraint $\|\eta_{0,Z}\|_\infty \leqslant 1$. Fortunately, [31] introduces a simpler proxy, which is defined by replacing this constraint by the linear one of having vanishing derivatives at the spikes positions

$$\eta_{V,Z} \stackrel{\text{def.}}{=} \Phi^* p_{V,Z} \quad \text{where} \quad p_{V,Z} \stackrel{\text{def.}}{=} \underset{p\in\mathcal{H}}{\operatorname{argmin}} \; \{\|p\|_{\mathcal{H}} \; ; \; \forall i, (\Phi^* p)(z_i) = 1, \nabla(\Phi^* p)(z_i) = \mathbf{0}_d\} \tag{1.2}$$

where $\nabla\eta(x) \in \mathbb{R}^d$ is the gradient vector of η at $x \in \mathcal{X}$. The interest of this vanishing derivatives pre-certificate $\eta_{V,Z}$ stems from the fact that

$$\eta_{V,Z} \text{ satisfies } (\text{ND}(Z)) \implies \eta_{0,Z} \text{ satisfies } (\text{ND}(Z)) \quad \text{and} \quad \eta_{0,Z} = \eta_{V,Z}.$$

(and the converse is also true), which means that one can simply check the non-degeneracy of $\eta_{V,Z}$ to guarantee support stability in the small noise regime.

Asymptotic analysis. As detailed in the previous works section 1.2, if the spikes amplitudes $(a_i)_i$ have arbitrary signs (and can even be complex numbers), ensuring that m_0 is a solution of ($\mathcal{P}_0(y)$) (and then ensuring noise stability of the solution to ($\mathcal{P}_\lambda(y)$)) requires the spike positions Z to obey a minimum separation condition, so that a super-resolution effect cannot be observed. To study super-resolution, we assume throughout this paper that the spikes are positive, i.e. we impose $a_i \geqslant 0$, and we consider positions $tZ = (tz_i)_i$ parameterized by $t > 0$. As $t \to 0$, spikes at positions tZ cluster near 0. Note that, up to a translation of the domain, one can consider clustering at an arbitrary position and it is convenient to assume that this point is 0. Our results can be extended (under suitable decay condition on $\varphi(x)$) to clustering of spikes around multiple points, provided that these clustering points are well-separated. Studying a linear scaling $t \mapsto tZ$ is natural since we are interested in super-resolution effects for small t. If one considers a smooth trajectory $t \mapsto (z_i(t))_i$ of positions with $z_i(0) = 0$, then setting $Z = (z_i'(0))$ allows one to study a first order expansion of this trajectory near the clustering point 0.

In 1-D, there is no loss of generality in studying super-resolution along a 1-D curve of spikes $t \mapsto tZ$, and t is proportional to a separation distance between the spikes. In higher dimension, this is of course restrictive and our theory depends on the clustering direction Z. It is actually an important finding of our theory that this limitation is inherent to the multi-dimensional setting and that the stability constants of our results (and how they scale with t) cannot be uniform in Z. In this setting, t cannot be interpreted anymore as a simple separation distance, and super-resolution cannot be analyzed solely through the prism of a separation between the spikes.

In order to understand the super-resolution capability of ($\mathcal{P}_\lambda(y)$) for small t, it makes sense to analyze the behavior of $\eta_{V,tZ}$ as $t \to 0$. When this limit exists, we denote the so-called limiting precertificate as

$$\eta_{W,Z} \stackrel{\text{def.}}{=} \lim_{t\to 0} \eta_{V,tZ} \quad \text{where} \quad \eta_{V,tZ} = \Phi^* p_{V,tZ}. \tag{1.3}$$

The question of support stability in the one-dimensional case is addressed in details in [26]. In this simpler setting, $\eta_{W,Z} = \eta_W$ does not depend on Z, which reflects the fundamental difference with the higher dimensional case. Furthermore, [26] shows that it can be computed by solving a higher order Hermite interpolation problem, $\eta_W \stackrel{\text{def.}}{=} \Phi^* p_W$ where

$$p_W \stackrel{\text{def.}}{=} \underset{p\in\mathcal{H}}{\operatorname{argmin}} \; \left\{\|p\|_{\mathcal{H}} \; ; \; (\Phi^* p)(0) = 1, \forall k = 1, \dots, 2N-1, (\Phi^* p)^{(k)}(0) = 0\right\}, \tag{1.4}$$

where $\eta^{(k)}$ denotes the k^{th} derivative of η.

The main result of [26] is that, if η_W is non-degenerate, in the sense that

$$\forall\, x \neq 0, \quad \eta_W(x) < 1, \quad \text{and} \quad \eta_W^{(2N)}(0) < 0$$

then for t small enough, $\eta_{V,tZ}$ is also non-degenerate, and one can compute a sharp estimate of the support stability constant involved as a function of t. More precisely, one should have $\|w\|_{\mathcal{H}}/\lambda = O(1)$ and $\lambda = O(t^{2N-1})$ in order for ($\mathcal{P}_\lambda(y)$) to recover the correct number N of spikes and to have smoothly converging positions and amplitudes as $\lambda \to 0$.

The simple expression (1.4) allows one to study easily η_W, and numerical computations show that it is indeed non-degenerate for many low pass filters [26]. It can even be computed in closed form in the case of the Gaussian filter. In Appendix C, we show that one can also compute η_W in closed form in the case of an ideal low-pass filter of cutoff frequency f_c, for the special case $N = f_c$, and that both $\eta_{V,Z}$ and η_W are non-degenerate. The general case of $f_c \neq N$ is still an open problem.

1.4. Contributions. The link between support stability and the limiting certificate is far more intricate in multiple-dimensions. In particular, in the multi-dimensional case, $\eta_{W,Z}$ does depend on Z. The geometric configuration of Z determines the behaviour of $\eta_{W,Z}$. In order to derive a closed form expression for $\eta_{W,Z}$ and to use it to study support stability, one needs to study multi-dimensional Hermite interpolation. Hermite interpolation imposes directional derivatives, which are conveniently expressed using polynomial notation that we now detail.

Notation. We denote by Π^d the space of polynomials in d variables $(X_1, \ldots, X_d)$, Π_n^d the space of n degree polynomials in d variables $(X_1, \ldots, X_d)$, and for a d-tuple $k = (k_1, \ldots, k_d)$, the associated monomial $X^k \stackrel{\text{def.}}{=} X_1^{k_1} \ldots X_d^{k_d}$. For a polynomial $P = \sum_k p_k X^k \in \Pi^d$, we denote $P(\partial)$ the differential operator

$$P(\partial) \stackrel{\text{def.}}{=} \sum_k p_k \partial_1^{k_1} \ldots \partial_d^{k_d}$$

where ∂_s is the derivative with respect to the s^{th} variable. We shall use ∇^j to denote the full gradient of order j, and given a multi-index $\alpha = (\alpha_j)_{j=1}^d \in \mathbb{N}_0^d$, let $\partial^\alpha \stackrel{\text{def.}}{=} \partial_1^{\alpha_1}\partial_2^{\alpha_2} \cdots \partial_d^{\alpha_d}$. Given $x, z \in \mathcal{X}$ and $f \in \mathscr{C}^j(\mathcal{X})$, let $\partial_z^j f(x) \stackrel{\text{def.}}{=} \frac{\mathrm{d}^j}{\mathrm{d}t^j}|_{t=0} f(x + tz)$. Given $f \in L^2(\mathcal{X}) \cap L^1(\mathcal{X})$, the Fourier transform of f at ξ is defined by

$$\hat{f}(\xi) \stackrel{\text{def.}}{=} \int_{\mathcal{X}} f(x) e^{-2\pi \mathrm{i} \langle x, \xi \rangle} \mathrm{d}x.$$

For $\mathcal{X} = \mathbb{T}^d$, this defines Fourier coefficients and one should restrict ξ to belong to $\xi \in \mathbb{Z}^d$. Given $N \in \mathbb{N}$, $a \in \mathbb{R}^N$ and $Z \stackrel{\text{def.}}{=} (z_j) \in \mathcal{X}^N$, let $m_{a,Z} \stackrel{\text{def.}}{=} \sum_{j=1}^N a_j \delta_{z_j}$. Given $N \in \mathbb{N}$, let 1_N be the vector of length N whose entries are all ones, 0_N be the vector of length N whose entries are all zeros, and let δ_N be the vector whose first entry is one, and all other entries are zeros.

Our first contribution is a characterisation of this limiting certificate $\eta_{W,Z}$ in the multi-dimensional setting (see Section 2.2.4). In particular, Theorem 2.11 shows that

$$\eta_{W,Z} \stackrel{\text{def.}}{=} \Phi^* p_{W,Z} \quad \text{where}$$

$$p_{W,Z} \stackrel{\text{def.}}{=} \underset{p \in \mathcal{H}}{\operatorname{argmin}} \; \left\{ \|p\|_{\mathcal{H}} \; ; \; (\Phi^* p)(0) = 1, \forall P \in \overline{\mathcal{S}}_Z, (P(\partial)[\Phi^* p])(0) = 0 \right\},$$

where $\overline{\mathcal{S}}_Z \subset \mathcal{S}_Z$ is the subspace of polynomials P such that $P(0) = 0$, and $\mathcal{S}_Z$ is the least interpolant space (introduced by de Boor and Ron in [22]) associated with Hermite interpolation at the points Z. Section 2.2.2 recall the construction of $\mathcal{S}_Z$ and Section (2.2.3) describes an algorithm to compute a basis of this space. Furthermore, we provide (see Proposition 2.18) a closed form expression for this limiting certificate in the case of the Gaussian filter. Due to the necessity of this certificate (see Theorem 2.16), our analysis thus sheds light on the behaviour of support stability in arbitrary dimensions.

Our second main contribution is a detailed analysis of the structure of the solution to $(\mathcal{P}_\lambda(y))$ in the 2-D case of $N = 2$ spikes. In particular, we consider the stability of solutions to $(\mathcal{P}_\lambda(y))$ to the "true" positions Z_0 and amplitudes a_0, when given measurements $y = \Phi m_0 + w$ with $m_0 = \sum_j a_{0,j}\delta_{z_{0,j}}$. This case is of fundamental importance, because it is the one that practitioners are facing most of the time, see for example [59] and the references within. In fact, clusterings of 3 or more spikes are so ill-posed (even in 1-D, as detailed in [26]) that stable super-resolution would require a noise level which is often impossible to meet in practice. It is important to note that while this result requires some a priori hypothesis on the number of spikes, the BLASSO method does not need to have access to this information. When $Z = (z_1, z_2) \in \mathcal{X}^2$, $\eta_{W,Z}$ is characterized by the constraints $\eta_{W,Z}(0) = 1$ and

$$\forall j = 1, 2, 3, \quad \partial^j_{d_Z}\eta_{W,Z}(0) = 0 \quad \text{and} \quad \forall k = 0, 1, \quad \partial^k_{d_Z}\partial_{d_Z^\perp}\eta_{W,Z}(0) = 0,$$

where $d_Z = (z_2 - z_1)/\left\|z_2 - z_1\right\|$. Let $\Psi_Z : \mathbb{R}^6 \to \mathcal{H}$ be defined by

$$a \in \mathbb{R}^6 \mapsto a_0\varphi(0) + \sum_{j=1}^{3} a_j \partial^j_{d_Z}\varphi(0) + \sum_{k=0}^{1} a_{4+k}\partial^k_{d_Z}\partial_{d_Z^\perp}\varphi(0).$$

Assume that Ψ_Z is full rank. Then, support stability is guaranteed provided that $\eta_{W,Z}$ is nondegenerate in the following sense.

Definition 1.1. *Let $Z \overset{\text{def.}}{=} \{z_1, z_2\} \in \mathcal{X}^2$, we say that $\eta_{W,Z}$ is non-degenerate if $\eta_{W,Z}(x) < 1$ for all $x \neq 0$ and*

$$\begin{pmatrix} \partial^2_{d_Z^\perp}\eta_{W,Z}(0) & \frac{1}{2}\partial_{d_Z^\perp}\partial^2_{d_Z}\eta_{W,Z}(0) \\ \frac{1}{2}\partial_{d_Z^\perp}\partial^2_{d_Z}\eta_{W,Z}(0) & \frac{1}{12}\partial^4_{d_Z}\eta_{W,Z}(0) \end{pmatrix} \prec 0.$$

More precisely, let $B(0, r) \subset \mathbb{R}\times\mathcal{H}$ be defined as $B(0, r) \overset{\text{def.}}{=} \{(\lambda, w)\ ;\ \lambda \in [0, r), \left\|w\right\|_{\mathcal{H}} < r\}$, then our result concerning the scaling of $\left\|w\right\|$ and λ in the 2-spikes case is as follows.

Theorem 1.2. *Let $Z_0 \in \mathcal{X}^2$ and $a_0 \in \mathbb{R}^2_+$. Suppose that Ψ_{Z_0} is full rank and that η_{W,Z_0} is non-degenerate. Then, there exists constants t_0, c_1, c_2, M, such that for all $t \in (0, t_0)$, all $(\lambda, w) \in B(0, c_1 t^4)$ and $\left\|w/\lambda\right\| \leqslant c_2$,*

- *for $y_t \overset{\text{def.}}{=} \Phi m_{a_0, tZ_0}$, $(\mathcal{P}_\lambda(y_t + w))$ has a unique solution.*
- *the solution has exactly 2 spikes and is of the form $m_{a,tZ}$ where $(a, Z) = g_t^*(\lambda, w)$, with g_t^* being a continuously differentiable function defined on $B(0, c_1 t^4)$.*
- *The following inequality holds:*

$$|(a, Z) - (a_0, Z_0)|_\infty \leqslant M\left(\frac{|\lambda| + \left\|w\right\|}{t^3}\right). \tag{1.5}$$

We make the following remarks on this result:

Remark 1.3.

- *The scaling of t^4 is inline with the statistical bounds computed in [60].*
- *The condition of non-degeneracy can be verified numerically, and theoretically, it holds when Φ is a Gaussian convolution operator (see Section 2.4.1). Moreover, as will become apparent later, our definition of nondegeneracy is the natural extension of the definition used in [26] for the one-dimensional case.*
- *The assumption that Ψ_Z is full rank is shown in Proposition 2.14 to be true when Φ is a convolutional operator provided that its kernel has a sufficiently large bandwidth.*

Section 3 is dedicated to the proof of Theorem 1.2. Lastly, Section 4 showcases numerical illustrations of these theoretical advances. The code to reproduce the results of this paper is available online[1].

2. Asymptotics of Dual Certificates. This section studies the limiting pre-certificate $\eta_{W,Z}$ defined in (1.3).

2.1. Correlation Kernel. In order to ease the computation of pre-certificates, we introduce the correlation kernel

$$\forall\, (x,x') \in \mathcal{X}^2, \quad C(x,x') \overset{\text{def.}}{=} \langle \varphi(x),\, \varphi(x') \rangle_{\mathcal{H}} \in \mathbb{R}. \tag{2.1}$$

It is important to note that $\eta_{V,Z}$ can be computed by solving a linear system of size $Q \times Q$ where $Q \overset{\text{def.}}{=} (d+1)N$ involving this kernel,

$$\eta_{V,Z}(x) = \sum_{i=1}^{N} \sum_{k=0}^{d} \alpha_{i,k} \partial_{1,k} C(z_i, x) \tag{2.2}$$

$$\text{where} \quad \alpha = M^{-1} u_{d,N} \quad \text{and} \quad M = (\partial_{1,k}\partial_{2,\ell} C(z_i,z_j))_{i,j=1,\ldots,N}^{k,\ell=0,\ldots,d},$$

where we denoted $\partial_{1,k}\partial_{2,\ell}C(x,x') \in \mathbb{R}$ the derivative at $(x,x') \in \mathcal{X}^2$ with respect to the k^{th} coordinate of x and the ℓ^{th} coordinate of x'. Note that if M is not invertible, following the definition (1.2) of $\eta_{V,Z}$, the corresponding linear system should be solved in a least squares sense, and thus M^{-1} should be replaced by the pseudo-inverse of M. The vector $u_{d,N} \in \mathbb{R}^Q$ is defined by

$$\forall\, (i,k) \in \{1,\ldots,N\} \times \{0,\ldots,d\}, \quad (u_{d,N})_{i,k} \overset{\text{def.}}{=} \begin{cases} 1 & \text{if} \quad k=0, \\ 0 & \text{otherwise.} \end{cases}$$

The matrix $M \in \mathbb{R}^{Q\times Q}$ operates as $M\alpha = (\sum_{j,\ell} M_{(i,k),(j,\ell)}\alpha_{j,\ell})_{(i,k)}$. This matrix M becomes singular at $t=0$, and is severely ill-posed for small t. In the following section, we study this limit by performing a desingularizing change of basis, which allows us to compute the limit as $t \to 0$ by solving a stable linear system.

This expression (2.2) for $\eta_{V,Z}$, which only requires the inversion of a $Q \times Q$ linear system, has been used in 1-D to show that $\eta_{V,Z}$ is non-degenerate, even for measure with arbitrary sign, under a minimum separation condition and for a class of convolution operator with some spatial decay, see [62, 49].

2.2. Asymptotic Pre-certificate. We now aim at generalizing the expression (1.4) to the case $d > 1$.

[1] https://github.com/gpeyre/2017-SIMA-super-resolution

2.2.1. General Definition of Vanishing Pre-Certicates. The first difficulty is that the limit of $\eta_{V,tZ}$ as $t \to 0$, provided that it exists, depends on the direction of convergence $Z \in \mathcal{X}^N$, as illustrated by Figure 2.1 and we thus denote it $\eta_{W,Z}$. Intuitively, the difficulty is to identify which derivatives of $\eta_{W,Z}$ should vanish in the limit. They should span a space of dimension $Q \stackrel{\text{def.}}{=} N(d+1)$, since this matches the number of constraints appearing in (1.2).

The construction of the limiting certificate requires to identify a linear subspace $\mathcal{S}_Z \subset \Pi^d$ which encodes the vanishing derivative constraints (which should have dimension Q) and solve $\eta_{W,Z} \stackrel{\text{def.}}{=} \Phi^* p_{W,Z}$ where

$$p_{W,Z} \stackrel{\text{def.}}{=} \underset{p \in \mathcal{H}}{\operatorname{argmin}} \left\{ \|p\|_{\mathcal{H}} \ ; \ (\Phi^* p)(0) = 1, \forall P \in \bar{\mathcal{S}}_Z, (P(\partial)[\Phi^* p])(0) = 0 \right\}, \tag{2.3}$$

where $\bar{\mathcal{S}}_Z \subset \mathcal{S}_Z$ is the linear subspace of polynomials P such that $P(0) = 0$. We show in Section 2.2.2 below that indeed such a space $\mathcal{S}_Z$ exists, and that it can be computed using a Gaussian elimination algorithm.

Note that in dimension $d = 1$, the expressions (2.3) is equivalent to (1.4) when using the canonical polynomial space of dimension $Q = 2N$, that is $\mathcal{S}_Z = \Pi^1_{2N-1} = \mathrm{Span}\{X_1^r\}_{r=0}^{2N-1}$.

Remark 2.1 (Computation of η_W). *Once again, $\eta_{W,Z}$ defined by (2.3) can be computed by solving a $Q \times Q$ linear system. Indeed, the linear space $\mathcal{S}_Z$ is described using a basis of polynomials*

$$\mathcal{S}_Z = \mathrm{Span}\left\{P_r \ ; \ r = 0, \dots, Q-1\right\}$$

to which we impose for notational convenience $P_0 = 1$ (so that $P_0(\partial)[\eta] = \eta$). Then, denoting the various derivatives of the correlation kernel as

$$\forall\, (r,s) \in \{0, \dots, Q-1\}^2, \quad C_{r,s} \stackrel{\text{def.}}{=} P_r^{[1]}(\partial) P_s^{[2]}(\partial)[C]$$

where here we have used the notations $P_r^{[1]}(\partial)$ and $P_s^{[2]}(\partial)$ to indicate whether the polynomial should be used to differentiate on the first variable x or the second variable x' of $C(x,x')$ (in particular $C_{0,0} = C$), one has

$$\eta_{W,Z}(x) = \sum_{r=0}^{Q-1} \beta_r C_{r,0}(0,x) \quad \textit{where} \quad \beta = R^{-1}\delta_Q \tag{2.4}$$

$$\textit{and} \quad R = \Big(C_{r,s}(0,0) \Big)_{r,s=0,\dots,Q-1} \in \mathbb{R}^{Q \times Q},$$

where δ_Q denotes the vector whose first entry is 1 and all other entries are 0.

2.2.2. The Least Interpolant Space $\mathcal{L}_{\mathcal{F}}$. The definition (1.2) of $\eta_{V,Z}$ involves a Hermite interpolation problem at nodes $Z = (z_i)_{i=1}^N$. Considering the asymptotic tZ with $t \to 0$ of the associated interpolation problem naturally leads to the analysis (through Taylor expansion) of the behavior of polynomial interpolation. Polynomial interpolation in arbitrary dimension is notoriously difficult, and we refer to the monographs [44, 55] for a detailed account on this topic. This is due in large part to the fact that finding suitable polynomial spaces so that the interpolation problem is *regular* (has a unique solution) is non trivial, and that, in contrary to the 1-D case, such a

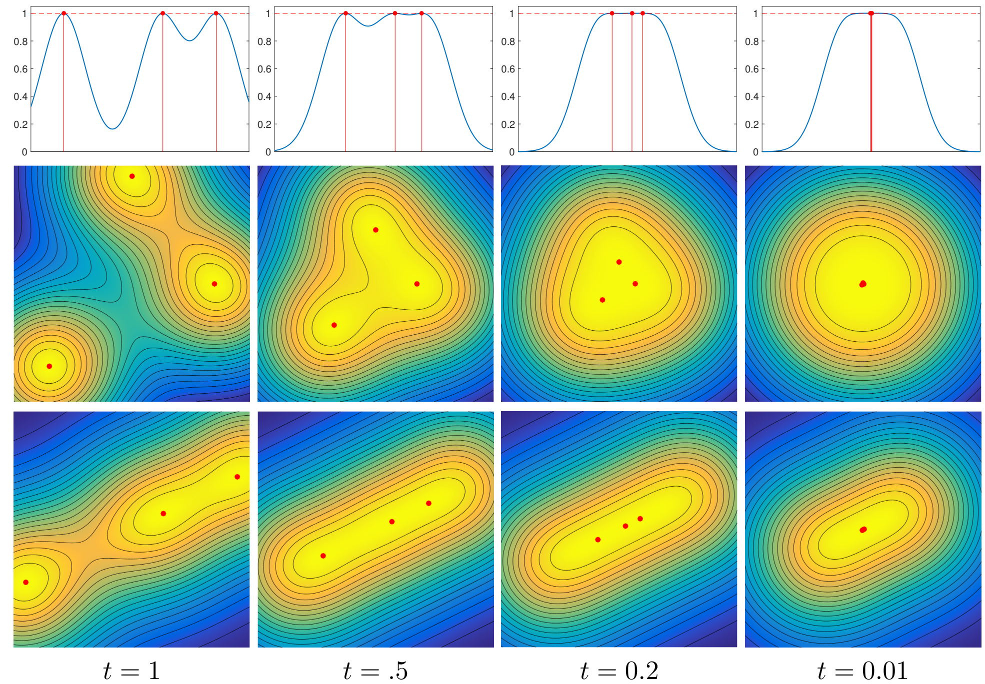


FIG. 2.1. *Display of the evolution of* $\eta_{V,tZ}$ *for* $t \to 0$. *Top: 1-D Gaussian convolution. Middle: 2-D Gaussian convolution, Z in generic positions. Bottom: same but with aligned positions Z.*

space depends on the interpolating positions Z. As we now explain, solving this issue is at the heart of the description of $\eta_{W,Z}$, and can be achieved in a canonical way using a construction of de Boor. There is a rich theory in polynomial interpolation, but since our interest is only in Hermite interpolation with first order derivatives only, for simplicity, we mention here only the relevant results for our specific context.

The Hermite interpolation (with first order derivatives) over a space $\mathcal{S} \subset \Pi^d$ looks for a polynomial $P \in \mathcal{S}$ solution of a system of equations

$$P(z_j) = c_{j,0} \quad \text{and} \quad \nabla P(z_j) = c_{j,1} \quad \text{for} \quad j = 1, \dots, N \tag{2.5}$$

for some $c_{j,0} \in \mathbb{R}$, $c_{j,1} \in \mathbb{R}^d$, $j = 1, \dots, N$ and positions $(z_i)_{i=1}^N$.

An important question is how one should choose the subspace $\mathcal{S}$ for this problem to be regular, i.e. have an unique solution for any choice of the values c in (2.5). In the univariate case $d = 1$, one can always choose the canonical polynomial space $\mathcal{S} = \Pi^1_{2N-1}$. However, the situation is much more complicated in the multivariate case because one cannot always choose $\mathcal{S} = \Pi^d_n$ for some $n \in \mathbb{N}_0$. To understand the issues here, first note that since $\dim \Pi^d_n = \binom{d+n}{d}$ and the number of interpolation conditions at z_i is $(d+1)$, we would need n to satisfy

$$\binom{d+n}{d} = N(d+1). \tag{2.6}$$

This may not be possible. For instance, if $N = 3$ and $d = 2$, then we have 9 interpolation constraints in total, but $\left|\Pi^2_3\right| = 6$ and $\left|\Pi^2_4\right| = 10$, so the condition (2.6) cannot be satisfied for any $n \in \mathbb{N}$. Furthermore, even when there exists n such that (2.6) holds, the choice of the canonical basis may not lead to a regular interpolation space, for instance, in the case $N = 2$ and $d = 2$, although choosing $\mathcal{S} = \Pi_2^2$ would satisfy (2.6), interpolation with this space is in fact singular for all choices of z_1 and z_2 [44].

In [22, 21], de Boor and Ron established a general technique for finding an appropriate polynomial space $\mathcal{S}$, so that the interpolation problem is regular, i.e. such that for any values c in (2.5), there exists a unique element $P \in \mathcal{S}$ such that (2.5) holds. This space is defined through the use of the least term in formal expansion of exponential forms.

DEFINITION 2.2 (Least term). *Let g be a real-analytic function on $\mathbb{R}^d$ (or at least analytic at $x = 0$, so that $g(x) = \sum_{|\alpha|=0}^{\infty} a_\alpha x^\alpha$). Let α_0 be the smallest integer $|\alpha|$ such that $a_\alpha \neq 0$. Then the least term $g_\downarrow$ of g is $g_\downarrow = \sum_{|\alpha|=\alpha_0} a_\alpha x^\alpha$.*

DEFINITION 2.3 (Least interpolant space). *For a linear functional $F \in (\Pi^d)'$ (the dual space of Π^d, so that $F : \Pi^d \to \mathbb{R}$ is linear), we define the formal power series*

$$g_F(x) \stackrel{\text{def.}}{=} F(e^{\langle \cdot,\, x\rangle}) \stackrel{\text{def.}}{=} \sum_{|\alpha|=0}^{\infty} \frac{F(p_\alpha)}{\alpha!} x^\alpha.$$

where $p_\alpha \stackrel{\text{def.}}{=} x \mapsto x^\alpha$. Given functionals $\mathcal{F} = \{F_q\}_q$ defined on Π^d, we define the space

$$\begin{aligned} \exp_{\mathcal{F}} &\stackrel{\text{def.}}{=} \operatorname{Span}\{g_F \ ;\ F \in \mathcal{F}\}, \\ \mathcal{L}_{\mathcal{F}} &\stackrel{\text{def.}}{=} \operatorname{Span}\{g_\downarrow \ ;\ g \in \exp_{\mathcal{F}}\}. \end{aligned}$$

The polynomial space $\mathcal{L}_{\mathcal{F}}$ is called the least interpolant space.

The main theorem of [22] asserts that this space $\mathcal{L}_{\mathcal{F}}$ defines regular interpolation problems. Note that of course $\mathcal{L}_{\mathcal{F}}$ depends on the positions Z. We recall below this theorem, stated as in [44].

THEOREM 2.4 ([44]). *Let $\mathcal{F} = \{F_q\}_{q=1}^{Q}$ be linear functionals defined on Π^d. Then, for any $c \in \mathbb{R}^Q$, there exists a unique $P \in \mathcal{L}_{\mathcal{F}}$ satisfying*

$$\forall\, r = 1, \ldots, Q, \quad F_r(P) = c_r.$$

EXAMPLE 2.5. *Let us consider an example presented in [44], which is also serves as an explanatory example throughout this article. Consider the problem of Hermite interpolation at $Z = \{z_1 \stackrel{\text{def.}}{=} (0,0), z_2 \stackrel{\text{def.}}{=} (0,1)\}$. The 6 linear functionals are*

$$\mathcal{F} = \{F_{0,i} : p \mapsto p(z_i),\ F_{1,i} : p \mapsto \partial_x p(z_i),\ F_{2,i} : p \mapsto \partial_y p(z_i)\}_{i=1,2}.$$

Then, the corresponding exponential functions are

$$\begin{aligned} &g_{F_{0,1}}(x,y) = 1, \quad g_{F_{1,1}}(x,y) = x, \quad g_{F_{2,1}}(x,y) = y, \\ &g_{F_{0,2}}(x,y) = e^y, \quad g_{F_{1,2}}(x,y) = xe^y, \quad g_{F_{2,2}}(x,y) = ye^y, \end{aligned}$$

and $\mathcal{L}_{\mathcal{F}} = \operatorname{Span}\{1, x, y, xy, y^2, y^3\}$.

REMARK 2.6. *The polynomial space constructed by de Boor and Ron preserves many properties of univariate interpolation, but notably, $\mathcal{L}_{\mathcal{F}}$ is of least degree in the sense that for any other space $\mathcal{S}$ leading to a regular interpolation problem,*

$$\dim(\mathcal{S} \cap \Pi_n^d) \leqslant \dim(\mathcal{L}_{\mathcal{F}} \cap \Pi_n^d), \qquad \forall n \in \mathbb{N}_0,$$

and in particular, if Π_n^d is regular, then we have $\mathcal{L}_{\mathcal{F}} = \Pi_n^d$. Furthermore, if $\mathcal{F}$ corresponds to Hermite interpolation at points Z, then writing $\mathcal{S}_Z \overset{\text{def.}}{=} \mathcal{L}_{\mathcal{F}}$ to be the least interpolant space associated to $\mathcal{F}$ (to highlight the dependency on Z) we have that for all invertible matrices $A \in \mathbb{R}^{d\times d}$ and $x \in \mathbb{R}^d$,

$$\mathcal{S}_{AZ+x} = \mathcal{S}_Z \circ A^\top = \left\{ P \circ A^\top \ ; \ P \in \mathcal{S}_Z \right\},$$

where we denoted $AZ + x = (Az_i + x)_i$, $A^\top$ the transpose matrix. This in particular implies that the least interpolant space is scale invariant with $\mathcal{S}_{tZ} = \mathcal{S}_Z$ for any $t > 0$.

2.2.3. Bases of $\mathcal{L}_{\mathcal{F}}$. In Section 2.2.4, we shall establish a precise link between the least interpolant space and $\eta_{W,Z}$ using Procedure 2.7. We emphasize that it is the space $\mathcal{L}_{\mathcal{F}}$, rather than the basis which determines $\eta_{W,Z}$. However, as highlighted already in Remark 2.1, to be useful from a computational point of view, it is important to describe an interpolation space $\mathcal{L}_{\mathcal{F}}$ using a basis of polynomials $(P_r)_{r=0,\ldots,Q-1}$. Algorithms for finding such a basis are presented in [21]. Although these algorithms are valid for the setting where the interpolation conditions are defined via general differential forms, since we are solely interested in the Hermite interpolation case with one derivative, for simplicity, we shall describe one of the algorithms in this setting for $d = 2$ (this extends verbatim to higher dimensions).

Gaussian elimination by monomials. The basic procedure of computing a basis of $\mathcal{S}_Z$ for $Z \in \mathcal{X}^N$ can be summarised as follows.

Procedure 2.7. *[21, Thm. 2.7]*

1. *By identifying each polynomial with its coefficients, define the Hermite intepolation operator $V_Z : \Pi^d \to \mathbb{R}^Q$ (where $Q = 3N$) as an infinite dimensional matrix (with infinitely many columns indexed by $\mathbb{N}_0^2$ columns indexed by $\{1, \ldots, Q\}$)*

$$V_Z : a \mapsto \left((P_a(z_i))_{i=1}^N, (\partial_{x_1} P_a(z_i))_{i=1}^N, (\partial_{x_2} P_a(z_i))_{i=1}^N \right) \in \mathbb{R}^Q,$$

   where $P_a = \sum_{\alpha\in\mathbb{N}_0^2} a_\alpha X^\alpha$.
2. *Perform Gaussian elimination with partial pivoting [64] to obtain the decomposition $V = LW$, where $L \in \mathbb{R}^{Q\times Q}$ is an invertible matrix and $W \in \mathbb{R}^{Q\times\mathbb{N}_0^2}$ is in row reduced echelon form.*
3. *For each row j of W (denoted by $W_{j,\cdot}$), let β_j be the first index of $W_{j,\cdot}$ such that $W_{j,\alpha} \neq 0$. Define*

$$\forall\, j = 0, \ldots, Q-1, \quad P_j(X) \overset{\text{def.}}{=} \sum_{|\alpha|=|\beta_j|} \frac{1}{\alpha!} W_{j,\alpha} X^\alpha.$$

Then, $\{P_j\}_{j=1}^{Q-1}$ defines a basis of $\mathcal{S}_Z$.

Remark 2.8. *It is in fact sufficient to restrict V_Z to the polynomial space Π_{2N-1} because Hermite interpolation on N nodes is always regular on Π_{2N-1} [44, Theorem 19]. Therefore, since $|\Pi_{2N-1}| = 2N^2 + N$, a basis of $\mathcal{S}_Z$ can be computed in $O(Q^2N^2)$ operations. In particular, we can replace V_Z by $\tilde{V}_Z \in \mathbb{R}^{Q\times(2N^2+N)}$ where*

$$\tilde{V}_Z \overset{\text{def.}}{=} \begin{pmatrix} (z_l^\alpha)_{l\in[N],|\alpha|\leqslant 2N-1} \\ \left(z_l^{\alpha-(1,0)}\right)_{l\in[N],|\alpha|\leqslant 2N-1} \\ \left(z_l^{\alpha-(0,1)}\right)_{l\in[N],|\alpha|\leqslant 2N-1} \end{pmatrix}.$$

Note that as a result, in Step 2, W is a $Q \times (2N^2 + N)$ matrix.

EXAMPLE 2.9. *Returning to Example 2.5 where $Z = \{(0,0),(0,1)\}$, we have that $\tilde{V}_Z = LW$, where*

$$L \stackrel{\text{def.}}{=} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1/2 & 0 & 1/6 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1/2 \end{pmatrix} \quad \text{and} \quad W : a \mapsto \begin{pmatrix} a_{0,0} \\ a_{0,1} \\ a_{1,0} \\ a_{0,2} + \sum_{j\geqslant 4} \frac{(6-2j)}{j!} a_{j,0} \\ a_{1,1} + \sum_{j\geqslant 2} \frac{1}{j!} a_{1,j} \\ a_{0,3} + \sum_{j\geqslant 4} \frac{(6j-12)}{j!} a_{0,j} \end{pmatrix}.$$

By Procedure 2.7, a basis of $\mathcal{L}_{\mathcal{F}}$ is therefore $\mathcal{B}_{\mathcal{F}} = \{1, y, x, y^2, xy, y^3\}$, where we write $x^j y^k$ for the polynomial $(x,y) \mapsto x^j y^k$

EXAMPLE 2.10 (Further examples).

- *When $Z = \{(a_j, 0)\ ;\ j = 1, \ldots, N\}$,*
$$\mathcal{B}_{\mathcal{F}} = \{1, x, x^2, \ldots, x^{2N-1}, y, xy, \ldots, x^{N-1}y\}$$
is a basis for $\mathcal{S}_Z$.
- *When $Z = \{(0,1),(0,0),(1,0)\}$,*
$$\mathcal{B}_{\mathcal{F}} = \{1, x, y, x^2, y^2, xy, x^3, y^3, x^2y - y^2x\}.$$
- *When $Z = \{(1,1),(1,-1),(-1,1),(-1,-1)\}$,*
$$\mathcal{B}_{\mathcal{F}} = \{1, x, y, x^2, y^2, xy, x^3, y^3, x^2y, y^2x, x^3y, y^3, x\}.$$

Gaussian elimination by degree. [20] also discussed another construction of a basis $\{g_j\}_j$ of $\exp_{\mathcal{F}}$, such that $\{(g_j)_\downarrow\}_j$ is a basis of $\mathcal{L}_{\mathcal{F}}$ which exhibit better numerical properties, and in particular satisfy the biorthogonality property that

$$i \neq j \iff \langle g_i,\, (g_j)_\downarrow \rangle_B = 0$$

where the *Bombieri inner product* is defined by $\langle P,\, Q\rangle_B \stackrel{\text{def.}}{=} (P(\partial)[Q])(0)$. The orthogonality property results in a basis which is more numerically stable and moreover provides an interpolation formula: given any function h,

$$I(h) \stackrel{\text{def.}}{=} \sum_j (g_j)_\downarrow \frac{\langle g_j,\, h\rangle_B}{\langle g_j,\, (g_j)_\downarrow\rangle_B}$$

is the unique polynomial in $\mathcal{L}_{\mathcal{F}}$ such that $F(h) = F(I(h))$ for all $F \in \mathcal{F}$. This basis can be constructed from any basis $\{f_j\}_j$ of $\exp_{\mathcal{F}}$, via a variant of the Gram-Schmidt process: assuming that $g_1, \ldots, g_{j-1}$ has been constructed, let

$$g_j \stackrel{\text{def.}}{=} f_j - \sum_{i<j} g_i \frac{\langle (g_i)_\downarrow,\, f_j\rangle_B}{\langle (g_i)_\downarrow,\, g_i\rangle_B}. \tag{2.7}$$

This ensures that $\langle (g_i)_\downarrow,\, g_j\rangle_B = 0$ for all $i < j$, and

$$\langle (g_j)_\downarrow,\, g_j\rangle_B \neq 0$$

since f_i are linearly independent and hence $g_j \neq 0$. A further correction of $g_i \stackrel{\text{def.}}{=} g_i - g_j \frac{\langle (g_j)_\downarrow,\, g_i\rangle_B}{\langle (g_j)_\downarrow,\, g_j\rangle_B}$ retains the biorthogonality established in (2.7) while ensuring that $\langle (g_j)_\downarrow,\, g_i\rangle_B = 0$ for all $i < j$. We remark also that one can equivalently construct this basis by appropriately grouping together columns of V_Z and performing Gaussian elimination degree by degree as opposed to monomial by monomial discussed above (see [21] for details).

2.2.4. The Limiting Certificate. With the construction of the least interpolant space at hand, we are now ready to explicitly define the limit $\eta_{W,Z}$ of $(\eta_{V,tZ})_{t>0}$ as $t \to 0$. We use for this the following interpolation space for spikes $Z = (z_i)_{i=1}^N$, $\mathcal{S}_Z \stackrel{\text{def.}}{=} \mathcal{L}_{\mathcal{F}}$ where

$$\mathcal{F} \stackrel{\text{def.}}{=} \{P \mapsto P(z_i), \quad P \mapsto \partial_1 P(z_i), \quad P \mapsto \partial_2 P(z_i), \quad i = 1, \dots, N\}. \tag{2.8}$$

THEOREM 2.11. *Let $\mathcal{S}_Z$ be the least interpolant space associated to the functionals in (2.8) and suppose that $\{h(\partial)\varphi(0)\ ;\ h \in \mathcal{S}_Z\}$ is of dimension $3N$. Then, for $Z \in \mathcal{X}^N$, one has*

$$\|p_{V,tZ} - p_{W,Z}\|_{\mathcal{H}} = \mathcal{O}(t) \quad \text{and} \quad \|\eta_{V,tZ} - \eta_{W,Z}\|_{L^\infty(\mathcal{X})} = \mathcal{O}(t)$$

where $p_{W,Z}$ and $\eta_{W,Z}$ are defined in (2.3) using $\mathcal{S}_Z$.

REMARK 2.12. *In this article, we are interested in the limit of $\eta_{V,tZ}$ which are defined using Hermite interpolation conditions at tZ. Note however that this result holds also for the limit of certificates defined via other differential forms. In particular, given any linear subspace of polynomials $\overline{\mathcal{S}}$ such that $P \in \overline{\mathcal{S}}$ implies that $P(0) = 0$, if*

$$\tilde{p}_{V,tZ} \stackrel{\text{def.}}{=} \left\{\|p\|\ ;\ (\Phi^* p)(tz_i) = 1,\ P(\partial)(\Phi^* p)(tz_i) = 0,\ \forall x \in Z,\ P \in \overline{\mathcal{S}}\right\},$$

then $\|\tilde{p}_{V,tZ} - \tilde{p}_{W,Z}\|_{\mathcal{H}} = \mathcal{O}(t)$, where $\tilde{p}_{W,Z}$ is defined through (2.3) using the least interpolation space associated with $\overline{\mathcal{S}} \cup \{x \mapsto 1\}$.

Proof. To begin with, let us define an operator which will be useful for this proof. Let

$$\Gamma_{tZ} : \mathbb{R}^{3N} \to \mathcal{H}, \qquad \Gamma_{tZ}\begin{pmatrix} a \\ b \\ c \end{pmatrix} \stackrel{\text{def.}}{=} \sum_{j=1}^N a_j \varphi(tz_j) + \sum_{j=1}^N b_j \partial_{x_1}\varphi(tz_j) + \sum_{j=1}^N c_j \partial_{x_2}\varphi(tz_j) \tag{2.9}$$

Note that the precertificates can be written as $p_{V,tZ} = \Gamma_{tZ}^{*,\dagger}\binom{1_N}{0_{2N}}$, where $A^\dagger$ is the pseudo-inverse of a matrix A. Moreover, observe that given $p \in \mathcal{H}$, by identifying p with the coefficients of the Taylor expansion of $(\Phi^* p)(t\cdot)$ around 0 up to degree $2N-1$, that is $P = \left(\langle p,\ \frac{t^{|\alpha|}}{\alpha!}\partial^\alpha\varphi(0)\rangle\right)_{\substack{\alpha \in \mathbb{N}_0^2 \\ |\alpha| \leqslant 2N-1}}$, we can associate Γ_{tZ} with the Hermite interpolation matrix V_Z (as defined in Procedure 2.7) via

$$\Gamma_{tZ}^* p = \operatorname{diag}((1_N^\top, t^{-1}1_{2N}^\top))\left(V_Z P + \mathcal{O}(t^{2N})\right). \tag{2.10}$$

Suppose that we decompose V_Z via Gaussian elimination so that $V_Z = LW$, where L is invertible and W is in row-reduced echelon form. Let $(\beta_j)_{j=0}^{Q-1}$ be as in Step 3 of Procedure 2.7. Then, using the representation of Γ_{tZ} from (2.10), we have that

$$\Gamma_{tZ}^* p = \begin{pmatrix} 1_N \\ 0_{2N} \end{pmatrix} \iff V_Z P + \mathcal{O}(t^{2N}) = \begin{pmatrix} 1_N \\ 0_{2N} \end{pmatrix} \iff WP + \mathcal{O}(t^{2N}) = L^{-1}\begin{pmatrix} 1_N \\ 0_{2N} \end{pmatrix}.$$

Note that since the first column of V_Z is $\binom{1_N}{0_{2N}}$ and the first column of W (since it is in row-reduced echelon form) is δ_{3N}, we have that $L^{-1}\binom{1_N}{0_{2N}} = \delta_{3N}$. By definition of the β_i's, we have that

$$WP = \left(t^{|\beta_i|}\sum_{|\alpha|=|\beta_i|}\frac{W_{z_i,\alpha}\langle\partial^\alpha\varphi(0),\, p\rangle}{\alpha!} + \mathcal{O}(t^{|\beta_i|+1})\right)_{i=1}^{3N}. \tag{2.11}$$

Let $\Psi_Z^* : \mathcal{H} \to \mathbb{R}^{3N}$ be defined by

$$\Psi_Z^* p = ((h_i(\partial)[\Phi^* p])(0))_{i=1}^{3N},$$

where

$$\{h_i\}_{i=1}^{3N} \overset{\text{def.}}{=} \left\{ X \mapsto \sum_{|\alpha|=|\beta_i|} \frac{W_{z_i,\alpha}}{\alpha!} X^\alpha \ ; \ i = 1, \ldots, 3N \right\}$$

is known to be a basis of $\mathcal{S}_Z \subseteq \Pi_{2N-1}$ from Procedure 2.7 and Remark 2.8. Therefore, from (2.11), there exists an operator $\tilde{\Psi}_t^* : \mathcal{H} \to \mathbb{R}^{3N}$ with $\tilde{\Psi}_t = \mathcal{O}(t)$ such that

$$WP + \mathcal{O}(t^{2N}) = \operatorname{diag}((t^{|\beta_i|})_{i=1}^{3N}) \left(\Psi_Z^* p + \tilde{\Psi}_t^* p \right). \tag{2.12}$$

Therefore,

$$\Gamma_{tZ}^* p = \begin{pmatrix} 1_N \\ 0_{2N} \end{pmatrix} \iff \Psi_Z^* p + \tilde{\Psi}_t^* p = \delta_{3N}$$

and $p_{V,tZ} = (\Psi_Z^* + \tilde{\Psi}_t^*)^\dagger \delta_{3N} = \Psi_Z^{*,\dagger} \delta_{3N} + \mathcal{O}(t)$ whenever Ψ_Z is full rank (which is true by assumption). Finally, since the first row of V_Z coincides with the first row of W (thanks to the fact that the top left entry of V_Z is 1), $h_1 \equiv 1$. Therefore, $p_{W,Z}$ is as defined in (2.3) using $\mathcal{S}_Z$. The final claim of this theorem is due to the following inequality:

$$\|\eta_{W,Z} - \eta_{V,tZ}\|_{L^\infty(\mathcal{X})} = \sup_{x\in\mathcal{X}} |\langle \varphi(x), p_{W,Z} - p_{V,tZ} \rangle| \leqslant \sup_{x\in\mathcal{X}} \|\varphi(x)\|_{\mathcal{H}} \|p_{W,Z} - p_{V,tZ}\|_{\mathcal{H}}.$$

□

REMARK 2.13 (Computation of $\eta_{W,Z}$). *With this theorem 2.11, it is now possible to compute $\eta_{W,Z}$ using the formula (2.4) in the scheme detailed in Remark 2.1 and an associated basis $\mathcal{B}_{\mathcal{F}} = \{P_r\}_{r=0}^{Q-1}$.*

A key assumption of Theorem 2.11 is that the dimension of the space

$$\{h(\partial)\varphi(0) \ ; \ h \in \mathcal{S}_Z\}$$

must be $3N$. The following proposition shows that this is satisfied for convolution kernels of sufficiently large bandwidth.

PROPOSITION 2.14. *Let Φ be a convolution operator with $\varphi(x) = \psi(\cdot - x)$ for some $\psi \in L^2(\mathcal{X})$. Let L be the smallest integer such that $\Pi_L^2 \supseteq \mathcal{S}_Z$. Suppose that the Fourier transform (as defined in Section 1.4) of ψ satisfies $\hat{\psi}(\alpha) \neq 0$ for all $\alpha \in \mathbb{N}_0^2$ such that $|\alpha| \leqslant L$. Then, given any $Z \in \mathcal{X}^N$, $\{h(\partial)\varphi(0) \ ; \ h \in \mathcal{S}_Z\}$ is of dimension $3N$. Furthermore, $L \leqslant 2N - 1$.*

REMARK 2.15 (Translation invariant correlation). *Note that this result extends to the more general case where $\Phi : \mathcal{M}(\mathcal{X}) \to \mathcal{H}$ is such that the correlation kernel defined in (2.1) satisfies $C(x, x') = K(x - x')$ for some positive definite function K. In this case, one can define $\tilde{\varphi}(x) = \psi(\cdot - x)$ where $\hat{\psi} = \sqrt{\hat{K}}$ and apply Proposition 2.14 to $\tilde{\varphi}$ in place of φ.*

Proof. **Step I.** Let us first show that elements of $\Psi_L \overset{\text{def.}}{=} \left\{ \partial^\alpha \varphi(0) \ ; \ |\alpha| \leqslant L, \ \alpha \in \mathbb{N}_0^2 \right\}$ are linearly independent provided that $\hat{\psi}(\alpha) \neq 0$ for all $\alpha \in \mathbb{N}_0^2$ with $|\alpha| \leqslant L$. Observe that $\partial^\alpha \varphi(0) = \partial^\alpha \psi$, so that $\Psi_L = \left\{ \partial^\alpha \psi \ ; \ |\alpha| \leqslant L, \ \alpha \in \mathbb{N}_0^2 \right\}$. On $\mathcal{X} = \mathbb{T}^d$, Ψ_L being independent is equivalent to the Fourier coefficients of the elements of Ψ_L being linearly independent. On $\mathcal{X} = \mathbb{T}^d$, Ψ_L being independent follows from these functions being independent when restricted to $[0,1]^d$, and thus we can restrict our attention to the Fourier transform sampled on $\mathbb{Z}^d$ and apply the same reasonning.

Let $m \in \mathbb{N}$ and let $\xi \stackrel{\text{def.}}{=} (\xi_j)_{j=1}^m \in \mathbb{R}^m$. The Fourier Transform of $\partial^\alpha \psi$ evaluated at frequencies ξ are $\left(\hat{\psi}(\xi_j)(2\mathrm{i}\pi\xi_j)^\alpha\right)_{j=1}^m$. Therefore, Ψ_L is linearly independent if the columns of the matrix $\mathrm{diag}((\hat{\psi}(\xi_j)_{1\leqslant j\leqslant N})M_\xi$ are linearly independent, where $M_\xi \stackrel{\text{def.}}{=} ((2\mathrm{i}\pi\xi_j)^\alpha)_{\substack{|\alpha|\leqslant L\\ 1\leqslant j\leqslant m}}$ is the Lagrange interpolation matrix, with evaluation at points ξ and using the polynomial basis $(X^\alpha)_{|\alpha|\leqslant L}$. From [44, Theorem 1], we know that M_ξ is invertible for almost every choice of ξ where $m = \left|\Pi_{\mathrm{L}}^2\right|$. Furthermore, one possible choice of ξ is $\left\{\alpha \in \mathbb{N}_0^2 \,;\, |\alpha| \leqslant L\right\}$. Therefore, to ensure linear independence of Ψ_L, it is enough to check that $\hat{\psi}(\alpha) \neq 0$ for all $\alpha \in \mathbb{N}_0^2$ such that $|\alpha| \leqslant L$.

Step II. We are now ready to show that $\{h(\partial)\varphi(0) \,;\, h \in \mathcal{S}_Z\}$ is of dimension $3N$. For this, it is sufficient to check that, for a given basis $\mathcal{B} \stackrel{\text{def.}}{=} \{g_i\}_{i=1}^{3N}$ of $\mathcal{S}_Z$, $\{g_i(\partial)\varphi(0) \,;\, i = 1, \ldots, 3N\}$ is linearly independent. Recall from Remark 2.8 that $\mathcal{S}_Z$ associated with Hermite interpolation at N points Z satisfies $\mathcal{S}_Z \subset \Pi_L^2$, where $L \stackrel{\text{def.}}{=} 2N-1$. Moreover, for each $i = 1, \ldots, 3N$, there exists coefficients $(c_{i,\alpha})_{|\alpha|\leqslant L}$ such that $g_i(X) = \sum_{|\alpha|\leqslant L} c_{i,\alpha} X^\alpha$. Let

$$B \stackrel{\text{def.}}{=} (c_{i,\alpha})_{\substack{i=1,\ldots,3N\\ |\alpha|\leqslant L}}$$

be the coefficient matrix such that $B(X^\alpha)_{|\alpha|\leqslant L} = (g_i(X))_{1\leqslant i\leqslant 3N}$. Note that since $\mathcal{B}$ is a basis, given any $a \in \mathbb{R}^{3N}$, $B^\top a = 0$ if and only if $a = 0$, since otherwise, there would be an $a \neq 0$ such that

$$\forall X \in \mathbb{R}^d, \qquad 0 = \langle a,\, B(X^\alpha)_{|\alpha|\leqslant L}\rangle,$$

so $\sum_{j=1}^{3N} a_j g_j \equiv 0$, which would contradict the assumption that $\mathcal{B}$ is a basis. Therefore, if the set of vectors $\{g_i(\partial)\varphi(0) \,;\, i = 1, \ldots, 3N\}$ is linearly dependent, then there exists $0 \neq a \in \mathbb{R}^{3N}$ such that

$$0 = \langle a,\, B(\partial^\alpha\varphi(0))_{|\alpha|\leqslant L}\rangle = \langle B^\top a,\, (\partial^\alpha\varphi(0))_{|\alpha|\leqslant L}\rangle.$$

This leads to the required contradiction, since we have shown in the first step that Ψ_L is linearly independent, and therefore, $B^\top a = 0$ and hence $a = 0$. ☐

2.3. Necessity of $\eta_{W,Z}$. The following result shows that it is necessary for $\eta_{W,Z}$ to be a valid certificate (that is, $\|\eta_{W,Z}\|_\infty \leqslant 1$) if support stability is possible for small t. In the following, given a measure m_{a_0,Z_0} where $a_0 \in \mathbb{R}^N$ and $Z_0 \in \mathcal{X}^N$, we say that m_{a_0,Z_0} is support stable if there exists a neighbourhood $V \subset \mathbb{R}_+ \times \mathcal{H}$ around $(0,0)$ and a continuous path $(\lambda, w) \in V \mapsto (a, Z) \in \mathbb{R}^N \times \mathcal{X}^N$ such that $m_{a,Z}$ is a solution to $\mathcal{P}_\lambda(\Phi m_{a_0,Z_0} + w)$.

Theorem 2.16. *Suppose that $\{h(\partial)\varphi(0) \,;\, h \in \mathcal{S}_{Z_0}\}$ is of dimension $3N$ and that $\varphi \in \mathscr{C}^2(\mathcal{X})$. Suppose that there exist sequences $(t_n)_{n\in\mathbb{N}} \subset \mathbb{R}_+$, $(a_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ and invertible affine maps $A_n : \mathcal{X} \to \mathcal{X}$ with $t_n \to 0$ and $\|A_n - \mathrm{Id}\| \to 0$ as $n \to \infty$ such that $m_{a_n, t_n A_n Z_0}$ is support stable. Here, $A_n Z_0 \stackrel{\text{def.}}{=} \{A_n z \,;\, z \in Z_0\}$. Then, $\|\eta_{W,Z_0}\|_{L^\infty} = 1$.*

Remark 2.17. *In particular, if $\eta_{W,Z_0}(x) > 1$ for some $x \in \mathcal{X}$, then for Z being a sufficiently small affine perturbation of Z_0, and t sufficiently small, under arbitrarily small noise w and regularization parameter λ, solving ($\mathcal{P}_\lambda(y)$) with $y = \Phi m_{a,tZ} + w$ produces solutions with additional small spikes (see Section 4).*

Proof. [Proof of Theorem 2.16] Let $Z_n \stackrel{\text{def.}}{=} A_n Z_0$. We first mention that from (2.12) in the proof of Theorem 2.11, $\Gamma_{tZ}^* = L(\Psi_Z^* + \mathcal{O}(t))$, where L is an invertible matrix. So, Γ_{tZ} (as defined in (2.9)) is full rank provided that t is sufficiently small and Ψ_Z is full rank. Note that since $\{h(\partial)\varphi(0)\ ;\ h \in \mathcal{S}_{Z_0}\}$ is of dimension $3N$, Ψ_{Z_0} is full rank. To see that Ψ_{Z_n} is also full rank, observe that as A_n is an affine mapping, it is of the form $\tilde{A}_n + b_n$, where $b_n \in \mathcal{X}$ and $\tilde{A}_n$ is an invertible matrix. Recall from Remark 2.6 that the least interpolant spaces satisfy $\mathcal{S}_{Z_n} = \mathcal{S}_{Z_0} \circ A_n^\top$. So, $\Psi_{Z_n} = \Psi_{Z_0} + \mathcal{O}(\|A_n - \mathrm{Id}\|)$ is full rank for all n sufficiently large. Therefore, there there exists $K \in \mathbb{N}$ such that $\Gamma_{t_n Z_n}$ is full rank for all $n \geqslant K$.

Fix $n \geqslant K$. Since $m_{a_n, t_n Z_n}$ is support stable, there exists a neighbourhood $V_n \subset \mathbb{R} \times \mathcal{H}$ of 0 and a continuous function $g_n : V_n \to \mathbb{R}^N \times \mathcal{X}^N$ such that given $(\lambda, w) \in V_n$ and $(a, Z) = g_n(\lambda, w)$, $m_{a,Z}$ solves $\mathcal{P}_\lambda(y_n + w)$ with $y_n = \Phi m_{a_n, t_n Z_n}$.

We first show that g_n coincides with a $\mathscr{C}^1$ function in a neighbourhood of 0: define for $u \stackrel{\text{def.}}{=} (a, Z) \in \mathbb{R}^N \times \mathcal{X}^N$ and $v \stackrel{\text{def.}}{=} (\lambda, w) \in \mathbb{R}_+ \times \mathcal{H}$, a function f_n : $(\mathbb{R}^N \times \mathcal{X}^N) \times (\mathbb{R}_+ \times \mathcal{H}) \to \mathbb{R}^{3N}$ by

$$f_n(u, v) \stackrel{\text{def.}}{=} \Gamma_{t_n Z}^*(\Phi_{t_n Z} a - \Phi_{t_n Z_n} a_n - w) - \lambda \begin{pmatrix} 1_N \\ 0_{2N} \end{pmatrix}.$$

For all $(\lambda, w) \in V_n$, optimality of $m_{a,Z}$ for $(a, Z) = g_n(\lambda, w)$ implies that

$$f_n(g_n(\lambda, w), (\lambda, w)) = 0.$$

Since f_n is $\mathscr{C}^1$, $f_n((a_n, Z_n), 0) = 0$, $\partial_u f_n((a_n, Z_n), 0)$ is invertible, we can apply the implicit function theorem to deduce that there exists $g_* \in \mathscr{C}^1$, a neighbourhood $V \subset \mathbb{R} \times \mathcal{H}$ of 0 and a neighbourhood $U \subset \mathbb{R}^N \times \mathcal{X}^N$ of (a_n, Z_n) such that $g_* : V \to U$ and $g_*(\lambda, w) = (a, Z)$ if and only if $f_n((a, Z), (\lambda, w)) = 0$. Now, by continuity of g_n, there exists $\tilde{V} \subset V_n \cap V$ such that $g_n(\tilde{V}) \subset U$. Moreover, since $f_n(g_n(\lambda, w), (\lambda, w)) = 0$ for all $(\lambda, w) \in \tilde{V}$, we have that $g_n(\lambda, w) = g_*(\lambda, w)$ for all $(\lambda, w) \in \tilde{V}$. Therefore, $m_{a_n, t_n Z_n}$ is support stable with a $\mathscr{C}^1$ function. We may now apply [31, Proposition 8] to conclude that $\|\eta_{V, t_n Z_n}\|_{L^\infty} \leqslant 1$. Finally,

$$\begin{aligned} \eta_{V, t_n Z_n} &= \Phi^* \Psi_{Z_n}^{*,\dagger} \delta_{3N} + \mathcal{O}(t_n) = \Phi^* \Psi_{A_n Z_0}^{*,\dagger} \delta_{3N} + \mathcal{O}(t_n) \\ &= \Phi^* \Psi_{Z_0}^{*,\dagger} \delta_{3N} + \mathcal{O}(\|A_n - \mathrm{Id}\|) + \mathcal{O}(t_n) = \eta_{W, Z_0} + \mathcal{O}(\|A_n - \mathrm{Id}\|) + \mathcal{O}(t_n), \end{aligned}$$

so $\eta_{V, t_n Z_n} = \Phi^* \Gamma_{t_n Z_n}^{*,\dagger} \binom{1_N}{0_{2N}} \to \eta_{W, Z_0}$ as $n \to \infty$ and hence, we must have $\|\eta_{W, Z_0}\|_{L^\infty} \leqslant 1$.
∎

2.4. Special cases.

2.4.1. Explicit formula of $\eta_{W,Z}$ for Gaussian convolution. We consider the Gaussian convolution measurement operator $\varphi(x) = \psi(x - \cdot) \in \mathcal{H} = L^2(\mathbb{R}^d)$ on $\mathcal{X} = \mathbb{R}^d$ where

$$\psi(x) \stackrel{\text{def.}}{=} e^{-\|x\|^2}$$

i.e. $\Phi m = m \star \psi$ is the convolution against kernel ψ. For the sake of simplicity, we only state the results below in 2-D, $d = 2$, but it extends to arbitrary dimensions.

In the following, we denote the exponential space associated to Hermite interpolation (of degree 1) at points Z by $\exp_Z$.

PROPOSITION 2.18. *Let $L \in \mathbb{N}$. Suppose that $\Pi_L^2 \subseteq \mathcal{S}_Z \subseteq \Pi_{L+1}^2$. Let $\{g_j\}_j$ be a basis of $\exp_Z$ so that $\{(g_j)_\downarrow\}_j$ is a basis of $\mathcal{S}_Z$ and $\langle g_j, (g_k)_\downarrow\rangle_B = 0$ for all $j \neq k$, where $\langle\cdot, \cdot\rangle_B$ is the Bombieri inner product (see Section 2.2.3). Then,*

$$\eta_{W,Z}(x) = \exp\left(-\frac{\|x\|^2}{2}\right)\left(\sum_j \frac{\langle (g_j)_\downarrow, \tilde{\psi}\rangle_B}{\langle g_j, (g_j)_\downarrow\rangle_B}(g_j)_\downarrow(x)\right) \quad \textit{where} \quad \tilde{\psi}(x) \stackrel{\text{def.}}{=} \exp\left(\frac{\|x\|^2}{2}\right).$$

In particular, if $\mathcal{S}_Z = \Pi_L^2$, then writing $x = (x_1, x_2)$,

$$\eta_{W,Z}(x) = \exp\left(-\frac{\|x\|^2}{2}\right) \sum_{0\leqslant 2\alpha_1+2\alpha_2\leqslant L} \frac{1}{\alpha!} x_1^{2\alpha_1} x_2^{2\alpha_2}.$$

In the case where Z consists of N points, all aligned along the first axis,

$$\eta_{W,Z}(x) = \exp\left(-\frac{\|x\|^2}{2}\right) \sum_{0\leqslant j\leqslant 2N-1} \frac{x_1^{2j}}{j!}.$$

Proof. Let $M \stackrel{\text{def.}}{=} |\mathcal{S}_Z|$. Since $\mathcal{S}_Z \supseteq \Pi_L$, we may assume that $\{g_j\}_{j=0}^{M-1}$ is such that

$$\mathcal{B} \stackrel{\text{def.}}{=} \{(g_j)_\downarrow\}_j \stackrel{\text{def.}}{=} \{X \mapsto X^\alpha;\ |\alpha| \leqslant L\} \cup \{P_j\}_{j\in J_{L+1}}$$

where P_j, $j \in J_{L+1}$ are homogeneous polynomials of degree $L+1$. Indeed, it is possible to write a basis of $\exp_Z$ which contains elements of the form $f_\alpha(X) = X^\alpha + \sum_{|\beta|>L} a_\beta X^\beta$ for $|\alpha| \leqslant L$. Since $\{f_\alpha\}$ satisfies $\langle f_\alpha, (f_\beta)_\downarrow\rangle_B = 0$ for all $\alpha \neq \beta$, the modified Gram Schmidt process in the construction of $\mathcal{B}$ would ensure that the first elements $\mathcal{B}$ are the monomials $\{X^\alpha\}_{|\alpha|\leqslant L}$. For simplicity of notation, let $h_j \stackrel{\text{def.}}{=} (g_j)_\downarrow$ and let g_0 be such that $h_0 \equiv 1$.

Recall that $\eta_{W,Z}$ is a function of the form

$$\eta_{W,Z}(x) = \sum_{|\alpha|\leqslant L} a_\alpha \partial^\alpha[\psi \star \psi](x) + \sum_{j\in J_{L+1}} a_\alpha (h_j(\partial)[\psi\star\psi])(x),$$

with $\eta_{W,Z}(0) = 1$, and $[h_j(\partial)]\eta_{W,Z}(0) = 0$ for $j \neq 0$. Note that

$$[\psi \star \psi](x_1, x_2) = \frac{\pi}{2} \exp\left(\frac{-({x_1}^2 + {x_2}^2)}{2}\right).$$

Moreover,

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n} \exp(-x^2/2) = (-1)^N \exp(-x^2/2) H_n(x)$$

where H_n is a monic polynomial of degree n (called the Hermite polynomial of degree n).

Since h_j is a homogeneous polynomial of degree $|\alpha|$, it follows that

$$(h_j(\partial)[\psi \star \psi])(x) = (-1)^{|\alpha|} \exp(-\|x\|^2/2)(h_j(x) + p_j(x))$$

where p_j is a polynomial of degree at most $\deg(h_j) - 1$. Therefore, there exists coefficients $\tilde{a}$ such that

$$F(x) \stackrel{\text{def.}}{=} \exp(\|x\|^2/2) \cdot \eta_{W,Z}(x) = \sum_{|\alpha| \leqslant L} \tilde{a}_\alpha x^\alpha + \sum_{j \in J_{L+1}} \tilde{a}_j h_j(x).$$

is a polynomial of degree $L+1$ and and it remains to determine the coefficients $\tilde{a}$.

Since $\langle g_j,\, h_k \rangle_B = 0$ for $j \neq k$, we have that

$$\langle g_j,\, F \rangle_B = g_j(\partial) F(0) = \tilde{a}_j \langle g_j,\, h_j \rangle_B \implies \tilde{a}_j = \frac{g_j(\partial)F(0)}{\langle g_j,\, h_j \rangle_B}.$$

Recall that $F(x) = \exp(\|x\|^2/2)\eta_{W,Z}(x)$, and since $\mathcal{S}_Z = \mathcal{S}_{tZ}$ for all $t > 0$, by writing $g_{j,t}$ as the basis element associated with $\exp_{tZ}$, we have that $(g_{j,t})_\downarrow = h_j$, we have $\lim_{t \to 0} g_{j,t}(\partial) F(0) = h_j(\partial) F(0) = [P_\alpha(\partial)\tilde{\psi}](0)$, where the final inequality is because $\partial^\alpha \eta_{W,Z}(0) = 0$ for all $|\alpha| \leqslant L$. Therefore,

$$\eta_{W,Z} = e^{-(x_1^2 + x_2^2)} \left(\sum_j \frac{\langle h_j,\, \tilde{\psi} \rangle_B}{\langle h_j,\, h_j \rangle_B} h_j(x) \right).$$

For the case where Z consists of aligned points, we can use the fact that

$$\{1, x, x^2, \ldots, x^{2N-1}, y, xy, \ldots, x^{N-1}y\}$$

is a basis of $\mathcal{S}_Z$ and proceed in a similar manner.

□

2.4.2. Convolution Operators and Vanishing Odd Derivatives. The following proposition shows that convolution operators enjoy the property that the odd derivatives of $\eta_{W,Z}$ vanish. This typically leads to better behaved (e.g. non-degenerate) certificates, as illustrated in Section 4. More generally, this proposition shows that, if the correlation kernel of Φ satisfies $C(x, x') = C(-x, -x')$, then the vanishing of consecutive derivatives up to some $N \in \mathbb{N}$ will imply the vanishing of all odd derivatives.

PROPOSITION 2.19. *Suppose that the correlation kernel $C : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined by $C(x, x') = \langle \varphi(x),\, \varphi(x') \rangle$ is such that $C(x, x') = C(-x, -x')$. Then, $\nabla^j \eta_{W,Z}(0) = 0$ for all odd integers $j \in \mathbb{N}$.*

Proof. Recall the notation from Remark 2.1 and in particular, the fact that given any basis $\{P_r \,;\, r = 0, \ldots, Q-1\}$ of the least interpolant space $\mathcal{S}_Z$, where we write for convenience $P_0 = 1$, $\eta_{W,Z}$ can be written as

$$\eta_{W,Z}(x) = \sum_{r=0}^{Q-1} \beta_r C_{r,0}(0, x),$$

where $\beta \stackrel{\text{def.}}{=} R^{-1}\delta_Q$ and $R \stackrel{\text{def.}}{=} \Big(C_{r,s}(0,0)\Big)_{r,s=0,\ldots,Q-1} \in \mathbb{R}^{Q \times Q}$. From 2.2.3, we can choose a basis such that each basis element P_r is a homogeneous polynomial. Since $C(x, x') = C(-x, -x')$, given multi-indices $\alpha, \alpha' \in \mathbb{N}_0^d$, we must have $\partial_x^\alpha \partial_{x'}^{\alpha'} C(0,0) = 0$ whenever $|\alpha + \alpha'|$ is odd. In particular, $C_{r,s}(0,0) = 0$ if P_r is a polynomial of odd degree and P_s is a polynomial of even degree. Let I_E index the basis polynomials which are of even degree and I_O index the basis polynomial which are of odd degree. Then, we can rearrange the row and columns of R so that

$$R = \begin{pmatrix} R_E & 0 \\ 0 & R_O \end{pmatrix}, \quad R_E = (C_{r,s}(0,0))_{r,s\in I_E}, \quad R_O = (C_{r,s}(0,0))_{r,s\in I_O}.$$

So,

$$R^{-1}(1,0,\cdots,0)^\top = \begin{pmatrix} R_E^{-1} & 0 \\ 0 & R_O^{-1} \end{pmatrix}(1,0,\cdots,0)^\top = R_E^{-1}(1,0,\cdots,0)^\top.$$

Therefore, there exists, for each $r \in I_E$, $b_r \in \mathbb{R}$, such that

$$\eta_{W,Z} = \sum_{r\in I_E} b_r C_{r,0}(0,x),$$

which is an even function, and therefore, all odd derivatives of $\eta_{W,Z}$ must vanish.
$\square$

3. Pair of Spikes. In this section, we consider the problem of recovering a superposition of two spikes at positions $tZ_0 \in \mathcal{X}^2$, $m_0 = a_{0,1}\delta_{tz_{0,1}} + a_{0,2}\delta_{tz_{0,2}}$, via BLASSO minimization ($\mathcal{P}_\lambda(y)$) with $y = \Phi m_0 + w$ for small noise $w \in \mathcal{H}$. We show that for small t, provided that the limiting certificate η_{W,Z_0} is non-degenerate (see Definition 1.1), the solution of ($\mathcal{P}_\lambda(y)$) is *support stable* with respect to m_0. By support stable, we mean that *the solution is unique, has exactly* 2 *spikes and that the positions and amplitude of the recovered measure converge to* a_0 *and* Z_0 *whenever* (λ, w) *converge to 0 sufficiently fast.* The purpose of this section is to prove Theorem 1.2, which establishes support stability, with precise bounds on how fast (λ, w) should converge to 0.

A key assumption in Theorem 1.2 is that $\eta_{W,Z}$ is nondegenerate in the sense of Definition1.1. Recall that in the 1-D case, the limiting certificate is said to be non-degenerate if $\eta_W(x) < 1$ for all $x \neq 0$, and its first derivative which has not been imposed to vanish at zero is negative at zero. In the case where η_W is defined by N spikes, this is the derivative of order $2N$. In 2-D, the behaviour of $\eta_{W,Z}$ is in general non-isotropic, and in general, full derivatives are not imposed to vanish completely. Given $Z = (z_1, z_2) \in \mathcal{X}^2$, let $d_Z \overset{\text{def.}}{=} (z_1 - z_2)/\|z_1 - z_2\| \overset{\text{def.}}{=} \binom{u}{v}$. Then, the least interpolant space associated to Hermite interpolation of degree 1 at Z is spanned by the polynomial basis:

$$\{1, x, y, (ux + vy)^2, (ux + vy)^3, (ux + vy)(vx - uy)\}.$$

So, $\eta_{W,Z}$ is characterized by the interpolation conditions $\eta_{W,Z}(0) = 1$,

$$\partial_{d_Z}^j \eta_{W,Z}(0) = 0,\ j = 1,2,3, \quad \partial_{d_Z^\perp}\eta_{W,Z}(0) = 0, \quad \partial_{d_Z}\partial_{d_Z^\perp}\eta_{W,Z}(0) = 0.$$

By considering the Taylor expansion of $\eta_{W,Z}$, one sees that Definition 1.1 is very natural to enforce that $\eta_{W,Z}$ saturates only at the origin. This nondegeneracy condition is numerically verifiable and is indeed satisfied in the Gaussian convolution case, thanks to the closed form expression given in Proposition 2.18.

Sketch of proof. We begin with some preliminary bounds in Section 3.1 to make precise the manner in which η_{V,tZ_0} converges to η_{W,Z_0}. In Section 3.2, we establish a non-degeneracy transfer result, namely that certificates which are sufficiently close to η_{W,Z_0} are necessarily nondegenerate. Note that support stability is then a direct consequence of this non-degeneracy transfer, thanks to Theorem 2.11 which shows that η_{V,tZ_0} converges to η_{W,Z_0} and the main result of [31] which shows that non-degeneracy of η_{V,tZ_0} implies support stability.

The remainder of this section is then devoted to establishing precisely *how fast* $(\lambda, \|w\|_{\mathcal{H}})$ need to converge to 0 to ensure support stability. In Section 3.3.1, we construct a $\mathscr{C}^1$ mapping $g : (\lambda, w) \to (a, Z)$, and show that the associated measure $m_{a,Z}$ is indeed a solution to ($\mathcal{P}_\lambda(y)$). Similarly to the approach of [26], this is achieved via the Implicit Function Theorem. Section 3.3.2 is then devoted to analysis of the size of the region for which this function g is defined, establishing bounds on the differential of g (which eventually leads to the convergence bounds of Theorem 1.2) and Section 3.3.3 proves that the measure $m_{a,Z}$ is indeed a solution of ($\mathcal{P}_\lambda(y)$).

3.1. Preliminaries. We have already seen from Theorem 2.11 that $\eta_{V,tZ}$ converges to $\eta_{W,Z}$ as $t \to 0$, with the proof providing an analysis of the asymptotic behaviour of Γ_{tZ} (as defined in (2.9)). For the purpose of deriving precise estimates on the speed of convergence of (λ, w) for support stability, we write explicitly in this section the asymptotic behaviour of Γ_{tZ} and the relationship between $\eta_{V,tZ}$ and $\eta_{W,Z}$.

LEMMA 3.1. *Let $Z = \{(u_1, u_2), (v_1, v_2)\} \in \mathcal{X}^2$. Define*

$$H_{tZ}^* \overset{\text{def.}}{=} \begin{pmatrix} 1 & tu_2 & tu_1 & t^2u_2^2/2 & t^2u_1u_2 & t^3u_2^3/6 \\ 1 & tv_2 & tv_1 & t^2v_2^2/2 & t^2v_1v_2 & t^3v_2^3/6 \\ 0 & 0 & 1 & 0 & tu_2 & 0 \\ 0 & 0 & 1 & 0 & tv_2 & 0 \\ 0 & 1 & 0 & tu_2 & tu_1 & t^2u_2^2/2 \\ 0 & 1 & 0 & tv_2 & tv_1 & t^2v_2^2/2 \end{pmatrix}.$$

Then, for $t > 0$,

$$\Gamma_{tZ} = \Psi_{tZ} H_{tZ}$$

where $\Psi_{tZ} \overset{\text{def.}}{=} \Psi_Z + \Lambda_{tZ}$,

$$\Psi_Z^* p \overset{\text{def.}}{=} \begin{pmatrix} a_{0,0} \\ a_{0,1} \\ a_{1,0} \\ a_{0,2} - a_{2,0}(u_1 - v_1)^2/(u_2 - v_2)^2 \\ a_{1,1} + (u_1 - v_1)a_{2,0}/(u_2 - v_2) \\ a_{0,3} + a_{3,0}\frac{(u_1-v_1)^3}{(u_2-v_2)^3} + 3a_{1,2}\frac{u_1-v_1}{u_2-v_2} + 3a_{2,1}\frac{(u_1-v_1)^2}{(u_2-v_2)^2} \end{pmatrix}, \qquad a_{j,k} = \langle \partial_x^j \partial_y^k \varphi(0),\, p\rangle,$$

and $\Lambda_{tZ}^* : \mathcal{H} \to \mathbb{R}^6$ satisfies the following properties:

- $\mathrm{diag}(t^{-2}, t^{-1}, t^{-1}, t^{-1}, t^{-1}, t^{-1})\Lambda_{tZ}^* = \mathcal{O}(1)$*, in particular, $\Lambda_{tZ} = \mathcal{O}(t)$,*
- *For any $Z_0 \in \mathcal{X}^2$, $\mathrm{diag}(t^{-2}, t^{-1}, t^{-1}, t^{-1}, t^{-1}, t^{-1})(\Lambda_{tZ}^* - \Lambda_{tZ_0}^*) = \mathcal{O}(\|Z - Z_0\|)$.*

Furthermore, given $Z_0 \in \mathcal{X}^2$, we have that

- $\Psi_Z = \Psi_{Z_0} + \mathcal{O}(\|Z - Z_0\|)$
- *If Ψ_{Z_0} is full rank, then $\Gamma_{tZ}^{*,\dagger}\binom{1_2}{0_4} = \Psi_{Z_0}^{*,\dagger}\delta_6 + \mathcal{O}(\|Z - Z_0\|) + \mathcal{O}(t)$. In particular, $p_{V,tZ} = p_{W,tZ_0} + \mathcal{O}(\|Z - Z_0\|) + \mathcal{O}(t)$.*

Proof. We first recall that for $p \in \mathcal{H}$, we can define a vector $a = (a_{j,k})_{(j,k)\in\mathbb{N}_0^2} \overset{\text{def.}}{=} (\langle \partial^\alpha \varphi(0),\, p\rangle)_{\alpha \in \mathbb{N}_0^2}$ and write

$$\Gamma_{tZ}^* p = V_{tZ} a,$$

where V_{tZ} is the Hermite interpolation matrix at tZ (see2.10). Then, by performing Gaussian elimination on the matrix V_{tZ}, we obtain the following decomposition:

$$\Gamma_{tZ}^* p = H_{tZ}^* \left(\underbrace{\begin{pmatrix} a_{0,0} \\ a_{0,1} \\ a_{1,0} \\ a_{0,2} - a_{2,0}(u_1 - v_1)^2/(u_2 - v_2)^2 \\ a_{1,1} + (u_1 - v_1) a_{2,0}/(u_2 - v_2) \\ a_{0,3} + a_{3,0} \frac{(u_1 - v_1)^3}{(u_2 - v_2)^3} + 3 a_{1,2} \frac{u_1 - v_1}{u_2 - v_2} + 3 a_{2,1} \frac{(u_1 - v_1)^2}{(u_2 - v_2)^2} \end{pmatrix}}_{\Psi_Z^* p} + \underbrace{\begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \end{pmatrix}}_{\Lambda_{tZ}^* p} \right).$$

Note that

$$H_{tZ}^{*,-1} = \operatorname{diag}(1, 1/t, 1/t, 1/t^2, 1/t^2, 1/t^3) H_Z^{*,-1} \operatorname{diag}(1, 1, t, t, t, t),$$

and, denoting $a_2 \stackrel{\text{def.}}{=} u_2 - v_2$, $a_1 \stackrel{\text{def.}}{=} u_1 - v_1$

$$A \stackrel{\text{def.}}{=} v_1 v_2 u_2^2 - 2 u_1 u_2 v_2^2 + u_1 v_2^3, \quad B \stackrel{\text{def.}}{=} u_1 u_2 v_2^2 - v_1 u_2^3 - 2 v_1 v_2 u_2^2,$$
$$C \stackrel{\text{def.}}{=} u_2^2 v_1 - u_1 v_2^2 - 4 u_1 u_2 v_2 + 2 u_2 v_1 v_2, \quad D \stackrel{\text{def.}}{=} u_1 v_2^2 + u_2^2 v_1 + 2 u_1 u_2 v_2 - 4 u_2 v_1 v_2,$$

$(u_2 - v_2)^3 H_Z^{*,-1}$ is equal to

$$\begin{pmatrix} -v_2^3 + 3u_2 v_2^2 & u_2^3 - 3 v_2 u_2^2 & A & B & -u_2 v_2^2 a_2 & -u_2^2 v_2 a_2 \\ -6 u_2 v_2 & 6 u_2 v_2 & C & D & (v_2^2 + 2 u_2 v_2) a_2 & (u_2^2 + 2 v_2 u_2 v_2) a_2 \\ 0 & 0 & -v_2 a_2^2 & u_2 a_2^2 & 0 & 0 \\ 6(u_2 + v_2) & -6(u_2 + v_2) & -2a_1(2u_2 + v_2) & -2a_1(u_2 + 2v_2) & -2(u_2 + 2v_2) a_2 & -2(2u_2 + v_2) a_2 \\ 0 & 0 & a_2^2 & -a_2^2 & 0 & 0 \\ -12 & 12 & 6a_1 & 6a_1 & 6a_2 & 6a_2 \end{pmatrix}.$$

By inspection of $H_{tZ}^{*,-1} V_{tZ}$ (although this calculation can be done by hand, a Matlab script for checking this computation can be found online at https://github.com/gpeyre/2017-SIMA-super-resolution), we see that

- $h_1 = \mathcal{O}(t^2)$ and $h_j = \mathcal{O}(t)$ for all $j = 2, \ldots, 6$.
- The terms h_1/t^2 and h_j/t for $j \geqslant 2$ are uniformly bounded in t for $|t| \leqslant t_0$, and when considered as functions of u_1, u_2, v_1, v_2, they are continuous and differentiable everywhere except at $u_2 = v_2$. So,

$$\operatorname{diag}(1/t^2, 1/t, \cdots, 1/t)(\Lambda_{tZ}^* - \Lambda_{tZ_0}^*) = \mathcal{O}(\|Z - Z_0\|)$$

  provided that $Z_0 = \{(a, b), (c, d)\}$ is such that $b \neq d$. Note that the case where $b = d$ can be dealt with similarly by changing the order of Gaussian elimination.

To see that $\Psi_Z^* = \Psi_{Z_0}^* + \mathcal{O}(\|Z - Z_0\|)$, observe that when considering $\Psi_Z^* p$ as a function of Z, it is differentiable everywhere except at $u_2 = v_2$ and provided that $Z_0 = \{(a, b), (c, d)\}$ is such that $b \neq d$. Again, the case where $b = d$ can be dealt with similarly by changing the order of Gaussian elimination.

For the last claim, note that $H_{tZ}^* \delta_6 = \binom{1_2}{0_4} = \delta_6$, which implies that $H_{tZ}^{*,-1} \binom{1_2}{0_4} = \delta_6$. So, $\Gamma_{tZ}^* p = \binom{1_2}{0_4}$ if and only if $(\Psi_Z^* + \Lambda_{tZ}^*) p = \delta_6$. From

$$\Psi_Z + \Lambda_{tZ} = \Psi_{Z_0} + \mathcal{O}(\|Z - Z_0\|) + \mathcal{O}(t),$$

we see that Ψ_Z is full rank whenever Ψ_{Z_0} is full rank and provided that $\|Z - Z_0\| + t$ is sufficiently small. Therefore,

$$(\Psi_Z^* + \Lambda_{tZ}^*)^\dagger = \Psi_{Z_0}^{*,\dagger} + \mathcal{O}(\|Z - Z_0\|) + \mathcal{O}(t).$$

□

In the case of $Z_0 = \{(0,0),(0,1)\}$, the de Boor basis associated with Z_0 is $\mathcal{B}_{Z_0} = \{1, y, x, y^2, xy, y^3\}$. Moreover, in this case, by writing $a_{j,k} = \langle \partial_x^j \partial_y^k \varphi(0),\, p\rangle$ for $p \in \mathcal{H}$,

$$\Psi_{Z_0}^* p = \begin{pmatrix} a_{0,0} \\ a_{0,1} \\ a_{1,0} \\ a_{0,2} \\ a_{1,1} \\ a_{0,3} \end{pmatrix}, \qquad \Lambda_{tZ_0}^* p = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \sum_{j\geqslant 4} \frac{t^{j-2}}{j!} a_{j,0}(6-2j) \\ \sum_{j\geqslant 2} \frac{t^{j-1}}{j!} a_{1,j} \\ \sum_{j\geqslant 4} \frac{t^{j-3}}{j!} a_{0,j}(6j-12) \end{pmatrix}, \tag{3.1}$$

and

$$H_{tZ}^{*,-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ -6/t^2 & 6/t^2 & 0 & 0 & -4/t & -2/t \\ 0 & 0 & -1/t & 1/t & 0 & 0 \\ 12/t^3 & -12/t^3 & 0 & 0 & 6/t^2 & 6/t^2 \end{pmatrix}, H_Z^* = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & t & 0 & t^2/2 & 0 & t^3/6 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & t & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & t & 0 & t^2/2 \end{pmatrix}.$$

In the following, let $\Pi_{tZ} \stackrel{\text{def.}}{=} P_{(\operatorname{Im}\Gamma_{tZ})^\perp} = \mathrm{Id} - \Gamma_{tZ}\Gamma_{tZ}^\dagger$ be the orthogonal projection of $(\operatorname{Im}\Gamma_{tZ})^\perp$.

LEMMA 3.2. *Assume that $\varphi \in \mathscr{C}^2(\mathcal{X})$ and let $N \in \mathbb{N}$. Let $Z, Z' \in \mathcal{X}^N$ and let $a \in \mathbb{R}^N$. Then, there exists some constant C which depends only on φ such that*

$$\left\| \Pi_{tZ}\Gamma_{tZ'} \begin{pmatrix} a \\ 0_{Nd} \end{pmatrix} \right\| \leqslant C \|a\|_1 t^2 \|Z - Z'\|_\infty^2 .$$

Proof. Let us assume without loss of generality that $t = 1$. Write $Z = (z_j)_{j=1}^N$ and $Z' = (z_j')_{j=1}^N$. By Taylor expanding $\varphi(z_j')$ about z_j, we obtain

$$\Gamma_{Z'} \begin{pmatrix} a \\ 0_{Nd} \end{pmatrix} = \sum_{j=1}^N a_j \varphi(z_j')$$
$$= \sum_{j=1}^N a_j \left(\varphi(z_j) + \langle \nabla\varphi(z_j),\, (z_j' - z_j)\rangle + \frac{1}{2}\langle \nabla^2\varphi(z_j)(z_j' - z_j),\, (z_j' - z_j)\rangle + o(\left\|z_j - z_j'\right\|^2) \right)$$

Recall that $\operatorname{Im}(\Gamma_Z) = \operatorname{Span}\{\varphi(z), \partial_i\varphi(z)\ ;\ z \in Z,\ i = 1,2\}$. So,

$$\Pi_Z\Gamma_{Z'} \begin{pmatrix} a \\ 0_{Nd} \end{pmatrix} = \sum_{j=1}^N a_j \Pi_Z \left(\frac{1}{2}\langle \nabla^2\varphi(z_j)(z_j' - z_j),\, (z_j' - z_j)\rangle + o(\left\|z_j - z_j'\right\|^2) \right).$$

Taking norms on both sides and noting that $\|\Pi_Z\| \leqslant 1$ yields the desired result, where the constant C depends only on $\sup_{x\in\mathcal{X}} \left\|\nabla^2\varphi(x)\right\|$. □

3.2. Non-degeneracy Transfer. We have already seen that η_{V,tZ_0} converges to η_{W,Z_0} as $t \to 0$. In this section, we show in Proposition 3.3 that non-degeneracy of η_{W,Z_0} in the sense of Definition 1.1 implies that any certificate which is defined via Hermite interpolation conditions at Z and is sufficiently close to η_{V,Z_0} will also be a valid certificate saturating only at Z. Furthermore, we show in Proposition3.5 that η_{V,tZ_0} is non-degenerate and as a direct consequence of the main result of [31], the solution of $(\mathcal{P}_\lambda(y))$ is stable with respect to m_0.

PROPOSITION 3.3. *Let $Z_0 \in \mathcal{X}^2$. Suppose that $\eta_{W,Z_0}(x)$ is non-degenerate (see Definition 1.1). Then, there exists $c_1, c_2, c_3 > 0$ such that given any $\eta \in \mathscr{C}^\infty(\mathcal{X})$, $t \in (0, c_2)$ and $Z = (z_1, z_2) \in \mathcal{X}^2$ such that $\|Z - Z_0\| \leqslant c_3$ satisfying*

(i) $\eta(tz_i) = 1$, $\nabla\eta(tz_i) = 0$ for $i = 1, 2$,

(ii) $\left\|\nabla^j\eta - \nabla^j\eta_{W,Z_0}\right\|_\infty \leqslant c_1$ for $|j| \leqslant 5$,

we have that $\eta(x) < 1$ for all $x \notin tZ$.

Proof. As usual, given $Z = (z_1, z_2) \in \mathcal{X}^2$ we let $d_Z = z_1 - z_2$. For a contradiction, suppose that for all $n > 0$, there exist $\eta_n \in \mathscr{C}^\infty(\mathcal{X})$, $t_n \in (0, 1/n)$, $Z_n = (z_{n,j})_{j=1,2} \in \mathcal{X}^2$, $x_n \notin t_n Z_n$ such that

(a) $\|Z_n - Z_0\| \leqslant 1/n$ and $\left\|\nabla^j\eta_n - \nabla^j\eta_{W,Z_0}\right\|_\infty \leqslant 1/n$ for all $|j| \leqslant 5$,

(b) for $j = 1, 2$, $\eta_n(t_n z_{n,j}) = 1$, $\nabla\eta_n(t_n z_{n,j}) = 0$,

(c) $\eta_n(x_n) = 1$.

Note that since $\eta_{W,Z_0}(x) < 1$ for all $x \neq 0$, we must have that $x_n \to 0$ as $n \to \infty$. Let $x_n = (f(t_n)u_n, g(t_n)v_n)$, where $\lim_{t\to 0} f(t) = 0$ and $\lim_{t\to 0} g(t) = 0$.

We first fix n and derive some equations satisfied by η_n and its derivatives. Without loss of generality, assume that $d_{Z_n} = (1, 0)$ (otherwise, we simply consider derivatives with respect to directions d_{Z_n} and $d_{Z_n}^\perp$ instead of the canonical directions). So, $Z_n = \{(\alpha_1, \beta), (\alpha_2, \beta)$ for some $\alpha_1, \alpha_2, \beta \in \mathbb{R}$. Let $c_n \stackrel{\text{def.}}{=} (\alpha_1 - \alpha_2)/2$ and $q_n \stackrel{\text{def.}}{=} ((\alpha_1 + \alpha_2)/2, \beta) \in \mathbb{R}^2$, such that $c_n(1, 0) + q_n = z_{n,1}$ and $c_n(-1, 0) + q_n = z_{n,2}$. Since $Z_n \to Z_0$, we must have $c_n \in [\underline{c}, \bar{c}]$ and $\|q_n\| \leqslant \bar{q}$ for some $\underline{c}, \bar{c}, \bar{q} > 0$. Let $\tilde{\eta}_n \stackrel{\text{def.}}{=} \eta_n(c_n \cdot + t_n q_n)$. Then, $\tilde{\eta}_n((t_n, 0)) = \tilde{\eta}_n((-t_n, 0)) = 1$, $\nabla\tilde{\eta}_n((t_n, 0)) = \nabla\tilde{\eta}((-t_n, 0)) = 0$, and $\tilde{\eta}_n(\tilde{x}_n) = 1$ for $\tilde{x}_n \stackrel{\text{def.}}{=} c_n^{-1}(\tilde{x}_n - t_n q_n)$. Note that $\tilde{x}_n \notin \{(t_n, 0), (-t_n, 0)\}$.

To simplify notation, let us drop the subscript n and write $\tilde{\eta}$, t, $\tilde{x}$ for $\tilde{\eta}_n$, t_n and $\tilde{x}_n$ in the following. By expanding $\tilde{\eta}$ about 0, we obtain

$$\begin{aligned}\eta(X) = \sum_{|\alpha|\leqslant 3} b_\alpha X^\alpha + R_{0,4}(X)X_2^4 + R_{1,3}(X)X_1X_2^3 + R_{2,2}(X)X_1^2X_2^2 + R_{3,1}(X)X_1^3X_2 \\ + b_{4,0}X_1^4 + R_{4,1}(X)X_1^4X_2 + R_{5,0}(X)X_1^5,\end{aligned}$$

where given $\alpha \in \mathbb{N}_0^2$,

$$b_\alpha \stackrel{\text{def.}}{=} \frac{\partial^\alpha\tilde{\eta}(0)}{\alpha!}, \qquad R_\alpha(X) \stackrel{\text{def.}}{=} \frac{|\alpha|}{\alpha!}\int_0^1 (1-s)^{|\alpha|-1}\partial^\alpha\tilde{\eta}(sX)\mathrm{d}s.$$

To simplify notation, in the following, we write $b_\alpha = R_\alpha$ and note that thanks to assumption (ii), each of these terms is uniformly bounded in n. Let

$$\begin{aligned}&\iota_0 \stackrel{\text{def.}}{=} \tilde{\eta}(\tilde{x}),\\ &\iota_1 \stackrel{\text{def.}}{=} \tilde{\eta}((t,0)), \qquad \partial_{X_1}\iota_1 \stackrel{\text{def.}}{=} \partial_{X_1}\tilde{\eta}((t,0)), \quad \partial_{X_2}\iota_1 \stackrel{\text{def.}}{=} \partial_{X_2}\tilde{\eta}((t,0)),\\ &\iota_2 \stackrel{\text{def.}}{=} \eta((-t,0)), \qquad \partial_{X_1}\iota_2 \stackrel{\text{def.}}{=} \partial_{X_1}\tilde{\eta}((-t,0)), \quad \partial_{X_2}\iota_2 \stackrel{\text{def.}}{=} \partial_{X_2}\tilde{\eta}((-t,0)).\end{aligned}$$

Then,

$$\begin{aligned}
\gamma_1 &\stackrel{\text{def.}}{=} \frac{\iota_1+\iota_2}{2} = b_{0,0} + t^2 b_{2,0} + t^4 b_{4,0} = 1,\\
\gamma_2 &\stackrel{\text{def.}}{=} \frac{\iota_1-\iota_2}{2t} = b_{1,0} + t^2 b_{3,0} + t^4 b_{5,0} = 0,\\
\gamma_3 &\stackrel{\text{def.}}{=} \frac{\partial_{X_2}\iota_1+\partial_{X_2}\iota_2}{2} = b_{0,1} + t^2 b_{2,1} + t^4 b_{4,1} = 0,\\
\gamma_4 &\stackrel{\text{def.}}{=} \frac{\partial_{X_2}\iota_1-\partial_{X_2}\iota_2}{2t} = b_{1,1} + t^2 b_{3,1} = 0,\\
\gamma_5 &\stackrel{\text{def.}}{=} \frac{\partial_{X_1}\iota_1-\partial_{X_1}\iota_2}{4t} = b_{2,0} + 2t^2 b_{4,0} = 0,\\
\gamma_6 &\stackrel{\text{def.}}{=} \frac{1}{2t^2}\left(\frac{\partial_{X_1}\iota_1+\partial_{X_1}\iota_2}{2} - \gamma_2\right) = b_{3,0} + 2t^2 b_{5,0} = 0.
\end{aligned}$$

By subtracting appropriate multiples of $\{\gamma_j\}_{j=1}^6$ from ι_0, we obtain

$$0 = \iota_0 - \gamma_1 - \tilde{x}_1\gamma_2 - \tilde{x}_2\gamma_3 - \tilde{x}_1\tilde{x}_2\gamma_4 - (\tilde{x}_1^2 - t^2)\gamma_5 - (\tilde{x}_1^3 - t^2\tilde{x}_1)\gamma_6,$$

and so,

$$0 = p_{t,1}^2(b_{0,2} + A_t) + p_{t,1}p_{t,2}(b_{2,1} + B_t) + p_{t,2}^2(b_{4,0} + C_t) \tag{3.2}$$

where

$$\begin{aligned}
&p_{t,1} \stackrel{\text{def.}}{=} \tilde{x}_2, \qquad p_{t,2} \stackrel{\text{def.}}{=} (\tilde{x}_1^2 - t^2)\\
&A_t \stackrel{\text{def.}}{=} \left(2b_{2,2}\tilde{x}_1^2 + b_{1,3}\tilde{x}_1\tilde{x}_2 + b_{0,4}\tilde{x}_2^2 + b_{0,3}\tilde{x}_2 + b_{1,2}\tilde{x}_1\right)\\
&B_t \stackrel{\text{def.}}{=} (\tilde{x}_1^2 + t^2)b_{4,1} + \tilde{x}_1 b_{3,1}\\
&C_t \stackrel{\text{def.}}{=} b_{5,0}\tilde{x}_1.
\end{aligned}$$

Note that $\lim_{t\to 0} A_t = \lim_{t\to 0} B_t = \lim_{t\to 0} C_t = 0$ and $p_{t,2} \neq 0$. Writing (3.2) in matrix form, we have

$$\langle \begin{pmatrix} (b_{0,2} + A_t) & \frac{1}{2}(b_{2,1} + B_t) \\ \frac{1}{2}(b_{2,1} + B_t) & (b_{4,0} + C_t) \end{pmatrix} \begin{pmatrix} p_{t,1} \\ p_{t,2} \end{pmatrix}, \begin{pmatrix} p_{t,1} \\ p_{t,2} \end{pmatrix} \rangle = 0 \tag{3.3}$$

Note that $\partial_{d_{Z_n}}^j \partial_{d_{Z_n}^\perp}^k \tilde{\eta}_n(0) = c_n^{j+k} \partial_{d_{Z_n}}^j \partial_{d_{Z_n}^\perp}^k \eta_n(0)$, c_n is uniformly bounded and since $Z_n \to Z_0$, we have $\lim_{n\to\infty} \partial_{d_{Z_n}}^j \partial_{d_{Z_n}^\perp}^k \eta_n(0) = \partial_{d_{Z_0}}^j \partial_{d_{Z_0}^\perp}^k \eta_{W,Z_0}(0)$. If we can extract a subsequence such that $\tilde{x}_{n,2} = 0$, then by dividing (3.3) by $\tilde{x}_{n,1}^2$ and taking the limit, we have that $\partial_{d_{Z_0}}^4 \eta_{W,Z_0}(0) = 0$ which is a contradition to nondegeneracy of η_{W,Z_0}. So, assume that $\tilde{x}_{n,2} \neq 0$ for some sequence of (t_n). We must have either $\liminf_{t\to 0} p_{t,1}/p_{t,2} \stackrel{\text{def.}}{=} p_{*,1} < \infty$ or $\liminf_{t\to 0} p_{t,2}/p_{t,1} \stackrel{\text{def.}}{=} p_{*,2} < \infty$. Therefore, extracting a subsequence such that $c_n \to c^*$, and by dividing (3.3) either by $p_{t,1}^2$ or $p_{t,2}^2$ and letting $t \to 0$, we have either

$$\langle \begin{pmatrix} \partial_{d_{Z_0}^\perp}^2 \eta_{W,Z_0}(0) & \frac{1}{2}\partial_{d_{Z_0}}^2 \partial_{d_{Z_0}^\perp} \eta_{W,Z_0}(0) \\ \frac{1}{2}\partial_{d_{Z_0}}^2 \partial_{d_{Z_0}^\perp} \eta_{W,Z_0}(0) & \frac{1}{12}\partial_{d_{Z_0}}^4 \eta_{W,Z_0}(0) \end{pmatrix} q_*, q_* \rangle = 0$$

for $q_* = (c_* p_{*,1}, c_*^2)$ or $q_* = (c_*, c_*^2 p_{*,2})$, which is a contradiction to the assumption that η_{W,Z_0} is non-degenerate. ◻

Remark 3.4. *From Proposition 3.3, it follows that η_{V,tZ_0} is a valid certificate for all t sufficiently small. The following result shows that η_{V,tZ_0} is in fact non-degenerate, and therefore, as a direct consequence of the main result of [31], the solution of ($\mathcal{P}_\lambda(y)$) is support stable with respect to m_0.*

In the following, for $a \in \mathbb{R}^2$ and $Z \in \mathcal{X}^2$ and $d_Z = (z_2 - z_1)/\left\|z_2 - z_1\right\|$, let $(\partial_{d_Z}^2 \Phi_{tZ})a \stackrel{\text{def.}}{=} \sum_{j=1}^2 a_j \partial_{d_Z}^2 \varphi(tz_j)$. The operator $\partial_{d_Z} \partial_{d_Z^\perp} \Phi_{tZ}$ is defined similarly.

Proposition 3.5. *Let $Z = (z_1, z_2) \in \mathcal{X}^2$. Then,*

$$\begin{aligned}
\partial_{d_Z}^2 \Phi_{tZ}^* p_{V,tZ} &= \frac{t^2}{12} \partial_{d_Z}^4 \eta_{W,Z}(0) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \mathcal{O}(t^3), \\
\partial_{d_Z^\perp} \partial_{d_Z} \Phi_{tZ}^* p_{V,tZ} &= \frac{t}{2} \partial_{d_Z}^2 \partial_{d_Z^\perp} \eta_{W,Z}(0) \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \mathcal{O}(t^2), \\
\partial_{d_Z^\perp}^2 \Phi_{tZ}^* p_{V,tZ} &= \partial_{d_Z^\perp}^2 \eta_{W,Z}(0) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \mathcal{O}(t).
\end{aligned}$$

Therefore, provided that $\eta_{W,Z}$ is non-degenerate, then for all t sufficiently small, $\nabla^2 \eta_{V,tZ}(tz) \prec 0$ for all $z \in Z$.

Proof. Without loss of generality, let $Z = \{(0,0),(0,1)\}$. We shall show that

$$\begin{aligned}
\partial_y^2 \Phi_{tZ}^* p_{V,tZ} &= \frac{t^2}{12} \partial_y^4 \eta_{W,Z}(0) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \mathcal{O}(t^3), \\
\partial_x \partial_y \Phi_{tZ}^* p_{V,tZ} &= \frac{t}{2} \partial_y^2 \partial_x \eta_{W,Z}(0) \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \mathcal{O}(t^2), \\
\partial_x^2 \Phi_{tZ}^* p_{V,tZ} &= \partial_x^2 \eta_{W,Z}(0) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \mathcal{O}(t).
\end{aligned}$$

Let $a \in \mathbb{R}^2$. By Taylor expanding about 0, we obtain

$$\begin{aligned}
\partial_y^2 \Phi_{tZ} a &= \Psi_Z \tilde{V}_{tZ}^{(1)} a + a_2 \sum_{j=2}^\infty \frac{t^j}{j!} \partial_y^{j+2} \varphi(0), \\
\partial_x \partial_y \Phi_{tZ} a &= \Psi_Z \tilde{V}_{tZ}^{(2)} a + a_2 \sum_{j \geqslant 1} \frac{t^j}{j!} \partial_y^{j+1} \partial_x \varphi(0),
\end{aligned}$$

where

$$\tilde{V}_{tZ}^{(1)} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 1 \\ 0 & 0 \\ 0 & t \end{pmatrix} \quad \text{and} \quad \tilde{V}_{tZ}^{(2)} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 1 \\ 0 & 0 \end{pmatrix},$$

and

$$\partial_x^2 \Phi_{tZ} a = a_1 \partial_x^2 \varphi(0) + a_2 \left(\sum_{j \geqslant 0} \frac{t^j}{j!} \partial_y^j \partial_x^2 \varphi(0) \right).$$

We first consider $\partial_y^2 \Phi_{tZ}^* p_{V,tZ}$. Recall the definition of Λ_{tZ} from (3.1) and observe that

$$\tilde{V}_{tZ}^{(1),*}(\Psi_Z^* + \Lambda_{tZ}^*)\Gamma_{tZ}^{*,\dagger}\begin{pmatrix}1_2\\0_4\end{pmatrix} = \tilde{V}_{tZ}^{(1),*}H_{tZ}^{*,-1}\Gamma_{tZ}^*\Gamma_{tZ}^{*,\dagger}\begin{pmatrix}1_2\\0_4\end{pmatrix} = \tilde{V}_{tZ}^{(1),*}\delta_6 = 0. \tag{3.4}$$

Moreover, using the fact that $p_{V,tZ} = p_{W,Z} + \mathcal{O}(t)$,

$$\begin{pmatrix}0\\ \sum_{j=2}^\infty \frac{t^j}{j!}\partial_y^{j+2}\varphi(0)\end{pmatrix} p_{V,tZ} = \begin{pmatrix}0\\ \frac{t^2}{2}\langle p_{V,tZ},\, \partial_y^4\varphi(0)\rangle + \mathcal{O}(t^3)\end{pmatrix} = \begin{pmatrix}0\\ \frac{t^2}{2}\partial_y^4\eta_{W,Z}(0) + \mathcal{O}(t^3)\end{pmatrix}. \tag{3.5}$$

Also,

$$\begin{aligned}
-\tilde{V}_{tZ}^{(1),*}\Lambda_{tZ}^* p_{V,tZ} &= -\tilde{V}_{tZ}^{(1,*)}\begin{pmatrix}0\\0\\0\\ \frac{-t^2}{12}\langle\partial_y^4\varphi(0),\, p_{V,tZ}\rangle + \mathcal{O}(t^3)\\ \frac{t}{2}\langle\partial_x\partial_y^2\varphi(0),\, p_{V,tZ}\rangle + \mathcal{O}(t^2)\\ \frac{t}{2}\langle\partial_y^4\varphi(0),\, p_{V,tZ}\rangle + \mathcal{O}(t^2)\end{pmatrix}\\
&= -\tilde{V}_{tZ}^{(1),*}\begin{pmatrix}0\\0\\0\\ \frac{-t^2}{12}\partial_y^4\eta_{W,Z}(0) + \mathcal{O}(t^3)\\ t\partial_y^2\partial_x\eta_{W,Z}(0) + \mathcal{O}(t^2)\\ \frac{t}{2}\partial_y^4\eta_{W,Z}(0) + \mathcal{O}(t^2)\end{pmatrix} = \begin{pmatrix}\frac{t^2}{12}\partial_y^4\eta_{W,Z}(0)\\ -\frac{5t^2}{12}\partial_y^4\eta_{W,Z}(0)\end{pmatrix} + \mathcal{O}(t^3).
\end{aligned} \tag{3.6}$$

Summing (3.4), (3.5) and (3.6) gives the required bound on $\partial_y^2\Phi_{tZ}^* p_{V,tZ}$.

For the second bound,

$$\begin{aligned}
\partial_x\partial_y\Phi_{tZ} &= \Psi_Z\tilde{V}_{tZ}^{(2)} + \begin{pmatrix}0\\ \sum_{j\geqslant 1}\frac{t^j}{j!}\partial_y^{j+1}\partial_x\varphi(0)\end{pmatrix}\\
&= (\Psi_z + \Lambda_{tZ})\tilde{V}_{tZ}^{(2)} - \Lambda_{tZ}\tilde{V}_{tZ}^{(2)} + \begin{pmatrix}0\\ \sum_{j\geqslant 1}\frac{t^j}{j!}\partial_y^{j+1}\partial_x\varphi(0)\end{pmatrix}.
\end{aligned}$$

As before, $\tilde{V}_{tZ}^{(2),*}(\Psi_z + \Lambda_{tZ})^* p_{V,tZ} = 0$,

$$\langle\sum_{j\geqslant 1}\frac{t^j}{j!}\partial_y^{j+1}\partial_x\varphi(0),\, p_{V,tZ}\rangle = t\partial_y^2\partial_x\eta_{W,Z}(0) + \mathcal{O}(t^2),$$

and

$$\tilde{V}_{tZ}^{(2),*}\Lambda_{tZ}^* p_{V,tZ} = \frac{t}{2}\partial_y^2\partial_x\eta_{W,Z}(0)\begin{pmatrix}1\\1\end{pmatrix} + \mathcal{O}(t^2).$$

So, $\partial_y\partial_x\Phi_{tZ}^* p_{V,tZ} = \frac{t}{2}\partial_y^2\partial_x\eta_{W,Z}(0)\binom{-1}{1} + \mathcal{O}(t^2)$.

The proof of the last bound follows because

$$\partial_x^2\Phi_{tZ}^* p_{V,tZ} = \partial_x^2\Phi_{tZ}^*(p_W + \mathcal{O}(t)) = \partial_x^2\eta_{W,Z}(0)\begin{pmatrix}1\\1\end{pmatrix} + \mathcal{O}(t).$$

Finally, note that for each $z \in Z$,

$$\nabla^2 \eta_{V,tZ}(tz) = \begin{pmatrix} \partial_{d_Z}^2 \eta_{V,tZ}(tz) & \partial_{d_Z}\partial_{d_Z^\perp}\eta_{V,tZ}(tz) \\ \partial_{d_Z}\partial_{d_Z^\perp}\eta_{V,tZ}(tz) & \partial_{d_Z^\perp}^2 \eta_{V,tZ}(tz) \end{pmatrix} \prec 0$$

if and only if

$$\begin{aligned} &\begin{pmatrix} \frac{1}{t^2}\partial_{d_Z}^2 \eta_{V,tZ}(tz) & \frac{1}{t}\partial_{d_Z}\partial_{d_Z^\perp}\eta_{V,tZ}(tz) \\ \frac{1}{t}\partial_{d_Z}\partial_{d_Z^\perp}\eta_{V,tZ}(tz) & \partial_{d_Z^\perp}^2 \eta_{V,tZ}(tz) \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{t} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \partial_{d_Z}^2 \eta_{V,tZ}(tz) & \partial_{d_Z}\partial_{d_Z^\perp}\eta_{V,tZ}(tz) \\ \partial_{d_Z}\partial_{d_Z^\perp}\eta_{V,tZ}(tz) & \partial_{d_Z^\perp}^2 \eta_{V,tZ}(tz) \end{pmatrix} \begin{pmatrix} \frac{1}{t} & 0 \\ 0 & 1 \end{pmatrix} \prec 0, \end{aligned}$$

which is true for all t sufficiently small if $\eta_{W,Z}$ is nondegenerate. $\square$

3.3. Proof of Theorem 1.2. First note that if Theorem1.2 is true for $Z_0 \stackrel{\text{def.}}{=} Z$ for some fixed $Z \in \mathcal{X}^2$, then given any $c \in \mathcal{X}$, the result is also true for $Z_0 \stackrel{\text{def.}}{=} c + Z$. Let $\eta^{\Phi,y,\lambda}$ be the solution to the dual formulation of ($\mathcal{P}_\lambda(y)$), and let $\eta_V^{\Psi,y}$ be the associated precertificate. Then, by letting $T : \mathcal{X} \to \mathcal{X}$, $z \mapsto z + tc$ and thanks to the reparametrization observations of AppendixA, we have

$$\eta_V^{\Phi,\Phi m_{a,t(Z_0+c)}} = \eta_V^{\Phi\circ T_\sharp,\Phi\circ T_\sharp m_{a,tZ_0}}(\cdot - tc).$$

Therefore $\eta^{\Phi}_{W,Z_0+c} = \eta^{\Phi\circ T_\sharp}_{W,Z_0}$. So, η^{Φ}_{W,Z_0+c} is non-degenerate if and only if $\eta^{\Phi\circ T_\sharp}_{W,Z_0}$ is non-degenerate. Moreover, since (provided that λ and w satisfy the conditions of Theorem1.2) non-degeneracy of $\eta^{\Phi\circ T_\sharp}_{W,Z_0}$ implies that $\eta^{\Phi\circ T_\sharp,\Phi\circ T_\sharp m_{a_0,tZ_0}+w,\lambda}$ saturates only at Z and $m_{a,Z}$ is the unique solution with a and Z satisfying (1.5), we know that $\eta^{\Phi,\tilde{y},\lambda}$ with $\tilde{y} = \Phi m_{a_0,t(Z_0+c)} + w$, saturates only at $Z + tc$ and $m_{a,Z+tc}$ is the unique solution. Therefore, without loss of generality, it suffices to prove Theorem1.2 for $Z_0 = \{(0,0),(a,b)\}$ for some $(a,b) \in \mathcal{X}$. Furthermore, we simply consider $Z_0 = \{(0,0),(0,1)\}$, since otherwise, in the following, we can simply consider derivatives with respect to d_{Z_0} and $d_{Z_0}^\perp$ instead of the canonical directions.

3.3.1. Implicit Function Theorem. From the first order optimality conditions of ($\mathcal{P}_\lambda(y)$), we have that $m_{a,Z}$ solves ($\mathcal{P}_\lambda(y)$) if and only if

$$\frac{\Phi^*(\Phi m_{a_0,Z_0} + w - \Phi m_{a,Z})}{\lambda} \in \partial\left|\cdot\right|(m_{a,Z}). \tag{3.7}$$

Therefore, we aim to construct a $\mathscr{C}^1$ mapping $g : (\lambda, w) \in \mathbb{R} \times \mathcal{H} \mapsto (a, Z) \in \mathbb{R}^2 \times \mathcal{X}^2$ such that (a,Z) satisfies (3.7). Furthermore, bounds on the derivatives of g will provide conditions on the required speed at which (λ, w) converge to 0. To this end, following [31] and [26], let $u = (a, Z)$ and $v = (\lambda, w)$, and define

$$f_t(u,v) \stackrel{\text{def.}}{=} \Gamma_{tZ}^*(\Phi_{tZ}a - \Phi_{tZ_0}a_0 - w) + \lambda \begin{pmatrix} 1_N \\ 0_{2N} \end{pmatrix}. \tag{3.8}$$

To construct candidate solutions to ($\mathcal{P}_\lambda(y)$), we search for parameters u and v for which $f_t(u,v) = 0$. This is achieved using the implicit function theorem as we show next.

For $a \in \mathbb{R}^2$ and $\alpha \in \mathbb{N}_0^2$, let $(\partial^\alpha \Phi_{tZ})a \stackrel{\text{def.}}{=} \sum_{j=1}^2 a_j \partial^\alpha \varphi(tz_j)$. Then, the derivatives of f_t are

$$\begin{aligned} \partial_u f_t(u,v) &= \Gamma_{tZ}^* \Gamma_{tZ} J_{ta} + t \begin{pmatrix} 0 & \operatorname{diag}(\partial_x \Phi_{tZ}^* A) & \operatorname{diag}(\partial_y \Phi_{tZ}^* A) \\ 0 & \operatorname{diag}(\partial_x^2 \Phi_{tZ}^* A) & \operatorname{diag}(\partial_y \partial_x \Phi_{tZ}^* A) \\ 0 & \operatorname{diag}(\partial_y \partial_x \Phi_{tZ}^* A) & \operatorname{diag}(\partial_y^2 \Phi_{tZ}^* A) \end{pmatrix} \\ \partial_v f_t(u,v) &= \begin{pmatrix} \binom{1_N}{0_{2N}} & -\Gamma_{tZ}^* \end{pmatrix} \end{aligned} \tag{3.9}$$

where

$$J_{ta} \stackrel{\text{def.}}{=} \begin{pmatrix} \mathrm{Id}_N & 0 & 0 \\ 0 & t \operatorname{diag}(a) & 0 \\ 0 & 0 & t \operatorname{diag}(a) \end{pmatrix}, \qquad A \stackrel{\text{def.}}{=} \Phi_{tZ} a - \Phi_{tZ_0} a_0 - w.$$

So, f_t is a continously diffferentiable function, $\partial_u f_t(u_0, 0) = \Gamma_{tZ_0}^* \Gamma_{tZ_0} J_{ta_0}$ is invertible by Proposition 3.1 (since Ψ_{Z_0} is assumed to be full rank) and $f_t(u_0, 0) = 0$. Therefore, we may apply the Implicit Function Theorem to deduce that there exists a neighbourhood V_t of 0 in $\mathbb{R} \times \mathcal{H}$, a neighbourhood U_t of u_0 in $\mathbb{R}^2 \times \mathcal{X}^2$ and a $\mathscr{C}^1$ function $g_t : V_t \to U_t$ such that for all $(u,v) \in U_t \times V_t$, $f_t(u,v) = 0$ if and only if $u = g_t(v)$. Furthermore, the derivative of g_t is

$$\mathrm{d}g_t(v) = -\left(\partial_u f_t(g_t(v), v)\right)^{-1} \partial_v f_t(g_t(v), v). \tag{3.10}$$

So, to prove Theorem 1.2, given (λ, w), for $(a, Z) = g((\lambda, w))$, we simply need to establish the following two facts.

1. g_t is well defined on a region V_t which contains a ball of radius on the order of t^4.
2. $m_{a,Z}$ is a solution of $(\mathcal{P}_\lambda(y))$ with $y = \Phi(m_{a_0, tZ_0}) + w$, i.e. it satisfies (3.7). To this end, we define the associated certificate as

$$p_{\lambda,t} \stackrel{\text{def.}}{=} \frac{(\Phi_Z a - \Phi_{tZ_0} a_0 - w)}{\lambda}, \qquad \eta_{\lambda,t} \stackrel{\text{def.}}{=} \Phi^* p_{\lambda,t} \tag{3.11}$$

   and show (Proposition 3.9) that $p_{\lambda,t}$ converges to p_{W,Z_0} as $(\lambda, w) \to 0$ and $t \to 0$. Therefore, by Theorem 3.3, $p_{\lambda,t} \in \partial\left|\cdot\right|(m_{a,Z})$.

We remark that although the key steps of this proof are the same as the 1-D proof presented in [26], the technical details differ due to the anisotropy of the limiting certificate η_{W,Z_0} and since some of the proofs in [26] rely on purely 1-D tools.

3.3.2. Bounds on V_t. For $r > 0$ and $a_0 \in \mathbb{R}^2$, let

$$B(a_0, r) \stackrel{\text{def.}}{=} \left\{ a \in \mathbb{R}^2 \; ; \; \|a_0 - a\| \leqslant r \right\}.$$

and given $Z_0 \in \mathcal{X}^2$,

$$B(Z_0, r) \stackrel{\text{def.}}{=} \left\{ Z \in \mathcal{X}^2 \; ; \; \|Z_0 - Z\| \leqslant r \right\}.$$

To show that we can construct a function g_t^* which is defined on a ball of radius t^4, let V_t^* be defined as follows: $V_t^* \stackrel{\text{def.}}{=} \bigcup_{v \in \mathcal{V}} V$, where $\mathcal{V}$ is the collections of all open sets $V \subset \mathbb{R} \times \mathcal{H}$ such that

- $0 \in V$,
- V is star-shaped with respect to 0,

- $V \subset B(0, c_* t^4)$ where $c_* > 0$ is the constant defined in Lemma 3.6,
- there exists a $\mathscr{C}^1$ function $g : V \to \mathbb{R}^2 \times \mathcal{X}^2$ such that $g(0) = u_0$, $f_t(g(v), v) = 0$ for all $v \in V$,
- $g(V) \subset B(a_0, c_*) \times B(Z_0, tc_*)$.

Note that the definition of this set V_t^* is the same as in [26], except for the last condition, where we require that $\|Z - Z_0\| \leqslant c_* t$ for all Z such that $(a, Z) \in g(V)$. This is natural, since we eventually require that the distance between $p_{\lambda,t}$ and p_{V,Z_0} is $\mathcal{O}(t)$. As explained in [26, Section 4.3], this set V_t^* is well defined and non-empty. We may therefore define a function $g_t^* : V_t^* \to \mathbb{R}^2 \times \mathcal{X}^2$ where

$$g_t^*(v) \stackrel{\text{def.}}{=} g(v), \quad \text{if } v \in V,\ V \in \mathcal{V}, \text{ and } g \text{ is the corresponding function.}$$

Indeed, given $V, V' \in \mathcal{V}$, $0 \in V \cap V'$ is nonempty and is an open set. Let g and g' be the functions associated to V and V' respectively. Note that by the implicit function theorem, $g = g'$ on a small neighbourhood around 0. Furthermore, since $V \cap V'$ is a connected set, we may apply the following lemma to deduce that $g = g'$ on the entire set $V \cap V'$.

Lemma 3.6. *For $\lambda > 0$, $w \in \mathcal{H}$, $Z \in \mathcal{X}^2$, $t > 0$ and $a \in \mathbb{R}^2$, let*

$$G_{tZ}(\lambda, w) \stackrel{\text{def.}}{=} \Psi_{tZ}^* \Psi_{tZ} + t H_{tZ}^{*,-1} F_{tZ} J_{ta}^{-1} H_{tZ}^{-1}$$

where

$$F_{tZ} \stackrel{\text{def.}}{=} -\begin{pmatrix} 0 & 0 & 0 \\ 0 & \operatorname{diag}(\partial_x^2 \Phi_{tZ}^* q_{tZ}) & \operatorname{diag}(\partial_y \partial_x \Phi_{tZ}^* q_{tZ}) \\ 0 & \operatorname{diag}(\partial_y \partial_x \Phi_{tZ}^* q_{tZ}) & \operatorname{diag}(\partial_y^2 \Phi_{tZ}^* q_{tZ}) \end{pmatrix}$$

and

$$q_{tZ} \stackrel{\text{def.}}{=} \lambda \Gamma_{tZ}^{*,\dagger} \begin{pmatrix} 1_N \\ 0_{2N} \end{pmatrix} + \Pi_{tZ} w + \Pi_{tZ} \Gamma_{tZ_0} \begin{pmatrix} a_0 \\ 0 \end{pmatrix}.$$

The following hold:

- *There exists $c_* > 0$ such that for all $Z \in \mathcal{X}^2$ and $\lambda \in \mathbb{R}$ and $w \in \mathcal{H}$ with $\|Z - Z_0\|_\infty \leqslant c_* t$, $\lambda \leqslant c_* t^4$ and $\|w\| \leqslant c_* t^4$, $G_{tZ}(\lambda, w)$ is invertible and has inverse bounded by $3 \left\|(\Psi_{Z_0}^* \Psi_{Z_0})^{-1}\right\|$.*
- *Recall the definition of f_t from (3.8). If for $u = (a, Z)$ and $v = (\lambda, w)$, $f_t(u, v) = 0$, then $\partial_u f_t(u, v) = H_{tZ}^* G_{tZ}(\lambda, w) H_{tZ} J_{ta}$.*

Proof. First observe that since for $|\alpha| = 2$, $\partial^\alpha \Phi^* \Gamma_{tZ}^{*,\dagger} \binom{1_N}{0_{2N}}$ converges uniformly to $\partial^\alpha \Phi^* p_{W,Z}$ as $t \to 0$ and $\|Z - Z_0\|_\infty \leqslant c_0 t$, we have that $\left|\partial^\alpha \Phi_{tZ}^* \Gamma_{tZ}^{*,\dagger} \binom{1_N}{0_{2N}}\right|$ is uniformly bounded. Therefore, from Proposition 3.2, for $|\alpha| = 2$,

$$|\partial^\alpha \Phi_{tZ}^* q_{tZ}|_\infty \lesssim \lambda + \|w\| + t^2 |Z - Z_0|_\infty^2 .$$

Therefore, $\left\|F_{tZ}^{-1}\right\| \leqslant C c_0 t^4$ for some constant C which depends only on φ.

Recalling the definition of H_{tZ} and J_{ta}, we have that

$$\begin{aligned} \left\| t H_{tZ}^{*,-1} F_{tZ} J_{ta}^{-1} H_{tZ}^{-1} \right\| &= \left\| t^2 \Delta_t H_Z^{*,-1} F_{tZ} J_a^{-1} H_Z^{-1} \Delta_t \right\| \\ &\leqslant \frac{\left\|H_Z^{-1}\right\|^2 \left\|J_a^{-1}\right\| \left\|F_{tZ}\right\|}{t^4} \leqslant C \left\|H_Z^{-1}\right\|^2 \left\|J_a^{-1}\right\| c_0, \end{aligned}$$

where we denoted $\Delta_t \overset{\text{def.}}{=} \operatorname{diag}(1, \frac{1}{t}, \frac{1}{t}, \frac{1}{t^2}, \frac{1}{t^2}, \frac{1}{t^3})$. Therefore, by the above bound and by Lemma 3.1, we have that

$$\left\|G_{tZ} - \Psi_{Z_0}^* \Psi_{Z_0}\right\| \leqslant C' \left|Z - Z_0\right|_\infty + C \left\|H_Z^{-1}\right\|^2 \left\|J_a^{-1}\right\| c_0,$$

where C' depends only on φ and the required result follows by choosing c_0 to be sufficiently small.

From (3.9) and since $\Gamma_{tZ} = \Psi_{tZ} H_{tZ}$ by Lemma3.1, if $f_t(u,v) = 0$, then $\partial_u f_t(u,v)$ is equal to

$$H_{tZ}^* \left(\Psi_{tZ} \Psi_{tZ} + t H_{tZ}^{*,-1} g \begin{pmatrix} 0 & 0 & 0 \\ 0 & \operatorname{diag}(\partial_x^2 \Phi_{tZ}^* A) & \operatorname{diag}(\partial_y \partial_x \Phi_{tZ}^* A) \\ 0 & \operatorname{diag}(\partial_y \partial_x \Phi_{tZ}^* A) & \operatorname{diag}(\partial_y^2 \Phi_{tZ}^* A) \end{pmatrix} J_{ta}^{-1} H_{tZ}^{-1} \right) H_{tZ} J_{ta}$$

where $A = \Phi_{tZ} a - \Phi_{tZ_0} a_0 - w$. Therefore, it is enough to show that $A = -q_{tZ}$.

Since $f_t(u,v) = 0$,

$$\Gamma_{tZ}^* (\Phi_{tZ} a - \Phi_{tZ_0} a_0 - w_n) + \lambda \begin{pmatrix} 1_N \\ 0_{2N} \end{pmatrix} = 0. \tag{3.12}$$

We can rewrite (3.12) as

$$-\Gamma_{t_n Z_n}^* \Gamma_{t_n Z_n} \begin{pmatrix} a_n \\ 0_{2N} \end{pmatrix} = -\Gamma_{t_n Z_n}^* \Gamma_{tZ_0} \begin{pmatrix} a_0 \\ 0_{2N} \end{pmatrix} - \Gamma_{t_n Z_n}^* w_n + \lambda_n \begin{pmatrix} 1_N \\ 0_{2N} \end{pmatrix}. \tag{3.13}$$

By applying $\Gamma_{tZ} (\Gamma_{tZ}^* \Gamma_{tZ})^\dagger$ to both sides, we obtain

$$-\Gamma_{tZ} \begin{pmatrix} a \\ 0_{2N} \end{pmatrix} = -\Gamma_{tZ} \Gamma_{tZ}^\dagger \Gamma_{tZ_0} \begin{pmatrix} a_0 \\ 0_{2N} \end{pmatrix} - \Gamma_{tZ} \Gamma_{tZ}^\dagger w + \lambda \Gamma_{tZ}^{*,\dagger} \begin{pmatrix} 1_N \\ 0_{2N} \end{pmatrix}.$$

Therefore,

$$A = \Gamma_{tZ} \begin{pmatrix} a \\ 0_{2N} \end{pmatrix} - \Gamma_{tZ_0} \begin{pmatrix} a_0 \\ 0_{2N} \end{pmatrix} - w = -\left(\lambda \Gamma_{tZ}^{*,\dagger} \begin{pmatrix} 1_N \\ 0_{2N} \end{pmatrix} + \Pi_{tZ} w + \Pi_{tZ} \Gamma_{tZ_0} \begin{pmatrix} a_0 \\ 0_{2N} \end{pmatrix} \right)$$

is equal to $-q_{tZ}$ as required.

□

The goal of the remainder of this subsection is to show that V_t^* contains a ball of radius on the order of t^4.

Corollary 3.7. *Let $c_0 \leqslant c_*$ where c_* is as in Lemma 3.6. Suppose that $|Z - Z_0|_\infty \leqslant c_0 t$, $\lambda \leqslant c_0 t^4$ and $\|w\| \leqslant c_0 t^4$. Then, there exists a constant dependent only on φ, a_0, Z_0 such that*

$$\|\mathrm{d}g_t^*(v)\| \leqslant \frac{C}{t^3}.$$

Proof. Thanks to Lemma3.6, we have

$$\begin{aligned} \mathrm{d}g_t^*(v) &= -J_{ta}^{-1} H_{tZ}^{-1} G_{tZ}(\lambda, w)^{-1} H_{tZ}^{*-1} \left(\begin{pmatrix} 1_N \\ 0_{2N} \end{pmatrix} \quad -H_{tZ}^* \Psi_{tZ}^* \right) \\ &= J_a^{-1} H_Z^{-1} \operatorname{diag}(1, t^{-1}, t^{-1}, t^{-2}, t^{-2}, t^{-3}) G_{tZ}(\lambda, w)^{-1} (\delta_{3N}, \Psi_{tZ}^*). \end{aligned}$$

The result then follows since $\left\|\operatorname{diag}(1, t^{-1}, t^{-1}, t^{-2}, t^{-2}, t^{-3})\right\| \leqslant t^{-3}$ and all the other terms above are bounded independently of t. □

We are finally ready to show that V_t^* contains a ball with radius on the order of t^4.

PROPOSITION 3.8. *There exists $C > 0$ such that for all $t \in (0, t_0)$,*

$$V_t^* \supset B(0, ct^4).$$

where $c \sim c_*$.

Proof. Let $v \in \mathbb{R} \times \mathcal{H}$ be such that $\max(\lambda, \|w\|) = 1$. Let

$$R_v = \sup \{r \geqslant 0 \; ; \; rv \in V_t^*\}.$$

First note that $R_v \in (0, Ct^4)$, and since g_t^* is uniformly continuous on V_t^*, $g_t^*(R_v v) \overset{\text{def.}}{=} \lim_{r \to R_v} g(rv)$ is well defined. Moreover, $f_t(g_t^*(R_v v), R_v v) = 0$.

By maximality of V_t^*, it is necessarily the case that $g_t^*(R_v v) \in \partial(B(a_0, c_*) \times B(Z_0, tc_*))$ (otherwise, we can apply the implicit function theorem to construct a neighbourhood $V \in \mathcal{V}$ such that $V_t^* \subsetneq V$).

Suppose that $g_t^*(R_v v) \in \overline{B(a_0, c_*)} \times \partial(B(Z_0, tc_*))$. Then, for $(a, Z) = g_t^*(R_v v)$,

$$c_* t = \|Z - Z_0\| \leqslant \int_0^1 |\mathrm{d}g_t^*(sR_v v) \cdot R_v v|_\infty \, \mathrm{d}s \leqslant \frac{C}{t^3} R_v \implies R_v \geqslant \frac{c_* t^4}{C},$$

where we have applied Corollary 3.7. Similarly, if $g_t^*(R_v v) \in \partial(B(a_0, c_*)) \times \overline{B(Z_0, tc_*)}$, then $R_v \geqslant \frac{c_* t^4}{C}$. Repeating this for all $v \in \mathbb{R} \times \mathcal{H}$ with unit norm yields the required result.
□

3.3.3. Use of Non-degeneracy. Throughout this section, we use the definition of $p_{\lambda,t}$ and $\eta_{\lambda,t}$ given in (3.11), where we set $(a, Z) = g_t(\lambda, w)$.

PROPOSITION 3.9. *Let $\varepsilon > 0$. Then, there exists $c_0 > 0$ and $t_0 > 0$ such that for all Z, λ, w, t with $0 < t < t_0$, $\lambda \leqslant c_0 t^4$ and $\|w\| \leqslant c_0 t^4$ and $\|w\| / \lambda \leqslant c_0$, we have that*

$$\|p_{\lambda,t} - p_{W,Z_0}\| \leqslant \varepsilon.$$

Proof. By Proposition3.8, there exists c such that for all $(\lambda, w) \in B(0, c_0 t^4)$, with $c_0 \leqslant c$, g_t^* is well defined. For $(a, Z) = g_t^*(\lambda, w)$,

$$\begin{aligned} p_{\lambda,t} &= \Gamma_{tZ}^{*,\dagger} \begin{pmatrix} 1_N \\ 0_{2N} \end{pmatrix} + \Pi_{tZ} \frac{w}{\lambda} + \frac{1}{\lambda} \Pi_{tZ} \Gamma_{tZ_0} \begin{pmatrix} a_0 \\ 0 \end{pmatrix} \\ &= p_{W,Z_0} + \mathcal{O}(t) + \mathcal{O}\left(\frac{\|w\|}{\lambda}\right) + \frac{1}{\lambda} \Pi_{tZ} \Gamma_{tZ_0} \begin{pmatrix} a_0 \\ 0 \end{pmatrix}. \end{aligned}$$

To bound the last term on the RHS,

$$\begin{aligned} \left\| \frac{1}{\lambda} \Pi_{tZ} \Gamma_{tZ_0} \begin{pmatrix} a_0 \\ 0 \end{pmatrix} \right\| &\leqslant \frac{C}{\lambda} t^2 \|Z - Z_0\|^2 \\ &\leqslant \frac{C}{\lambda} \frac{\|w\|^2 + \lambda^2}{t^4} \leqslant Cc_0 \left(\frac{\|w\|}{\lambda} + 1\right) \end{aligned}$$

where the first inequality follows from Proposition 3.2, the second inequality follows from Corollary 3.7 and the third inequality uses $\|w\|, \lambda \leqslant c_0 t^4$. The result now follows by choosing c_0 sufficiently small. □

Proof. [Proof of Theorem 1.2] By Proposition 3.9, if $(a, Z) = g_t^*(\lambda, w)$, then since $p_{\lambda,t}$ can be made arbitrarily close to p_{W,Z_0}, we can apply Proposition3.3 to conclude that $p_{\lambda,t}$ is a valid certificate and hence the (unique) solution to the dual problem of ($\mathcal{P}_\lambda(y)$) with $y = \Phi m_{a_0,tZ_0}$. Moreover, $\eta_{\lambda,t}$ attains the value 1 only at the points in Z. Therefore, the support of any solution of ($\mathcal{P}_\lambda(y)$) is contained in Z and by invertibility of $\Phi_Z^*\Phi_Z$, it follows that $m_{a,Z}$ is the unique solution of ($\mathcal{P}_\lambda(y)$). Finally, the bounds on $\|(a, Z) - (a_0, Z_0)\|$ is a direct consequence on the bounds on the differential $\mathrm{d}g_t^*$. □

3.4. Limitations. The key idea behind the stability result of Theorem 1.2 is Proposition 3.3: any certificate which is sufficiently close to η_{W,Z_0} is also a valid certificate. We have only proved this result in the case of a pair of spikes, although a similar proof technique can be applied to the case where Z_0 consists of N aligned points in a direction d_{Z_0}, with the natural extension of the non-degeneracy condition (c.f. the construction of η_{W,Z_0} from Example 2.10) being

$$\begin{pmatrix} \partial^2_{d_{Z_0}^\perp} \eta_{W,Z_0}(0) & \frac{1}{N!}\partial_{d_{Z_0}^\perp}\partial^N_{d_{Z_0}}\eta_{W,Z_0}(0) \\ \frac{1}{N!}\partial_{d_{Z_0}^\perp}\partial^N_{d_{Z_0}}\eta_{W,Z_0}(0) & \frac{2}{(2N)!}\partial^{2N}_{d_{Z_0}}\eta_{W,Z_0}(0) \end{pmatrix} \prec 0.$$

However, Proposition 3.3 is in general not valid and therefore, the question of whether there is support stability in the case of more than 2 spikes remains open. The purpose of this section is to present some examples to illustrate this phenomenon. Note also that there exists examples (see the Gaussian mixture example from Section 4) where one can numerically observe support stability when recovering a pair of spikes, but not in the case of 3 or more spikes.

In the following examples, we consider a convolution operator Φ, i.e. $\varphi(x) = \tilde{\varphi}(x - \cdot) \in L^2(\mathcal{X})$. In the $N = 3$ case, the following proposition gives an example where Proposition 3.3 fails. In particular, for $Z \in \mathcal{X}^3$, it is possible to define a sequence of precertificates on Z which saturate at an additional point and, and yet converge to $\eta_{W,Z}$.

Proposition 3.10 (Case $N = 3$). *Let $Z = \{z_1, z_2, z_3\} \in \mathcal{X}^3$ be 3 points which are not collinear. Let x is any point in the interior of the convex hull of Z. Let*

$$p_t \stackrel{\text{def.}}{=} \operatorname{argmin}\left\{\|p\| \ ; \ (\Phi^*p)(tv) = 1, \nabla(\Phi^*p)(tv) = 0, \ \forall v \in Z, \ \Phi^*p(tx) = 1\right\}.$$

Then, $\lim_{t\to 0}\|p_t - p_{W,Z}\| = 0$.

Proof. The least interpolant space associated to Hermite interpolation at Z contains the polynomial space of degree 2, Π_2^2. Moreover, by Lemma 2.19, since Φ is a convolution operator, $\nabla^k\eta_{W,Z}(0) = 0$ for all odd integers k. Therefore, $\nabla^3\eta_{W,Z}(0) = 0$ for $k = 1, 2, 3$. On the other hand, the least interpolant space associated to Hermite interpolation at Z plus Lagrange interpolation at x is Π_3^2. Therefore, $p_{W,Z} = p_t + \mathcal{O}(t)$. □

Similar to the previous proposition, we give an example in the case of $Z \in \mathcal{X}^N$ with $N = 4$ where Proposition 3.3 fails.

Proposition 3.11. *Let $\tilde{\varphi}$ be the Gaussian kernel. Let*

$$Z = \{(1, 1), (-1, 1), (1, -1), (-1, -1)\}.$$

Let $(u, v) \in \mathbb{R}^2$ be such that $u^2 + v^2 = 1$ and let $\tilde{Z} = \{(u, v)\} \cup Z$. Let

$$p_t \stackrel{\text{def.}}{=} \operatorname{argmin}\left\{\|p\| \ ; \ (\Phi^*p)(tx) = 1, \nabla(\Phi^*p)(tx) = 0, \ \forall x \in Z, \ \Phi^*p(t(u, v)) = 1\right\}.$$

Then, $\lim_{t\to 0}\|p_t - p_{W,Z}\| = 0$.

Proof. First note that the least interpolant space associated with Hermite interpolation at Z is spanned by the following basis:

$$\mathcal{B}_Z = \{X^\alpha \; ; \; |\alpha| \leqslant 3\} \cup \left\{X^\beta \; ; \; \beta \in \{(1,3),(3,1)\}\right\}. \tag{3.14}$$

Let $\tilde{Z} = \{(1,1),(-1,1),(1,-1)\}$. Then, we have that $\nabla^j \eta_{W,\tilde{Z}}(0) = 0$ for $j = 1,2,3$ and $\partial_x^3\partial_y \eta_{W,\tilde{Z}}(0) = \partial_y^3\partial_x \eta_{W,\tilde{Z}}(0) = 0$. Therefore, $\eta_{W,\tilde{Z}} = \eta_{W,Z}$.

Observe now that the de Boor basis associated with Hermite interpolation on Z and Lagrange interpolation on (u,v) is

$$\mathcal{B}_Z \cup \left\{ p(x,y) = y^4 + 6x^2y^2\left(\frac{u^2-1}{v^2-1}\right) + x^4\left(\frac{u^2-1}{v^2-1}\right)^2 \right\}$$

Moreover, by the explicit formula given in Proposition 2.18, we have that $\partial_y^4 \eta_{W,Z}(0) = \partial_x^4 \eta_{W,Z}(0) = -48$ and $\partial_y^2\partial_x^2 \eta_{W,Z}(0) = -16$. Therefore, $p(\partial_x,\partial_y)\eta_{W,Z}(0) = 0$ whenever

$$-48 - 96\left(\frac{u^2-1}{v^2-1}\right) - 48\left(\frac{u^2-1}{v^2-1}\right)^2 = 0.$$

i.e. $u^2+v^2 = 2$. So, provided that $u^2+v^2=2$, then $p_{W,Z} = \lim_{t\to 0} p_t$. ☐

4. Numerical Study.

4.1. Considered Setups. We consider three different imaging operators Φ, intended to be representative of three different setups routinely encountered in imaging or machine learning. For each setup, in order to perform the computations of $\eta_{V,Z}$, $\eta_{W,Z}$ and to implement the Frank-Wolfe algorithm detailed in Section 4.3, the only requirement is to be able to evaluate the correlation kernel C defined in (2.1) and its derivatives.

In these examples, we consider the clustering of the spikes positions at a fixed point $z_0 \in \mathcal{X}$, i.e. consider for $t > 0$ the positions $Z_t = (z_0 + t(z_i - z_0))_{i=1}^N \in \mathcal{X}^N$. For the purpose of simplifying notation, the previous sections detailed only the case of $z_0 = 0$, i.e. $Z_t = tZ$, however, all previous results also hold in this more general setting by a change of variable $x \in \mathcal{X} \to x - z_0 \in \mathcal{X}$. Note that if $\mathcal{X}$ is not translation invariant, one should restrict the translation around z_0 and extend it into a smooth diffeomorphism on $\mathcal{X}$, see Appendix A for a proof of the reparametrization invariance of $\eta_{V,Z}$.

- *Gaussian convolution:* this corresponds to a translation invariant setup, which is typical in the modelling of acquisition blur in image processing. We consider $\varphi(x) = e^{-\frac{\|x-\cdot\|^2}{2\sigma^2}} \in \mathcal{H} = L^2(\mathbb{R}^2)$ on $\mathcal{X} = \mathbb{R}^2$, and one has
$$C(x,x') = e^{-\frac{\|x-x'\|^2}{4\sigma^2}}. \tag{4.1}$$
In this case, the clustering point is set to be $z_0 = 0$.
- *Gaussian mixture estimation:* in machine learning, an important problem is to estimate the parameters $(z_i)_{i=1}^N \in \mathcal{X}^N$ of a mixture $\sum_{i=1}^N a_i\varphi(z_i)$ of N elementary distributions parameterized by φ from samples or moments observations, see [35] for an overview of this problem. This problem can be recast as a super-resolution problem, where one seeks to recover the measure $m_0 = \sum_i a_i \delta_{z_i}$ from observations of the form (1.1) where the noise w

accounts for the sampling scheme (in a real-life machine learning setup, the operator Φ itself is noisy to account for the sampling scheme). We consider here a classical instance of this setup, where one looks for a mixture of 1-D Gaussians, parameterized by mean $m \in \mathbb{R}$ and standard deviation $s \in \mathbb{R}_+^*$, i.e. $x = (m, s) \in \mathcal{X} = \mathbb{R} \times \mathbb{R}_+^*$, so that $\varphi(x) = \frac{1}{s} e^{-\frac{(\cdot - m)^2}{2s^2}} \in \mathcal{H} = L^2(\mathbb{R})$ and the correlation operator reads

$$C((m,s),(m',s')) = \frac{1}{\sqrt{s^2 + s'^2}} e^{-\frac{(m-m')^2}{2(s^2+s'^2)}}. \tag{4.2}$$

In this case, the clustering point is set to be $z_0 = (m_0, s_0) = (0, 2)$.

- *Neuro-imaging:* for medical and neuroscience imaging applications, a standard goal is to estimate pointwise sources inside some domain $\mathcal{X} \subset \mathbb{R}^d$ (where $d = 2$ or 3) from measurements on the boundary $\partial\mathcal{X}$. The operator is thus of the form $\varphi(x) = (\psi(x,u))_{u \in \partial\mathcal{X}} \in \mathcal{H} = L^2(\partial\mathcal{X})$ (equipped with the uniform measure on the boundary) where the kernel $\psi(x,u)$ corresponds to the impulse response of the measurement operator. To model MEG or EEG acquisition [34], we consider a singular kernel $\psi(x,u) = \|x - u\|^{-2}$ which accounts for the decay of the electric or magnetic field in a stationary regime. We consider a disk domain $\mathcal{X} = \left\{ x \in \mathbb{R}^2 \; ; \; \|x\| < 1 \right\}$ which could model a slice of a head. The correlation function associated to this problem is

$$C(x,x') = 2\pi \frac{1 - \|x\|^2 \|x'\|^2}{(1-\|x\|^2)(1-\|x'\|^2)((1 - \langle x, x' \rangle)^2 + |x \wedge x'|^2)}, \tag{4.3}$$

see Appendix B for a proof. In this case, the clustering point is set to be $z_0 = (0.4, 0.3) \in \mathcal{X}$.

As it is customary for sparse regularization, we perform the BLASSO recovery using an L^2 normalized operator, i.e. perform the replacement

$$\varphi(x) \leftarrow \frac{\varphi(x)}{\|\varphi(x)\|_{\mathcal{H}}} \quad \Longrightarrow \quad C(x,x') \leftarrow \frac{C(x,x')}{\sqrt{C(x,x)C(x',x')}}.$$

Note that for translation invariant operators (i.e. convolutions), the kernels are already normalized.

4.2. Asymptotic Certificate $\eta_{W,Z}$. Figure 4.1 explores the behaviour of $\eta_{W,Z}$ in the three considered cases:

- *Gaussian convolution* (4.1)*:* we found numerically that $\eta_{W,Z}$ is always non-degenerate, for any N and spikes configuration Z. This is inline with the theoretical results of Section 2.4.1. This implies that one can hope (and provably do so for $N = 2$ according to Theorem 1.2) to achieve super-resolution for Gaussian deconvolution (provided, of course, that the signal-to-noise ratio is large enough).
- *Neuro-imaging* (4.3)*:* we observed numerically that $\eta_{W,Z}$ is always non-degenerate for $N = 2$ and more generally for aligned spikes. In contrast, for three non-aligned spikes, $\eta_{W,Z}$ is not a valid certificate ($\|\eta_{W,Z}\|_\infty > 1$) which means that in the presence of noise, one cannot stably super-resolve 3 close spikes.
- *Gaussian mixture estimation* (4.2)*:* here, the situation is more complicated, and for $N = 2$ spikes, $\eta_{W,Z}$ is non degenerate if $|m_2 - m_1| \leqslant |s_2 - s_1|$. This

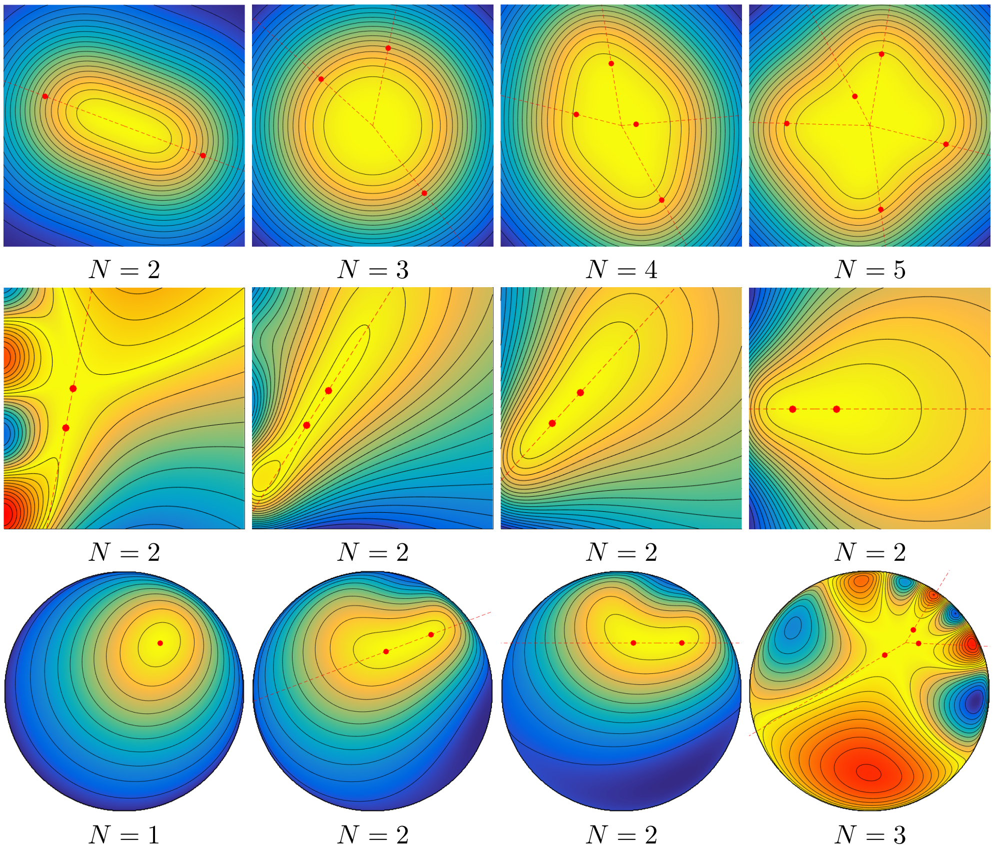
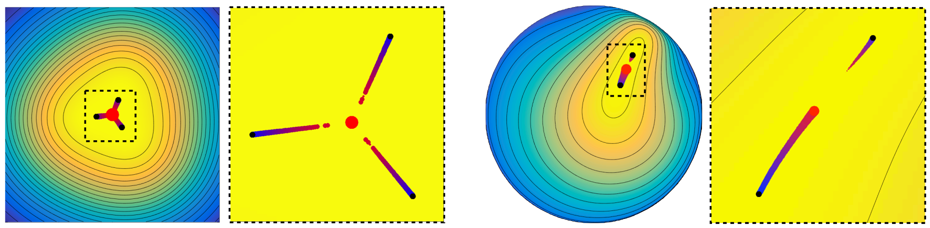


FIG. 4.1. *Display of the evolution of $\eta_{W,Z}$ for the three different operators Φ. The dashed red line shows the directions $(Z_t)_{t>0}$ along which the spikes are converging. Red color indicates regions where $\eta_{W,Z}(x) > 1$, i.e. it is degenerated. Top: Gaussian convolution (4.1). Middle: Gaussian mixture (4.2), here the horizontal axis is the standard deviation $s \in [0.5, 6]$ and the vertical axis is the mean $m \in [-3, 3]$. Bottom: neuro-imaging like (4.3).*

means that one can super-resolve with BLASSO a mixture of two Gaussians provided that the variation in the means is not too large with respect to the variation in standard deviations. Note also that in the special 1-D case where either the means or the standard deviations are equal and known (which leads to a 1-D super resolution problem along the m or s axis) then the resulting 1-D η_W is non-degenerate. It is the interplay between means and standard deviation that makes the super-resolution possibly problematic.

An important aspect to consider, which explains partly the above observations, is that, as explained in Section 2.4.2, convolution operators tend to have much better behaved $\eta_{W,Z}$ than arbitrary operators (such as the neuro-imaging and the Gaussian mixture), because their odd derivatives always vanish. In contrast, the vanishing of odd derivatives for a generic operator only occur for particular values of N and spikes configuration (e.g. aligned spikes). Without having its odd derivatives vanishing, $\eta_{W,Z}$ cannot be expected to be smaller than 1 near the spikes position $(z_i)_i$.

4.3. Spikes Recovery with Frank-Wolfe. In order to solve numerically the BLASSO problem ($\mathcal{P}_\lambda(y)$), we follow [10, 9] and use the Frank-Wolfe algorithm (also



$\eta_{V,Z}$ and m_λ Zoom $\eta_{V,Z}$ and m_λ Zoom

Gaussian $N=3$ Neuro-imaging $N=2$

Fig. 4.2. *Display of the evolution of the solution m_λ of ($\mathcal{P}_\lambda(y)$) (computed using Frank-Wolfe algorithm) as a function of λ for two different operators Φ, in cases where $\eta_{V,Z}$ is non-degenerated. The settings are the same as for Figure 4.1, and the bottom row is a zoom in the dashed rectangular region indicated on the top row. A spike $a\delta_x$ of m_λ is indicated with a disk centered at x of radius proportional to a, and the color ranges between blue for $\lambda = 0$ and red for $\lambda = \lambda_{\max}$. The background color image shows $\eta_{V,Z}$ where z are the spikes possitions of m_0 (plotted in black).*

known as conditional gradient) with improved non-convex updates. The algorithm starts with the initial zero measure $m^{(0)} = 0$, and alternates between a "matching pursuit" step which generates a new spike location

$$\tilde{x} \overset{\text{def.}}{=} \underset{x\in\mathcal{X}}{\operatorname{argmax}}\, |\eta^{(\ell)}(x)| \quad \text{where} \quad \eta^{(\ell)}(x) \overset{\text{def.}}{=} \frac{1}{\lambda}\langle \varphi(x),\, y - \Phi m^{(\ell)}\rangle_{\mathcal{H}}, \tag{4.4}$$

with associated amplitude $\tilde{a} \overset{\text{def.}}{=} \lambda\eta^{(\ell)}(\tilde{x}_{\ell+1})$, and a local non-convex minimization step, initialized with $r \leftarrow (x_1^{(\ell)}, \ldots, x_\ell^{(\ell)}, \tilde{x}) \in \mathcal{X}^{\ell+1}$ and $b \leftarrow (a_1^{(\ell)}, \ldots, a_\ell^{(\ell)}, \tilde{a}) \in \mathbb{R}^{\ell+1}$

$$(x^{(\ell+1)}, a^{(\ell+1)}) \overset{\text{def.}}{=} \underset{(r,b)\in\mathcal{X}^{\ell+1}\times\mathbb{R}^{\ell+1}}{\operatorname{argmin}} \frac{1}{2\lambda}\left\| y - \sum_{i=1}^{\ell+1} b_i\varphi(r_i)\right\|^2 + \|b\|_1\,. \tag{4.5}$$

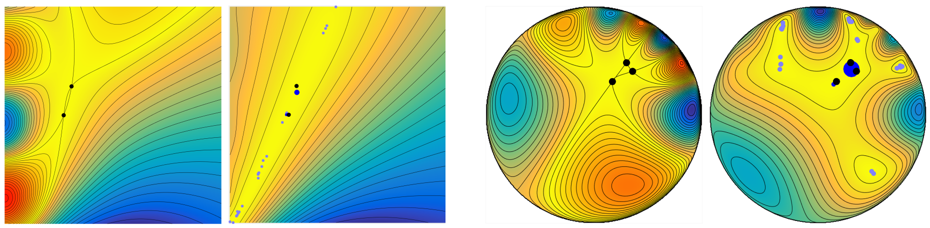
After each iteration, the measure is updated as

$$m^{(\ell+1)} \overset{\text{def.}}{=} \sum_{i=1}^{\ell+1} a_i^{(\ell+1)}\delta_{x_i^{(\ell+1)}}.$$

The termination criterion is $|\eta^{(\ell)}(\tilde{x})| \leqslant 1$, which means that $m^{(\ell)}$ is a solution to ($\mathcal{P}_\lambda(y)$) because $\eta^{(\ell)}$ is a valid dual certificate of optimality for $m^{(\ell)}$. The algorithm is known to converge in the sense of the weak topology of measures to a solution of ($\mathcal{P}_\lambda(y)$), see [10]. Without the non-convex update, convergence is slow (the rate on is only $O(1/\ell)$ on the BLASSO functional being minimized [38]). However, as we illustrate next, empirical observations suggest that by applying the non-convex update (4.5), convergence is often reached in a finite number of iteration.

Numerically, the low-dimensional optimization problems (4.4) and (4.5) are solved using a quasi-Newton (L-BFGS) solver. Computing the gradient of the involved functionals only require the evaluation of the correlation operator C and its derivative, assuming the measure $m^{(\ell)}$ are stored using a list of (positions, amplitudes).

Figure 4.2 explores the behaviour of the solution m_λ of ($\mathcal{P}_\lambda(y)$) as $(\lambda, w) \to 0$, in cases where $\eta_{V,Z}$ is non-degenerate, so that support is stable in this low-noise regime. Inline with support stability theorems, we scale the noise linearly with λ, $y = \Phi m_0 + \lambda w$, and set the noise w to be of the form $w = \Phi\bar{m}$ where $\bar{m}$ is a random



$\eta_{V,Z}$ and m_0 $\quad$ $\eta^{(\ell)}$ and $m^{(\ell)}$

Gaussian mixture $N = 2$

$\eta_{V,Z}$ and m_0 $\quad$ $\eta^{(\ell)}$ and $m^{(\ell)}$

Neuro-imaging $N = 3$

FIG. 4.3. *Display of the solution $m^{(\ell)}$ computed using $\ell = 40$ Frank-Wolfe iterations, in cases where $\eta_{V,Z}$ is degenerated (as indicated by red regions). The settings are the same as for Figure 4.1. The light blue dots indicate the support of $m^{(\ell)}$ (which thus allows one to locate spikes with very small amplitude) while blue dots are displayed with a size propositional to the amplitude of the corresponding spike.*

measure $\sum_j b_j \delta_{u_j}$ of $Q = 20$ random points $(u_j)_{j=1}^Q \in \mathcal{X}^Q$ where $(b_j)_j$ is white noise with standard deviation 10^{-3}. Numerically, we found that in these cases where $\eta_{V,Z}$ is non-degenerate, Frank-Wolfe with non-convex update converges in a finite number of steps. The color (from blue to red) allows to track the evolution with λ of the solution, which highlight the smoothness of the solution path.

Figure 4.3 shows, in contrast, cases where $\eta_{V,Z}$ is degenerate. According to Section 2.3, in this case, the support of the solution m_λ is not stable for small λ, and one expects this solution to be composed of more than N diracs. Numerically, in these cases, Frank-Wolfe does not converge in a finite number of steps, and it keeps creating new spikes of very small amplitudes. The figure shows how these additional spikes are added to force $|\eta^{(\ell)}|$ to be smaller than 1, while $\eta_{V,Z}$ is not.

Acknowlegements. We would like to thank Vincent Beck stimulating discussions about polynomial interpolation. We would also like to thank Vincent Duval for helpful discussions and for suggesting a more elegant proof of Theorem C.1, which is the one given in Appendix C.

5. Conclusion. This article presented a study of the multivariate BLASSO problem in order to recover clusters of positive spikes. In particular, we focussed on the question of support stability. Previous studies [31, 26] highlighted the importance of precertificates for this question, and as a first contribution, we presented a procedure for computing the limit $\eta_{W,Z}$ of the associated precertificates as the spikes converge towards a limit point. Since a necessary condition for support stability is that this certificate is valid (it is uniformly bounded by 1), one can check whether this certificate is valid before proceeding with more detailed analysis. Our second main contribution is a detailed analysis in the case of recovering a superposition of 2 spikes. Here, we showed that under a non-degeneracy condition on $\eta_{W,Z}$, support stability can be achieved provided that the norm of the additive noise $\|w\|$ and the regularization parameter λ decay like t^4, where t is the spacing between the 2 spikes. The question of which conditions are necessary for support stability when recovering more than 2 spikes remains open. The final part of this paper presented numerical examples related to 3 different imaging situations. In addition to verifying the results of this article, we observed the breakdown of support stability in the Gaussian mixture case and the neuro-imaging case in certain settings.

Appendix A. Reparameterization Invariance.

In the following, given a Borel map $T : \mathcal{X} \to \mathcal{Y}$, and a Borel measure μ defined on $\mathcal{X}$, $T_\sharp\mu$ is the pushforward measure of μ, so that for all integrable $f \in L^1(\mathcal{Y})$,

$$\int_{\mathcal{Y}} f(x)\mathrm{d}(T_\sharp\mu)(x) = \int_{\mathcal{X}} f(T(x))\mathrm{d}\mu(x).$$

Proposition A.1. *Let $T : \mathcal{X} \to \mathcal{Y}$ be a bijection such that the Jacobian of T^{-1} is invertible. Consider the following minimization problems:*

$$\min_m \frac{1}{2}\|y - \Phi m\|_{\mathcal{H}}^2 + \lambda |m|(\mathcal{X}), \tag{A.1}$$

$$\min_m \frac{1}{2}\left\|y - (\Phi \circ T_\sharp^{-1})m\right\|_{\mathcal{H}}^2 + \lambda |m|(\mathcal{Y}). \tag{A.2}$$

If μ solve (A.1), then $\nu \overset{\text{def.}}{=} T_\sharp\mu$ solves (A.2). Let $\eta_D^{\Phi,y}$ and $\eta_D^{\Phi\circ T_\sharp^{-1},y}$ be the dual certificates and let $\eta_V^{\Phi,y}$ and $\eta_V^{\Phi\circ T_\sharp^{-1},y}$ be the precertificates associated to (A.1) and (A.2) respectively. Then, $\eta_D^{\Phi\circ T_\sharp^{-1},y} = \eta_D^{\Phi,y} \circ T^{-1}$ and $\eta_V^{\Phi\circ T_\sharp^{-1},y} = \eta_V^{\Phi,y} \circ T^{-1}$.

Proof. The dual certificate of (A.2) is

$$\eta_D^{\Phi\circ T_\sharp^{-1},y} = \frac{1}{\lambda}(\Phi \circ T_\sharp^{-1})^*(\Phi \circ T_\sharp^{-1}\nu - y) = \frac{1}{\lambda}(\Phi \circ T_\sharp^{-1})^*(\Phi\mu - y) = \eta_D^{\Phi,y} \circ T^{-1}.$$

For the precertificates, suppose that $y = \Phi m_{a,Z} = (\Phi \circ T_\sharp^{-1})m_{a,TZ}$. Then,

$$\begin{aligned} p_V^{\Phi\circ T_\sharp^{-1},y} &= \operatorname{argmin}\left\{\|p\| \ ; \ [(\Phi \circ T_\sharp^{-1})^*p](Tz_j) = 1,\ [\nabla((\Phi \circ T_\sharp^{-1})^*p)](Tz_j) = 0\right\} \\ &= \operatorname{argmin}\left\{\|p\| \ ; \ [\Phi^*p](z_j) = 1,\ [\nabla(\Phi^*p)](z_j) = 0\right\} = p_V^{\Phi,y} \end{aligned}$$

where we have used the fact that, by letting $J_{T^{-1}}$ denote the Jacobian of T^{-1},

$$0 = [\nabla((\Phi \circ T_\sharp^{-1})^*p)](Tz_j) = \nabla[\Phi^*p(T^{-1}\cdot)](Tz_j) = [J_{T^{-1}}(Tz_j)]\nabla(\Phi^*p)(z_j)$$

implies that $\nabla(\Phi^*p)(z_j)$ since the Jacobian of T^{-1} is invertible. Therefore,

$$\eta_V^{\Phi\circ T_\sharp^{-1}} = (\Phi \circ T_\sharp^{-1})^* p_V^{\Phi\circ T_\sharp^{-1},y} = (\Phi \circ T_\sharp^{-1})^* p_V^{\Phi,y} = \eta_V^{\Phi,y} \circ T^{-1}.$$

□

Remark A.2. *Let T be as in Proposition A.1, let $m_{a,Z} = \sum_j a_j\delta_{z_j}$ and $y = \Phi m_{a,Z}$. Let $\tilde{\Phi} \overset{\text{def.}}{=} (\Phi \circ T_\sharp^{-1})$ and $\tilde{y} \overset{\text{def.}}{=} \tilde{\Phi} m_{a,TZ}$. Then, $\tilde{y} = \Phi m_{a,Z}$. So, (A.2) can be rewritten as*

$$\min_m \frac{1}{2}\left\|\tilde{y} - \tilde{\Phi}m\right\|^2 + \lambda |m|(\mathcal{Y}).$$

Therefore, to check that the dual certificate of this problem $\eta_V^{\tilde{\Phi},\tilde{y}}$ is nondegenerate, it is enough to show that $\eta_V^{\Phi,y}$ is nondegenerate.

Corollary A.3. *Let T be as in Proposition A.1. Then,*

$$\eta_V^{\Phi,\Phi m_{a,TZ}} = \eta_V^{\Phi,\Phi m_{a,Z}} + \mathcal{O}(\|\mathrm{Id} - T\|).$$

Proof. Given $\varphi : \mathcal{X} \to \mathcal{H}$, let $\Gamma_{\varphi,Z} : \mathbb{R}^{3N} \to \mathcal{H}$ be defined as in (2.9), where the subscript φ makes explicit the associated kernel. By Proposition A.1, we have that

$$\eta_V^{\Phi,\Phi m_{a,TZ}} = \eta_V^{\Phi\circ T_\sharp,\Phi\circ T_\sharp m_{a,Z}} \circ T^{-1} = \Phi^*\Gamma_{\psi,Z}^{*,\dagger}\begin{pmatrix} 1_N \\ 0_{2N} \end{pmatrix},$$

where $\psi = \varphi \circ T$. On the other hand, $\eta_V^{\Phi,\Phi m_{a,Z}} = \Phi^*\Gamma_{\varphi,z}^{*,\dagger}\binom{1_N}{0_{2N}}$. Therefore,

$$\eta_V^{\Phi,\Phi m_{a,TZ}} = \eta_V^{\Phi,\Phi m_{a,Z}} + \mathcal{O}(\|\mathrm{Id} - T\|).$$

□

Appendix B. Proof of Correlation Function (4.3).

Let $\mathcal{X} \subset \mathbb{R}^2$ denote the open unit disc. Then, for $x = (x_1, x_2), x' = (x_1', x_2') \in \mathcal{X}$, we let $p = M_1 + \mathrm{i}M_2$, $p_1 = A_1 + \mathrm{i}A_2$ and $z = e^{it}$. Interpreting X as the unit disc on the complex plane, one has

$$\begin{aligned} C(x,x') = \langle \varphi(x), \varphi(x')\rangle &= \int_{\partial X} \frac{\mathrm{d}z}{\mathrm{i}z\,|z-p|^2\,|z-p_1|^2} \\ &= \int_{\partial X} \frac{-\mathrm{i}z}{(z-p)(1-z\overline{p})(z-p_1)(1-z\overline{p}_1)}. \end{aligned}$$

When $p \neq p_1$, there are 2 poles inside $\mathcal{X}$: $z = p, p_1$, so by the Cauchy residue theorem,

$$C(x,x') = 2\pi\left(\frac{p}{(1-|p|^2)(p-p_1)(1-p\overline{p}_1)} + \frac{p_1}{(p_1-p)(1-p_1\overline{p})(1-|p_1|^2)}\right)$$

When $p = p_1$, there is 1 pole inside X: $z = p$, so,

$$C(x,x') = 2\mathrm{i}\pi\left(\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{\mathrm{i}z}{(1-z\overline{p})^2}\right)\Big|_{z=p}\right) = 2\pi\left(\frac{1}{(1-|p|^2)^2} + \frac{2\,|p|^2}{(1-|p|^2)^3}\right)$$

One can then check that both expressions simplify to

$$C(x,x') = 2\pi\frac{1-|p|^2|p_1|^2}{(1-|p|^2)(1-|p_1|^2)|1-p\overline{p}_1|^2}.$$

Appendix C. η_W in 1-D for the Ideal Low Pass Filter.

The following proposition is new and thus a contribution of our paper.

THEOREM C.1. *Let $\mathcal{X} = \mathbb{Z}/\mathbb{R}$. Let $\tilde{\varphi}_D(x) = \sum_{|k|\leqslant f_c} e^{2\pi ikx}$ where $f_c \in \mathbb{N}$ is the cutoff frequency. Let $\varphi(x) = \tilde{\varphi}_D(\cdot - x)$. If $f_c = N$ and $Z \in \mathcal{X}^N$, then*

$$\eta_{V,Z}(x) = 1 - C_Z\prod_{j=1}^N \sin^2(\pi(z_j - x)) \quad \textit{and} \quad \eta_{W,Z}(x) = 1 - C_W\sin^{2N}(\pi x),$$

where

$$C_Z \stackrel{\text{def.}}{=} \frac{\int\prod_{j=1}^N \sin^2(\pi(z_j - z))\mathrm{d}z}{\int\prod_{j=1}^N \sin^4(\pi(z_j - z))\mathrm{d}z} \quad \textit{and} \quad C_W \stackrel{\text{def.}}{=} \frac{\int_0^1 \sin^{2N}(\pi z)\mathrm{d}z}{\int_0^1 \sin^{4N}(\pi z)\mathrm{d}z}.$$

In particular, $\eta_{V,Z}(x) < 1$ *for all* $x \in \mathcal{X} \setminus Z$ *and* $\eta_W(x) < 1$ *for all* $x \in \mathcal{X} \setminus \{0\}$.

Proof. We recall that in 1-D, η_V is of the form

$$\begin{aligned}\eta_V(x) &= \sum_{j=1}^N \alpha_j \langle \varphi(z_j),\, \varphi(x)\rangle + \sum_{j=1}^N \beta_j \langle \varphi'(z_j),\, \varphi(x)\rangle \\ &= \sum_{j=1}^N \alpha_j \langle v(z_j),\, v(x)\rangle + \sum_{j=1}^N \beta_j \langle v'(z_j),\, v(x)\rangle,\end{aligned}$$

where $v : \mathbb{T} \to \mathbb{C}^{2f_c+1}$ is defined by $v(x) \stackrel{\text{def.}}{=} (e^{2\pi i k x})_{|k|\leqslant f_c}$. We are interested in the special case of $f_c = N$. Note that η_V is a trigonometric polynomial of degree N and $\|\eta_V\|^2 = \sum_{|k|\leqslant N} \left|\langle \eta_V,\, e^{2\pi i k\cdot}\rangle\right|^2$, so, we can write $\eta_V = \operatorname{argmin}\{\|\eta\| \ ; \ \eta \in S\} = P_S(0)$, where P_S is the projection onto the space of trigonometric polynomials S defined by

$$S \stackrel{\text{def.}}{=} \left\{ p = \sum_{|k|\leqslant N} a_k e^{2\pi i k\cdot} \ ; \ p(z_j) = 1,\ p'(z_j) = 0, \quad j = 1, \ldots, N \right\}.$$

Note also that $1 \in S$.

Since $\eta_V - 1$ is a trigonometric polynomial of degree N with double roots at z_j for $j = 1, \ldots, N$, we can write

$$\eta_V(z) - 1 = f(z) \prod_{j=1}^N \sin^2(\pi(z_l - z)),$$

for some trigonometric polynomial f. However, since $\eta_V - 1$ is of degree N, we must have that $f \equiv C$ for some constant C. Observe now that because $\eta_V = P_S(0)$, we simply need to find C which minimizes

$$\int \left|1 + C \prod_{j=1}^N \sin^2(\pi(z_l - z))\right|^2 \mathrm{d}z,$$

which is

$$C = -\frac{\int \prod_{j=1}^N \sin^2(\pi(z_j - z))\mathrm{d}z}{\int \prod_{j=1}^N \sin^4(\pi(z_j - z))\mathrm{d}z} < 0.$$

So η_V is non-degenerate. Letting $z_1 = z_2 = \cdots = z_N = 0$ yields the required formula for η_W. ☐

REFERENCES

[1] Fredrik Andersson and Marcus Carlsson, *Espirit for multidimensional general grids*, arXiv preprint arXiv:1705.07892, (2017).

[2] Jean-Marc Azais, Yohann De Castro, and Fabrice Gamboa, *Spike detection from inaccurate samplings*, Applied and Computational Harmonic Analysis, 38 (2015), pp. 177–195.

[3] Sylvain Baillet, John C Mosher, and Richard M Leahy, *Electromagnetic brain mapping*, IEEE Signal processing magazine, 18 (2001), pp. 14–30.

[4] Tamir Bendory, *Robust recovery of positive stream of pulses*, IEEE Transactions on Signal Processing, 65 (2017), pp. 2114–2122.

[5] Eric Betzig, George H Patterson, Rachid Sougrat, O Wolf Lindwasser, Scott Olenych, Juan S Bonifacino, Michael W Davidson, Jennifer Lippincott-Schwartz, and Harald F Hess, *Imaging intracellular fluorescent proteins at nanometer resolution*, Science, 313 (2006), pp. 1642–1645.

[6] Badri Narayan Bhaskar, Gongguo Tang, and Benjamin Recht, *Atomic norm denoising with applications to line spectral estimation*, IEEE Transactions on Signal Processing, 61 (2013), pp. 5987–5999.

[7] Thierry Blu, Pier-Luigi Dragotti, Martin Vetterli, Pina Marziliano, and Lionel Coulot, *Sparse sampling of signal innovations: Theory, algorithms and performance bounds*, IEEE Signal Processing Magazine, 25 (2008), pp. 31–40.

[8] J Frédéric Bonnans and Alexander Shapiro, *Perturbation analysis of optimization problems*, Springer Science & Business Media, 2013.

[9] Nicholas Boyd, Geoffrey Schiebinger, and Benjamin Recht, *The alternating descent conditional gradient method for sparse inverse problems*, SIAM Journal on Optimization, 27 (2017), pp. 616–639.

[10] Kristian Bredies and Hanna Katriina Pikkarainen, *Inverse problems in spaces of measures*, ESAIM: Control, Optimisation and Calculus of Variations, 19 (2013), pp. 190–218.

[11] James A Cadzow, *Signal enhancement-a composite property mapping algorithm*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 36 (1988), pp. 49–62.

[12] Emmanuel J. Candès and Carlos Fernandez-Granda, *Super-resolution from noisy data*, Journal of Fourier Analysis and Applications, 19 (2013), pp. 1229–1254.

[13] ———, *Towards a mathematical theory of super-resolution*, Communications on Pure and Applied Mathematics, 67 (2014), pp. 906–956.

[14] Emmanuel J Candès, Justin Romberg, and Terence Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Transactions on information theory, 52 (2006), pp. 489–509.

[15] Scott Shaobing Chen, David L Donoho, and Michael A Saunders, *Atomic decomposition by basis pursuit*, SIAM review, 43 (2001), pp. 129–159.

[16] Jon F Claerbout and Francis Muir, *Robust modeling with erratic data*, Geophysics, 38 (1973), pp. 826–844.

[17] Michael P Clark and Louis L Scharf, *Two-dimensional modal analysis based on maximum likelihood*, IEEE Transactions on Signal Processing, 42 (1994), pp. 1443–1452.

[18] H Clergeot, Sara Tressens, and A Ouamri, *Performance of high resolution frequencies estimation methods compared to the cramer-rao bounds*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 37 (1989), pp. 1703–1720.

[19] Laurent Condat and Akira Hirabayashi, *Cadzow denoising upgraded: A new projection method for the recovery of dirac pulses from noisy linear measurements*, Sampling Theory in Signal and Image Processing, 14 (2015), pp. p–17.

[20] Carl De Boor and Amos Ron, *On multivariate polynomial interpolation*, Constructive Approximation, 6 (1990), pp. 287–302.

[21] ———, *Computational aspects of polynomial interpolation in several variables*, Mathematics of Computation, 58 (1992), pp. 705–727.

[22] ———, *The least solution for the polynomial interpolation problem*, Mathematische Zeitschrift, 210 (1992), pp. 347–378.

[23] Yohann De Castro and Fabrice Gamboa, *Exact reconstruction using Beurling minimal extrapolation*, Journal of Mathematical Analysis and applications, 395 (2012), pp. 336–354.

[24] Yohann De Castro, Fabrice Gamboa, Didier Henrion, and J-B Lasserre, *Exact solutions to super resolution on semi-algebraic domains in higher dimensions*, IEEE Transactions on Information Theory, 63 (2017), pp. 621–630.

[25] Laurent Demanet and Nam Nguyen, *The recoverability limit for superresolution via sparsity*, arXiv preprint arXiv:1502.01385, (2015).

[26] Quentin Denoyelle, Vincent Duval, and Gabriel Peyré, *Support recovery for sparse super-resolution of positive measures*, to appear in Journal of Fourier Analysis and Applications, (2017).

[27] David L Donoho, *Superresolution via sparsity constraints*, SIAM journal on mathematical analysis, 23 (1992), pp. 1309–1331.

[28] ———, *Compressed sensing*, IEEE Transactions on information theory, 52 (2006), pp. 1289–1306.

[29] David L Donoho, Iain M Johnstone, Jeffrey C Hoch, and Alan S Stern, *Maximum entropy and the nearly black object*, Journal of the Royal Statistical Society. Series B (Methodological), (1992), pp. 41–81.

[30] Vincent Duval, *A characterization of the non-degenerate source condition in super-resolution*, arXiv preprint arXiv:1712.06373, (2017).
[31] Vincent Duval and Gabriel Peyré, *Exact support recovery for sparse spikes deconvolution*, Foundations of Computational Mathematics, 15 (2015), pp. 1315–1355.
[32] Vincent Duval and Gabriel Peyré, *Sparse spikes super-resolution on thin grids I: the LASSO*, Inverse Problems, 33 (2017), p. 055008.
[33] Jean-Jacques Fuchs, *Sparsity and uniqueness for some specific under-determined linear systems*, in Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on, vol. 5, IEEE, 2005, pp. v–729.
[34] Alexandre Gramfort, Daniel Strohmeier, Jens Haueisen, Matti S Hämäläinen, and Matthieu Kowalski, *Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations*, NeuroImage, 70 (2013), pp. 410–422.
[35] Rémi Gribonval, Gilles Blanchard, Nicolas Keriven, and Yann Traonmilin, *Compressive statistical learning with random feature moments*, arXiv preprint arXiv:1706.07180, (2017).
[36] Yingbo Hua and Tapan K Sarkar, *Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 38 (1990), pp. 814–824.
[37] Laurent Jacques and Christophe De Vleeschouwer, *A geometrical study of matching pursuit parametrization*, IEEE Transactions on Signal Processing, 56 (2008), pp. 2835–2848.
[38] Martin Jaggi, *Revisiting frank-wolfe: Projection-free sparse convex optimization*, in ICML (1), 2013, pp. 427–435.
[39] Tao Jiang, Nicholas D Sidiropoulos, and Jos MF ten Berge, *Almost-sure identifiability of multidimensional harmonic retrieval*, IEEE Transactions on Signal Processing, 49 (2001), pp. 1849–1859.
[40] Hamid Krim and Mats Viberg, *Two decades of array signal processing research: the parametric approach*, IEEE signal processing magazine, 13 (1996), pp. 67–94.
[41] Stefan Kunis, Thomas Peter, Tim Römer, and Ulrich von der Ohe, *A multivariate generalization of Prony's method*, Linear Algebra and its Applications, 490 (2016), pp. 31–47.
[42] Shlomo Levy and Peter K Fullagar, *Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution*, Geophysics, 46 (1981), pp. 1235–1243.
[43] Wenjing Liao and Albert Fannjiang, *Music for single-snapshot spectral estimation: Stability and super-resolution*, Applied and Computational Harmonic Analysis, 40 (2016), pp. 33–67.
[44] Rudolph A Lorentz, *Multivariate hermite interpolation by algebraic polynomials: a survey*, Journal of computational and applied mathematics, 122 (2000), pp. 167–201.
[45] Stéphane G Mallat and Zhifeng Zhang, *Matching pursuits with time-frequency dictionaries*, IEEE Transactions on signal processing, 41 (1993), pp. 3397–3415.
[46] Ankur Moitra, *The threshold for super-resolution via extremal functions*, arXiv preprint arXiv:1408.1681, 2 (2014).
[47] Veniamin I Morgenshtern and Emmanuel J Candes, *Super-resolution of positive sources: The discrete setup*, SIAM Journal on Imaging Sciences, 9 (2016), pp. 412–444.
[48] Thomas Peter, Gerlind Plonka, and Robert Schaback, *Reconstruction of multivariate signals via Prony's method*, Proc. Appl. Math. Mech., to appear, (2017).
[49] Clarice Poon, Nicolas Keriven, and Gabriel Peyré, *A dual certificates analysis of compressive off-the-grid recovery*, arXiv preprint arXiv:1802.08464, (2018).
[50] Gaspard de Prony, *Essai expérimental et analytique: sur les lois de la dilatabilité de fluides élastique et sur celles de la force expansive de la vapeur de l'alkool, à différentes températures*, J. de l'Ecole Polytechnique, 1 (1795), pp. 24–76.
[51] Richard Roy and Thomas Kailath, *ESPRIT-estimation of signal parameters via rotational invariance techniques*, IEEE Transactions on acoustics, speech, and signal processing, 37 (1989), pp. 984–995.
[52] Michael J Rust, Mark Bates, and Xiaowei Zhuang, *Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM).*, Nature methods, 3 (2006), pp. 793–795.
[53] Joseph J Sacchini, William M Steedly, and Randolph L Moses, *Two-dimensional prony modeling and parameter estimation*, IEEE Transactions on signal processing, 41 (1993), pp. 3127–3137.
[54] Fadil Santosa and William W Symes, *Linear inversion of band-limited reflection seismograms*, SIAM Journal on Scientific and Statistical Computing, 7 (1986), pp. 1307–1330.

[55] Tomas Sauer, *Polynomial interpolation in several variables: lattices, differences, and ideals*, in Studies in Computational Mathematics, vol. 12, Elsevier, 2006, pp. 191–230.

[56] ———, *Prony's method in several variables*, Numerische Mathematik, 136 (2017), pp. 411–438.

[57] Geoffrey Schiebinger, Elina Robeva, and Benjamin Recht, *Superresolution without separation*, arXiv preprint arXiv:1506.03144, (2015).

[58] Ralph Schmidt, *Multiple emitter location and signal parameter estimation*, IEEE transactions on antennas and propagation, 34 (1986), pp. 276–280.

[59] Morteza Shahram and Peyman Milanfar, *Imaging below the diffraction limit: A statistical analysis*, IEEE Transactions on image processing, 13 (2004), pp. 677–689.

[60] ———, *Statistical and information-theoretic analysis of resolution in imaging*, IEEE Transactions on Information Theory, 52 (2006), pp. 3411–3437.

[61] Petre Stoica, Randolph L Moses, et al., *Spectral analysis of signals*, vol. 452, Pearson Prentice Hall Upper Saddle River, NJ, 2005.

[62] Gongguo Tang and Benjamin Recht, *Atomic decomposition of mixtures of translation-invariant signals*, IEEE CAMSAP, (2013).

[63] Robert Tibshirani, *Regression shrinkage and selection via the Lasso*, Journal of the Royal Statistical Society. Series B (Methodological), (1996), pp. 267–288.

[64] Lloyd N Trefethen and David Bau III, *Numerical linear algebra*, vol. 50, Siam, 1997.