Sparse non-negative super-resolution — simplified and stabilised

Armin Eftekhari[1,2], Jared Tanner[1,3], Andrew Thompson[3], Bogdan Toader[3], and Hemant Tyagi[1,2]

[1]Alan Turing Institute
[2]University of Edinburgh
[3]University of Oxford

April 4, 2018

To David L. Donoho, a uniquely positive gentleman,
in celebration of his 60th birthday and with thanks for his inspiration and support.

Abstract

The convolution of a discrete measure, $x = \sum_{i=1}^{k} a_i \delta_{t_i}$, with a local window function, $\phi(s-t)$, is a common model for a measurement device whose resolution is substantially lower than that of the objects being observed. Super-resolution concerns localising the point sources $\{a_i, t_i\}_{i=1}^{k}$ with an accuracy beyond the essential support of $\phi(s-t)$, typically from m samples $y(s_j) = \sum_{i=1}^{k} a_i \phi(s_j - t_i) + \eta_j$, where η_j indicates an inexactness in the sample value. We consider the setting of x being non-negative and seek to characterise all non-negative measures approximately consistent with the samples. We first show that x is the unique non-negative measure consistent with the samples provided the samples are exact, i.e. $\eta_j = 0$, $m \geqslant 2k+1$ samples are available, and $\phi(s-t)$ generates a Chebyshev system. This is independent of how close the sample locations are and *does not rely on any regulariser beyond non-negativity*; as such, it extends and clarifies the work by Schiebinger et al. in [1] and De Castro et al. in [2], who achieve the same results but require a total variation regulariser, which we show is unnecessary.

Moreover, we characterise non-negative solutions $\hat{x}$ consistent with the samples within the bound $\sum_{j=1}^{m} \eta_j^2 \leqslant \delta^2$. Any such non-negative measure is within $\mathcal{O}(\delta^{1/7})$ of the discrete measure x generating the samples in the generalised Wasserstein distance. Similarly, we show using somewhat different techniques that the integrals of $\hat{x}$ and x over $(t_i - \epsilon, t_i + \epsilon)$ are similarly close, converging to one another as ϵ and δ approach zero. We also show how to make these general results, for windows that form a Chebyshev system, precise for the case of $\phi(s-t)$ being a Gaussian window. The main innovation of these results is that non-negativity alone is sufficient to localise point sources beyond the essential sensor resolution and that, while regularisers such as total variation might be particularly effective, they are not required in the non-negative setting.

1 Introduction

Super-resolution concerns recovering a resolution beyond the essential size of the point spread function of a sensor. For instance, a particularly stylised example concerns multiple point sources which, because of the finite *resolution* or *bandwidth* of the sensor, may not be visually distinguishable. Various instances of this problem exist in applications such as astronomy [3], imaging in chemistry, medicine and neuroscience [4, 5, 6, 7, 8, 9, 10, 11], spectral estimation [12, 13], geophysics [14], and system identification [15]. Often in these application much is known about the point spread function of the sensor, or can be estimated and, given such model information, it is possible to identify point source locations with accuracy substantially below the essential width of the sensor point spread function. Recently there has been substantial interest from the mathematical community in posing algorithms and proving super-resolution guarantees in this setting, see for instance [16, 17, 18, 19, 20, 21, 22, 23]. Typically these approaches borrow notions from compressed sensing [24, 25, 26]. In particular, the aforementioned contributions to super-resolution consider what is known as the Total Variation norm minimisation over measures which are consistent with the samples. In this manuscript we show first that, for suitable point spread functions, such as the Gaussian, any discrete non-negative measure composed of k point sources is uniquely defined from $2k+1$ of its samples, and moreover that this uniqueness is independent of the separation between the point sources. We then show that by simply imposing non-negativity, which is typical in many applications, any non-negative measure suitably consistent with the samples is similarly close to the discrete non-negative measure which would generate the noise free samples. These results substantially simply results by [1, 2] and show that, while regularisers such as Total Variation may be particularly effective, in the setting of non-negative point sources such regularisers are not necessary to achieve stability.

1.1 Problem setup

Throughout this manuscript we consider non-negative measures in relation to discrete measures. To be concrete, let x be a k-discrete *non-negative* Borel measure supported on the interval $I = [0,1] \subset \mathbb{R}$, given by

$$x = \sum_{i=1}^{k} a_i \cdot \delta_{t_i} \quad \text{with } \; a_i > 0 \quad \text{ and } t_i \in int(I) \quad \text{ for all } i. \tag{1}$$

Consider also real-valued and continuous functions $\{\phi_j\}_{j=1}^m$ and let $\{y_j\}_{j=1}^m$ be the possibly noisy measurements collected from x by convolving against sampling functions $\phi_j(t)$:

$$y_j = \int_I \phi_j(t) x(\mathrm{d}t) + \eta_j = \sum_{i=1}^{k} a_i \phi_j(t_i) + \eta_j, \tag{2}$$

where η_j with $\|\eta\|_2 \leqslant \delta$ can represent additive noise. Organising the m samples from (2) in matrix notation by letting

$$y := [y_1 \; \cdots \; y_m]^T \in \mathbb{R}^m, \qquad \Phi(t) := [\phi_1(t) \; \cdots \; \phi_m(t)]^T \in \mathbb{R}^m \tag{3}$$

allows us to state the program we investigate:

$$\text{find } z \geqslant 0 \text{ subject to } \left\| y - \int_I \Phi(t) z(\mathrm{d}t) \right\|_2 \leqslant \delta', \tag{4}$$

with $\delta \leqslant \delta'$. Herein we characterise non-negative measures consistent with measurements (2) in relation to the discrete measure (1). That is, we consider any non-negative Borel measure z from the Program (4) [1] and show that any such z is close to x given by (1) in an appropriate metric, see Theorems 4, 5, 11, 12 and 13. Moreover, we show that the x from (1) is the unique solution to Program (4) when $\delta' = 0$; e.g. in the setting of exact samples, $\eta_i = 0$ for all i. Program (4) is particularly notable in that there is no regulariser of z beyond imposing non-negativity and, rather than specify an algorithm to select a z which satisfies Program (4), we consider all admissible solutions. The admissible solutions of Program (4) are determined by the source and sample locations, which we denote as

$$T = \{t_i\}_{i=1}^k \subset \mathrm{int}(I) \quad \text{ and } \quad S = \{s_j\}_{j=1}^m \subseteq I \tag{5}$$

respectively, as well as the particular functions $\phi_j(t)$ used to sample the k-sparse non-negative measure x from (1). Lastly, we introduce the notions of minimum separation and sample proximity, which we use to characterise solutions of Program (4).

Definition 1. ***(Minimum separation and sample proximity)*** *For finite $\tilde{T} = T \cup \{0,1\} \subset I$, let $\Delta(T) > 0$ be the minimum separation between the points in T along with the endpoints of I, namely*

$$\Delta(T) = \min_{T_i, T_j \in \tilde{T}, i \neq j} |T_i - T_j|. \tag{6}$$

[1] An equivalent formulation of Program (4) minimises $\|y - \int_I \Phi(t) z(\mathrm{d}t)\|_2$ over all non-negative measures on I (without any constraints). In this context, however, we find it somewhat more intuitive to work with Program (4), particularly considering the importance of the case $\delta = 0$.

We define the sample proximity to be the number $\lambda \in (0, \frac{1}{2})$ such that, for each source location t_i, there exists a closest sample location $s_{l(i)} \in S$ to t_i with

$$|t_i - s_{l(i)}| \leqslant \lambda \Delta(T). \tag{7}$$

We describe the nearness of solutions to Program (4) in terms of an additional parameter ϵ associated with intervals around the sources T; that is we let $\epsilon \leqslant \Delta(T)/2$ and define intervals as:

$$T_{i,\epsilon} := \left\{t : |t - t_i| \leqslant \epsilon\right\} \cap I, \qquad i \in [k], \quad T_\epsilon := \bigcup_{i=1}^{k} T_{i,\epsilon}, \tag{8}$$

where $[k] = 1, 2, \ldots, k$, and set $T_{i,\epsilon}^C$ and T_ϵ^C to be the complements of these sets with respect to I. In order to make the most general result of Theorems 11 and 12 more interpretable, we turn to presenting them in Section 1.2 for the case of $\phi_j(t)$ being shifted Gaussians.

1.2 Main results simplified to Gaussian window

In this section we consider $\phi_j(t)$ to be shifted Gaussians with centres at the source locations s_j, specifically

$$\phi_j(t) = g(t - s_j) = e^{-\frac{(t-s_j)^2}{\sigma^2}}. \tag{9}$$

We might interpret (9) as the "point spread function" of the sensing mechanism being a Gaussian window and s_j the sample locations in the sense that

$$\int_I \phi_j(t) x(\mathrm{d}t) = \int_I g(t - s_j) x(\mathrm{d}t) = (g \star x)(s_j), \qquad \forall j \in [m], \tag{10}$$

evaluates the "filtered" copy of x at locations s_j where $\star$ denotes convolution.

As an illustration, Figure 1 shows the discrete measure x in blue for $k = 3$, the continuous function $y(s) = (g \star x)(x)$ in red, and the noisy samples $y(s_j)$ at the sample locations S represented as the black circles.

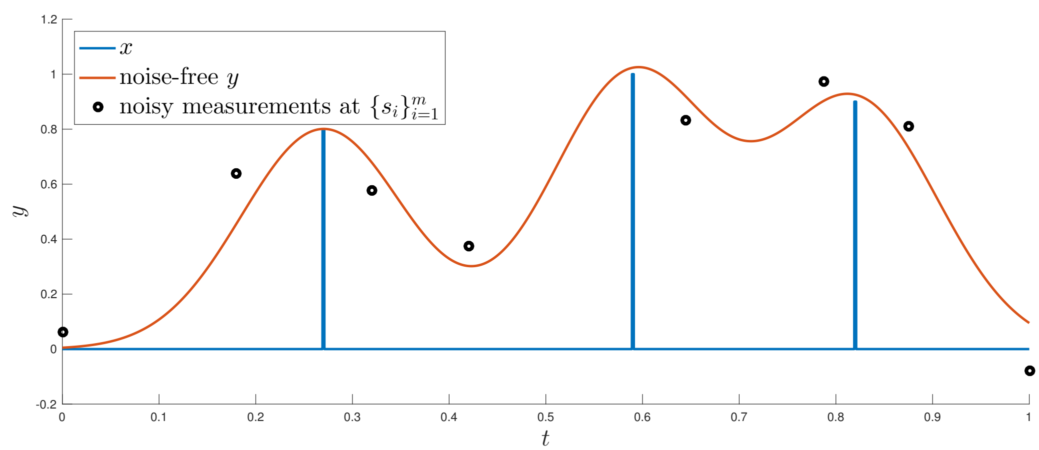


Figure 1: Example of discrete measure x and measurements where $\phi_j(t) = \phi(t - s_j)$ for $s_j \in S$ and the Gaussian kernel $\phi(t) = e^{-\frac{t^2}{\sigma^2}}$.

The conditions we impose to ensure stability of Program (4) for $\phi(t)$ Gaussian as in (9) are as follows:

Conditions 2. ***(Gaussian window conditions)*** *When the window function is a Gaussian $\phi(t) = e^{-\frac{t^2}{\sigma^2}}$, we require its width σ and the source and sampling locations from (5) to satisfy the following conditions:*

1. *Samples define the interval boundaries:* $s_1 = 0$ *and* $s_m = 1$,

2. *Samples near sources: for every* $i \in [k]$, *there exists a pair of samples* $s, s' \subset S$, *one on each side of* t_i, *such that* $|s - t_i| \leqslant \eta$ *and* $s' - s = \eta$, *for* $\eta \leqslant \sigma^2$ *small enough; which is quantified in Lemma 24.*

3. *Sources away from the boundary:* $\sigma\sqrt{\log(1/\eta^3)} \leqslant t_i, s_j \leqslant 1 - \sigma\sqrt{\log(1/\eta^3)}$ *for every* $i \in [k]$ *and* $j \in [2 : m-1]$,

4. *Minimum separation of sources:* $\sigma \leqslant \sqrt{2}$ *and* $\Delta(T) > \sigma\sqrt{\log\left(3 + \frac{4}{\sigma^2}\right)}$, *where the minimum separation* $\Delta(T)$ *of the sources is defined in Definition 1.*

The four properties in Conditions 2 can be interpreted as follows: Property 1 imposes that the sources are within the interval defined by the minimum and maximum sample; Property 2 ensures that there is a pair of samples near each source which translates into a sampling density condition in relation to the minimum separation between sources and in particular requires the number of samples $m \geqslant 2k + 2$; Property 3 is a technical condition to ensure sources are not overly near the sampling boundary; and Property 4 relates the minimum separation between the sources to the width σ of the Gaussian window.

We can now present our main results on the robustness of Program (4) as they apply to the Gaussian window; these are Theorem 4, which follows from Theorem 11, and Theorem 5, which follows from Theorem 12. However, before stating the stability results, it is important to note that, in the setting of exact samples, $\eta_i = 0$, the solution of Program (4) is unique when $\delta' = 0$.

Proposition 3. (Uniqueness of exactly sampled sparse non-negative measures for $\phi(t)$ Gaussian) *Let* x *be a non-negative* k*-sparse discrete measure supported on* I, *see (1). If* $\delta = 0$, $m \geqslant 2k+1$ *and* $\{\phi_j\}_{j=1}^m$ *are shifted Gaussians as in* (9), *then* x *is the unique solution of Program (4) with* $\delta' = 0$.

Proposition 3 states that Program (4) successfully localises the k impulses present in x given only $2k+1$ measurements when $\phi_j(t)$ are shifted Gaussians whose centres are in I. Theorems 4 and 5 extend this uniqueness condition to show that any solution to Program (4) with $\delta' > 0$ is proportionally close to the unique solution when $\delta' = 0$.

Theorem 4. (Wasserstein stability of Program (4) **for $\phi(t)$ Gaussian)** *Let* $I = [0, 1]$ *and consider a* k*-sparse non-negative measure* x *supported on* $T \subset \mathrm{int}(I)$. *Consider also an arbitrary increasing sequence* $\{s_j\}_{j=1}^m \subset \mathbb{R}$ *and, for positive* σ, *let* $\{\phi_j(t)\}_{j=1}^m$ *be defined in* (9), *which form* Φ *according to (3). If* $m \geqslant 2k+2$ *and Conditions 2 hold, then Program (4) with* $\delta' = \delta$ *is stable in the sense that*

$$d_{GW}(x, \widehat{x}) \leqslant F_1 \cdot \delta + \|x\|_{TV} \cdot \epsilon \tag{11}$$

for all $\epsilon \leqslant \Delta(T)/2$ *where* d_{GW} *is the generalised Wasserstein distance as defined in* (19) *and the exact expression of* $F_1 = F_1(k, \Delta(T), \frac{1}{\sigma}, \frac{1}{\epsilon}, \eta)$ *is given in the proof (see* (64) *in Section 3.4.2). In particular, for* $\sigma < \frac{1}{\sqrt{3}}$ *and* $\Delta(T) > \sigma\sqrt{\log\frac{5}{\sigma^2}}$, *we have:*

$$F_1(k, \Delta(T), \frac{1}{\sigma}, \frac{1}{\epsilon}, \eta) < \frac{c_1 k C_1(\frac{1}{\epsilon})}{\eta\sigma^2}\left[\frac{c_2}{\sigma^6(1-3\sigma^2)^2}\right]^k, \tag{12}$$

if

$$\eta \leqslant \min\left\{\frac{c_3\sigma^6(1-3\sigma^2)}{(k+1)^{\frac{3}{2}}}, \frac{c_4\bar{C}^{\frac{1}{6}}\sigma^{\frac{2}{3}}}{(k+1)^{\frac{1}{3}}}\right\}, \tag{13}$$

where c_1, c_2, c_3, c_4 *are universal constants and* $C_1\left(\frac{1}{\epsilon}\right)$ *is given by* (59) *in Section 3.4.1*

The central feature of Theorem 4 is that the proportionality to δ and ϵ of the Wasserstein distance between any solution to Program (4) and the unique solution for $\delta' = 0$ is of the form (11). The particular form of $F_1(\cdot)$ is not believed to be sharp; in particular, the exponential dependence on k in (12) follows from bounding the determinant of a matrix similar to Φ (see (128)) by a lower bound on the minimum eigenvalue to the k^{th} power. The scaling with respect to σ^{-2} is a feature of δ' in Program (4) not being normalized with respect to $\|y\|_2$ which, for T and S fixed, decays with σ due to the increased localisation of the Gaussian. Note that the ϵ dependence is a feature of the proof and the ϵ which minimises the bound in (11) is proportional to δ to some power as determined by $C_1(\epsilon^{-1})$ from (12). Theorem 4 follows from the more general result of Theorem 11, whose proof is given in Section 3 and the appendices.

As an alternative to showing stability of Program (4) in the Wasserstein distance, we also prove in Theorem 5 that any solution to Program (4) is locally consistent with the discrete measure in terms of local averages over intervals $T_{i,\epsilon}$ as given in (8). Moreover, for Theorem 5, we make Property 2 of Conditions 2 more transparent by using the sample proximity $\lambda\Delta(T)$ from Definition 1; that is, η defined in Conditions 2 is related to the sample proximity from Definition 1 by $\lambda\Delta(T) \leqslant \eta/2$.

Theorem 5. (Average stability of Program (4) for $\phi(t)$ Gaussian: source proximity dependence) *Let $I = [0,1]$ and consider a k-sparse non-negative measure x supported on T and sample locations S as given in (5) and for positive σ, let $\{\phi_j(t)\}_{j=1}^m$ as defined in (9). If the Conditions 2 hold, then, in the presence of additive noise, Program (4) is stable and it holds that, for any solution $\hat{x}$ of Program (4) with $\delta' = \delta$:*

$$\left| \int_{T_{i,\epsilon}} \hat{x}(\mathrm{d}t) - a_i \right| \leqslant \left[(c_1 + F_2) \cdot \delta + c_2 \frac{\|\hat{x}\|_{TV}}{\sigma^2} \cdot \epsilon \right] F_3, \tag{14}$$

$$\int_{T_\epsilon^C} \hat{x}(\mathrm{d}t) \leqslant F_2 \cdot \delta, \tag{15}$$

where the exact expressions of $F_2 = F_2(k, \Delta(T), \frac{1}{\sigma}, \frac{1}{\epsilon})$ and $F_3 = F_3(\Delta(T), \sigma, \lambda)$ are given in the proof (see (70) in Section 3.4.3), provided that λ, $\Delta(T)$ and σ satisfy (27). In particular, for $\sigma < \frac{1}{\sqrt{3}}$, $\Delta(T) > \sigma\sqrt{\log \frac{5}{\sigma^2}}$ and $\lambda < 0.4$, we have $F_3(\Delta(T), \sigma, \lambda) < c_5$ and:

$$F_2(k, \Delta(T), \frac{1}{\sigma}, \frac{1}{\epsilon}) < c_3 \frac{kC_2(\frac{1}{\epsilon})}{\sigma^2} \left[\frac{c_4}{\sigma^6(1-3\sigma^2)^2} \right]^k. \tag{16}$$

Above, c_1, c_2, c_3, c_4, c_5 are universal constants and $C_2(\frac{1}{\epsilon})$ is given by (60) in Section 3.4.1.

The bounds in Theorems 4 and 5 are intentionally similar, and though their proofs make use of the same bounds, they have some fundamental differences. While both (11) and (14) have the same proportionality to δ and ϵ, the role of ϵ in particular differs substantially in that Theorem 5 considers averages of $\hat{x}$ over $T_{i,\epsilon}$. Also different in their form is the dependence on $\|x\|_{TV}$ and $\|\hat{x}\|_{TV}$ in Theorems 4 and 5 respectively. The presence of $\|\hat{x}\|_{TV}$ in Theorem 5 is a feature of the proof which we expect can be removed and replaced with $\|x\|_{TV}$ by proving any solution of Program (4) is necessarily bounded due to the sampling proximity condition of Definition 1. It is also worth noting that (14) avoids an unnatural η^{-1} dependence present in (11). Theorem 5 follows from the more general result of Theorem 12, whose proof is given in Section 3.4.3.

Lastly, we give a corollary of Theorems 4 and 5 where we show that, for $\delta > 0$ but sufficiently small, one can equate the δ and ϵ dependent terms in Theorems 4 and 5 to show that their respective errors approach zero as δ goes to zero.

Corollary 6. *Under the conditions in Theorems 4 and 5 and for $\sigma < \frac{1}{\sqrt{3}}$, $\Delta(T) > \sigma\sqrt{\log \frac{5}{\sigma^2}}$ and $\lambda < 0.4$, there exists δ_0 such that:*

$$d_{GW}(x, \hat{x}) \leqslant \bar{C}_1 \cdot \delta^{\frac{1}{7}}, \tag{17}$$

$$\left| \int_{T_{i,\epsilon}} \hat{x}(\mathrm{d}t) - a_i \right| \leqslant \bar{C}_2 \cdot \delta^{\frac{1}{6}}, \tag{18}$$

for all $\delta \in (0, \delta_0)$, where $\bar{C}_1$ and $\bar{C}_2$ are given in the proof in Section 3.4.4.

1.3 Organisation and summary of contributions

Organisation: The majority of our contributions were presented in the context of Gaussian windows in Section 1.2. These are particular examples of a more general theory for windows that form a *Chebyshev* *system*, commonly abbreviated as *T-system*, see Definition 7. A T-system is a collection of continuous functions that loosely behave like algebraic monomials. It is a general and widely-used concept in classical approximation theory [27, 28, 29] that has also found applications in modern signal processing [1, 2]. The framework we use for these more general results is presented in Section 2.1, the results presented in Section 2.2, and their proof sketched in Section 3. Proofs of the lemmas used to develop the results are deferred to the appendices.

Summary of contributions: We begin discussing results for general window function ϕ with Proposition 8, which establishes that for exact samples, namely $\delta = 0$, $\{\phi_j\}_{j=1}^m$ a T-system, and from $m \geqslant 2k+1$ measurements, the unique solution to Program (4) with $\delta' = 0$ is the k-sparse measure x given in (1). In other words, we show that the measurement operator Φ in (3) is an injective map from k-sparse non-negative measures on I to $\mathbb{R}^m$ when $\{\phi_j\}_{j=1}^m$ form a T-system. No minimum separation between impulses is necessary here and $\{\phi_j\}_{j=1}^m$ need only to be continuous. As detailed in Section 1.4, Proposition 8 is more general and its derivation is far simpler and more intuitive than what the current literature offers. Most importantly, no explicit regularisation is needed in Program (4) to encourage sparsity: the solution is unique.

Our main contributions are given in Theorems 11 and 12, namely that solutions to Program (4) with $\delta' > 0$ are proportionally close to the unique solution (1) with $\delta' = 0$; these theorems consider nearness in terms of the Wasserstein distance and local averages respectively. Furthermore, Theorem 11 allows x to be a general non-negative measure, and shows that solutions to Program (4) must be proportional to both how well x might be approximated by a k-sparse measure, χ, with minimum source separation 2ϵ, and a δ proportional distance between χ and solutions to Program (4). These theorems require $m \geqslant 2k+2$ and loosely-speaking the measurement apparatus forms a T*-system, which is an extension of a T-system to allow the inclusion of an additional function which may be discontinuous, and enforcing certain properties of minors of Φ. To derive the bounds in Theorems 4 and 5 we show that shifted Gaussians as given in (9) augmented with a particular piecewise constant function form a T*-system.

Lastly, in Section 2.2.1, we consider an extension of Theorem 12 where the minimum separation between sources $\Delta(T)$ is smaller than ϵ. We extend the intervals $T_{i,\epsilon}$ from (8) to $\tilde{T}_{i,\epsilon}$ in (31), where intervals $T_{i,\epsilon}$ which overlap are combined. The resulting Theorem 13 establishes that, while sources closer than ϵ may not be identifiable individually by Program (4), the local average over $\tilde{T}_{i,\epsilon}$ of both x in (1) and any solution to Program (4) will be proportionally within δ of each other.

To summarise, the results and analysis in this work simplify, generalise and extend the existing results for grid-free and non-negative super-resolution. These extensions follow by virtue of the non-negativity constraint in Program (4), rather than the common approach based on the TV norm as a sparsifying penalty. We further put these results in the context of existing literature in Section 1.4.

1.4 Comparison with other techniques

We show in Proposition 8 that a *non-negative* k-sparse discrete measure can be exactly reconstructed from $m \geqslant 2k+1$ samples (provided that the atoms form a T-system, a property satisfied by Gaussian windows for example) by solving a feasibility problem. This result is in contrast to earlier results in which a TV norm minimisation problem is solved. De Castro and Gamboa [2] proved exact reconstruction using TV norm minimisation, provided the atoms form a homogeneous T-system (one which includes the constant function). An analysis of TV norm minimisation based on T-systems was subsequently given by Schiebinger et al. in [1], where it was also shown that Gaussian windows satisfy the given conditions. We show in this paper that the TV norm can be entirely dispensed with in the case of non-negative super-resolution. Moreover, analysis of Program (4) is substantially simpler than its alternatives. In particular, Proposition 8 for noise-free super-resolution immediately follows from the standard results in the theory of T-systems. The fact that Gaussian windows form a T-system is immediately implied by well-known results in the T-system theory, in contrast to the heavy calculations involved in [1].

While neither of the above works considers the noisy setting or model mismatch, Theorems 11 and 12 in our work show that solutions to the non-negative super-resolution problem which are both stable to measurement noise and model inaccuracy can also be obtained by solving a feasibility program. The most closely related prior work is by Doukhan and Gamboa [30], in which the authors bound the maximum distance between a sparse measure and any other measure satisfying noise-corrupted versions of the same measurements. While [30] does not explicitly consider reconstruction using the TV norm, the problem is posed over probability measures, that is those with TV norm equal to one. Accuracy is captured according to the Prokhorov metric. It is shown that, for sufficiently small noise the Prokhorov distance between the measures is bounded by δ^c, where δ is the noise level and c depends upon properties of the window function. In contrast, we do not make any total variation restrictions on the underlying sparse measure, we extend to consider model inaccuracy and we consider different error metrics (the generalised Wasserstein distance and the local averaged error).

More recent results on noisy non-negative super-resolution all assume that an optimisation problem involving the TV norm is solved. Denoyelle et al. [21] consider the non-negative super-resolution problem with a minimum separation t between source locations. They analyse a TV norm-penalized least squares problem and show that a k-sparse discrete measure can be stably approximated provided the noise scales with t^{2k-1}, showing that the minimum separation condition exhibits a certain stability to noise. In the gridded setting, stability results for noisy non-negative super-resolution were obtained in the case of Fourier convolution kernels in [31] under the assumption that the spike locations satisfy a Rayleigh regularity property, and these results were extended to the case of more general convolution kernels in [32].

Super-resolution in the more general setting of *signed* measures has been extensively studied. In this case, the story is rather different, and stable identification is only possible if sources satisfy some separation condition. The required minimum separation is dictated by the resolution of the sensing system, e.g., the Rayleigh limit of the optical system or the bandwidth of the radar receiver. Indeed, it is impossible to resolve extremely close sources with equal amplitudes of opposite signs; they nearly cancel out, contributing virtually nothing to the measurements. A non-exhaustive list of references is [33, 17, 18, 19, 20, 22, 23].

In Theorem 12 we give an explicit dependence of the error on the sampling locations. This result relies on local windows, hence it requires samples near each source, and we give a condition that this distance must satisfy. The condition that there are samples near each source in order to guarantee reconstruction also appears in a recent manuscript on sparse deconvolution [34]. However, this work relies on the minimum separation and differentiability of the convolution kernel, which we overcome in Theorem 12.

2 Stability of Program (4) to inexact samples for $\phi_j(t)$ T-systems

The main results stated in the introduction, Theorems 4 and 5, are for Gaussian windows, which allows the results to omit technical details of the more general results of Theorems 11-13. These more general results apply to windows that form Chebyshev systems, see Definition 7, and an extension to T^*-systems, see Definition 9, which allows for explicit control of the stability of solutions to Program (4). These Chebyshev systems and other technical notions needed are introduced in Section 2.1 and our most general contributions are presented using these properties in Section 2.2.

2.1 Chebyshev systems and sparse measures

Before establishing stability of Program (4) to inexact samples, we show that solutions to Program (4) with $\delta' = 0$, that is with y_i in (2) having $\eta_i = 0$, has x from (1) as its unique solution once $m \geqslant 2k+1$. This result relies on $\phi_j(t)$ forming a Chebyshev system, commonly abbreviated T-system [27].

Definition 7. ***(Chebyshev, T-system [27])*** *Real-valued and continuous functions $\{\phi_j\}_{j=1}^m$ form a T-system on the interval I if the $m \times m$ matrix $[\phi_j(\tau_l)]_{l,j=1}^m$ is nonsingular for any increasing sequence $\{\tau_l\}_{l=1}^m \subset I$.*

Example of T-systems include the monomials $\{1, t, \cdots, t^{m-1}\}$ on any closed interval of the real line. In fact, T-systems generalise monomials and in many ways preserve their properties. For instance, any "polynomial" $\sum_{j=1}^m b_j\phi_j$ of a T-system $\{\phi_j\}_{j=1}^m$ has at most $m-1$ distinct zeros on I. Or, given m distinct points on I, there exists a unique polynomial in $\{\phi_j\}_{j=1}^m$ that interpolates these points. Note also that linear independence of $\{\phi_j\}$ is a necessary condition for forming a T-system, but not sufficient. Let us emphasise that T-system is a broad and general concept with a range of applications in classical approximation theory and modern signal processing. In the context of super-resolution for example, translated copies of the Gaussian window, as given in (9), and many other measurement windows form a T-system on any interval.

We refer the interested reader to [27, 29] for the role of T-systems in classical approximation theory and to [35] for their relationship to *totally positive kernels*.

2.1.1 Sparse non-negative measure uniqueness from exact samples

Our analysis based on T-Systems has been inspired by the work by Schiebinger et al. [1], where the authors use the property of T-Systems to construct the dual certificate for the spike deconvolution problem and to show uniqueness of the solution to the TV norm minimisation problem without the need of a minimum separation. The theory of T-Systems has also been used in the same context by De Castro and Gamboa in [2]. However, both [1] and [2] focus on the noise-free problem exclusively, while we will extend the T-Systems approach to the noisy case as well, as we will see later.

Our work, in part, simplifies the prior analysis considerably by using readily available results on T-Systems and we go one step further to show uniqueness of the solution of the feasibility problem, which removes the need for TV norm regularisation in the results of Schiebinger et al. [1]; this simplification in the presence of exact samples is given in Proposition 8.

Proposition 8. (Uniqueness of exactly sampled sparse non-negative measures) *Let x be a non-negative k-sparse discrete measure supported on I as given in* (1)*. Let $\{\phi_j\}_{j=1}^m$ form a T-system on I, and given $m \geqslant 2k+1$ measurements as in* (2)*, then x is the unique solution of Program (4) with $\delta' = 0$.*

Proposition 8 states that Program (4) successfully localises the k impulses present in x given only $2k+1$ measurements when $\{\phi_j\}_{j=1}^m$ form a T-system on I. Note that $\{\phi_j\}_{j=1}^m$ only need to be continuous and *no* minimum separation is required between the impulses. Moreover, as discussed in Section 1.4, the noise-free analysis here is substantially simpler as it avoids the introduction of the TV norm minimisation and is more insightful in that it shows that it is not the sparsifying property of TV minimisation which implies the result, but rather it follows from the non-negativity constraint and the T-system property, see Section 3.1.

2.1.2 T*-systems in terms of source and sample configuration

While Proposition 8 implies that T-systems ensure unique non-negative solutions, more is needed to ensure stability of these results to inexact samples; that is $\delta > 0$. This is to be expected as T-systems imply invertibility of the linear system Φ in (3) for any configuration of sources and samples as given in (5), but doe not limit the condition number of such a system. We control the condition number of Φ by imposing further conditions on the source and sample configuration, such as those stated in Conditions 2, which is analogous to imposing conditions that there exists a dual polynomial which is sufficiently bounded away from zero in regions away from sources, see Section 2.2. In particular, we extend the notion of T-system in Definition 7 to a T*-system which includes conditions on samples at the boundary of the interval, additional conditions on the window function, and a condition ensuring that there exist samples sufficiently near sources as given by the notation (8) but stated in terms of a new variable ρ so as to highlight its different role here.

Definition 9. ***(T*-system)*** *For an even integer m, real-valued functions $\{\phi_j\}_{j=0}^m$ form a T*-system on $I=[0,1]$ if the following holds for every $T=\{t_1,t_2,\ldots,t_k\}\subset I$ when $\rho>0$ is sufficiently small. For any increasing sequence $\tau=\{\tau_l\}_{l=0}^m \subset I$ such that*

- $\tau_0 = 0$, $\tau_m = 1$,
- *except exactly three points, namely τ_0, τ_m, and say $\tau_{\underline{l}} \in \operatorname{int}(I)$, the other points belong to T_ρ,*
- *every $T_{i,\rho}$ contains an even number of points,*

we have that

1. *the determinant of the $(m+1)\times(m+1)$ matrix $M_\rho := [\phi_j(\tau_l)]_{l,j=0}^m$ is positive, and*
2. *the magnitudes of all minors of M_ρ along the row containing $\tau_{\underline{l}}$ approach zero at the same rate*[2] *when $\rho \to 0$.*

[2]A function $u:\mathbb{R}\to\mathbb{R}^+$ approaches zero at the rate ρ^P when $u(\rho)=\Theta(\rho^P)$. See, for example [36], page 44.

Let us briefly discuss T*-systems as an alternative to T-systems in Definition 7. The key property of a T-system to our purpose is that an arbitrary polynomial $\sum_{j=0}^{m} b_j\phi_j$ of a T-system $\{\phi_j\}_{j=0}^{m}$ on I has at most m zeros. Polynomials of a T*-system may not have such a property as T-systems allow arbitrary configurations of points τ while T*-systems only ensure the determinant in condition 1 of Definition 9 be positive for configurations where the majority of points in τ are paired in T_ρ. However, as the analysis later shows, condition 1 in Definition 9 is designed for constructing dual certificates for Program (4). We will also see later that condition 2 in Definition 9 is meant to exclude trivial polynomials that do not qualify as dual certificates. Lastly, rather than any increasing sequence $\{\tau_l\}_{l\in[0:m]} \subset I$, Definition 9 only considers subsets τ that mainly cluster around the support T, whereas in our use all but one entry in τ is taken from the set of samples S; this is only intended to simplify the burden of verifying whether a family of functions form a T*-system. While the first and third bullet points in Definition 9 require that there need to be at least two samples per interval $T_{i,\rho}$ as well as samples which define the interval endpoints which gives a sampling complexity $m = 2k+2$, we typically require S to include additional samples, $m > 2k+2$, due to the location of T being unknown. In fact, as T is unknown, the third bullet point imposes a sampling density of m being proportional to the inverse of the minimum separation of the sources $\Delta(T)$. The additional point $\tau_{\underline{l}}$ is not taken from the set S, it instead acts as a free parameter to be used in the dual certificate. In Figure 2, we show an example of points $\{\tau_l\}_{l=0}^{10}$ which satisfy the conditions in Definition 9 for $k = 3$ sources.

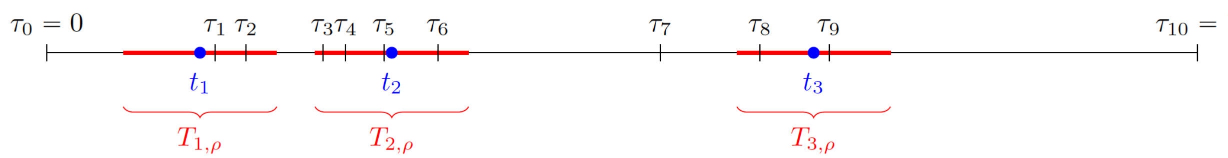


Figure 2: Example of $\{\tau_l\}_{l=1}^{m}$ that satisfy the conditions in Definition 9 for $m = 10$ and $k = 3$.

We will state some of our more general stability results for solutions of Program (4) in terms of the generalised Wasserstein distance [37] between x_1 and x_2, both non-negative measures supported on I, defined as

$$d_{GW}\left(x_1, x_2\right) = \inf_{z_1,z_2} \left(\|x_1 - z_1\|_{TV} + d_W\left(z_1, z_2\right) + \|x_2 - z_2\|_{TV}\right), \tag{19}$$

where the infimum is over all non-negative Borel measures z_1, z_2 on I such that $\|z_1\|_{TV} = \|z_2\|_{TV}$. Here, $\|z\|_{TV} = \int_I |z(\mathrm{d}t)|$ is the *total variation* of measure z, akin to the ℓ_1-norm in finite dimensions, and d_W is the standard Wasserstein distance, namely

$$d_W\left(z_1, z_2\right) = \inf_{\gamma} \int_I |\tau_1 - \tau_2| \cdot \gamma\left(\mathrm{d}\tau_1, \mathrm{d}\tau_2\right), \tag{20}$$

where the infimum is over all measures γ on $I \times I$ that produce z_1 and z_2 as marginals. In a sense, d_{GW} extends d_W to allow for calculating the distance between measures with different masses. [3]

Moreover, in some of our most general results we consider the extension to where x need not be a discrete measure, see Theorem 11. In that setting, we introduce an intermediate k-discrete measure which approximates x in the d_{GW} metric. That is, given an integer k and positive ϵ, let $x_{k,\epsilon}$ be a k-sparse 2ϵ-separated measure supported on $T_{k,\epsilon} \subset \mathrm{int}(I)$ of size k and with $\Delta(T_{k,\epsilon}) \geqslant 2\epsilon$ such that, for $\beta \geqslant 1$,

$$R(x, k, \epsilon) := d_{GW}(x, x_{k,\epsilon}) \leqslant \beta \inf_{\chi} d_{GW}(x, \chi), \tag{21}$$

where the infimum is over all k-sparse 2ϵ-separated non-negative measures supported on $\mathrm{int}(I)$ and the parameter β allows for near projections of x onto the space of k-sparse 2ϵ-separated measures.

Lastly, we also assume that the measurement operator Φ in (3) is Lipschitz continuous, namely there exists $L \geqslant 0$ such that

$$\left\|\int_I \Phi(t)(x_1(\mathrm{d}t) - x_2(\mathrm{d}t))\right\|_2 \leqslant L \cdot d_{GW}(x_1, x_2), \tag{22}$$

for every pair of measures x_1, x_2 supported on I.

[3] In [37], the authors consider the p-Wasserstein distance, where popular choices of p are 1 and 2. In our work, we only use the 1-Wasserstein distance.

2.2 Stability of Program (4)

Equipped with the definitions of T and T*-systems, Definitions 7 and 9 respectively, we are able to characterise any solution to Program (4) for $\phi_j(t)$ which form a T-system and suitable source and sample configurations (5). We control the stability to inexact measurements by introducing two auxiliary functions in Definition 10, which quantify the dual polynomials $q(t)$ and $q^\pi(t)$ associated with Program (4) to be at least $\bar{f}$ away from the necessary constraints for all values of t at least ϵ away from the sources. Specifically, for F and F^π defined below, we will require that $q(t) \geqslant F(t)$ and $q^\pi(t) \geqslant F^\pi(t)$ for all $t \in [0,1]$.

Definition 10. ***(Dual polynomial separators)*** *Let $f : \mathbb{R} \to \mathbb{R}_+$ be a bounded function with $f(0) = 0$, $\bar{f}, f_0, f_1$ be positive constants, and $\{T_{i,\epsilon}\}_{i=1}^k$ the neighbourhoods as defined in (8). We then define*

$$F(t) := \begin{cases} f_0, & t = 0, \\ f_1, & t = 1, \\ f(t - t_i), & \text{when there exists } i \in [k] \text{ such that } t \in T_{i,\epsilon}, \\ \bar{f}, & \text{elsewhere on } int(I). \end{cases} \tag{23}$$

Moreover, let $\pi \in \{\pm 1\}^k$ be an arbitrary sign pattern. We define F^π as

$$F^\pi(t) := \begin{cases} f_0, & t = 0, \\ f_1, & t = 1, \\ \pm 1 - f\left(t - t_i\right), & \text{when there exists } i \in [k] \text{ such that } t \in T_{i,\epsilon} \text{ and } \pi_i = \pm 1, \\ -\bar{f}, & \text{everywhere else on } int(I). \end{cases} \tag{24}$$

We defer the introduction of dual polynomials q and q^π and the precise role of the above dual polynomial separators to Section 3, but state our most general results characterising the solutions to Program (4) in terms of these separators.

Theorem 11. (Wasserstein stability of Program (4) for $\phi_j(t)$ a T-system) *Consider a non-negative measure x supported on $\mathrm{int}(I) = (0,1)$ and assume that the measurement operator Φ is L-Lipschitz, see (3) and (22). Consider a k-sparse non-negative discrete measure χ supported on $T = \{t_i\}_{i=1}^k \subset \mathrm{int}(I)$ and fix $\epsilon \leqslant \Delta(T)/2$, see (6), and consider functions $F(t)$ and $F^\pi(t)$ as defined in Definition 10. For $m \geqslant 2k+2$, suppose that*

- *$\{\phi_j\}_{j=1}^m$ form a T-system on I,*
- *$\{F\} \cup \{\phi_j\}_{j=1}^m$ form a T*-system on I, and*
- *$\{F^\pi\} \cup \{\phi_j\}_{j=1}^m$ form a T*-system on I for any sign pattern π.*

Let $\widehat{x}$ be a solution of Program (4) with

$$\delta' = \delta + L \cdot d_{GW}\left(x, \chi\right). \tag{25}$$

Then there exist vectors $b, \{b^\pi\}_\pi \subset \mathbb{R}^m$ such that

$$d_{GW}\left(x, \widehat{x}\right) \leqslant \left(\left(6 + \frac{2}{\bar{f}}\right) \|b\|_2 + 6 \min_\pi \|b^\pi\|_2\right) \delta' + \epsilon \|\chi\|_{TV} + d_{GW}(x, \chi). \tag{26}$$

where the minimum is over all sign patterns π and the vectors $b, \{b^\pi\}_\pi \subset \mathbb{R}^m$ above are the vectors of coefficients of the dual polynomials q and q^π associated with Program (4), see Lemmas 16 and 17 in Section 3 for their precise definitions.

Theorem 4 follows from Theorem 11 by considering $\phi_j(t)$ Gaussian as stated in (9) which is known to be a T-system [27], and introducing Conditions 2 on the source and sample configuration (5) such that the conditions of Theorem 11 can be proved and the dual coefficients b and b^π bounded; the details of these proofs and bounds are deferred to Section 3 and the appendices.

The particular form of F and $\{F^\pi\}_\pi$ in Theorem 11, constant away from the support T of $x_{k,\epsilon}$, is purely to simplify the presentation and proofs. Note also that the error $d_{GW}(x, \widehat{x})$ depends both on the noise level δ and the residual $R(x, k, \epsilon)$, not unlike the standard results in finite-dimensional sparse recovery and compressed sensing [24, 38]. In particular, when $\delta, \epsilon, R(x, k, \epsilon) \to 0$, we approach the setting of Proposition 8, where we have uniqueness of k-sparse non-negative measures from exact samples.

Note that the noise level δ and the residual $R(x, k, \epsilon)$ are not independent; that is, δ specifies confidence in the samples and the model for how the samples are taken while $R(x, k, \epsilon)$ reflects nearness to the model of k-discrete measures. Corollary 6 show that the parameter ϵ can be removed, for $\phi_j(t)$ shifted Gaussians, in the setting where x is k-discrete, that is $R(x, k, \epsilon) = 0$, in which case $d_{GW}(x, \hat{x})$ is bounded by $\mathcal{O}(\delta^{1/7})$.

The more general variant of Theorem 5 follows from Theorem 12 by introducing alternative conditions on the source and sample configuration and omitting the need for the functions F^π, which is the cause of the unnatural η^{-1} dependence in Theorem 4.

Theorem 12. (Average stability for Program (4) **for $\phi_j(t)$ a T-system)** *Let $\hat{x}$ be a solution of Program* (4) *and consider the function $F(t)$ as defined in Definition 10. Suppose that:*

- $\{\phi_j\}_{j=1}^m$ *form a T-system on I,*
- $\{F\} \cup \{\phi_j\}_{j=1}^m$ *form a T*-system on I, and*
- $\Delta = \Delta(T)$ *and $\lambda = \lambda_0 \in (0, 1/2)$ from Definition 1 satisfy*

$$\phi(\lambda\Delta) = \phi(\Delta - \lambda\Delta) + \phi(\Delta + \lambda\Delta) + \frac{1}{\Delta}\int_{\Delta-\lambda\Delta}^{1/2-\lambda\Delta} \phi(x)\,\mathrm{d}x + \frac{1}{\Delta}\int_{\Delta+\lambda\Delta}^{1/2+\lambda\Delta} \phi(x)\,\mathrm{d}x. \tag{27}$$

Then, for any $\epsilon \in (0, \Delta/2)$ and for all $i \in [k]$,

$$\left|\int_{T_{i,\epsilon}} \hat{x}(\mathrm{d}t) - a_i\right| \leqslant \left(2\left(1 + \frac{\phi^\infty \|b\|_2}{\bar{f}}\right) \cdot \delta + L\|\hat{x}\|_{TV} \cdot \epsilon\right) \sum_{j=1}^{k} (A^{-1})_{ij}, \tag{28}$$

$$\int_{T_\epsilon^C} \hat{x}(\mathrm{d}t) \leqslant \frac{2\|b\|_2\delta}{\bar{f}}, \tag{29}$$

where:

- $b \in \mathbb{R}^m$ *is the same vector of coefficients of the dual certificate q as in Theorem 11 and $\bar{f}$ is given in Definition 10, which is used to construct the dual certificate q, as described in Lemma 16 in Section 3,*
- $\phi^\infty = \max_{s,t \in I} |\phi(s - t)|$,
- *L is the Lipschitz constant of ϕ,*
- $A \in \mathbb{R}^{k \times k}$ *is the matrix*

$$A = \begin{bmatrix} |\phi_1(t_1)| & -|\phi_1(t_2)| & \dots & -|\phi_1(t_k)| \\ -|\phi_2(t_1)| & |\phi_2(t_2)| & \dots & -|\phi_2(t_k)| \\ \vdots & \vdots & \ddots & \vdots \\ -|\phi_k(t_1)| & -|\phi_k(t_2)| & \dots & |\phi_k(t_k)| \end{bmatrix}, \tag{30}$$

with $\phi_i(t_i) = \phi(t_i - s_{l(i)})$ evaluated at $s_{l(i)}$ as defined in (7).

Theorem 12 bounds the difference between the average over the interval $T_{i,\epsilon}$ of any solution to Program (4) and the discrete measure whose average is simply a_i. The condition on λ to satisfy (27) is used to ensure the matrix from (30) is strictly diagonally dominant. It relies on the windows $\phi_j(t)$ being sufficiently localised about zero. Though Theorem 12 explicitly states that the location of the closest samples to each source is less than $\lambda_0\Delta(T)$, this can be achieved without knowing the locations of the sources by placing the samples uniformly at interval $2\lambda_0\Delta(T)$ which gives a sampling complexity of $m = (2\lambda_0\Delta(T))^{-1}$. Lastly, a similar bound on the integral of $\hat{x}$ over T_ϵ^C is given by Lemma 16 in Section 3.

2.2.1 Clustering of indistinguishable sources

Theorems 11 and 12 give uniform guarantees for all sources in terms of the minimum separation condition $\Delta(T)$, which measures the worst proximity of sources. One might imagine that, for example, if all but two sources are sufficiently well separated, then Theorem 12 might hold for the sources that are well separated; moreover, assuming ϵ is fixed, then if two sources t_i and t_{i+1} with magnitudes a_i and a_{i+1} are closer than 2ϵ, namely $|t_i - t_{i+1}| < 2\epsilon$, we might imagine that a variant of Theorem 12 might hold but with sources t_i and t_{i+1} approximated with source t_ξ near t_i and t_{i+1} and with $a_\xi = a_i + a_{i+1}$.

In this section we extend Theorem 12 to this setting by considering ϵ fixed and alternative intervals $\{\tilde{T}_{i,\epsilon}\}_{i=1}^{\tilde{k}}$ a partition of T_ϵ such that each $\tilde{T}_{i,\epsilon}$ contains a group of consecutive sources $t_{i1}, \dots, t_{ir_i}$ (with weights $a_{i1}, \dots, a_{ir_i}$ respectively) which are within at most 2ϵ of each other. Define

$$\tilde{T}_{i,\epsilon} = \bigcup_{l=1}^{k_i} T_{il,\epsilon}, \quad \text{where } t_{il} \in T_{il,\epsilon} \quad \text{and} \quad |t_{il+1} - t_{il}| < 2\epsilon, \quad \forall l \in [k_i - 1], \tag{31}$$

for $\sum_{i=1}^{\tilde{k}} k_i = k$, so that we have

$$T_\epsilon = \bigcup_{i=1}^{\tilde{k}} \tilde{T}_{i,\epsilon} \quad \text{and} \quad \tilde{T}_{i,\epsilon} \bigcap \tilde{T}_{j,\epsilon} \neq \varnothing, \quad \forall i \neq j. \tag{32}$$

Theorem 13. (Average stability for Program (4): grouped sources) *Let $\hat{x}$ be a solution of Program (4) and $I = [0,1]$ be partitioned as described by (31). If the samples are placed uniformly at interval $2\lambda_0\epsilon$ where $\lambda = \lambda_0$ satisfies (27) with $\Delta = 2\epsilon$, then there exist $\{\xi_i\}_{i \in [\tilde{k}]}$ with $\xi_i \in \tilde{T}_{i,\epsilon}$ such that*

$$\left| \int_{\tilde{T}_{i,\epsilon}} \hat{x}(\mathrm{d}t) - \sum_{r=1}^{k_i} a_{ir} \right| \leqslant \left(2\left(1 + \frac{\phi^\infty \|b\|_2}{\bar{f}}\right) \cdot \delta + (2k-1)L\|\hat{x}\|_{TV} \cdot \epsilon \right) \sum_{j=1}^{k} (\tilde{A}^{-1})_{ij}, \tag{33}$$

where the constants are the same as in (12) and the matrix $\tilde{A} \in \mathbb{R}^{\tilde{k} \times \tilde{k}}$ is

$$\tilde{A} = \begin{bmatrix} |\phi_1(\xi_1)| & -|\phi_1(\xi_2)| & \dots & -|\phi_1(\xi_{\tilde{k}})| \\ -|\phi_2(\xi_1)| & |\phi_2(\xi_2)| & \dots & -|\phi_2(\xi_{\tilde{k}})| \\ \vdots & \vdots & \ddots & \vdots \\ -|\phi_{\tilde{k}}(\xi_1)| & -|\phi_{\tilde{k}}(\xi_2)| & \dots & |\phi_{\tilde{k}}(\xi_{\tilde{k}})| \end{bmatrix}.$$

Note that Lemma 16 still holds if we replace any group of sources from an interval $\tilde{T}_{i,\epsilon}$ with some $\xi_i \in \tilde{T}_{i,\epsilon}$, so the bound from Lemma 16 on T_ϵ^C remains valid without modification.

As an exemplar source location where Theorem 13 might be applied, consider the situation where the k source locations comprising T are drawn uniformly at random in $(0,1)$, where we have that (from [39] page 42, Exercise 22)

$$P(\Delta(T) > \theta) = [1 - (k+1)\theta]^k, \quad \theta \in \left[0, \frac{1}{k+1}\right].$$

Then, the cumulative distribution function is

$$F(\theta) = P(\Delta(T) \leqslant \theta) = 1 - [1 - (k+1)\theta]^k,$$

and so the distribution of $\Delta(T)$ is

$$f(\theta) = F'(\theta) = (k+1)k[1 - (k+1)\theta]^{k-1},$$

with an expectation of

$$E(\Delta(T)) = \int_0^{\frac{1}{k+1}} P(\Delta(T) > \theta)\,\mathrm{d}\theta = \frac{1}{(k+1)^2}. \tag{34}$$

That is, for x from (1) with sources T drawn uniformly at random in $(0,1)$, the expected value of $\Delta(T)$ is given by (34) and, in Theorems 11 and 12, the corresponding number of samples m would scale quadratically with the number of sources k due to the scaling of $m \sim \Delta(T)^{-1}$. Alternatively, Theorem 13 allows meaningful results for m proportional to k by grouping the sources that are within k^{-2} of one another.

3 Dual polynomials for stability of non-negative measures

The results in Section 2 are developed by establishing dual polynomials of Program (4) which are non-negative except at the source locations, which implies that the solution to Program (4) is unique when $\delta' = 0$, see Proposition 8, and then showing that the dual polynomials are sufficiently non-negative away from the source locations and using this property to develop Theorems 11, 12, and 13. In this section we state the key lemmas used to prove the aforementioned results and then bound the quantities involving the dual polynomials in order to establish Theorems 4 and 5 for the case of Gaussian windows.

3.1 Uniqueness of non-negative sparse measures from exact samples: proof of Proposition 8

Proposition 8 states that if x is a non-negative k-sparse discrete measure supported on I, see (1), provided $m \geqslant 2k+1$ and $\{\phi_j\}_{j=1}^m$ are a T-system, then x is the unique non-negative solution to Program (4) with $\delta' = 0$. This follows from the existence of a dual polynomial as stated in Lemma 14, the proof of which is given in Appendix A.

Lemma 14. (Dual polynomial and uniqueness of non-negative sparse measure equivalence) *Let x be a non-negative k-sparse discrete measure supported on I, see (1). Then, x is the unique solution of Program (4) with $\delta' = 0$ if*

- *the $k \times m$ matrix $[\phi_j(t_i)]_{i=1,j=1}^{i=k,j=m}$ is full rank, and*
- *there exist real coefficients $\{b_j\}_{j=1}^m$ and $q(t) = \sum_{j=1}^m b_j \phi_j(t)$ such that $q(t)$ is non-negative on I and vanishes only on T.*

Figure 3 shows an example of such a dual certificate using Gaussian ϕ_j as defined in (9). It remains to show that such a dual polynomial exists. To do this, we employ the concept of T-system introduced in Definition 7. Of particular interest to us is Theorem 5.1 in [27], slightly simplified below, which immediately proves Proposition 8.

Lemma 15. (Dual polynomial existence for T-systems) [27, Theorem 5.1, pp. 28] *With $m \geqslant 2k+1$, suppose that $\{\phi_j\}_{j=1}^m$ form a T-system on I. Then there exists a polynomial $q(t) = \sum_{j=1}^m b_j \phi_j(t)$ that is non-negative on I and vanishes only on T.*

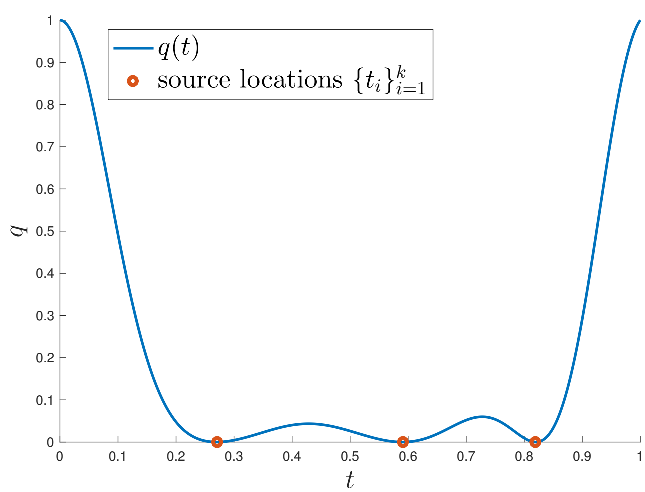


Figure 3: Example of dual certificate $q(t)$ required in Lemma 14. Here, we have $k = 3$ and $t_1 = 0.27, t_2 = 0.59$ and $t_3 = 0.82$.

3.2 Stabilising dual polynomials for non-negative sparse measures: proof of Theorem 11

We develop the proof of Theorem 11 by using a dual polynomial analogous to that in Lemma 14, but with further guarantees that away from the sources the dual polynomial must be sufficiently bounded away from the constraint bounds. However, first, let us bring the generality of Theorem 11 to discrete measures by introducing an intermediate measure χ which is k-discrete and whose support is 2ϵ separated. Noting that the measurement operator Φ is L-Lipschitz, see (22), and using the triangle inequality, it follows that

$$\begin{aligned}\left\| y - \int_I \Phi(t)\chi(\mathrm{d}t) \right\|_2 &\leqslant \left\| y - \int_I \Phi(t)x(\mathrm{d}t) \right\|_2 + \left\| \int_I \Phi(t)(x(\mathrm{d}t) - \chi(\mathrm{d}t)) \right\|_2 \\ &\leqslant \delta + L \cdot d_{GW}(x, \chi) =: \delta'. \end{aligned} \tag{35}$$

Therefore, a solution $\widehat{x}$ of Program (4) with δ' specified above can be considered as an estimate of χ. In the rest of this section, we first bound the error $d_{GW}(\chi, \widehat{x})$ and then use the triangle inequality to control $d_{GW}(x, \widehat{x})$.

To control $d_{GW}(\chi, \widehat{x})$ in turn, we will first show that the existence of certain dual certificates leads to stability of Program (4). Then we see that these certificates exist under certain conditions on the measurement operator Φ. Turning now to the details, the following result is slightly more general than the one in [40] and guarantees the stability of Program (4) if a prescribed dual certificate q exists. The proof is provided in Appendix B.

Lemma 16. (Error away from the support) *Let $\widehat{x}$ be a solution of Program (4) with δ' specified in (35) and set $h = \widehat{x} - \chi$ to be the error. Consider $F(t)$ given in Definition 10 and suppose that there exist a positive $\epsilon \leqslant \Delta(T)/2$, real coefficients $\{b_j\}_{j=1}^m$, and a polynomial $q = \sum_{j=1}^m b_j\phi_j$ such that*

$$q(t) \geqslant F(t),$$

where the equality holds on T. Then we have that

$$\bar{f}\int_{T_\epsilon^C} h(\mathrm{d}t) + \sum_{i=1}^{k}\int_{T_{i,\epsilon}} f(t-t_i)\, h(\mathrm{d}t) \leqslant 2\|b\|_2\delta', \tag{36}$$

where $b \in \mathbb{R}^m$ is the vector formed by the coefficients $\{b_j\}_{j=1}^m$.

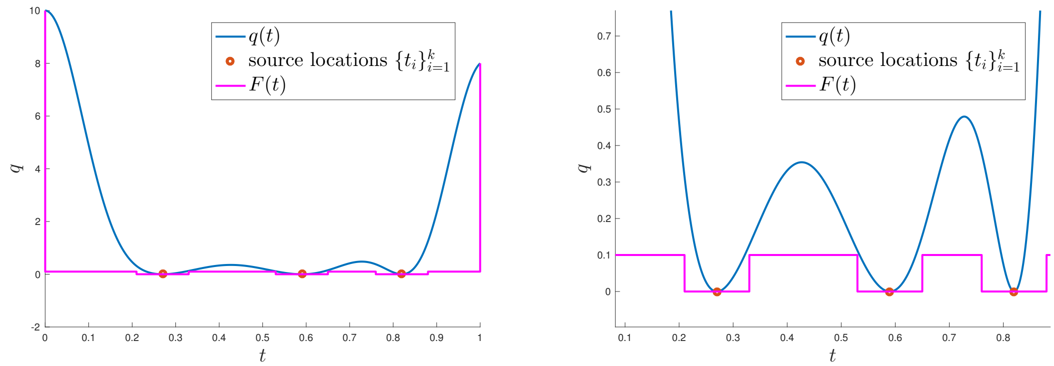
There is a natural analogy here with the case of exact samples. In the setting where $\eta_i = 0$ in (2), the dual certificate q in Lemma 14 was required to be positive off the support T. In the presence of inexact samples however, Lemma 16 loosely-speaking requires the dual certificate to be bounded[4] *away* from zero (see example in Figure 4) for $t \in T_\epsilon^C$.

Note also that Lemma 16 controls the error h away from the support T, as it guarantees that

$$\int_{T_\epsilon^C} h(\mathrm{d}t) \leqslant \frac{2\|b\|_2\delta'}{\bar{f}}, \tag{37}$$

if the dual certificate q exists. Indeed, (37) follows directly from (36) because the sum in (36) is non-negative. This is in turn the case because $f(0) = 0$ and the error h is non-negative off the support T. Another key observation is that Lemma 16 is almost silent about the error near the impulses in χ. Indeed, because $f(0) = 0$ by assumption, (36) completely fails to control the error on the support T. However, as the next result states, Lemma 16 can be strengthened near the support provided that an additional dual certificate q^0 exists. The proof, given in Appendix C, is not unsimilar to the analysis in [41].

[4]Note the scale invariance of (36) under scaling of f and $\bar{f}$. Indeed, by changing $f, \bar{f}$ to $\alpha f, \alpha\bar{f}$ for positive α, the proof dictates that b changes to αb and consequently α cancels out from both sides of (36). Similarly, if we change Φ to $\alpha\Phi$ in (3), the proof dictates that b changes to b/α and α again cancels out, leaving (36) unchanged.



(a) (b)

Figure 4: Example of dual certificate $q(t)$ that satisfies the conditions in Lemma 16 where the window function is the Gaussian kernel $\phi(t) = e^{-t^2/\sigma^2}$. We take $k = 3$ and $t_i \in \{0.27, 0.59, 0.82\}$ and the function $F(t)$ such that $q(t) \geqslant F(t)$.

Lemma 17. (Error near the support) *Suppose that the dual certificate in Lemma 16 exists. Consider a function $F^0(t)$ defined like $F^\pi(t)$ in Definition 10 for the sign pattern π^0 such that*

$$\pi_i^0 = \begin{cases} 1, & \text{when there exists } i \in [k] \text{ such that } t \in T_{i,\epsilon} \text{ and } \int_{T_{i,\epsilon}} h(\mathrm{d}t) > 0, \\ -1, & \text{when there exists } i \in [k] \text{ such that } t \in T_{i,\epsilon} \text{ and } \int_{T_{i,\epsilon}} h(\mathrm{d}t) \leqslant 0, \end{cases}$$

and suppose also that there exist real coefficients $\{b_j^0\}_{j\in[m]}$ and a polynomial $q^0 = \sum_{j=1}^m b_j^0 \phi_j$ such that

$$q^0(t) \geqslant F^0(t),$$

where the equality holds on T. Then we have that

$$\sum_{i=1}^k \left| \int_{T_{i,\epsilon}} h(\mathrm{d}t) \right| \leqslant 2 \left(\|b\|_2 + \|b^0\|_2 \right) \delta'. \tag{38}$$

In words, Lemma 17 controls the error h near the support T, provided that a certain dual certificate q^0 exists (see example in Figure 5). Note that (38) does not control the *mass* of the error, namely $\sum_i \int_{T_{i,\epsilon}} |h(\mathrm{d}t)| = \int_{T_\epsilon} |h(\mathrm{d}t)|$, but rather it controls $\sum_i |\int_{T_{i,\epsilon}} h(\mathrm{d}t)|$. Of course, the latter is always bounded by the former, that is

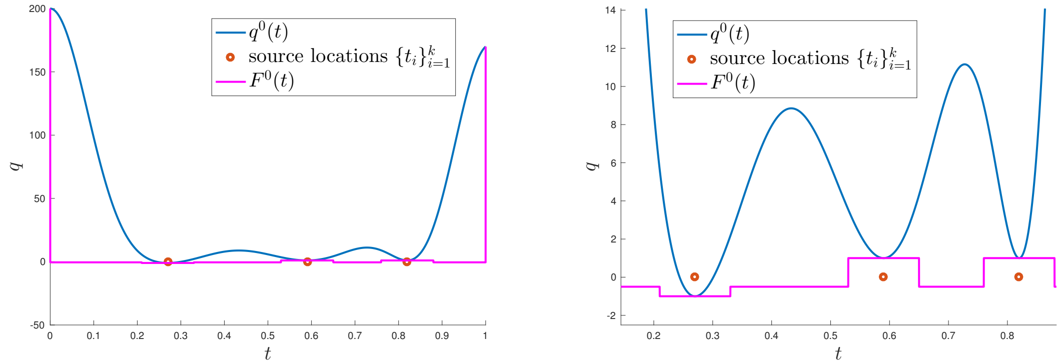
$$\sum_{i=1}^k \left| \int_{T_{i,\epsilon}} h(\mathrm{d}t) \right| \leqslant \int_{T_\epsilon} |h(\mathrm{d}t)|. \tag{39}$$

However, the two sides of (39) might differ significantly. For instance, it might happen that the solution $\widehat{x}$ returns a slightly incorrect impulse at $t_i + \epsilon/4$ (rather than t_i) but with the correct amplitude of a_i. As a result, the mass of the error is large in this case ($\int_{T_{i,\epsilon}} |h(\mathrm{d}t)| = 2a_i$) but the left-hand side of (39) vanishes, namely $|\int_{T_{i,\epsilon}} h(\mathrm{d}t)| = 0$. Note that we cannot hope to strengthen (38) by replacing its left-hand side with the mass of the error, namely $\int_{T_\epsilon} |h(\mathrm{d}t)|$. This is the case mainly because the total variation is not the appropriate error metric for this context.

Indeed, while the mass of the error $\int_I |h(\mathrm{d}t)|$ might not be small in general, we can instead control the generalised Wasserstein distance between the true and estimated measures, namely x and $\widehat{x}$, see (19) and (20). The following result is proved by combining Lemmas 16 and 17, see Appendix D.

Lemma 18. (Stability of Program (4) in the Generalised Wasserstein distance) *Suppose that the dual certificates in Lemmas 16 and 17 exist. Then it holds that*

$$d_{GW}(\chi, \widehat{x}) \leqslant \left(\left(6 + \frac{2}{\bar{f}} \right) \|b\|_2 + 6\|b^0\|_2 \right) \delta' + \epsilon \|\chi\|_{TV}. \tag{40}$$



(a) (b)

Figure 5: Example of dual certificate $q^0(t)$ that satisfies the conditions in Lemma 17 where the window function is the Gaussian kernel $\phi(t) = e^{-t^2/\sigma^2}$. We take $k = 3$ and $t_i \in \{0.27, 0.59, 0.82\}$ and the function $F^0(t)$ for the sign pattern $\pi^0 = \{-1, 1, 1\}$ such that $q^0(t) \geqslant F^0(t)$. For the Gaussian kernel, the existence of $q^\pi(t)$ for any sign pattern π guarantees the existence of $q^0(t)$ in Lemma 17.

An application of triangle inequality now yields that

$$\begin{aligned} d_{GW}(x, \widehat{x}) &\leqslant d_{GW}(x, \chi) + d_{GW}(\chi, \widehat{x}) \\ &\leqslant d_{GW}(x, \chi) + \left(\left(6 + \frac{2}{\bar{f}} \right) \|b\|_2 + 6\|b^0\|_2 \right) \delta' + \epsilon \|\chi\|_{TV}. \qquad \text{(see Lemma 18)} \end{aligned} \tag{41}$$

In words, Program (4) is stable x if the certificates q and q^0 exist. Let us now study the existence of these certificates. Proposition 19, proved in Appendix E, guarantees the existence of the dual certificate q required in Lemma 16 and heavily relies on the concept of T*-system in Definition 9. We remark that the proof benefits from the ideas in [27]. Similarly, Proposition 20, stated without proof, ensures the existence of the certificate q^0 required in Lemma 17.

Proposition 19. (Existence of q) *For $m \geqslant 2k+2$, suppose that $\{\phi_j\}_{j=1}^m$ form a T-system on I and that $\{F\} \cup \{\phi_j\}_{j=1}^m$ form a T*-system on I, where $F(t)$ is the function given in Definition 10. Then the dual certificate q in Lemma 16 exists and consequently Program (4) is stable in the sense that (36) holds.*

Note that to ensure the success of Program (4), it suffices that there exists a polynomial $q = \sum_{j=1}^m b_j \phi_j$ such that $q(t) \geqslant F(t)$ with the equality met on the support T, see Lemma 16. Equivalently, it suffices that there exists a non-negative polynomial $\dot{q} = -b_0 F + \sum_{j=1}^m b_j \phi_j$ that vanishes on T such that $b_0 > 0$ and at least one other coefficient, say b_{j_0}, is nonzero. This situation is reminiscent of Lemma 15. In contrast to Lemma 15, however, such $\dot{q}$ exists when $\{F\} \cup \{\phi_j\}_{j=1}^m$ is a T*-system rather than a T-system. The more subtle T*-system requirement is to avoid trivial or unbounded polynomials.

Proposition 20. (Existence of q^0) *For $m \geqslant 2k+2$, suppose that $\{\phi_j\}_{j=1}^m$ form a T-system on I and that $\{F^0\} \cup \{\phi_j\}_{j=1}^m$ form a T*-system on I, where $F^0(t)$ is the function defined in Lemma 17. Then the dual certificate q^0 in Lemma 17 exists and consequently Program (4) is stable in the sense that (38) holds.*

Having constructed the necessary dual certificates in Propositions 19 and 20, the proof of Theorem 11 is now complete in light of (41).

3.3 Proof of Theorem 12 (Average stability for Program (4))

In this section we give an overview of the main ideas involved in proving Theorem 12. To start with, let $A \in \mathbb{R}^{k \times k}$ be defined as in (30):

$$A = \begin{bmatrix} |\phi_1(t_1)| & -|\phi_1(t_2)| & \dots & -|\phi_1(t_k)| \\ -|\phi_2(t_1)| & |\phi_2(t_2)| & \dots & -|\phi_2(t_k)| \\ \vdots & \vdots & \ddots & \vdots \\ -|\phi_k(t_1)| & -|\phi_k(t_2)| & \dots & |\phi_k(t_k)| \end{bmatrix},$$

where $\phi_i(t_i) = \phi(t_i - s_{l(i)})$ is evaluated at the source t_i and the closest sample to it, as defined in (7).

The proof of Theorem 12 consists of two steps. We first show that we can bound the error if the matrix A is strictly diagonally dominant. It is easy to see that, if the window function ϕ is localised, then the entries on the main diagonal are larger in absolute value than the off-diagonal entries. If, moreover, we choose the sampling locations $\{s_i\}_{i\in[m]}$ such that A is strictly diagonally dominant (which means that for each source, there is a sampling location that is "close enough" to it), then the bound (28) is guaranteed.

Proposition 21. *For each source t_i, select $s_{l(i)}$ to be the closest sample as defined in* (7)*, and define the matrix A in* (30) *using the sequences $\{t_i\}_{i=1}^k$, $\{s_{l(i)}\}_{i=1}^k$. If A is strictly diagonally dominant, then the error around the support is bounded according to* (28).

Then, we want to go further and see what it means exactly for A to be strictly diagonally dominant, so the second step in the proof of Theorem 12 is to give an upper bound for the distance between the sources $\{t_i\}_{i\in[k]}$ and the closest sampling locations $\{s_{l(i)}\}_{i\in[k]}$ such that A is strictly diagonally dominant.

Given an even positive function ϕ that is localised at 0 and with fast decay, let Δ and λ as given in Definition 1, so

$$|t_i - s_{l(i)}| \leqslant \lambda\Delta \tag{42}$$

We want to find λ_0 such that

$$\phi(s_{l(i)} - t_i) \geqslant \sum_{j\neq i} \phi(s_{l(i)} - t_j), \quad \forall\lambda \in (0, \lambda_0), \quad \forall i \in [k], \tag{43}$$

namely, we want the matrix A to be strictly diagonally dominant. From the conditions (43), we can obtain a more general equality, depending on ϕ and Δ, that λ_0 must satisfy such that, for $\lambda < \lambda_0$, A is strictly diagonally dominant. The equality is given by (27):

$$\phi(\lambda_0\Delta) = \phi(\Delta - \lambda_0\Delta) + \phi(\Delta + \lambda_0\Delta) + \frac{1}{\Delta}\int_{\Delta-\lambda_0\Delta}^{1/2-\lambda_0\Delta} \phi(x)\,\mathrm{d}x + \frac{1}{\Delta}\int_{\Delta+\lambda_0\Delta}^{1/2+\lambda_0\Delta} \phi(x)\,\mathrm{d}x.$$

Proposition 22. *Let $\lambda_0 \in (0, \frac{1}{2})$ such that $\left|t_i - s_{l(i)}\right| \leqslant \lambda_0\Delta$ for all $i \in [k]$. If λ_0 satisfies* (27)*, then the matrix A defined in* (30) *is strictly diagonally dominant.*

Finally, we note that the proof of Theorem 13 involves the same ideas as the ones discussed in this section, with a few modifications. The detailed proofs of Proposition 21 and Proposition 22 are given in Appendices F and G respectively. The proof of Theorem 13 is similar to the proof presented in the current section, so we only show the differences in Appendix H.

3.4 Proofs of Theorems 4 and 5 (Gaussian with sparse measure)

In this section we give the main steps taken to obtain the explicit bounds in Theorems 4 and 5 for the Gaussian window function. These are particular cases of the more general Theorems 11 and 12 respectively, where the window function is taken to be the Gaussian $\phi_j(t) = e^{-(t-s_j)^2/\sigma^2}$, given in (9), and the true measure x is a k-discrete non-negative measure as in (1).

3.4.1 Bounds on the coefficients of the dual certificates for Gaussian window

We will now give explicit bounds on the vectors of coefficients $\|b\|_2$ and $\|b^\pi\|_2$ of the dual certificates q and q^π from Lemmas 16 and 17 in terms of the parameters of the problem k, T, S and σ (the width of the Gaussian window).

Firstly, we introduce a more specific form of the dual polynomial separators $F(t)$, $F^\pi(t)$ from Definition 10. Here, we take $f(t) = 0$ for $t \in (-\epsilon, \epsilon)$ and $\bar{f}, f_1$ positive constants with $\bar{f} < 1$. Then, f_0 is defined to be greater that both $\bar{f}$ and f_1, with the exact relationship between f_0 and $\bar{f}$ given in the proof of Lemma 23. Therefore, for $\epsilon > 0$ and a sign pattern $\pi \in \{\pm 1\}^k$, $F(t)$ and $F^\pi(t)$ are:

$$F(t) = \begin{cases} f_0, & t = 0, \\ f_1, & t = 1, \\ 0, & \text{when there exists } i \in [k] \text{ such that } t \in T_{i,\epsilon}, \\ \bar{f}, & \text{elsewhere on } I. \end{cases} \tag{44}$$

$$F^\pi(t) = \begin{cases} f_0, & t = 0, \\ f_1, & t = 1, \\ \pm 1, & \text{when there exists } i \in [k] \text{ such that } t \in T_{i,\epsilon} \text{ and } \pi_i = \pm 1, \\ -\bar{f}, & \text{elsewhere on } I. \end{cases} \tag{45}$$

With the above definitions, both Theorem 11 and Theorem 12 require that $\{F\} \cup \{\phi_j\}_{j=1}^m$ form a T*-system on I. Likewise, Theorem 11 requires that $\{F^\pi\} \cup \{\phi_j\}_{j=1}^m$ form a T*-system for any sign pattern π. We show in Lemma 23 that both these requirements are satisfied for the choice in (9) of $\phi_j(t) = e^{-(t-s_j)^2/\sigma^2}$. The proof is given in Appendix J.

Lemma 23. *Consider the function $F(t)$ defined in (44) and suppose that $m \geqslant 2k+2$. Then $\{F\} \cup \{\phi_j\}_{j=1}^m$ form a T*-system on I, with ϕ extended totally positive, even Gaussian and ϕ_j defined as in (9), provided that $f_0 \gg \bar{f}$, $f_0 \gg f_1$ and $\bar{f}, f_0, f_1 \gg 0$. These requirements are made precise in the proof and are dependent on ϵ. Moreover, for an arbitrary sign pattern π and F^π as defined in (45), $\{F^\pi\} \cup \{\phi_j\}_{j=1}^m$ form a T*-system on I when, in addition, $f_0 \gg 1$.*

In this setting, consider a subset of $m = 2k+2$ samples $\{s_j\}_{j=1}^m \subset S$ (since in the proof of Lemma 23 we select the $2k+2$ samples that are the closest to the sources) such that they satisfy Conditions 2. Therefore, we have that $s_1 = 0$, $s_m = s_{2k+2} = 1$, and

$$|s_{2i} - t_i| \leqslant \eta, \qquad s_{2i+1} - s_{2i} = \eta, \qquad \forall i \in [k], \tag{46}$$

for a small $\eta \leqslant \sigma^2$, see (9). That is, we collect two samples close to each impulse t_i in x, one on each side of t_i. Suppose also that

$$\sigma \leqslant \sqrt{2}, \quad \Delta > \sigma\sqrt{\log\left(3 + \frac{4}{\sigma^2}\right)}, \quad \eta \leqslant \sigma^2, \tag{47}$$

namely the width of the Gaussian is much smaller than the separation of the impulses in x. Lastly, assume that the impulse locations $T = \{t_i\}_{i=1}^k$ and sampling points $\{s_j\}_{j=2}^{m-1}$ are away from the boundary of I, namely

$$\begin{aligned} \sigma\sqrt{\log(1/\eta^3)} \leqslant t_i \leqslant 1 - \sigma\sqrt{\log(1/\eta^3)}, \qquad \forall i \in [k], \\ \sigma\sqrt{\log(1/\eta^3)} \leqslant s_j \leqslant 1 - \sigma\sqrt{\log(1/\eta^3)}, \qquad \forall j \in [2:m-1]. \end{aligned} \tag{48}$$

We can now give explicit bounds on $\|b\|_2$ and $\|b^\pi\|_2$ for the Gaussian window function, as required by Theorems 11 and 12. The following result is proved in Appendix K.

Lemma 24. *Suppose that the window function ϕ is Gaussian, as defined in* (9)*, the assumptions* (46)*,* (47) *and* (48) *(namely Conditions 2) hold and η satisfies:*

$$\eta \leqslant \min \left\{ \frac{8F_{\min}(\Delta, \frac{1}{\sigma})}{34(2k+2)\left(80k+8+kP\left(\frac{1}{\sigma}\right)\frac{3}{1-e^{-\frac{\Delta^2}{\sigma^2}}}\right)^{\frac{1}{2}}}, \frac{\bar{C}(f_0,f_1)^{\frac{1}{6}}}{\left(4k+4+\frac{4k}{\sigma^2}\right)^{\frac{1}{3}}} \right\}. \tag{49}$$

Then we have the following bounds:

$$\|b\|_2 \leqslant \frac{\sqrt{(2k+2)\left(4k+5+\frac{4k}{\sigma^4}\right)}}{1-\frac{\sqrt{e}}{2}} \bar{C}(f_0,f_1)^{\frac{5}{4}} \left[\frac{F_{\max}\left(\Delta, \frac{1}{\sigma}\right)}{F_{\min}\left(\Delta, \frac{1}{\sigma}\right)^2}\right]^k, \tag{50}$$

$$\|b^\pi\|_2 \leqslant \frac{\sqrt{2k+2}}{\eta\left(1-\frac{\sqrt{e}}{2}\right)} \left(\bar{C}(f_0,f_1)+2k\right)^{\frac{3}{2}} \left[\frac{F_{\max}\left(\Delta, \frac{1}{\sigma}\right)}{F_{\min}\left(\Delta, \frac{1}{\sigma}\right)^2}\right]^k, \tag{51}$$

where

$$\bar{C}(f_0,f_1) = f_0^2 + f_1^2 + 2f_0 + 2f_1 + 2, \tag{52}$$

$$P\left(\frac{1}{\sigma}\right) = \frac{4}{\sigma^4} + \frac{13}{4}\left(\frac{2}{\sigma^2}+\frac{4}{\sigma^4}\right)^2 + \frac{9}{4}\left(\frac{12}{\sigma^4}+\frac{8}{\sigma^6}\right)^2. \tag{53}$$

$$F_{\max}\left(\Delta, \frac{1}{\sigma}\right) = \left(8 + \left(1+\frac{4}{\sigma^4}\right)\frac{2}{1-e^{-\frac{\Delta^2}{\sigma^2}}}\right)^{\frac{1}{2}} \left(32 + \left(\frac{1}{\sigma^4}+\frac{2}{\sigma^6}+\frac{2}{\sigma^8}\right)\frac{24}{1-e^{-\frac{\Delta^2}{\sigma^2}}}\right)^{\frac{1}{2}}, \tag{54}$$

$$F_{\min}\left(\Delta, \frac{1}{\sigma}\right) = 1 - \left(1+\frac{2}{\sigma^2}\right)\frac{2e^{-\frac{\Delta^2}{\sigma^2}}}{1-e^{-\frac{\Delta^2}{\sigma^2}}}. \tag{55}$$

To obtain the final bounds for the Gaussian window function, we will substitute the above bounds in the right hand side of (26) in Theorem 11 and in (28) and (29) in Theorem 12. We will then obtain F_1 in Theorem 4 (see (64)) and F_2 in Theorem 5 (see (71)).

For more clarity, in the following lemma we simplify F_1 and F_2 further, in the case when stronger conditions apply to σ, $\Delta(T)$ and λ.

Lemma 25. *If the conditions in Lemma 24 hold and, in addition, $\sigma < \frac{1}{\sqrt{3}}$, $\Delta > \sigma\sqrt{\log\frac{5}{\sigma^2}}$ and $\bar{f} < 1$, then*

$$\frac{F_{\max}\left(\Delta, \frac{1}{\sigma}\right)}{F_{\min}\left(\Delta, \frac{1}{\sigma}\right)^2} < \frac{c_1}{\sigma^6(1-3\sigma^2)^2}, \tag{56}$$

$$\left((6+\frac{2}{\bar{f}})\sqrt{4k+5+\frac{4k}{\sigma^4}}\bar{C}(f_0,f_1)^{\frac{5}{4}} + \frac{6}{\eta}(\bar{C}(f_0,f_1)+2k)^{\frac{3}{2}}\right)\frac{\sqrt{2k+2}}{1-\frac{\sqrt{e}}{2}} < c_2 \cdot \frac{kC_1(\frac{1}{\epsilon})}{\eta\sigma^2}, \tag{57}$$

$$\frac{\sqrt{(2k+2)\left(4k+5+\frac{4k}{\sigma^4}\right)}}{1-\frac{\sqrt{e}}{2}} \frac{\bar{C}(f_0,f_1)^{\frac{5}{4}}}{\bar{f}} < c_3 \cdot \frac{kC_2(\frac{1}{\epsilon})}{\sigma^2}, \tag{58}$$

where

$$C_1\left(\frac{1}{\epsilon}\right) = \frac{(\bar{C}(f_0,f_1)+2k)^{\frac{3}{2}}}{\bar{f}}, \tag{59}$$

$$C_2\left(\frac{1}{\epsilon}\right) = \frac{\bar{C}(f_0,f_1)^{\frac{5}{4}}}{\bar{f}}, \tag{60}$$

and, similarly, the condition (49) *is simplified to condition* (13). *Moreover, if* λ *in Theorem 12 satisfies* $\lambda < \frac{2}{4}$, *then*

$$\frac{1}{e^{-\frac{\Delta^2\lambda^2}{\sigma^2}} - e^{-\frac{\Delta^2\lambda^2}{\sigma^2}} \cdot \frac{e^{-\frac{\Delta^2}{\sigma^2}} + e^{-\frac{2\Delta^2}{\sigma^2}}}{1-e^{-\frac{\Delta^2}{\sigma^2}}} - e^{-\frac{\Delta^2(1-\lambda)^2}{\sigma^2}}} < c_4. \tag{61}$$

Above, c_1, c_2, c_3, c_4 *are universal constants.*

More specifically, (56), (57) will be used to bound F_1 to obtain (12) and (56), (58), will be used to bound F_2 to obtain (16). Lastly, (61) will be used to bound F_3, which appears in the bound given by Theorem 5.

The proof of Lemma 25 is given in Appendix N. Note that we give C_1 and C_2 as functions of $\frac{1}{\epsilon}$ because, as $\epsilon \to 0$, C_1 and C_2 grow at a rate dependent on ϵ, as indicated in the Lemma 27 in Section 3.4.4.

3.4.2 Proof of Theorem 4 (Wasserstein stability of Program (4) for $\phi(t)$ Gaussian)

We consider Theorem 11 and restrict ourselves to the case $x = \chi$, a k-discrete non-negative measure as in (1) with support T. Let $\Delta = \Delta(T) \geqslant 2\epsilon$, with T as the support of x. We begin with estimating the Lipschitz constant of the measurement operator Φ according to (22), see Appendix I for the proof of the next result.

Lemma 26. *Consider* $S = \{s_j\}_{j=1}^m \subset \mathbb{R}$ *and* $\{\phi_j\}_{j=1}^m$ *specified in* (9). *Then the operator* $\Phi : I \to \mathbb{R}^m$ *defined in (3) is* $\frac{2\sqrt{m}}{\sigma\sqrt{2e}}$*-Lipschitz with respect to the generalised Wasserstein distance, namely (22) holds with* $L = \frac{2\sqrt{m}}{\sigma\sqrt{2e}}$.

We may now invoke Theorem 11 to conclude that, for an arbitrary k-sparse 2ϵ-separated non-negative measure x and arbitrary sampling points $\{s_j\}_{j=1}^m \subset \mathbb{R}$, Program (4) with $\delta' = \delta$ is stable in the sense that there exist vectors $b, \{b^\pi\}_\pi \subset \mathbb{R}^m$ such that

$$d_{GW}\left(x, \widehat{x}\right) \leqslant \left(\left(6 + \frac{2}{\bar{f}}\right)\|b\|_2 + 6\min_\pi \|b^\pi\|_2\right) \cdot \delta + \|x\|_{TV} \cdot \epsilon \tag{62}$$

provided that $m \geqslant 2k+2$ and $f_0 \gg f_1 \gg \bar{f} \gg 0$. The exact relationships between $\bar{f}, f_0, f_1$ are given in the proof of Lemma 23 in Appendix J.

Combining Lemma 24 with (62) yields a final bound on the stability of Program (4) with Gaussian window and completes the proof of the first part of Theorem 4:

$$d_{GW}(x, \widehat{x}) \leqslant F_1(k, \Delta(T), \frac{1}{\sigma}, \frac{1}{\epsilon}, \eta) \cdot \delta + \|x\|_{TV} \cdot \epsilon, \tag{63}$$

where

$$F_1(k, \Delta(T), \frac{1}{\sigma}, \frac{1}{\epsilon}, \eta) = \left((6 + \frac{2}{\bar{f}})\sqrt{4k + 5 + \frac{4k}{\sigma^4}}\bar{C}^{\frac{5}{4}} + \frac{6}{\eta}(\bar{C} + 2k)^{\frac{3}{2}}\right)\frac{\sqrt{2k+2}}{1 - \frac{\sqrt{e}}{2}}\left(\frac{F_{\max}\left(\Delta, \frac{1}{\sigma}\right)}{F_{\min}\left(\Delta, \frac{1}{\sigma}\right)^2}\right)^k, \tag{64}$$

with $\bar{C}, F_{\max}, F_{\min}$ given in (52), (54), (55) respectively and $f_0 = f_0(\frac{1}{\epsilon})$ depends on ϵ (see the proof of Lemma 27).

Finally, to show that (12) holds when $\sigma < \frac{1}{\sqrt{3}}$, and $\Delta > \sigma\sqrt{\log \frac{5}{\sigma^2}}$, we apply the first part of Lemma 25 with $\bar{f} < 1$ and the proof of Theorem 4 is complete.

Remark. In particular, f_0 increases as $\epsilon \to 0$ and f_0 also depends on the other parameters of the problem, namely $\eta, \sigma, \Delta(T), k$. See Section 3.5 for a detailed discussion. Furthermore, f_1 and $\bar{f}$ are considered fixed positive constants with $f_1 < f_0$.

3.4.3 Proof of Theorem 5 (Average stability of Program (4) for $\phi(t)$ Gaussian: source proximity dependence)

We now apply Theorem 12 with $\phi(t) = g(t) = e^{-t^2/\sigma^2}$. We have that $\phi^\infty = 1$, the Lipschitz constant L of g on $[-1,1]$ is $L = \frac{2}{\sigma^2}$, and

$$\sum_{j=1}^{k} (A^{-1})_{ij} \leqslant \|A^{-1}\|_\infty < \frac{1}{\min_j \left(g(s_j - t_j) - \sum_{i \neq j} g(s_j - t_i) \right)}. \tag{65}$$

The last inequality comes from the definition of A in (30) with $\phi(t) = g(t)$ and $s_j := s_{l(j)}$ as given in Definition 1, and [42]. Then, by assumption, $|s_j - t_j| \leqslant \lambda\Delta$ for an arbitrary $j \in [k]$ and g is decreasing, so

$$g(s_j - t_j) \geqslant g(\lambda\Delta) = e^{-\frac{\lambda^2\Delta^2}{\sigma^2}}.$$

We now assume without loss of generality that $s_j < t_j$. Then, it follows that

$$|s_j - t_i| \geqslant |j-i|\Delta - \lambda\Delta, \text{ if } i < j \qquad \text{and} \qquad |s_j - t_i| \geqslant |j-i|\Delta + \lambda\Delta, \text{ if } i > j.$$

This, in turn, leads to

$$\begin{aligned}
\sum_{i \neq j} g(s_j - t_i) &= \sum_{i=1}^{j-1} g(s_j - t_i) + \sum_{i=j+1}^{k} g(s_j - t_i) \\
&\leqslant \sum_{i=1}^{j-1} g((j-i)\Delta - \lambda\Delta) + \sum_{i=j+1}^{k} g((i-j)\Delta + \lambda\Delta) \\
&\leqslant \sum_{i=1}^{\infty} g((i-\lambda)\Delta) + \sum_{i=1}^{\infty} g((i+\lambda)\Delta).
\end{aligned} \tag{66}$$

We now bound each sum in (66) as follows

$$\begin{aligned}
\sum_{i=1}^{\infty} g((i-\lambda)\Delta) &= \sum_{i=1}^{\infty} e^{-\frac{(i-\lambda)^2\Delta^2}{\sigma^2}} \\
&= e^{-\frac{\Delta^2(1-\lambda)^2}{\sigma^2}} + e^{-\frac{\Delta^2\lambda^2}{\sigma^2}} \sum_{i=2}^{\infty} \left(e^{-\frac{\Delta^2}{\sigma^2}} \right)^{i^2 - 2i\lambda} \\
&\leqslant e^{-\frac{\Delta^2(1-\lambda)^2}{\sigma^2}} + e^{-\frac{\Delta^2\lambda^2}{\sigma^2}} \sum_{i=2}^{\infty} \left(e^{-\frac{\Delta^2}{\sigma^2}} \right)^{i} \qquad (i^2 - 2i\lambda > i \text{ for } i \geqslant 2) \\
&= e^{-\frac{\Delta^2(1-\lambda)^2}{\sigma^2}} + e^{-\frac{\Delta^2\lambda^2}{\sigma^2}} \cdot \frac{e^{-\frac{2\Delta^2}{\sigma^2}}}{1 - e^{-\frac{\Delta^2}{\sigma^2}}},
\end{aligned} \tag{67}$$

and similarly, we have that

$$\sum_{i=1}^{\infty} g((i+\lambda)\Delta) = \sum_{i=1}^{\infty} e^{-\frac{(i+\lambda)^2\Delta^2}{\sigma^2}} \leqslant e^{-\frac{\Delta^2\lambda^2}{\sigma^2}} \cdot \frac{e^{-\frac{\Delta^2}{\sigma^2}}}{1 - e^{-\frac{\Delta^2}{\sigma^2}}}. \tag{68}$$

By combining (66), (67) and (68), we obtain:

$$g(s_j - t_j) - \sum_{i \neq j} g(s_j - t_i) \geqslant e^{-\frac{\Delta^2\lambda^2}{\sigma^2}} - e^{-\frac{\Delta^2\lambda^2}{\sigma^2}} \cdot \frac{e^{-\frac{\Delta^2}{\sigma^2}} + e^{-\frac{2\Delta^2}{\sigma^2}}}{1 - e^{-\frac{\Delta^2}{\sigma^2}}} - e^{-\frac{\Delta^2(1-\lambda)^2}{\sigma^2}}. \tag{69}$$

The above inequality also holds when we take the minimum over $j \in [k]$ and, inserting it in (65) and using this result and the bound on $\|b\|_2$ from Lemma 24 in (28), we obtain (14):

$$\left| \int_{T_{i,\epsilon}} \hat{x}(\mathrm{d}t) - a_i \right| \leqslant \left[(c_1 + F_2) \cdot \delta + c_2 \frac{\|\hat{x}\|_{TV}}{\sigma^2} \cdot \epsilon \right] F_3, \tag{70}$$

where

$$F_2(k, \Delta(T), \frac{1}{\sigma}, \frac{1}{\epsilon}) = \frac{\sqrt{(2k+2)\left(4k+5+\frac{4k}{\sigma^4}\right)}}{1-\frac{\sqrt{e}}{2}} \frac{\bar{C}(f_0, f_1)^{\frac{5}{4}}}{\bar{f}} \left[\frac{F_{\max}\left(\Delta, \frac{1}{\sigma}\right)}{F_{\min}\left(\Delta, \frac{1}{\sigma}\right)^2}\right]^k, \tag{71}$$

$$F_3(\Delta(T), \sigma, \lambda) = \frac{1}{e^{-\frac{\Delta^2\lambda^2}{\sigma^2}} - e^{-\frac{\Delta^2\lambda^2}{\sigma^2}} \cdot \frac{e^{-\frac{\Delta^2}{\sigma^2}} + e^{-\frac{2\Delta^2}{\sigma^2}}}{1-e^{-\frac{\Delta^2}{\sigma^2}}} - e^{-\frac{\Delta^2(1-\lambda)^2}{\sigma^2}}}, \tag{72}$$

and $\bar{C}, F_{\max}, F_{\min}$ are given in (52), (54), (55) respectively. The error bound away from the sources (15) is obtained by applying Lemma 16 with the same bounds on $\|b\|_2$.

Then, by using Lemma 25 with $\bar{f} < 1$, we obtain (16). Note that we can apply Lemma 25 because, for $\sigma < \frac{1}{\sqrt{3}}$, we have that $\frac{5}{\sigma^2} > 3 + \frac{4}{\sigma^2}$ and, therefore, $\Delta > \sigma\sqrt{\log \frac{5}{\sigma^2}} > \sigma\sqrt{\log\left(3 + \frac{4}{\sigma^2}\right)}$.

3.4.4 Proof of Corollary 6

First we give an explicit dependence of $C_1(\frac{1}{\epsilon})$ and $C_2(\frac{1}{\epsilon})$ on ϵ for small $\epsilon > 0$ in the following lemma, proved in Appendix O.

Lemma 27. *If $f_1 < f_0$, $1 < f_0$ and $\bar{f} < 1$, then there exists $\epsilon_0 > 0$ such that:*

$$C_1\left(\frac{1}{\epsilon}\right) < \frac{(\bar{c}_1 C_\epsilon^2 + 2k\epsilon^4)^{\frac{3}{2}}}{\bar{f}} \cdot \frac{1}{\epsilon^6} \tag{73}$$

$$C_2\left(\frac{1}{\epsilon}\right) < \bar{c}_2 C_\epsilon^{\frac{5}{2}} \cdot \frac{1}{\epsilon^5}, \tag{74}$$

for all $\epsilon \in (0, \epsilon_0)$, where $\bar{c}_1$ and $\bar{c}_2$ are universal constants and C_ϵ is defined in the proof, see (236).

To prove the first part of the corollary, we first let $\epsilon = \delta^{\frac{1}{7}}$ in the bound on $C_1(\frac{1}{\epsilon})$ in Lemma 27:

$$C_1\left(\frac{1}{\epsilon}\right) < \frac{(\bar{c}_1 C_\epsilon^2 + 2k\epsilon^4)^{\frac{3}{2}}}{\bar{f}} \cdot \frac{1}{\delta^{\frac{6}{7}}}, \quad \forall \delta < \epsilon_0^7, \tag{75}$$

and we substitute the above inequality in the bound (12) in Theorem 4 to obtain:

$$d_{GW}(x, \hat{x}) < \bar{C}_1 \cdot \delta^{\frac{1}{7}}, \quad \forall \delta < \epsilon_0^7,$$

where

$$\bar{C}_1 = \frac{c_1 k(\bar{c}_1 C_\epsilon^2 + 2k\epsilon^4)^{\frac{3}{2}}}{\eta\sigma^2 \bar{f}} \left[\frac{c_2}{\sigma^6(1-3\sigma^2)^2}\right]^k + \|x\|_{TV}. \tag{76}$$

Similarly, let $\epsilon = \delta^{\frac{1}{6}}$ in the bound on $C_2(\frac{1}{\epsilon})$ in Lemma 27:

$$C_2\left(\frac{1}{\epsilon}\right) < \bar{c}_2 C_\epsilon^{\frac{5}{2}} \cdot \frac{1}{\delta^{\frac{5}{6}}}, \quad \forall \delta < \epsilon_0^6, \tag{77}$$

which we substitute in the bound (16) in Theorem 5 to obtain:

$$\left|\int_{T_{i,\epsilon}} \hat{x}(\mathrm{d}t) - a_i\right| \leqslant \bar{C}_2 \cdot \delta^{\frac{1}{6}}, \quad \forall \delta < \epsilon_0^6,$$

where

$$\bar{C}_2 = c_1 + c_3 \frac{k\bar{c}_2 C_\epsilon^{\frac{5}{2}}}{\sigma^2} \left[\frac{c_4}{\sigma^6(1-3\sigma^2)^2}\right]^k + \frac{c_2\|\hat{x}\|_{TV}}{\sigma^2}. \tag{78}$$

Note that we apply Lemma 27 with $\bar{f} < 1$ and that both inequalities in the corollary hold for $\delta < \delta_0 = \epsilon_0^7$, where ϵ_0 is given by Lemma 27.

3.5 Discussion

In this section, we discuss a few issues regarding the robustness of our construction of the dual certificate from Appendix E. There are two points that need to be raised: the construction itself and the proof that we indeed have a T*-System.

At the moment, we do not use any samples that are away from sources in the the construction of the dual certificate. If the sources are close enough compared to σ, then this is not an issue. However, for σ small relative to the distance between samples, in light of the proof of Lemma 23 (see Appendix J), if we consider the dual certificate as the expansion of the determinant N of M^ρ in (122) along the $\tau_{\underline{l}}$ row:

$$N = -F(\tau_l)\beta_0 + \sum_{j=1}^{m}(-1)^{j+1}\beta_j g(\tau_l - s_j), \tag{79}$$

then the terms $g(\tau_{\underline{l}})$ become exponentially small (as $\tau_{\underline{l}}$ is far from all samples s_j) and, therefore, the value of N is close to $-F(\tau_l)$ (which is $-\bar{f}$ if $\tau_l \in T_\epsilon^C$). This is problematic, as we require that $N > 0$. We can overcome this by adding "fake" sources at intervals η^{-1} so that they cover the regions where we have no true sources, together with two close samples for each extra source. The determinant N becomes:

$$N = \begin{vmatrix} f_0 & f(s_1) & \cdots & g(s_m) \\ 0 & g(t_1 - s_1) & \cdots & g(t_1 - s_m) \\ 0 & g'(t_1 - s_1) & \cdots & g'(t_1 - s_m) \\ \vdots & \vdots & & \vdots \\ \bar{f} & g(\tau_j - s_1) & \cdots & g(\tau_j - s_m) \\ \bar{f} & g'(\tau_j - s_1) & \cdots & g'(\tau_j - s_m) \\ \vdots & \vdots & & \vdots \\ F(\tau_{\underline{l}}) & g(\tau_{\underline{l}} - s_1) & \cdots & g(\tau_{\underline{l}} - s_m) \\ \vdots & \vdots & \cdots & \vdots \\ 0 & g(t_k - s_1) & \cdots & g(t_k - s_m) \\ 0 & g'(t_k - s_1) & \cdots & g'(t_k - s_m) \\ f_1 & g(1 - s_1) & \cdots & g(1 - s_m) \end{vmatrix}. \tag{80}$$

Here, the rows are ordered according to the ordering of the set containing $t_i, \tau_j, \tau_{\underline{l}}$. The terms in the expansion of (80) along the row with $\tau_{\underline{l}}$ do not approach 0 exponentially with this construction, since for any $\tau_{\underline{l}}$ there exists s_i close enough so that $g(\tau_{\underline{l}} - s_i) > f^*$ for some $f^* > 0$.

More specifically, consider also the expansion of N along the first column:

$$N = f_0 N_{1,1} + f_1 N_{m+1,1} - F(\tau_{\underline{l}})N_{\tau_{\underline{l}},1} - \bar{f}\sum_{j<\tau_{\underline{l}}}(N_{j,1} - N_{j+1,1}) + \bar{f}\sum_{j>\tau_{\underline{l}}}(N_{j,1} - N_{j+1,1}). \tag{81}$$

We use this expansion in the proof of Lemma 23 in Appendix J to show that the functions $F \cup \{g_j\}_{j=1}^m$ form a T*-System. For $\tau_{\underline{l}} \in T_\epsilon^C$, $F(\tau_{\underline{l}}) = \bar{f}$ and the setup in Lemma 23, we require that (see (127)):

$$\frac{f_0}{\bar{f}} \geqslant \frac{N_{\tau_{\underline{l}},1}}{\min_{\tau_{\underline{l}} \in T_\epsilon^C} N_{1,1}}. \tag{82}$$

In the construction (80), if we upper bound the pairs $N_{j,1} - N_{j+1,1}$ in the two sums in (81) (a separate problem by itself), then we can impose a similar condition to (82) for f_0 and $\bar{f}$. From here, we obtain that

$f_0 = C\bar{f}$ where finding $C \geqslant \frac{N_{\tau_{\underline{l}},1}}{\min_{\tau_{\underline{l}} \in T_\epsilon^C} N_{1,1}}$ involves finding a lower bound on $N_{1,1}$:

$$N = \begin{vmatrix} g(t_1 - s_1) & \cdots & g(t_1 - s_m) \\ g'(t_1 - s_1) & \cdots & g'(t_1 - s_m) \\ \vdots & & \vdots \\ g(\tau_j - s_1) & \cdots & g(\tau_j - s_m) \\ g'(\tau_j - s_1) & \cdots & g'(\tau_j - s_m) \\ \vdots & & \vdots \\ g(\tau_{\underline{l}} - s_1) & \cdots & g(\tau_{\underline{l}} - s_m) \\ \vdots & & \vdots \\ g(t_k - s_1) & \cdots & g(t_k - s_m) \\ g'(t_k - s_1) & \cdots & g'(t_k - s_m) \\ g(1 - s_1) & \cdots & g(1 - s_m) \end{vmatrix}. \tag{83}$$

The structure of the above determinant is similar to the denominator in Appendix K but only up to the row with $\tau_{\underline{l}}$. The rows after it do not preserve the diagonally dominant structure of the matrix, as each source becomes associated with one close sample to it and the first sample corresponding to the next source. This is an issue in both the construction described in the proof of Proposition 19 in Appendix E (and detailed in the proof of Lemma 23) and the construction described in the current section (which would result from considering a determinant with "fake" sources like (80)). However, by adding extra "fake" sources, one could argue that the determinant (80) is better behaved, as the distance between a source and the first sample corresponding to the next source is smaller, which we leave for further work.

Acknowledgements

AE is grateful to Gongguo Tang and Tamir Bendory for enlightening discussions and helpful comments. This publication is based on work supported by the EPSRC Centre For Doctoral Training in Industrially Focused Mathematical Modelling (EP/L015803/1) in collaboration with the National Physical Laboratory and by the Alan Turing Institute under the EPSRC grant EP/N510129/1 and the Turing Seed Funding grant SF019. The authors thank Stephane Chretien, Alistair Forbes and Peter Harris from the National Physical Laboratory for the support and insight regarding the associated applications of super-resolution.

References

[1] Geoffrey Schiebinger, Elina Robeva, and Benjamin Recht. Superresolution without separation. *Information and Inference: a journal of the IMA*, 00:1–30, 2017.

[2] Yohann De Castro and Fabrice Gamboa. Exact reconstruction using beurling minimal extrapolation. *Journal of Mathematical Analysis and applications*, 395(1):336–354, 2012.

[3] Klaus G. Puschmann and Franz Kneer. On super-resolution in astronomical imaging. *Astronomy & Astrophysics*, 436(1):373–378, 2005.

[4] 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(5793):1642–1645, 2006.

[5] Samuel T. Hess, Thanu P. K. Girirajan, and Michael D. Mason. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical journal*, 91(11):4258–4272, 2006.

[6] Michael J. Rust, Mark Bates, and Xiaowei Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm). *Nature methods*, 3(10):793–796, 2006.

[7] Charles W. Mccutchen. Superresolution in microscopy and the abbe resolution limit. *JOSA*, 57(10):1190–1192, 1967.

[8] Hayit Greenspan. Super-resolution in medical imaging. *The Computer Journal*, 52(1):43–63, 2009.

[9] Chaitanya Ekanadham, Daniel Tranchina, and Eero P. Simoncelli. Neural spike identification with continuous basis pursuit. *Computational and Systems Neuroscience (CoSyNe), Salt Lake City, Utah*, 2011.

[10] Stefan Hell. Primer: fluorescence imaging under the diffraction limit. *Nature Methods*, 6(1):19, 2009.

[11] Ronen Tur, Yonina C. Eldar, and Zvi Friedman. Innovation rate sampling of pulse streams with application to ultrasound imaging. *IEEE Transactions on Signal Processing*, 59(4):1827–1842, 2011.

[12] David J. Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096, 1982.

[13] Gongguo Tang, Badri Narayan Bhaskar, and Benjamin Recht. Near minimax line spectral estimation. *IEEE Transactions on Information Theory*, 61(1):499–512, 2015.

[14] Valery Khaidukov, Evgeny Landa, and Tijmen Jan Moser. Diffraction imaging by focusing-defocusing: An outlook on seismic superresolution. *Geophysics*, 69(6):1478–1490, 2004.

[15] Parikshit Shah, Badri Narayan Bhaskar, Gongguo Tang, and Benjamin Recht. Linear system identification via atomic norm regularization. In *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pages 6265–6270. IEEE, 2012.

[16] Emmanuel J. Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.

[17] Gongguo Tang, Badri Narayan Bhaskar, Parikshit Shah, and Benjamin Recht. Compressed sensing off the grid. *IEEE Transactions on Information Theory*, 59(11):7465–7490, 2013.

[18] Karsten Fyhn, Hamid Dadkhahi, and Marco F. Duarte. Spectral compressive sensing with polar interpolation. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013.

[19] Laurent Demanet, Deanna Needell, and Nam Nguyen. Super-resolution via superset selection and pruning. *Proceedings of the 10th International Conference on Sampling Theory and Applications (SampTA)*, 2013.

[20] Vincent Duval and Gabriel Peyré. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, pages 1–41, 2015.

[21] Quentin Denoyelle, Vincent Duval, and Gabriel Peyré. Support recovery for sparse super-resolution of positive measures. *Journal of Fourier Analysis and Applications*, 23(5):1153–1194, 2017.

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

[23] Tamir Bendory, Shai Dekel, and Arie Feuer. Robust recovery of stream of pulses using convex optimization. *Journal of Mathematical Analysis and Applications*, 442(2):511–536, 2016.

[24] David L. Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

[25] 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(2):489–509, 2006.

[26] Emmanuel J. Candès and Terence Tao. Decoding by linear programming. *Information Theory, IEEE Transactions on*, 12(51):4203–4215, Dec 2005.

[27] Samuel Karlin and William J. Studden. *Tchebycheff systems: with applications in analysis and statistics.* Pure and applied mathematics. Interscience Publishers, 1966.

[28] Samuel Karlin. *Total Positivity, Volume 1.* Total Positivity. Stanford University Press, 1968.

[29] Mark G. Krein, Adolf A. Nudelman, and David Louvish. *The Markov Moment Problem And Extremal Problems.*

[30] Paul Doukhan and Fabrice Gamboa. Superresolution rates in Prokhorov metric. *Canad. Math. Bull.*, 48(2):316–329, 1996.

[31] Veniamin I. Morgenshtern and Emmanuel J. Candès. Stable super-resolution of positive sources: the discrete setup. *SIAM J. on Imaging Sciences*, 9(1):412–444, 2016.

[32] Tamir Bendory. Robust recovery of positive stream of pulses. *IEEE Transactions on Signal Processing*, 65(8):2114–2122, 2017.

[33] David L. Donoho. Superresolution via sparsity constraints. *SIAM J. Math. Anal.*, 23(5):1309–1331, 1992.

[34] Brett Bernstein and Carlos Fernandez-Granda. Deconvolution of Point Sources: A Sampling Theorem and Robustness Guarantees. (July), 2017.

[35] Allan Pinkus. Spectral properties of totally positive kernels and matrices. *Total positivity and its applications*, 359:1–35, 1996.

[36] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms (3rd ed.).* MIT Press, 2009.

[37] Benedetto Piccoli and Francesco Rossi. Generalized Wasserstein distance and its application to transport equations with source. *Archive for Rational Mechanics and Analysis*, 211(1):335–358, 2014.

[38] Emmanuel J. Candès and Michael B. Wakin. An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30, 2008.

[39] William Feller. *An Introduction to Probability Theory and its Applications, Vol 2.* Wiley, 1968.

[40] Emmanuel J. Candès and Carlos Fernandez-Granda. Super-resolution from noisy data. *Journal of Fourier Analysis and Applications*, 19(6):1229–1254, 2013.

[41] Carlos Fernandez-Granda. Support detection in super-resolution. *Proceedings of the 10th International Conference on Sampling Theory and Applications (SampTA)*, pages 145–148, 2013.

[42] James M. Varah. A lower bound for the smallest singular value of a matrix. *Linear Algebra and its Applications*, 11(1):3–5, 1975.

[43] Robert J. Plemmons. M-Matrix Characterizations. I-Nonsingular M-Matrices. *Linear Algebra and its Applications*, 18(2):175–188, 1977.

[44] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling.* Progress in Nonlinear Differential Equations and Their Applications. Springer International Publishing, 2015.

[45] Roger A. Horn and Charles R. Johnson. *Matrix Analysis.* Cambridge University Press, 1990.

[46] Izrail S. Gradshteyn and Iosif M. Ryzhik. *Tables of Integrals, Series, and Products (6th ed).* San Diego, CA: Academic Press, 2000.

A Proof of Lemma 14

Let $\hat{x}$ be a solution of Program 4 with $\delta = 0$ and let $h = \hat{x} - x$ be the error. Then, by feasibility of both x and $\widehat{x}$ in Program (4), we have that

$$\int_I \phi_j(t) h(\mathrm{d}t) = 0, \qquad j \in [m]. \tag{84}$$

Let T^C be the complement of $T = \{t_i\}_{i=1}^k$ with respect to I. By assumption, the existence of a dual certificate allows us to write that

$$\begin{aligned}
\int_{T^C} q(t) h(\mathrm{d}t) &= \int_I q(t) h(\mathrm{d}t) - \int_T q(t) h(\mathrm{d}t) \\
&= \int_I q(t) h(\mathrm{d}t) \qquad \big(q(t_i) = 0,\ i \in [k]\big) \\
&= \sum_{j=1}^m b_j \int_I \phi_j(t) h(\mathrm{d}t) \\
&= 0. \qquad \text{(see (84))}
\end{aligned}$$

Since $x = 0$ on T^C, then $h = \hat{x}$ on T^C, so the last equality is equivalent to

$$\int_{T^C} q(t) \hat{x}(\mathrm{d}t) = 0. \tag{85}$$

But q is strictly positive on T^C, so it must be that $h = \hat{x} = 0$ on T^C and, therefore, $h = \sum_{i=1}^k c_i \delta_{t_i}$ for some coefficients $\{c_i\}$. Now (84) reads $\sum_{i=1}^k c_i \phi_j(t_i) = 0$ for every $j \in [m]$. This gives $c_i = 0$ for all $i \in [k]$ because $[\phi_j(t_i)]_{i,j}$ is, by assumption, full rank. Therefore $h = 0$ and $\hat{x} = x$ on I, which completes the proof of Lemma 14.

B Proof of Lemma 16

Let $\hat{x}$ be a solution of Program (4) and set $h = \hat{x} - \chi$ to be the error. Then, by feasibility of both χ and $\widehat{x}$ in Program (4) and using the triangle inequality, we have that

$$\left\| \int_I \Phi(t) h(\mathrm{d}t) \right\|_2 \leqslant 2\delta'. \tag{86}$$

Next, the existence of the dual certificate q allows us to write that

$$\begin{aligned}
&\bar{f} \int_{T_\epsilon^C} h(\mathrm{d}t) + \sum_{i=1}^k \int_{T_{i,\epsilon}} f(t - t_i)\, h(\mathrm{d}t) \\
&\leqslant \int_{T_\epsilon^C} q(t) h(\mathrm{d}t) + \sum_{i=1}^k \int_{T_{i,\epsilon}} q(t) h(\mathrm{d}t) \\
&= \int_{T_\epsilon^C} q(t) h(\mathrm{d}t) + \int_{T_\epsilon} q(t) h(\mathrm{d}t) \qquad \left(T_\epsilon = \cup_{i=1}^k T_{i,\epsilon}\right) \\
&= \int_I q(t) h(\mathrm{d}t) \\
&= \sum_{j=1}^m b_j \int_I \phi_j(t) h(\mathrm{d}t) \\
&\leqslant \|b\|_2 \cdot \left\| \int_I \Phi(t) h(\mathrm{d}t) \right\|_2 \qquad \text{(Cauchy-Schwarz inequality)} \\
&\leqslant \|b\|_2 \cdot 2\delta', \qquad \text{(see (86))}
\end{aligned}$$

which completes the proof of Lemma 16.

C Proof of Lemma 17

The existence of the dual certificate q^0 allows us to write that

$$\begin{aligned}
&\sum_{i=1}^{k}\left|\int_{T_{i,\epsilon}} h(\mathrm{d}t)\right| \\
&= \sum_{i=1}^{k}\int_{T_{i,\epsilon}} s_i h(\mathrm{d}t) \qquad s_i = \operatorname{sign}\left(\int_{T_{i,\epsilon}} h(\mathrm{d}t)\right) \\
&= \sum_{i=1}^{k}\int_{T_{i,\epsilon}} \left(s_i - q^0(t)\right) h(\mathrm{d}t) + \sum_{i=1}^{k}\int_{T_{i,\epsilon}} q^0(t)h(\mathrm{d}t) \\
&= \sum_{i=1}^{k}\int_{T_{i,\epsilon}} \left(s_i - q^0(t)\right) h(\mathrm{d}t) + \int_I q^0(t)h(\mathrm{d}t) - \int_{T_\epsilon^C} q^0(t)h(\mathrm{d}t) \\
&= \sum_{s_i=1}\int_{T_{i,\epsilon}} \left(1 - q^0(t)\right) h(\mathrm{d}t) + \sum_{s_i=-1}\int_{T_{i,\epsilon}} \left(-1 - q^0(t)\right) h(\mathrm{d}t) \\
&\quad + \int_I q^0(t)h(\mathrm{d}t) - \int_{T_\epsilon^C} q^0(t)h(\mathrm{d}t) \\
&\leqslant \sum_{s_i=1}\int_{T_{i,\epsilon}} f\left(t-t_i\right) h(\mathrm{d}t) + \sum_{s_i=-1}\int_{T_{i,\epsilon}} f\left(t-t_i\right) h(\mathrm{d}t) + \int_I q^0(t)h(\mathrm{d}t) + \bar{f}\int_{T_\epsilon^C} h(\mathrm{d}t) \\
&= \sum_{i=1}^{k}\int_{T_{i,\epsilon}} f\left(t-t_i\right) h(\mathrm{d}t) + \bar{f}\int_{T_\epsilon^C} h(\mathrm{d}t) + \int_I q^0(t)h(\mathrm{d}t) \\
&\leqslant 2\|b\|_2\delta' + \int_I q^0(t)h(\mathrm{d}t) \qquad \text{(see Lemma 16)} \\
&= 2\|b\|_2\delta' + \sum_{j=1}^{m} b_i^0 \int_I \phi_j(t)h(\mathrm{d}t) \\
&\leqslant 2\|b\|_2\delta' + \|b^0\|_2 \cdot 2\delta', \qquad \text{(see (86))}
\end{aligned}$$

which completes the proof of Lemma 17.

D Proof of Lemma 18

Our strategy is as follows. We first argue that

$$d_{GW}\left(\chi, \widehat{x}\right) \approx d_{GW}\left(\chi, \widetilde{x}\right),$$

where $\widetilde{x}$ is the restriction of $\widehat{x}$ to the ϵ-neighbourhood of the support of χ, namely T_ϵ defined in (8). This, loosely speaking, reduces the problem to that of computing the distance between two discrete measures supported on T. We control the latter distance using a particular suboptimal choice of measure γ in (20). Let us turn to the details.

Let $\widetilde{x}$ be the restriction of $\widehat{x}$ to T_ϵ, namely $\widetilde{x} = \widehat{x}|_{T_\epsilon}$, or more specifically

$$\widetilde{x}(\mathrm{d}t) = \begin{cases} \widehat{x}(\mathrm{d}t) & t \in T_\epsilon, \\ 0 & t \in T_\epsilon^C. \end{cases}$$

Then, using the triangle inequality, we observe that

$$d_{GW}\left(\chi, \widehat{x}\right) \leqslant d_{GW}\left(\chi, \widetilde{x}\right) + d_{GW}\left(\widetilde{x}, \widehat{x}\right). \tag{87}$$

The last distance above is easy to control: We write that

$$
\begin{aligned}
d_{GW}\left(\widetilde{x},\widehat{x}\right) &= \inf_{\widetilde{z},\widehat{z}}\left(\|\widetilde{x}-\widetilde{z}\|_{TV}+\|\widehat{x}-\widehat{z}\|_{TV}+d_W\left(\widetilde{z},\widehat{z}\right)\right), \qquad \text{(see (19))}\\
&\leqslant \|\widetilde{x}-\widehat{x}\|_{TV}+\|\widehat{x}-\widehat{x}\|_{TV}+d_W\left(\widehat{x},\widehat{x}\right) \qquad (\widetilde{z}=\widehat{z}=\widehat{x})\\
&= \|\widetilde{x}-\widehat{x}\|_{TV}\\
&= \left\|\widehat{x}|_{T_\epsilon^C}\right\|_{TV}\\
&= \int_{T_\epsilon^C}\widehat{x}(\mathrm{d}t) \qquad (\widehat{x}\text{ is non-negative})\\
&= \int_{T_\epsilon^C}h(\mathrm{d}t) \qquad \left(h(\mathrm{d}t)=\widehat{x}(\mathrm{d}t)-\chi(\mathrm{d}t)=\widehat{x}(\mathrm{d}t)\text{ when }t\in T^C\right)\\
&\leqslant \frac{2\|b\|_2\delta'}{\bar{f}}. \qquad \text{(see Lemma 16)}
\end{aligned}
\tag{88}
$$

We next control the term $d_{GW}\left(\chi,\widetilde{x}\right)$ in (87) by writing that

$$
\begin{aligned}
d_{GW}\left(\chi,\widetilde{x}\right) &= \inf_{\widetilde{z},\widehat{z}}\left(\|\chi-z\|_{TV}+\|\widetilde{x}-\widetilde{z}\|_{TV}+d_W\left(z,\widetilde{z}\right)\right) \qquad \text{(see (19))}\\
&\leqslant \|\chi-\chi\|_{TV}+\left\|\widetilde{x}-\widetilde{x}\cdot\frac{\|\chi\|_{TV}}{\|\widetilde{x}\|_{TV}}\right\|_{TV}+d_W\left(\chi,\widetilde{z}\right) \qquad \left(z=\chi,\ \widetilde{z}=\widetilde{x}\cdot\frac{\|\chi\|_{TV}}{\|\widetilde{x}\|_{TV}}\right)\\
&= \left|\|\widetilde{x}\|_{TV}-\|\chi\|_{TV}\right|+d_W\left(\chi,\widetilde{z}\right)\\
&= \left|\left\|\widetilde{x}|_{T_\epsilon^C}\right\|_{TV}+\|\widetilde{x}|_{T_\epsilon}\|_{TV}-\|\chi|_{T_\epsilon}\|_{TV}\right|+d_W\left(\chi,\widetilde{z}\right)\\
&= \left|\|\widehat{x}|_{T_\epsilon}\|_{TV}-\|\chi|_{T_\epsilon}\|_{TV}\right|+d_W\left(\chi,\widetilde{z}\right) \qquad \left(\widetilde{x}=\widehat{x}|_{T_\epsilon}\right)\\
&= \left|\int_{T_\epsilon}\widehat{x}(\mathrm{d}t)-\int_{T_\epsilon}\chi(\mathrm{d}t)\right|+d_W\left(\chi,\widetilde{z}\right) \qquad (\chi\text{ and }\widehat{x}\text{ are non-negative})\\
&= \left|\int_{T_\epsilon}h(\mathrm{d}t)\right|+d_W\left(\chi,\widetilde{z}\right)\\
&\leqslant \sum_{i=1}^{k}\left|\int_{T_{i,\epsilon}}h(\mathrm{d}t)\right|+d_W\left(\chi,\widetilde{z}\right) \qquad \text{(triangle inequality)}\\
&\leqslant 2\left(\|b\|_2+\|b^0\|_2\right)\delta'+d_W\left(\chi,\widetilde{z}\right). \qquad \text{(see Lemma 17)}
\end{aligned}
\tag{89}
$$

For future use, we record an intermediate result that is obvious from studying (89), namely

$$
\left|\|\widetilde{x}\|_{TV}-\|\chi\|_{TV}\right| \leqslant 2\left(\|b\|_2+\left\|b^0\right\|_2\right)\delta'. \tag{90}
$$

It remains to control $d_W(\chi,\widetilde{z})$ above where

$$
d_W\left(\chi,\widetilde{z}\right)=\inf\int_{I^2}|\tau-\widetilde{\tau}|\,\gamma\left(\mathrm{d}\tau,\mathrm{d}\widetilde{\tau}\right), \tag{91}
$$

is the Wasserstein distance between the measures χ and $\widetilde{z}$. The infimum above is over all measures γ on $I^2=I\times I$ that produce χ and $\widetilde{z}$ as marginals, namely we have:

$$
\int_{A\times I}\gamma(\mathrm{d}\tau,\mathrm{d}\widetilde{\tau})=\chi(A) \quad\text{and}\quad \int_{I\times B}\gamma(\mathrm{d}\tau,\mathrm{d}\widetilde{\tau})=\widetilde{z}(B),\quad \forall A,B\subset I \tag{92}
$$

For every $i\in[k]$, let also $\widetilde{x}_i=\widetilde{x}|_{T_{i,\epsilon}}$ and $\widetilde{z}_i=\widetilde{z}|_{T_{i,\epsilon}}$ be the restrictions of $\widetilde{x}$ and $\widetilde{z}$ to $T_{i,\epsilon}$, respectively. Because of our choice of $z,\widetilde{z}$ in the second line of (89), note that $\widetilde{z}_i=\widetilde{x}_i\cdot\|z\|_{TV}/\|\widetilde{x}\|_{TV}$. Recalling that χ is supported on $T = \{t_i\}_{i=1}^k$, we write that $\chi = \sum_{i=1}^k a_i \delta_{t_i}$ for non-negative amplitudes $\{a_i\}_{i=1}^k$. Then, noting that χ is supported on T and $\tilde{z}$ is supported on T_ϵ, then any feasible γ in (91) is supported on $T \times T_\epsilon$ and we can construct a feasible but suboptimal $\gamma(\mathrm{d}\tau, \mathrm{d}\widetilde{\tau})$ in a two-step approach, where we first extract up to a_i weight of $\widetilde{z}_i$ on each δ_{t_i}, for example let

$$\gamma_1 = \sum_{i=1}^k \widetilde{z}_i(\mathrm{d}\widetilde{\tau}) 1_{T_i} \quad \text{where} \quad 1_{T_i} = \begin{cases} T_{i,\epsilon} & \text{if } \int_{T_{i,\epsilon}} z_i(\mathrm{d}\tau) \leqslant a_i, \\ [t_i - \xi_i, t_i + \xi_i] & \text{if otherwise,} \end{cases} \tag{93}$$

where ξ_i is defined such that $\int_{t_i-\xi_i}^{t_i+\xi_i} z_i(\mathrm{d}\tau) = a_i$. As a result, γ_1 has $\mathrm{d}\tau$ marginal equal to $\widetilde{z}_i$ on the support of γ_1 and the $\mathrm{d}\widetilde{\tau}$ marginal no more than the desired a_i. We then construct γ_2 by partitioning the remaining $\widetilde{z}_i$ into the $\mathrm{d}\widetilde{\tau}$ subsets in order to make up the $\tilde{z}$ marginal, which is exactly achievable using all of $\tilde{z}$ due to $\int \tilde{z}(\mathrm{d}\tau) = \sum_{i=1}^k a_i$. Then, we take $\gamma = \gamma_1 + \gamma_2$.

Intuitively, this is a transport plan according to which we move as much mass as possible inside each $T_{i,\epsilon}$ (from δ_{a_i} to $\widetilde{z}_i$, the minimum of the masses the two) and the remaining mass is moved outside $T_{i,\epsilon}$. Therefore, for this choice of γ, we have that

$$\begin{aligned} d_W(\chi, \widetilde{z}) &\leqslant \int_{I^2} |\tau - \widetilde{\tau}| \gamma(\mathrm{d}\tau, \mathrm{d}\widetilde{\tau}) \\ &= \sum_{i=1}^k \int_{\{t_i\} \times T_{i,\epsilon}} |\tau - \widetilde{\tau}| \gamma(\mathrm{d}\tau, \mathrm{d}\widetilde{\tau}) + \sum_{i=1}^k \int_{\{t_i\} \times T_{i,\epsilon}^C} |\tau - \widetilde{\tau}| \gamma(\mathrm{d}\tau, \mathrm{d}\widetilde{\tau}) \\ &\leqslant \epsilon \sum_{i=1}^k \int_{\{t_i\} \times T_{i,\epsilon}} \gamma_1(\mathrm{d}\tau, \mathrm{d}\widetilde{\tau}) + \sum_{i=1}^k \int_{\{t_i\} \times T_{i,\epsilon}^C} \gamma_2(\mathrm{d}\tau, \mathrm{d}\widetilde{\tau}) \end{aligned} \tag{94}$$

The third line above uses the fact that $\tau = t_i$ and, if $\widetilde{\tau} \in T_{i,\epsilon}$, then $|\tau - \widetilde{\tau}| \leqslant \epsilon$ and $|\tau - \widetilde{\tau}| \leqslant 1$ otherwise. By evaluating the integrals in the last line above, we find that

$$\begin{aligned} d_W(\chi, \widetilde{z}) &\leqslant \epsilon \sum_{i=1}^k \min\{a_i, \|\widetilde{z}_i\|_{TV}\} + \sum_{i=1}^k \left(a_i - \min\{a_i, \|\widetilde{z}_i\|_{TV}\}\right) \\ &\leqslant \epsilon \sum_{i=1}^k a_i + \sum_{i=1}^k \left|a_i - \|\widetilde{z}_i\|_{TV}\right| \\ &= \epsilon \|\chi\|_{TV} + \sum_{i=1}^k \left|a_i - \|\widetilde{z}_i\|_{TV}\right| \qquad \left(\sum_{i=1}^k a_i = \|\chi\|_{TV}\right). \end{aligned} \tag{95}$$

Then, we have that

$$\begin{aligned} d_W(\chi, \widetilde{z}) &\leqslant \epsilon \|\chi\|_{TV} + \sum_{i=1}^k \left| \int_{T_{i,\epsilon}} \chi(\mathrm{d}t) - \frac{\|\chi\|_{TV}}{\|\widetilde{x}\|_{TV}} \int_{T_{i,\epsilon}} \widetilde{x}(\mathrm{d}t) \right| \qquad \big(\text{see the second line of (89)}\big) \\ &\leqslant \epsilon \|\chi\|_{TV} + \sum_{i=1}^k \left| \int_{T_{i,\epsilon}} \chi(\mathrm{d}t) - \widetilde{x}(\mathrm{d}t) \right| + \sum_{i=1}^k \|\widetilde{x}_i\|_{TV} \left| 1 - \frac{\|\chi\|_{TV}}{\|\widetilde{x}\|_{TV}} \right| \qquad \text{(triangle inequality)} \\ &= \epsilon \|\chi\|_{TV} + \sum_{i=1}^k \left| \int_{T_{i,\epsilon}} h(\mathrm{d}t) \right| + \left| \|\widetilde{x}\|_{TV} - \|\chi\|_{TV} \right| \qquad \big(\widetilde{x} = \widehat{x}|_{T_\epsilon}\big) \\ &\leqslant \epsilon \|\chi\|_{TV} + \sum_{i=1}^k \left| \int_{T_{i,\epsilon}} h(\mathrm{d}t) \right| + 2\left(\|b\|_2 + \|b^0\|_2\right)\delta' \qquad \text{(see (90))} \\ &\leqslant \epsilon \|\chi\|_{TV} + 2\left(\|b\|_2 + \|b^0\|_2\right)\delta' + 2\left(\|b\|_2 + \|b^0\|_2\right)\delta'. \qquad \text{(see Lemma 17)} \end{aligned} \tag{96}$$

Substituting the above bound back into (89), we find that

$$d_{GW}(\chi, \widetilde{x}) \leqslant 6\left(\|b\|_2 + \|b^0\|_2\right)\delta' + \epsilon \|\chi\|_{TV}. \tag{97}$$

Then combining (88) and (97) yields

$$\begin{aligned}d_{GW}\left(\chi,\widehat{x}\right) &\leqslant d_{GW}\left(x,\widetilde{x}\right)+d_{GW}\left(\widetilde{x},\widehat{x}\right)\\ &\leqslant \frac{2\|b\|_2\delta'}{\bar{f}}+6\left(\|b\|_2+\|b^0\|_2\right)\delta'+\epsilon\|\chi\|_{TV},\end{aligned}$$

which completes the proof of Lemma 18.

E Proof of Proposition 19

Without loss of generality and for better clarity, suppose that $T=\{t_i\}_{i=1}^k$ is an increasing sequence. Consider a positive scalar ρ such that $\rho\leqslant\epsilon\leqslant\Delta/2$. Consider also an increasing sequence $\{\tau_l\}_{l=1}^m\subset I=[0,1]$ such that $\tau_1=0$, $\tau_m=1$, and every $T_{i,\rho}$ contains an even and nonzero number of the remaining points. Let us define the polynomial

$$q^\rho(t)=\begin{vmatrix}-F(t) & \phi_1(t) & \cdots & \phi_m(t)\\ -F(\tau_1) & \phi_1(\tau_1) & \cdots & \phi_m(\tau_1)\\ -F(\tau_2) & \phi_1(\tau_2) & \cdots & \phi_m(\tau_2)\\ \vdots & \vdots & \vdots & \vdots\\ -F(\tau_m) & \phi_1(\tau_m) & \cdots & \phi_m(\tau_m)\end{vmatrix},\qquad t\in I. \tag{98}$$

Note that $q^\rho(t)=0$ when $t\in\{\tau_l\}_{l=1}^m$. By assumption, $\{F\}\cup\{\phi_j\}_{j=1}^m$ form a T*-system on I. Therefore, invoking the first part of Definition 9, we find that q^ρ is non-negative on T_ρ^C. We represent this polynomial with $q^\rho=-\beta_0^\rho F+\sum_{j=1}^m(-1)^j\beta_j^\rho\phi_j$ and note that $\beta_0^\rho=|\phi_j(\tau_i)|_{i,j=1}^m$. By assumption also $\{\phi_j\}_{j=1}^m$ form a T-system on I and therefore $\beta_0^\rho>0$. This observation allows us to form the normalized polynomial

$$\dot{q}^\rho:=\frac{q^\rho}{\beta_0^\rho}=-F+\sum_{j=1}^m(-1)^j\frac{\beta_j^\rho}{\beta_0^\rho}\phi_j=:-F+\sum_{j=1}^m(-1)^j b_j^\rho\phi_j.$$

Note also that the coefficients $\{\beta_j^\rho\}_{j\in[0:m]}$ correspond to the minors in the second part of Definition 9. Therefore, for each $j\in[0:m]$, we have that $|\beta_j^\rho|$ approaches zero at the same rate, as $\rho\to 0$. So for sufficiently small ρ_0, every b_j^ρ is bounded in magnitude when $\rho\leqslant\rho_0$; in particular, $|b_j^\rho|=\Theta(1)$. This means that for sufficiently small ρ_0, $\{\dot{q}^\rho:\rho\leqslant\rho_0\}$ is bounded. Therefore, we can find a subsequence $\{\rho_l\}_l\subset[0,\rho_0]$ such that $\rho_l\to 0$ and the subsequence $\{\dot{q}^{\rho_l}\}_l$ converges to the polynomial

$$\dot{q}:=-F+\sum_{j=1}^m b_j\phi_j.$$

Note that $b_j\neq 0$ for every $j\in[m]$; in particular, $|b_j|=\Theta(1)$. Hence the polynomial $\sum_{j=1}^m b_j\phi_j$ is nontrivial, namely does not uniformly vanish on I. (It would have sufficed to have some nonzero coefficient, say b_{j_0}, rather than requiring all $\{b_j\}_j$ to be nonzero. However that would have made the statement of Definition 9 more cumbersome.) Lastly observe that $\dot{q}$ is non-negative on I and vanishes on T (as well as on the boundary of I). This completes the proof of Proposition 19.

F Proof of Proposition 21

In this proof, we will use the following result for strictly diagonally dominant matrices from [43]:

Lemma 28. *If A is a strictly diagonally dominant matrix with positive entries on the main diagonal and negative entries otherwise, then A is invertible and A^{-1} has non-negative entries.*

Proof of Proposition 21

Let $\hat{x}$ be a solution of (4) and $h = x - \hat{x}$. Then, with $\phi_j(t) = \phi(t - s_j)$ for some j, by reverse triangle inequality we have

$$\begin{aligned}\delta \geqslant \left(\sum_{j=1}^{m}\left(y(s_j) - \int_0^1 \phi_j(t)\hat{x}(\mathrm{d}t)\right)^2\right)^{1/2} &\geqslant \left(\sum_{j=1}^{m}\big(\phi_j(t)h(\mathrm{d}t) + \eta_j\big)^2\right)^{1/2} \\ &\geqslant \left(\sum_{j=1}^{m}\big(\phi_j(t)h(\mathrm{d}t)\big)^2\right)^{1/2} - \|\eta\|_2 \\ &\geqslant \left(\sum_{j=1}^{m}\big(\phi_j(t)h(\mathrm{d}t)\big)^2\right)^{1/2} - \delta,\end{aligned}$$

and so

$$\sum_{j=1}^{m}\left(\int_0^1 \phi_j(t)h(\mathrm{d}t)\right)^2 \leqslant 4\delta^2 \implies \left|\int_0^1 \phi_j(t)h(\mathrm{d}t)\right| \leqslant 2\delta, \quad \forall j \in [m]. \tag{100}$$

We apply the reverse triangle inequality again to find a lower bound of the left-hand side term in (100):

$$\begin{aligned}\left|\int_0^1 \phi_j(t)h(\mathrm{d}t)\right| \geqslant &\left|\int_{T_{i,\epsilon}} \phi_j(t)\hat{x}(\mathrm{d}t) - a_i\phi_j(t_i)\right| &\text{(101a)}\\ &- \sum_{l\neq i}\left|\int_{T_{l,\epsilon}} \phi_j(t)\hat{x}(\mathrm{d}t) - a_l\phi_j(t_l)\right| &\text{(101b)}\\ &- \left|\int_{T_\epsilon^C} \phi_j(t)\hat{x}(\mathrm{d}t)\right|. &\text{(101c)}\end{aligned}$$

We now need to lower bound the term in (101a) and upper bound the terms in (101b), (101c). For the first one, we obtain:

$$\begin{aligned}\left|\int_{T_{i,\epsilon}} \phi_j(t)\hat{x}(\mathrm{d}t) - a_i\phi_j(t_i)\right| &= \left|\int_{T_{i,\epsilon}} \phi_j(t)\hat{x}(\mathrm{d}t) - a_i\phi_j(t_i) + \phi_j(t_i)\int_{T_{i,\epsilon}}\hat{x}(\mathrm{d}t) - \phi_j(t_i)\int_{T_{i,\epsilon}}\hat{x}(\mathrm{d}t)\right| \\ &\geqslant |\phi_j(t_i)|\left|\int_{T_{i,\epsilon}}\hat{x}(\mathrm{d}t) - a_i\right| - \int_{T_{i,\epsilon}}|\phi_j(t) - \phi_j(t_i)|\,\hat{x}(\mathrm{d}t) \\ &\geqslant |\phi_j(t_i)|\left|\int_{T_{i,\epsilon}}\hat{x}(\mathrm{d}t) - a_i\right| - L\int_{T_{i,\epsilon}}|t - t_i|\,\hat{x}(\mathrm{d}t) \\ &\geqslant |\phi_j(t_i)|\left|\int_{T_{i,\epsilon}}\hat{x}(\mathrm{d}t) - a_i\right| - L\epsilon\int_{T_{i,\epsilon}}\hat{x}(\mathrm{d}t).\end{aligned} \tag{102}$$

Therefore, from (102), we obtain:

$$\left|\int_{T_{i,\epsilon}} \phi_j(t)\hat{x}(\mathrm{d}t) - a_i\phi_j(t_i)\right| \geqslant |\phi_j(t_i)|\left|\int_{T_{i,\epsilon}}\hat{x}(\mathrm{d}t) - a_i\right| - L\epsilon\|\hat{x}\|_{TV}. \tag{103}$$

For the term (101b), we have:

$$\begin{aligned}\left|\int_{T_{l,\epsilon}} \phi_j(t)\hat{x}(\mathrm{d}t) - a_l\phi_j(t_l)\right| &= \left|\int_{T_{l,\epsilon}} \phi_j(t)\hat{x}(\mathrm{d}t) - a_l\phi_j(t_l) + \phi_j(t_l)\int_{T_{l,\epsilon}} \hat{x}(\mathrm{d}t) - \phi_j(t_l)\int_{T_{l,\epsilon}} \hat{x}(\mathrm{d}t)\right| \\ &\leqslant \left|\phi_j(t_l)\right|\left|\int_{T_{l,\epsilon}} \hat{x}(\mathrm{d}t) - a_l\right| + \int_{T_{l,\epsilon}} \left|\phi_j(t) - \phi_j(t_l)\right|\hat{x}(\mathrm{d}t) \\ &\leqslant \left|\phi_j(t_l)\right|\left|\int_{T_{l,\epsilon}} \hat{x}(\mathrm{d}t) - a_l\right| + L\int_{T_{l,\epsilon}} |t - t_l|\,\hat{x}(\mathrm{d}t) \\ &\leqslant \left|\phi_j(t_l)\right|\left|\int_{T_{l,\epsilon}} \hat{x}(\mathrm{d}t) - a_l\right| + L\epsilon\int_{T_{l,\epsilon}} \hat{x}(\mathrm{d}t),\end{aligned}$$

so

$$\sum_{l\neq i}\left|\int_{T_{l,\epsilon}} \phi_j(t)\hat{x}(\mathrm{d}t) - a_l\phi_j(t_l)\right| \leqslant \sum_{l\neq i}\left(\left|\phi_j(t_l)\right|\left|\int_{T_{l,\epsilon}} \hat{x}(\mathrm{d}t) - a_l\right|\right) + L\epsilon\sum_{l\neq i}\int_{T_{l,\epsilon}} \hat{x}(\mathrm{d}t). \tag{104}$$

Finally, for the term (101c), we have:

$$\left|\int_{T_\epsilon^C} \phi_j(t)\hat{x}(\mathrm{d}t)\right| \leqslant \max_{t\in T_\epsilon^C}\left|\phi_j(t)\right|\int_{T_\epsilon^C} \hat{x}(\mathrm{d}t) \leqslant \phi^\infty\left(\frac{2\|b\|_2\delta}{\bar{f}}\right). \tag{105}$$

Let us denote

$$z_i = \left|\int_{T_{i,\epsilon}} \hat{x}(\mathrm{d}t) - a_i\right|,$$

for all $i \in [k]$. Then, by combining (101) with the bounds (102),(104) and (105), we obtain the j-th row of a linear system:

$$2\delta \geqslant \left|\phi_j(t_i)\right| z_i - L\epsilon\int_{T_{i,\epsilon}} \hat{x}(\mathrm{d}t) - \sum_{l\neq i}\left|\phi_j(t_l)\right| z_l - L\epsilon\sum_{l\neq i}\int_{T_{l,\epsilon}} \hat{x}(\mathrm{d}t) - \phi^\infty\frac{2\|b\|_2}{\bar{f}}\delta. \tag{106}$$

By using (106) along with

$$\int_{T_{i,\epsilon}} \hat{x}(\mathrm{d}t) + \sum_{l\neq i}\int_{T_{l,\epsilon}} \hat{x}(\mathrm{d}t) = \int_{T_\epsilon} \hat{x}(\mathrm{d}t) \leqslant \int_I \hat{x}(\mathrm{d}t) = \|\hat{x}\|_{TV},$$

we obtain:

$$2\left(1 + \frac{\phi^\infty\|b\|_2}{\bar{f}}\right)\delta + \epsilon L\|\hat{x}\|_{TV} \geqslant \left|\phi_j(t_i)\right| z_i - \sum_{l\neq i}\left|\phi_j(t_l)\right| z_l. \tag{107}$$

Now, for all i, we select the $j = l(i)$, the index corresponding to the closest sample as defined in Definition 1. The inequalities in (107) can be written as

$$Az \leqslant v, \tag{108}$$

where A, z and v are defined as

$$A = \begin{bmatrix} |\phi_1(t_1)| & -|\phi_1(t_2)| & \dots & -|\phi_1(t_k)| \\ -|\phi_2(t_1)| & |\phi_2(t_2)| & \dots & -|\phi_2(t_k)| \\ \vdots & \vdots & \ddots & \vdots \\ -|\phi_k(t_1)| & -|\phi_k(t_2)| & \dots & |\phi_k(t_k)| \end{bmatrix}, \quad z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_k \end{bmatrix}, \quad v = \left(2\left(1 + \frac{\phi^\infty\|b\|_2}{\bar{f}}\right)\delta + \epsilon L\|\hat{x}\|_{TV}\right)\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

Because A is strictly diagonally dominant, Lemma 28 holds and therefore A^{-1} exists and has non-negative entries, so when we multiply (108) by A^{-1}, the sign does not change:

$$z \leqslant A^{-1}v, \tag{109}$$

where we can bound the entries of A^{-1} [42]:

$$\|A^{-1}\|_\infty < \frac{1}{\min_j(\phi_j(t_j) - \sum_{i\neq j}\phi_j(t_i))}.$$

The proof of Proposition 21 is now complete, since (109) is equivalent to our error bound (28).

G Proof of Proposition 22

For the sake of simplicity, let s_i be the closest sample $s_{l(i)}$ to the source t_i, as defined in Definition 1. For a fixed i, assume (without loss of generality, as we will see later) that $s_i < t_i$. It follows that

$$\begin{aligned}
|s_i - t_l| = s_i - t_l \geqslant (i-l)\Delta - \lambda\Delta, \quad \forall l < i,\\
|s_i - t_l| = t_l - s_i \geqslant (l-i)\Delta + \lambda\Delta, \quad \forall l > i,
\end{aligned}$$

and so

$$\begin{aligned}
\phi(|s_i - t_l|) \leqslant \phi((i-l)\Delta - \lambda\Delta), \quad \forall l < i,\\
\phi(|s_i - t_l|) \leqslant \phi((l-i)\Delta + \lambda\Delta), \quad \forall l > i.
\end{aligned}$$

Then we have

$$\begin{aligned}
\sum_{l\neq i}\phi(|s_i - t_l|) &= \sum_{l=1}^{i-1}\phi(|s_i - t_l|) + \sum_{l=i+1}^{k}\phi(|s_i - t_l|)\\
&\leqslant \sum_{l=1}^{i-1}\phi((i-l)\Delta - \lambda\Delta) + \sum_{l=i+1}^{k}\phi((l-i)\Delta + \lambda\Delta)\\
&= \sum_{l=1}^{i-1}\phi(l\Delta - \lambda\Delta) + \sum_{l=1}^{k-i}\phi(l\Delta + \lambda\Delta). \qquad (110)
\end{aligned}$$

We now want to find upper bounds for each of the two sums in (110). We will derive the bound for the first term, as the second one is similar. We have that

$$\sum_{l=1}^{i-1}\phi(l\Delta - \lambda\Delta) = \phi(\Delta - \lambda\Delta) + \frac{1}{\Delta}\sum_{l=2}^{i-1}\phi(l\Delta - \lambda\Delta)\Delta, \qquad (111)$$

and the sum in the previous equation is a lower Riemann sum (note that ϕ is decreasing in $[0,1]$)

$$S = \sum_{l=2}^{i-1}\phi(x_l^*)(x_l - x_{l-1})$$

of $\phi(x)$ over $[\Delta - \lambda\Delta, (i-1)\Delta - \lambda\Delta]$, with partition and x_l^* chosen as follows:

$$\begin{aligned}
[x_l, x_{l-1}] &= [l\Delta - \lambda\Delta, (l-1)\Delta - \lambda\Delta], \qquad && l = 2\ldots, i-1,\\
x_l^* &= l\Delta - \lambda\Delta, && l = 2, \ldots, i-1.
\end{aligned}$$

Therefore, the sum S is less than or equal to the integral:

$$\sum_{l=2}^{i-1}\phi(l\Delta - \lambda\Delta)\Delta \leqslant \int_{\Delta-\lambda\Delta}^{(i-1)\Delta-\lambda\Delta}\phi(x)\,\mathrm{d}x. \qquad (112)$$

By substituting (112) into (111), we obtain

$$\sum_{l=1}^{i-1}\phi(l\Delta - \lambda\Delta) \leqslant \phi(\Delta - \lambda\Delta) + \frac{1}{\Delta}\int_{\Delta-\lambda\Delta}^{(i-1)\Delta-\lambda\Delta}\phi(x)\,\mathrm{d}x.$$

We can obtain a similar upper bound for the second sum in (110) and then

$$\sum_{l\neq i}\phi(|s_i-t_l|) \leqslant \phi(\Delta-\lambda\Delta)+\phi(\Delta+\lambda\Delta)+\frac{1}{\Delta}\int_{\Delta-\lambda\Delta}^{(i-1)\Delta-\lambda\Delta}\phi(x)\,\mathrm{d}x+\frac{1}{\Delta}\int_{\Delta+\lambda\Delta}^{(k-i)\Delta+\lambda\Delta}\phi(x)\,\mathrm{d}x, \quad \forall i \in [k]. \tag{113}$$

We can further upper bound the right hand side over all $i \in [k]$ and this bound corresponds to the case when the source t_i is in the middle of the unit interval (at $\frac{1}{2}$) and the sources t_1 and t_k are at 0 and 1 respectively:

$$\int_{\Delta-\lambda\Delta}^{(i-1)\Delta-\lambda\Delta}\phi(x)\,\mathrm{d}x+\int_{\Delta+\lambda\Delta}^{(k-i)\Delta+\lambda\Delta}\phi(x)\,\mathrm{d}x \leqslant \int_{\Delta-\lambda\Delta}^{1/2-\lambda\Delta}\phi(x)\,\mathrm{d}x+\int_{\Delta+\lambda\Delta}^{1/2+\lambda\Delta}\phi(x)\,\mathrm{d}x, \quad \forall i \in [k],$$

and therefore we have

$$\sum_{l\neq i}\phi(|s_i-t_l|) \leqslant \phi(\Delta-\lambda\Delta)+\phi(\Delta+\lambda\Delta)+\frac{1}{\Delta}\int_{\Delta-\lambda\Delta}^{1/2-\lambda\Delta}\phi(x)\,\mathrm{d}x+\frac{1}{\Delta}\int_{\Delta+\lambda\Delta}^{1/2+\lambda\Delta}\phi(x)\,\mathrm{d}x, \quad \forall i \in [k].$$

In order to find λ_0, we solve (27) since $|s_i - t_i| \leqslant \lambda\Delta$ implies $\phi(|s_i - t_i|) \geqslant \phi(\lambda\Delta)$.

We note that if we only have three sources, then the integral terms should not be included, and if we have four sources, then the last integral term should not be included.

H Proof of Theorem 13

The proof of Theorem 13 involves the same ideas as Theorem 12. The differences are in the analysis of Proposition 21. We continue this analysis from (101), where we lower bound the left-hand side term of (100):

$$2\delta \geqslant \left|\int_0^1 \phi_j(t)h(\mathrm{d}t)\right| \geqslant \left|\int_{\tilde{T}_{i,\epsilon}}\phi_j(t)\hat{x}(\mathrm{d}t)-\sum_{r\in[k_i]}a_{ir}\phi_j(t_{ir})\right| \tag{114a}$$

$$-\sum_{l\neq i}\left|\int_{\tilde{T}_{l,\epsilon}}\phi_j(t)\hat{x}(\mathrm{d}t)-\sum_{r\in[k_l]}a_{lr}\phi_j(t_{lr})\right| \tag{114b}$$

$$-\left|\int_{T_\epsilon^C}\phi_j(t)\hat{x}(\mathrm{d}t)\right|, \tag{114c}$$

where, in each term, the sum from $r = 1$ to k_i is over all the true sources in $\tilde{T}_{i,\epsilon}$ (for all $i \in [\tilde{k}]$). In order to obtain bounds for the terms (114a) and (114b), we need the following fact:

$$\exists \xi_i \in [\operatorname*{argmin}_{r\in[k_i]}\phi_j(t_{ir}), \operatorname*{argmax}_{r\in[k_i]}\phi_j(t_{ir})] \quad \text{such that} \quad \phi_j(\xi_i)\sum_{r=1}^{k_i}a_{ir} = \sum_{r=1}^{k_r}a_{rk}\phi_j(t_{ir}), \quad \forall r \in [\tilde{k}]. \tag{115}$$

This comes from the continuity of ϕ_j and intermediate value theorem, since:

$$\min_k \phi_j(t_{ir}) \leqslant \frac{\sum_{r=1}^{k_i}a_{ir}\phi_j(t_{ir})}{\sum_{r=1}^{k_i}a_{ir}} \leqslant \max_k \phi_j(t_{ir}).$$

We proceed as before to find a lower bound for (114a) and an upper bound for (114b), while the upper bound for (114c) is the same. For (114a):

$$\begin{aligned}
\left|\int_{\tilde{T}_{i,\epsilon}} \phi_j(t)\hat{x}(\mathrm{d}t) - \sum_{r=1}^{k_i} a_{ir}\phi_j(t_{ir})\right| &= \left|\int_{\tilde{T}_{i,\epsilon}} \phi_j(t)\hat{x}(\mathrm{d}t) - \phi_j(\xi_i)\sum_{r=1}^{k_i} a_{ir}\right| \\
&= \left|\int_{\tilde{T}_{i,\epsilon}} \phi_j(t)\hat{x}(\mathrm{d}t) - \phi_j(\xi_i)\sum_{r=1}^{k_i} a_{ir} + \phi_j(\xi_i)\int_{\tilde{T}_{i,\epsilon}} \hat{x}(\mathrm{d}t) - \phi_j(\xi_i)\int_{\tilde{T}_{i,\epsilon}} \hat{x}(\mathrm{d}t)\right| \\
&\geqslant \left|\phi_j(\xi_i)\right|\left|\int_{\tilde{T}_{i,\epsilon}} \hat{x}(\mathrm{d}t) - \sum_{r=1}^{k_i} a_{ir}\right| - \int_{\tilde{T}_{i,\epsilon}} \left|\phi_j(t)-\phi_j(\xi_i)\right|\hat{x}(\mathrm{d}t) \\
&\geqslant \left|\phi_j(\xi_i)\right|\left|\int_{\tilde{T}_{i,\epsilon}} \hat{x}(\mathrm{d}t) - \sum_{r=1}^{k_i} a_{ir}\right| - L\int_{\tilde{T}_{i,\epsilon}} |t-\xi_i|\,\hat{x}(\mathrm{d}t) \\
&\geqslant \left|\phi_j(\xi_i)\right|\left|\int_{\tilde{T}_{i,\epsilon}} \hat{x}(\mathrm{d}t) - \sum_{r=1}^{k_i} a_{ir}\right| - L(2k_i-1)\epsilon\int_{\tilde{T}_{i,\epsilon}} \hat{x}(\mathrm{d}t),
\end{aligned}$$

where the width of $\tilde{T}_{i,\epsilon}$ is at most $2k_i\epsilon$ and $\xi_i \in \tilde{T}_{i,\epsilon}$ is chosen according to (115), so the distance $|t-\xi_i|$ for $t \in \tilde{T}_{i,\epsilon}$ is at most $(2k_i-1)\epsilon$. For the second term (114b):

$$\begin{aligned}
\left|\int_{\tilde{T}_{l,\epsilon}} \phi_j(t)\hat{x}(\mathrm{d}t) - \sum_{r=1}^{k_l} a_{lr}\phi_j(t_{lr})\right| &= \left|\int_{\tilde{T}_{l,\epsilon}} \phi_j(t)\hat{x}(\mathrm{d}t) - \phi_j(\xi_l)\sum_{r=1}^{k_l} a_{lr}\right| \\
&= \left|\int_{\tilde{T}_{l,\epsilon}} \phi_j(t)\hat{x}(\mathrm{d}t) - \phi_j(\xi_l)\sum_{r=1}^{k_l} a_{lr} + \phi_j(\xi_l)\int_{\tilde{T}_{l,\epsilon}} \hat{x}(\mathrm{d}t) - \phi_j(\xi_l)\int_{\tilde{T}_{l,\epsilon}} \hat{x}(\mathrm{d}t)\right| \\
&\leqslant \left|\phi_j(\xi_l)\right|\left|\int_{\tilde{T}_{l,\epsilon}} \hat{x}(\mathrm{d}t) - \sum_{r=1}^{k_l} a_{lr}\right| + \int_{\tilde{T}_{l,\epsilon}} \left|\phi_j(t)-\phi_j(\xi_l)\right|\hat{x}(\mathrm{d}t) \\
&\leqslant \left|\phi_j(t_l)\right|\left|\int_{\tilde{T}_{l,\epsilon}} \hat{x}(\mathrm{d}t) - \sum_{r=1}^{k_l} a_{lr}\right| + L\int_{\tilde{T}_{l,\epsilon}} |t-\xi_l|\,\hat{x}(\mathrm{d}t) \\
&\leqslant \left|\phi_j(t_l)\right|\left|\int_{\tilde{T}_{l,\epsilon}} \hat{x}(\mathrm{d}t) - \sum_{r=1}^{k_l} a_{lr}\right| + L(2k_l-1)\epsilon\int_{\tilde{T}_{l,\epsilon}} \hat{x}(\mathrm{d}t)
\end{aligned}$$

so

$$\sum_{l\neq i}\left|\int_{\tilde{T}_{l,\epsilon}} \phi_j(t)\hat{x}(\mathrm{d}t) - \sum_{r=1}^{k_l} a_{lr}\phi_j(t_{lr})\right| \leqslant \sum_{l\neq i}\left(\left|\phi_j(\xi_l)\right|\left|\int_{\tilde{T}_{l,\epsilon}} \hat{x}(\mathrm{d}t) - \sum_{r=1}^{k_l} a_{lr}\right|\right) + L\epsilon\sum_{l\neq i}(2k_l-1)\int_{\tilde{T}_{l,\epsilon}} \hat{x}(\mathrm{d}t).$$

Let

$$\tilde{z}_i = \left|\int_{\tilde{T}_{i,\epsilon}} \hat{x}(\mathrm{d}t) - \sum_{r=1}^{k_i} a_{ir}\right|$$

and we obtain an inequality as before:

$$2\left(1+\frac{\phi^\infty\|b\|_2}{\bar{f}}\right)\delta + (2k-1)\epsilon L\|\hat{x}\|_{TV} \geqslant \left|\phi_j(\xi_i)\right|\tilde{z}_i - \sum_{l\neq i}\left|\phi_j(\xi_l)\right|\tilde{z}_l,$$

where we obtained the second constant as follows:

$$L\epsilon(2k_i-1)\int_{\tilde{T}_{i,\epsilon}}\hat{x}(\mathrm{d}t)+L\epsilon\sum_{l\neq i}(2k_l-1)\int_{\tilde{T}_{l,\epsilon}}\hat{x}(\mathrm{d}t) = L\epsilon\sum_{l=1}^{\tilde{k}}(2k_l-1)\int_{\tilde{T}_{l,\epsilon}}\hat{x}(\mathrm{d}t) \leqslant L\epsilon\|\hat{x}\|_{TV}\sum_{l=1}^{\tilde{k}}(2k_l-1) \leqslant (2k-1)L\epsilon\|\hat{x}\|_{TV}$$

and, for all $i \in [\tilde{k}]$, we select $j(i) = \operatorname{argmin}_j |s_j - \xi_i|$. The linear system is

$$\tilde{A}\tilde{z} \leqslant \tilde{v},$$

with:

$$\tilde{A} = \begin{bmatrix} |\phi_1(\xi_1)| & -|\phi_1(\xi_2)| & \dots & -|\phi_1(\xi_{\tilde{k}})| \\ -|\phi_2(\xi_1)| & |\phi_2(\xi_2)| & \dots & -|\phi_2(\xi_{\tilde{k}})| \\ \vdots & \vdots & \ddots & \vdots \\ -|\phi_{\tilde{k}}(\xi_1)| & -|\phi_{\tilde{k}}(\xi_2)| & \dots & |\phi_{\tilde{k}}(\xi_{\tilde{k}})| \end{bmatrix}, \quad \tilde{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_{\tilde{k}} \end{bmatrix}, \quad \tilde{v} = \left(2\left(1+\frac{\phi^\infty\|b\|_2}{\bar{f}}\right)\delta + (2k-1)L\epsilon\|\hat{x}\|_{TV}\right)\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

In Appendix G we discuss what the choice of λ should be so that, if $|t_i - s_i| \leqslant \lambda\Delta$, the matrix A is strictly diagonally dominant. Here, the matrix $\tilde{A}$ is similar to A except that we evaluate ϕ at $|\xi_i - s_i|$, where ξ_i corresponds to a group of sources in $\tilde{T}_{i,\epsilon}$ that are located within distances smaller than 2ϵ and s_i is the closest sample to ξ_i. Given that the minimum separation between ξ_i sources is 2ϵ, the analysis in Appendix G is the same, so $\tilde{A}$ is strictly diagonally dominant if

$$|\xi_i - s_i| \leqslant 2\lambda\epsilon, \quad \forall i \in [\tilde{k}]$$

and λ is chosen to satisfy (27) where we take $\Delta = 2\epsilon$. For the value of λ found this way, we select the sampling locations uniformly at intervals of $2\lambda\epsilon$.

I Proof of Lemma 26

For real s, let us first study the Lipschitz constant of the operator that takes a non-negative measure x supported on I to $\int_I g(s-t)x(\mathrm{d}t)$, where $g(t) = e^{-\frac{t^2}{\sigma^2}}$. To that end, consider a pair of non-negative measures z_1, z_2 supported on I such that $\|z_1\|_{TV} = \|z_2\|_{TV}$. Their Wasserstein distance $d_W(z_1, z_2)$ was defined in (20). The dual of Program (20) is in fact (see, for example [44], Chapter 5)

$$\max_{\omega}\{\int_I \omega(t)(z_1 - z_2)(\mathrm{d}t) : \omega \text{ is 1-Lipschitz}\}. \tag{116}$$

where the supremum is over all (measurable) functions $\omega : I \to \mathbb{R}$. Consider the particular choice of $\omega(t) = \alpha \cdot g(s-t)$ for positive α to be set shortly. Let us see if this choice of ω is feasible for Program (116). For $t_1, t_2 \in I$, we write that

$$\begin{aligned} \omega(t_1) - \omega(t_2) &= \alpha\left(g(s-t_1) - g(s-t_2)\right) \\ &= \alpha\int_{t_1}^{t_2} g'(s-t)\ dt \\ &\leqslant \alpha \max_{t\in[-1,1]} |g'(t)| \cdot |t_2 - t_1| \\ &= \frac{\alpha}{\sigma}\sqrt{\frac{2}{e}} \cdot |t_2 - t_1| \\ &= |t_2 - t_1|, \end{aligned} \tag{117}$$

where we set $\alpha = \frac{\sigma\sqrt{2e}}{2}$ in the last line above, for small enough σ (specifically, for $\sigma < \sqrt{2}$). Therefore, with this choice of α, the function ω specified above is feasible for Program (116). From this observation and the strong duality between Programs (20) and (116), it follows that

$$\frac{\sigma\sqrt{2e}}{2}\int_I g(s-t)(z_1(\mathrm{d}t) - z_2(\mathrm{d}t)) \leqslant d_W(z_1, z_2). \tag{118}$$

Combined with the other direction, we find that

$$\frac{\sigma\sqrt{2e}}{2}\left|\int_I g(s-t)(z_1(\mathrm{d}t)-z_2(\mathrm{d}t))\right| \leqslant d_W(z_1,z_2), \tag{119}$$

namely that the map $z \to \int_I g(s-t)z(\mathrm{d}t)$ is $(2/\sigma\sqrt{2e})$-Lipschitz with respect to the Wasserstein distance.

It is easy to extend the above conclusion to the generalised Wasserstein distance. Consider a pair of non-negative measures x_1, x_2 supported on I, and a pair of non-negative measures z_1, z_2 on I such that $\|z_1\|_{TV} = \|z_2\|_{TV}$. Then, using the triangle inequality, we write that

$$\begin{aligned}
&\left|\int_I g(s-t)(x_1(\mathrm{d}t)-x_2(\mathrm{d}t))\right| \\
&\leqslant \left|\int_I g(s-t)(x_1(\mathrm{d}t)-z_1(\mathrm{d}t))\right| + \left|\int_I g(s-t)(z_1(\mathrm{d}t)-z_2(\mathrm{d}t))\right| + \left|\int_I g(s-t)(z_2(\mathrm{d}t)-x_2(\mathrm{d}t))\right| \\
&\leqslant \int_I \left|x_1(\mathrm{d}t)-z_1(\mathrm{d}t)\right| + \left|\int_I g(s-t)(z_1(\mathrm{d}t)-z_2(\mathrm{d}t))\right| + \int_I \left|z_2(\mathrm{d}t)-x_2(\mathrm{d}t)\right| \qquad \big(g(t)\leqslant 1\big) \\
&= \|x_1-z_1\|_{TV} + \left|\int_I g(s-t)(z_1(\mathrm{d}t)-z_2(\mathrm{d}t))\right| + \|z_2-x_2\|_{TV} \\
&\leqslant \|x_1-z_1\|_{TV} + \frac{2}{\sigma\sqrt{2e}} d_W(z_1,z_2) + \|z_2-x_2\|_{TV} \qquad \text{(see (119))} \\
&\leqslant \frac{2}{\sigma\sqrt{2e}}\big(\|x_1-z_1\|_{TV} + d_W(z_1,z_2) + \|z_2-x_2\|_{TV}\big) \qquad \left(\text{for } \sigma \leqslant \frac{2}{\sqrt{2e}}\right).
\end{aligned} \tag{120}$$

The choice of z_1, z_2 above was arbitrary and therefore, recalling the definition of d_{GW} in (19), we find that

$$\left|\int_I g(s-t)(x_1(\mathrm{d}t)-x_2(\mathrm{d}t))\right| \leqslant \frac{2}{\sigma\sqrt{2e}} d_{GW}(x_1,x_2), \tag{121}$$

namely the map $x \to \int_I g(s-t)x(\mathrm{d}t)$ is a $\frac{2}{\sigma\sqrt{2e}}$-Lipschitz operator. It immediately follows that the operator Φ that was formed using the sampling points $S = \{s_j\}_{j=1}^m$ in (3) is $\frac{2\sqrt{m}}{\sigma\sqrt{2e}}$-Lipschitz. This completes the proof of Lemma 26.

J Proof of Lemma 23

Following the definition of T*-systems in Definition 9, consider an increasing sequence $\{\tau_l\}_{l=0}^m \subset I$ such that $\tau_0 = 0$, $\tau_m = 1$, and except one more point (say $\tau_{\underline{l}}$), the rest of points belong to T_ρ, the ρ-neighbourhood of the support $T \subset \mathrm{int}(I)$.

We also select the subset of samples S of size $2k+2$ that is closest to the support T (in Hausdorff distance), so without loss of generality, we set $m = 2k+2$. With this assumption, the setup in Definition 9 forces that every neighbourhood $T_{i,\rho}$ contain exactly two points, say t_i and $t_{i,\rho} := t_i + \rho$ to simplify the presentation. Then the determinant in part 1 of Definition 9 can be written as

$$
M^\rho = \begin{vmatrix}
F(0) & g(s_1) & \cdots & g(s_m) \\
F(t_1) & g(t_1 - s_1) & \cdots & g(t_1 - s_m) \\
F\left(t_{1,\rho}\right) & g\left(t_{1,\rho} - s_1\right) & \cdots & g\left(t_{1,\rho} - s_m\right) \\
\vdots & & & \\
F\left(\tau_{\underline{l}}\right) & g\left(\tau_{\underline{l}} - s_1\right) & \cdots & g\left(\tau_{\underline{l}} - s_m\right) \\
\vdots & & & \\
F(t_k) & g(t_k - s_1) & \cdots & g(\tau - s_m) \\
F\left(t_{k,\rho}\right) & g\left(t_{k,\rho} - s_1\right) & \cdots & g\left(t_{k,\rho} - s_m\right) \\
F(1) & g(1 - s_1) & \cdots & g(1 - s_m)
\end{vmatrix}
$$

$$
= \begin{vmatrix}
f_0 & g(s_1) & \cdots & g(s_m) \\
0 & g(t_1 - s_1) & \cdots & g(t_1 - s_m) \\
0 & g\left(t_{1,\rho} - s_1\right) & \cdots & g\left(t_{1,\rho} - s_m\right) \\
\vdots & & & \\
F\left(\tau_{\underline{l}}\right) & g\left(\tau_{\underline{l}} - s_1\right) & \cdots & g\left(\tau_{\underline{l}} - s_m\right) \\
\vdots & & & \\
0 & g(t_k - s_1) & \cdots & g(\tau - s_m) \\
0 & g\left(t_{k,\rho} - s_1\right) & \cdots & g\left(t_{k,\rho} - s_m\right) \\
f_1 & g(1 - s_1) & \cdots & g(1 - s_m)
\end{vmatrix}. \qquad \text{(evaluating } F(t) \text{ as in (44))} \tag{122}
$$

We will now need the following lemma in order to simplify the determinant above. The result is proved in Appendix L.

Lemma 29. *Let $A, B \in \mathbb{R}^{m\times m}$ with $m \geqslant 2$ and $\det(A) > 0$. If $0 \leqslant \epsilon \leqslant \frac{8}{34m\rho(A^{-1}B)}$, then*

$$
\det(A)\left(1 - \frac{17\sqrt{e}}{8} m\epsilon\rho(A^{-1}B)\right) \leqslant det(A + \epsilon B) \leqslant \det(A)\left(1 + \frac{17\sqrt{e}}{8} m\epsilon\rho(A^{-1}B)\right),
$$

where $\rho(X)$ is the spectral radius of the matrix X. In particular, for the stated choice of ϵ,

$$
\det(A)\left(1 - \frac{\sqrt{e}}{2}\right) \leqslant \det(A + \epsilon B) \leqslant \det(A)\left(1 + \frac{\sqrt{e}}{2}\right).
$$

As $\rho \to 0$, note that $g(t_{i,\rho} - s_j) = g(t_i - s_j) + \rho \cdot g'(t_i - s_j) + \frac{\rho^2}{2} g''(\xi)$, for some $\xi \in [t_i - s_j, t_i - s_j + \rho]$. After applying this expansion, we subtract the rows with $g(t_i - s_j)$ from the rows with $g(t_{i,\rho} - s_j)$, take ρ^k outside of the determinant and we can write M^ρ as:

$$
M^\rho = \rho^k \det(M_N + \rho M_P),
$$

M_N is a matrix with entries independent of ρ and with determinant:

$$
N = \det(M_N) = \begin{vmatrix}
f_0 & g(s_1) & \cdots & g(s_m) \\
0 & g(t_1 - s_1) & \cdots & g(t_1 - s_m) \\
0 & g'(t_1 - s_1) & \cdots & g'(t_1 - s_m) \\
\vdots & & & \\
F\left(\tau_{\underline{l}}\right) & g\left(\tau_{\underline{l}} - s_1\right) & \cdots & g\left(\tau_{\underline{l}} - s_m\right) \\
\vdots & & & \\
0 & g(t_k - s_1) & \cdots & g(\tau - s_m) \\
0 & g'(t_k - s_1) & \cdots & g'(t_k - s_m) \\
f_1 & g(1 - s_1) & \cdots & g(1 - s_m)
\end{vmatrix}. \tag{123}
$$

Moreover, while the entries of M_P depend on ρ, the magnitude of each entry can be bounded from above independently of ρ. Consequently, $\|M_P\|_F$ is bounded from above independently of ρ. Let us assume for the moment that $N > 0$ and so M_N is invertible. Since $\rho(M_N^{-1}M_P) \leqslant \left\|M_N^{-1}\right\|_2 \|M_P\|_F$, this implies that $\rho(M_N^{-1}M_P)$ is bounded from above independently of ρ. We can then apply the stronger result of Lemma 29 to M^ρ and obtain:

$$0 < (1-\rho C_N)\rho^k N \leqslant M^\rho \leqslant (1+\rho C_N)\rho^k N, \tag{124}$$

where $C_N > 0$ is a constant that does not depend on ρ. Note that we do not need write the condition on ρ required by Lemma 29 explicitly because $\rho \to 0$ and also that (124) applies to the minors of M^ρ and N along the row containing $\tau_{\underline{l}}$. That N is indeed positive (and therefore we can apply Lemma 29) is established below By its definition in (44), $F(\tau_{\underline{l}})$ can take two values above. Either

- $F(\tau_{\underline{l}}) = 0$, which happens when there exists $i_0 \in [k]$ such that $\tau_{\underline{l}} \in T_{i_0,\epsilon}$, namely when $\tau_{\underline{l}}$ is close to the support T. In this case, by applying the Laplace expansion to N, we find that

$$N = f_0 \cdot N_{1,1} + f_1 \cdot N_{m+1,1}, \tag{125}$$

where $N_{1,1}$ are $N_{m+1,1}$ are the corresponding minors in (123). Note that both $N_{1,1}$ and $N_{m+1,1}$ are positive because the Gaussian window is *extended totally positive*, see Example 5 in [27]. Recalling that $f_0, f_1 > 0$, we conclude that N is positive. Therefore, when ρ is sufficiently small, (123) implies that M^ρ is non-negative when $F(\tau_{\underline{l}}) = 0$. Or

- $F(\tau_{\underline{l}}) = \bar{f}$, which happens when $\tau_{\underline{l}} \in T_\epsilon^C$, namely when $\tau_{\underline{l}}$ is away from the support T. Suppose that $f_0 \gg \bar{f}$ so that N is dominated by its first minor, namely $N_{1,1}$. More precisely, by applying the Laplace expansion to N in (123), we find that

$$N = f_0 \cdot N_{1,1} - \bar{f} \cdot N_{\underline{l},1} + f_1 \cdot N_{m+1,1}, \tag{126}$$

in which all three minors are positive because the Gaussian window is extended totally positive. Also, note that $N_{\underline{l},1}$ does *not* depend on $\tau_{\underline{l}}$ and recall also that $f_0, \bar{f}, f_1$ are all positive. Therefore N in (126) is positive if

$$\frac{f_0}{\bar{f}} > \frac{N_{\underline{l},1}}{\min_{\tau_{\underline{l}}} N_{1,1}}, \tag{127}$$

where the minimum is over $\tau_{\underline{l}} \in T_\epsilon^C$. The right-hand side above is well-defined because $N_{1,1} = N_{1,1}(\tau_{\underline{l}})$ is positive for every $\tau_{\underline{l}} \in I$, $N_{1,1}(\tau_{\underline{l}})$ is a continuous function of $\tau_{\underline{l}}$, and I is compact. Indeed, $N_{1,1}(\tau_{\underline{l}})$ is positive because the Gaussian window is extended totally positive. As before, N being positive implies that M^ρ is non-negative when ρ is sufficiently small, see (123).

By combining both cases above, we conclude that M^ρ is non-negative for sufficiently small ρ provided that (127) holds, thereby verifying part 1 of Definition 9. To verify part 2 of that definition, consider the minors along the row containing $\tau_{\underline{l}}$ in M^ρ, see (123). Starting with the first minor along this row and applying the same arguments as before for M^ρ, we observe, after applying Lemma 29, that

$$M_{\underline{l},1}^\rho \geqslant (1-\rho C_{\underline{l},1})\rho^k \begin{vmatrix} g(s_1) & \cdots & g(s_m) \\ g(t_1 - s_1) & \cdots & g(t_1 - s_m) \\ g'(t_1 - s_1) & \cdots & g'(t_1 - s_m) \\ \vdots & & \vdots \\ g(t_k - s_1) & \cdots & g(t_1 - s_m) \\ g'(t_k - s_1) & \cdots & g'(t_k - s_m) \\ g(1 - s_1) & \cdots & g(1 - s_m) \end{vmatrix} =: (1-\rho C_{\underline{l},1})\rho^k \cdot N_{\underline{l},1}, \tag{128}$$

and also $M_{\underline{l},1}^\rho \leqslant (1+\rho C_{\underline{l},1})\rho^k \cdot N_{\underline{l},1}$ as $\rho \to 0$. Here $C_{\underline{l},1} > 0$ is a constant that does not depend on ρ. Moreover, $N_{\underline{l},1}$ does not depend on ρ and is positive because the Gaussian window is extended totally positive. Therefore $M_{\underline{l},1}^\rho$ in (128) approaches zero at the rate ρ^k.

Consider next the $(j+1)$th minor along the row containing $\tau_{\underline{l}}$ of M^ρ in (123), namely $M^\rho_{\underline{l},j+1}$ with $j = 1, \ldots, m$. Using the same arguments as before, we obtain after applying Lemma 29 that

$$
\begin{aligned}
M^\rho_{\underline{l},j+1} &\geqslant (1-\rho C_{\underline{l},j+1})\rho^k \begin{vmatrix}
f_0 & g(s_1) & \cdots & g(s_{j-1}) & g(s_{j+1}) & \cdots & g(s_m) \\
0 & g(t_1-s_1) & \cdots & g(t_1-s_{j-1}) & g(t_1-s_{j+1}) & \cdots & g(t_1-s_m) \\
0 & g'(t_1-s_1) & \cdots & g'(t_1-s_{j-1}) & g'(t_1-s_{j+1}) & \cdots & g'(t_1-s_m) \\
\vdots & & & & & & \\
0 & g(t_k-s_1) & \cdots & g(t_k-s_{j-1}) & g(t_k-s_{j+1}) & \cdots & g(t_k-s_m) \\
0 & g'(t_k-s_1) & \cdots & g'(t_k-s_{j-1}) & g'(t_k-s_{j+1}) & \cdots & g'(t_k-s_m) \\
f_1 & g(1-s_1) & \cdots & g(1-s_{j-1}) & g(1-s_{j+1}) & \cdots & g(1-s_m)
\end{vmatrix} \\
&= (1-\rho C_{\underline{l},j+1})\rho^k f_0 \begin{vmatrix}
g(t_1-s_1) & \cdots & g\left(t_1-s_{j-1}\right) & g\left(t_1-s_{j+1}\right) & \cdots & g(t_1-s_m) \\
g'(t_1-s_1) & \cdots & g'\left(t_1-s_{j-1}\right) & g'\left(t_1-s_{j+1}\right) & \cdots & g'(t_1-s_m) \\
\vdots & & & & & \\
g(t_k-s_1) & \cdots & g\left(t_k-s_{j-1}\right) & g\left(t_k-s_{j+1}\right) & \cdots & g(t_k-s_m) \\
g'(t_k-s_1) & \cdots & g'\left(t_k-s_{j-1}\right) & g'\left(t_k-s_{j+1}\right) & \cdots & g'(t_k-s_m) \\
g(1-s_1) & \cdots & g\left(1-s_{j-1}\right) & g\left(1-s_{j+1}\right) & \cdots & g(1-s_m)
\end{vmatrix} \\
&\quad - (1-\rho C_{\underline{l},j+1})\rho^k f_1 \begin{vmatrix}
g(s_1) & \cdots & g\left(s_{j-1}\right) & g\left(s_{j+1}\right) & \cdots & g(s_m) \\
g(t_1-s_1) & \cdots & g\left(t_1-s_{j-1}\right) & g\left(t_1-s_{j+1}\right) & \cdots & g(t_1-s_m) \\
g'(t_1-s_1) & \cdots & g'\left(t_1-s_{j-1}\right) & g'\left(t_1-s_{j+1}\right) & \cdots & g'(t_1-s_m) \\
\vdots & & & & & \\
g(t_k-s_1) & \cdots & g\left(t_k-s_{j-1}\right) & g\left(t_k-s_{j+1}\right) & \cdots & g(t_k-s_m) \\
g'(t_k-s_1) & \cdots & g'\left(t_k-s_{j-1}\right) & g'\left(t_k-s_{j+1}\right) & \cdots & g'(t_k-s_m)
\end{vmatrix} \\
&=: (1-\rho C_{\underline{l},j+1})\rho^k \left(f_0 \cdot N_{\underline{l},j+1,0} - f_1 \cdot N_{\underline{l},j+1,1}\right) =: (1-\rho C_{\underline{l},j+1})\rho^k \cdot N_{\underline{l},j+1},
\end{aligned}
\tag{129}
$$

and also $M^\rho_{\underline{l},j+1} \leqslant (1+\rho C_{\underline{l},j+1})\rho^k \cdot N_{\underline{l},j+1}$ as $\rho \to 0$, provided $N_{\underline{l},j+1} > 0$. Here, $N_{\underline{l},j+1}$ is the determinant on the first line of (129) and $C_{\underline{l},j+1} > 0$ is a constant independent of ρ. Note that $N_{\underline{l},j+1,0}$ and $N_{\underline{l},j+1,1}$ are both positive because the Gaussian window is extended totally positive. To ensure $N_{\underline{l},j+1} > 0$, we require $f_0 \gg f_1$, or more precisely, the following to hold.

$$
\frac{f_0}{f_1} > \frac{N_{\underline{l},j+1,1}}{N_{\underline{l},j+1,0}}. \tag{130}
$$

It then follows that $M^\rho_{\underline{l},j+1}$ approaches zero at the rate ρ^k for every j, thereby verifying part 2 in Definition 9 for $\tau_{\underline{l}} \in int(I)$. In conclusion, we find that $\{F\} \cup \{\phi_j\}_{j=1}^m$ form a T*-system on I with $\phi_j(t) = g(t-s_j) = e^{-\frac{(t-s_j)^2}{\sigma^2}}$ as in (9), provided that (127) and (130) hold.

To establish that $\{F^\pi\} \cup \{\phi_j\}_{j=1}^m$ form a T*-system on I with the Gaussian window, we note that $F(\tau_{\underline{l}})$ is replaced by by $F^\pi(\tau_{\underline{l}})$, which takes values ± 1 when $\tau_{\underline{l}} \in T_{i,\epsilon}$ for some i, as indicated in (45). The previous argument, thus, goes through similarly, showing that $\{F^\pi\} \cup \{\phi_j\}_{j=1}^m$ is a T*-system on I for arbitrary sign pattern π when (127) and (130) hold and $f_0 \gg 1$. The extra condition $f_0 \gg 1$ comes from the only difference between the two proofs, namely that, instead of (125), we have

$$
N = f_0 \cdot N_{1,1} \pm 1 \cdot N_{\underline{l},1} + f_1 \cdot N_{m+1,1}. \tag{131}
$$

Therefore to ensure $N > 0$, we now require the following to hold.

$$
f_0 > \frac{N_{\underline{l},1}}{\min_{\tau_{\underline{l}}} N_{1,1}}. \tag{132}
$$

We leave out the mostly repetitive details. Hence if (127), (130) and (132) hold, then $\{F^\pi\} \cup \{\phi_j\}_{j=1}^m$ form a T*-system on I with the Gaussian window. This completes the proof of Lemma 23.

Remark 1. While we do not require that $f_1 \gg \bar{f}$, if we impose that both $f_0 \gg \bar{f}$ and $f_1 \gg \bar{f}$, then (126) holds for smaller $\frac{f_0}{\bar{f}}$, so it is useful in practice. Similarly, from (131) we want that also $f_1 \gg 1$.

Remark 2. From this proof and in light of the first remark, we see that we only need to specify one end point rather than both, so we only need $m = 2k+1$. However, the dual polynomial is better in practice if we have conditions at both end points (if we specify both f_0 and f_1).

K Proof of Lemma 24

Recall our assumptions that

$$m = 2k+2, \qquad s_1 = 0, \quad s_m = s_{2k+2} = 1, \tag{133}$$

and that

$$|s_{2i} - t_i| \leqslant \eta, \qquad s_{2i+1} - s_{2i} = \eta, \qquad \forall i \in [k], \tag{134}$$

That is, we collect two samples near each impulse in x, supported on T. In addition, we make the following assumptions on η and σ:

$$\sigma \leqslant \sqrt{2}, \quad \Delta > \sigma\sqrt{\log\left(3 + \frac{4}{\sigma^2}\right)}, \quad \eta \leqslant \sigma^2. \tag{135}$$

After studying Appendix E, it becomes clear that the entries of $b \in \mathbb{R}^m$ are specified as

$$b_j = \lim_{\rho\to 0}(-1)^{j+1}\frac{M^\rho_{\underline{l},j+1}}{M^\rho_{\underline{l},1}} = (-1)^{j+1}\lim_{\rho\to 0}\frac{M^\rho_{\underline{l},j+1}}{M^\rho_{\underline{l},1}}, \qquad j \in [m], \tag{136}$$

where the numerator and the denominator are the minors $\{M^\rho_{\underline{l},j}\}_{j=1}^{m+1}$ of M^ρ in (122) along the row containing $\tau_{\underline{l}}$. Using the upper and lower bounds on these quantities derived earlier, we obtain:

$$\frac{(1-\rho C_{\underline{l},j+1})N_{\underline{l},j+1}}{(1+\rho C_{\underline{l},1})N_{\underline{l},1}} \leqslant \frac{M^\rho_{\underline{l},j+1}}{M^\rho_{\underline{l},1}} \leqslant \frac{(1+\rho C_{\underline{l},j+1})N_{\underline{l},j+1}}{(1-\rho C_{\underline{l},1})N_{\underline{l},1}}, \qquad j \in [m], \tag{137}$$

which in turn implies the following expression for b_j:

$$b_j = (-1)^{j+1}\frac{N_{\underline{l},j+1}}{N_{\underline{l},1}}, \qquad j \in [m]. \tag{138}$$

Recall that $N_{\underline{l},1}, N_{\underline{l},j+1} > 0$ and so, $|b_j| = \frac{N_{\underline{l},j+1}}{N_{\underline{l},1}}$ for $j \in [m]$. Therefore, in order to upper bound each $|b_j|$, we will respectively lower bound the denominator $N_{\underline{l},1}$ and upper bound the numerators $N_{\underline{l},j+1}$.

K.1 Bound on the Denominator of (138)

We now find a lower bound for the first minor, namely $N_{\underline{l},1}$ in (138). Let us conveniently assume that the spike locations $T = \{t_i\}_{i=1}^k$ and the sampling points $S = \{s_j\}_{j=2}^{m-1}$ are away from the boundary of interval $I = [0,1]$, namely

$$\sigma\sqrt{\log(1/\eta^3)} \leqslant t_i \leqslant 1 - \sigma\sqrt{\log(1/\eta^3)}, \qquad \forall i \in [k],$$
$$\sigma\sqrt{\log(1/\eta^3)} \leqslant s_j \leqslant 1 - \sigma\sqrt{\log(1/\eta^3)}, \qquad \forall j \in [2:m-1]. \tag{139}$$

In particular, (139) implies that

$$g(t_i) \leqslant \eta^3, \quad g(1-t_i) \leqslant \eta^3, \qquad i \in [k],$$
$$g(s_j) \leqslant \eta^3, \quad g(1-s_j) \leqslant \eta^3, \qquad j \in [2:m-1]. \tag{140}$$

For the derivatives, we have that:

$$|g'(t_i)| = \frac{2t_i}{\sigma^2}g(t_i) \leqslant \frac{2\eta^3}{\sigma^2},$$

where we used the fact that $0 \leqslant t_i \leqslant 1$ and (140). Similarly, for the $1-t_i$, s_j and $1-s_j$:

$$\begin{aligned} |g'(t_i)| \leqslant \frac{2\eta^3}{\sigma^2}, \quad |g'(1-t_i)| \leqslant \frac{2\eta^3}{\sigma^2}, \qquad i \in [k], \\ |g'(s_j)| \leqslant \frac{2\eta^3}{\sigma^2}, \quad |g'(1-s_j)| \leqslant \frac{2\eta^3}{\sigma^2}, \qquad j \in [2:m-1]. \end{aligned} \tag{141}$$

With the assumptions in (133), (134), (140), (141) and the fact that $g(0)=1$, we have that the determinant $N_{\underline{l},1}$ in (128) is equal to

$$N_{\underline{l},1} = \begin{vmatrix} 1 & \cdots & O(\eta^3) & O(\eta^3) & \cdots & O(\eta^3) \\ \vdots & & \vdots & \vdots & & \vdots \\ O(\eta^3) & \cdots & g(t_i - s_{2u}) & g(t_i - s_{2u+1}) & \cdots & O(\eta^3) \\ O(\eta^3/\sigma^2) & \cdots & g'(t_i - s_{2u}) & g'(t_i - s_{2u+1}) & \cdots & O(\eta^3/\sigma^2) \\ \vdots & & \vdots & \vdots & & \vdots \\ O(\eta^3) & \cdots & O(\eta^3) & O(\eta^3) & \cdots & 1 \end{vmatrix}, \tag{142}$$

where we wrote

$$g(t_i) = O(\eta^3) \text{ and } g'(t_i) = O(\frac{\eta^3}{\sigma^2}) \iff |g(t_i)| \leqslant M_1 \eta^3 \text{ and } |g'(t_i)| \leqslant M_2 \frac{\eta^3}{\sigma^2}, \tag{143}$$

for t_i within the bounds defined in (139) and some $M_1, M_2 > 0$. Here, we can take $M_1 = M_2 = 2$ and we use the same notation for $1-t_i$, s_j and $1-s_j$ with the same constants $M_1 = M_2 = 2$. We then take the Taylor expansion of $g(t_i - s_{2u+1})$ and $g'(t_i - s_{2u+1})$ around $t_i - s_{2u}$, subtract the columns with $t_i - s_{2u}$ from the columns where we performed the expansion, take η^k outside of the determinant and we obtain:

$$N_{\underline{l},1} = \eta^k |C + \eta C'|, \tag{144}$$

where

$$C = \begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & g(t_i - s_{2u}) & -g'(t_i - s_{2u}) & \cdots & 0 \\ 0 & \cdots & g'(t_i - s_{2u}) & -g''(t_i - s_{2u}) & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{145}$$

$$C' = \begin{bmatrix} 0 & \cdots & O(\eta^2) & O(2\eta) & \cdots & O(\eta^2) \\ \vdots & & \vdots & \vdots & & \vdots \\ O(\eta^2) & \cdots & 0 & \frac{1}{2} g''(\xi_{i,u}) & \cdots & O(\eta^2) \\ O(\eta^2/\sigma^2) & \cdots & 0 & \frac{1}{2} g'''(\xi'_{i,u}) & \cdots & O(\eta^2/\sigma^2) \\ \vdots & & \vdots & \vdots & & \vdots \\ O(\eta^2) & \cdots & O(\eta^2) & O(2\eta) & \cdots & 0 \end{bmatrix}, \tag{146}$$

for some $\xi_{i,u}, \xi'_{i,u} \in [t_i - s_{2u} - \eta, t_i - s_{2u}]$ for all $i, u = 1, \dots, k$. Note that, using the notation in (143), if we subtract a function that is $O(\eta)$ with constant $M_1 > 0$ from another function that is $O(\eta)$ with constant $M_2 > 0$, we obtain a function that is $O(\eta)$ with constant $M_1 + M_2$, which is why we wrote $O(2\eta)$ on the first and last columns where we subtracted two functions $O(\eta)$ with the same constant M_1 (so $O(2\eta)$ implies $\leqslant 2M_1\eta$). Next, we apply Taylor expansion around $t_i - t_u$ in the terms with $t_i - s_{2u}$ in C as follows:

$$g(t_i - s_{2u}) = g(t_i - t_u + t_u - s_{2u}) = g(t_i - t_u) + (t_u - s_{2u}) g'(\xi^*_{i,u}), \tag{147}$$

for some $\xi^*_{i,u} \in [t_i - t_u - |t_u - s_{2u}|, t_i - t_u + |t_u - s_{2u}|]$ for all $i, u = 1, \ldots, k$. Note that $|t_u - s_{2u}| \leqslant \eta$ according to (134). By applying a similar Taylor expansion to g' and g'', we can write

$$C = A + \eta A', \tag{148}$$

where

$$A = \begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & g(t_i - t_u) & -g'(t_i - t_u) & \cdots & 0 \\ 0 & \cdots & g'(t_i - t_u) & -g''(t_i - t_u) & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{149}$$

and

$$A' = \begin{bmatrix} 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & \frac{t_u - s_{2u}}{\eta} g'(\xi^*_{i,u}) & -\frac{t_u - s_{2u}}{\eta} g''(\xi^{*\prime}_{i,u}) & \cdots & 0 \\ 0 & \cdots & \frac{t_u - s_{2u}}{\eta} g''(\xi^{*\prime}_{i,u}) & -\frac{t_u - s_{2u}}{\eta} g'''(\xi^{*\prime\prime}_{i,u}) & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{150}$$

for some $\xi^*_{i,u}, \xi^{*\prime}_{i,u}, \xi^{*\prime\prime}_{i,u} \in [t_i - t_u - |t_u - s_{2u}|, t_i - t_u + |t_u - s_{2u}|]$ for all $i, u = 1, \ldots, k$. We now substitute (148) into (144) and we obtain:

$$N_{\underline{l},1} = \eta^k |A + \eta(A' + C')|, \tag{151}$$

Assuming for the moment that $|A| > 0$ holds, we obtain via Lemma 29 the following bound:

$$\eta^k \left(1 - \frac{\sqrt{e}}{2}\right) \det(A) \leqslant N_{\underline{l},1} \leqslant \eta^k \left(1 + \frac{\sqrt{e}}{2}\right) \det(A) \tag{152}$$

if

$$\eta \leqslant \frac{8}{34(2k+2)\rho(A^{-1}(A' + C'))}. \tag{153}$$

We look closer at the condition (153) on η in Section K.3. That $|A| > 0$ (and therefore our application of Lemma 29 above is valid) is established below. We can in fact write A more compactly as follows. For scalar t, let

$$H(t) := \begin{bmatrix} g(t) & -g'(t) \\ g'(t) & -g''(t) \end{bmatrix} \in \mathbb{R}^{2 \times 2}, \tag{154}$$

which allows us to rewrite A as

$$A = \begin{bmatrix} 1 & 0_{1 \times 2k} & 0 \\ 0_{2k \times 1} & B & 0_{2k \times 1} \\ 0 & 0_{1 \times 2k} & 1 \end{bmatrix} \in \mathbb{R}^{m \times m}, \tag{155}$$

$$B := \begin{bmatrix} H(0) & H(t_1 - t_2) & H(t_1 - t_3) & \cdots & H(t_1 - t_k) \\ H(t_2 - t_1) & H(0) & H(t_2 - t_3) & \cdots & H(t_2 - t_k) \\ \vdots & & & & \\ H(t_k - t_1) & H(t_k - t_2) & H(t_k - t_3) & \cdots & H(0) \end{bmatrix} \in \mathbb{R}^{2k \times 2k}, \tag{156}$$

where $0_{a \times b}$ is the matrix of zeros of size $a \times b$. It follows from (155) that $|A| = |B|$ by Laplace expansion of $|A|$. In particular, the eigenvalues of A are: $1, 1$ and the eigenvalues of B. Let us note that B is a symmetric matrix[5], since $H(-t) = H(t)^T$; hence, A is also symmetric. We now proceed to lower bound $|B|$, the details of which are given in Appendix M. The main observation is that $H(0)$ is a diagonal matrix while the entries of $H(t_i - t_j)$ for $i \neq j$ decay with the separation of sources Δ.

[5] Indeed, g, g'' are even functions and g' is an odd function.

Lemma 30. *Let $\sigma \leqslant \sqrt{2}$, $\Delta > \sigma\sqrt{\log\left(3+\frac{4}{\sigma^2}\right)}$ and*

$$0 < \quad F_{\min}\left(\Delta, \frac{1}{\sigma}\right) = 1 - \left(1 + \frac{2}{\sigma^2}\right)\frac{2e^{-\frac{\Delta^2}{\sigma^2}}}{1 - e^{-\frac{\Delta^2}{\sigma^2}}} \quad < 1.$$

Then, for each $i = 1, \dots, 2k$, it holds that

$$\lambda_i(B) \geqslant F_{\min}\left(\Delta, \frac{1}{\sigma}\right).$$

Since $|A| = |B|$, we obtain via Lemma 30 that $|A| \geqslant F_{\min}\left(\Delta, \frac{1}{\sigma}\right)^{2k} > 0$. Using this in (152), leads to the following bound:

$$N_{\underline{l},1} \geqslant \eta^k \left(1 - \frac{\sqrt{e}}{2}\right) F_{\min}\left(\Delta, \frac{1}{\sigma}\right)^{2k}. \tag{157}$$

K.2 Bound on the Numerator of (138)

Since (124) holds for all the minors $M^{\rho}_{\underline{l},j}$ and $N_{\underline{l},j}$, let us now upper bound the $N_{\underline{l},j}$ for $j = 2, \dots, m+1$ in the numerator of (138). Note that we distinguish two cases: $j \in \{3, \dots, m\}$ and $j \in \{2, m+1\}$. To simplify the presentation for the first case, suppose, for example, that $j = 3$. Using the assumptions in (133), (134), (140), (141) and the fact that $g(0) = 1$,

$$N_{\underline{l},3} = \begin{vmatrix} f_0 & 1 & O(\eta^3) & \cdots & O(\eta^3) & O(\eta^3) & \cdots & O(\eta^3) \\ \vdots & & & & & & & \\ 0 & O(\eta^3) & g(t_i - s_3) & \cdots & g(t_i - s_{2u}) & g(t_i - s_{2u+1}) & \cdots & O(\eta^3) \\ 0 & O(\eta^3/\sigma^2) & g'(t_i - s_3) & \cdots & g'(t_i - s_{2u}) & g'(t_i - s_{2u+1}) & \cdots & O(\eta^3/\sigma^2) \\ \vdots & & & & & & & \\ f_1 & O(\eta^3) & O(\eta^3) & \cdots & O(\eta^3) & O(\eta^3) & \cdots & 1 \end{vmatrix}. \tag{158}$$

We now expand $g(t_i - s_{2u+1})$ around $g(t_i - s_{2u})$, subtract the columns and take η out of the determinant as before:

$$N_{\underline{l},3} = \eta^{k-1} \begin{vmatrix} f_0 & 1 & O(\eta^3) & \cdots & O(\eta^3) & O(2\eta^2) & \cdots & O(\eta^3) \\ \vdots & & & & & & & \\ 0 & O(\eta^3) & g(t_i - s_3) & \cdots & g(t_i - s_{2u}) & -g'(\xi_{i,u}) & \cdots & O(\eta^3) \\ 0 & O(\eta^3/\sigma^2) & g'(t_i - s_3) & \cdots & g'(t_i - s_{2u}) & -g''(\xi'_{i,u}) & \cdots & O(\eta^3/\sigma^2) \\ \vdots & & & & & & & \\ f_1 & O(\eta^3) & O(\eta^3) & \cdots & O(\eta^3) & O(2\eta^2) & \cdots & 1 \end{vmatrix} =: \eta^{k-1}\det(\tilde{N}_3), \tag{159}$$

where $\xi_{i,u}, \xi'_{i,u} \in [t_i - s_{2u} - \eta, t_i - s_{2u}]$ and denote the matrix in (159) by $\tilde{N}_3$. Note the η^{k-1} since we only perform column operations on $k-1$ columns and $\det(\tilde{N}_3) > 0$ (due to the choice of f_0 and f_1). Next let $C \in \mathbb{R}^{m\times 3}$ consist of the first, second, and last columns of $\tilde{N}_3$ and $D \in \mathbb{R}^{m\times(m-3)}$ consist of the rest of the columns of $\tilde{N}_3$. Then, we may write that

$$\det([CD])^2 = \left(\begin{bmatrix} C^* \\ D^* \end{bmatrix}[CD]\right) = \det\left(\begin{bmatrix} C^*C & C^*D \\ D^*C & D^*D \end{bmatrix}\right) \leqslant \det(C^*C)\det(D^*D), \tag{160}$$

where we note that swapping columns only changes the sign in a determinant (here, $\det([CD]) = -\det(\tilde{N}_3)$) and in the last inequality we applied Fischer's inequality (see, for example, Theorem 7.8.3 in [45]), which works because the matrix $[CD]^*[CD]$ is Hermitian positive definite. Therefore, we have that

$$\det(\tilde{N}_3) = |\det([CD])| \leqslant \det(C^*C)^{\frac{1}{2}}\det(D^*D)^{\frac{1}{2}}, \tag{161}$$

and it suffices to bound the determinants on the right-hand side above.

K.2.1 Bounding $\det(C^*C)$

We now write C as follows:

$$C = \begin{bmatrix} f_0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ f_1 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & O(\eta^3) \\ \vdots & \vdots & \vdots \\ 0 & O(\eta^3) & O(\eta^3) \\ 0 & O(\eta^3/\sigma^2) & O(\eta^3/\sigma^2) \\ \vdots & \vdots & \vdots \\ 0 & O(\eta^3) & 0 \end{bmatrix} =: X + \widetilde{Z}, \tag{162}$$

where we denote the first matrix by X and the second matrix by $\widetilde{Z}$. We have that

$$\det(C^*C) = \det((X+\widetilde{Z})^*(X+\widetilde{Z})) = \det(X^*X + \widetilde{Z}'), \tag{163}$$

with $\widetilde{Z}' = X^*\widetilde{Z} + \widetilde{Z}^*X + \widetilde{Z}^*\widetilde{Z}$ and we apply Weyl's inequality to C^*C to obtain:

$$\det(C^*C) \leqslant (\lambda_1(X^*X) + \lambda_{\max}(\widetilde{Z}'))(\lambda_2(X^*X) + \lambda_{\max}(\widetilde{Z}'))(\lambda_3(X^*X) + \lambda_{\max}(\widetilde{Z}')). \tag{164}$$

Then

$$X^*X = \begin{bmatrix} f_0^2 + f_1^2 & f_0 & f_1 \\ f_0 & 1 & 0 \\ f_1 & 0 & 1 \end{bmatrix} \quad \text{with} \quad \lambda_3(X^*X) = 0, \tag{165}$$

and

$$\|X^*X\|_F \leqslant \bar{C}(f_0, f_1), \quad \text{where} \quad \bar{C}(f_0, f_1) = f_0^2 + f_1^2 + 2f_0 + 2f_1 + 2, \tag{166}$$

so

$$\|X\|_2^2 = \|X^*X\|_2 \leqslant \|X^*X\|_F \leqslant \bar{C}(f_0, f_1) \quad \text{so} \quad \|X\|_2 = \sqrt{\bar{C}(f_0, f_1)} \tag{167}$$

and

$$\begin{aligned} \|\widetilde{Z}'\|_2 &= \|X^*\widetilde{Z} + \widetilde{Z}^*X + \widetilde{Z}^*\widetilde{Z}\|_2 \\ &\leqslant 2\|X\|_2\|\widetilde{Z}\|_2 + \|\widetilde{Z}\|_2^2 = \left(2\sqrt{\bar{C}(f_0, f_1)} + \|\widetilde{Z}\|_2\right)\|\widetilde{Z}\|_2 \\ &\leqslant 3\sqrt{\bar{C}(f_0, f_1)}\|\widetilde{Z}\|_2, \end{aligned} \tag{168}$$

where the last inequality holds if

$$\|\widetilde{Z}\|_2 \leqslant \sqrt{\bar{C}(f_0, f_1)}. \tag{169}$$

Noting that $\widetilde{Z}'$ is symmetric and $\lambda_{\max}(\widetilde{Z}') \leqslant \|\widetilde{Z}'\|_2$, we substitute (165) and (168) into (164) and obtain:

$$\det(C^*C) \leqslant \left(\bar{C}(f_0, f_1) + 3\sqrt{\bar{C}(f_0, f_1)}\|\widetilde{Z}\|_2\right)^2 3\sqrt{\bar{C}(f_0, f_1)}\|\widetilde{Z}\|_2, \tag{170}$$

if (169) holds. Further applying (169) in the parentheses, we obtain:

$$\det(C^*C) \leqslant 48\bar{C}(f_0, f_1)^{\frac{5}{2}}\|\widetilde{Z}\|_2. \tag{171}$$

Now, using (143) with $M_1 = M_2 = 2$, we are able to upper bound $\|\widetilde{Z}\|_F$, which is also an upper bound for $\|\widetilde{Z}\|_2$:

$$\begin{aligned} \|\widetilde{Z}\|_2 \leqslant \|\widetilde{Z}\|_F &\leqslant \sqrt{(2k+2)(2\eta^3)^2 + 2k\left(\frac{2\eta^3}{\sigma^2}\right)^2} \\ &\leqslant (2k+2)2\eta^3 + 2k\frac{2\eta^3}{\sigma^2} = \eta^3\left(4k + 4 + \frac{4k}{\sigma^2}\right). \end{aligned} \tag{172}$$

Therefore, to satisfy (169), it is sufficient to find η such that:

$$\eta^3 \left(4k + 4 + \frac{4k}{\sigma^2}\right) \leqslant \sqrt{\bar{C}(f_0, f_1)}. \tag{173}$$

With this choice of η, by substituting (172) into (171), we obtain:

$$\det(C^*C) \leqslant 48\bar{C}(f_0, f_1)^{\frac{5}{2}} \left(4k + 4 + \frac{4k}{\sigma^2}\right) \eta^3. \tag{174}$$

Note that all our calculations so far will be used for bounding $\|b\|_2$. In case of $\|b^\pi\|_2$, everything is the same except that we have:

$$X_\pi^* X_\pi = \begin{bmatrix} f_0^2 + f_1^2 + 2k & f_0 & f_1 \\ f_0 & 1 & 0 \\ f_1 & 0 & 1 \end{bmatrix}, \tag{175}$$

which does not have a zero eigenvalue, so we will not obtain the η^3 factor. We omit the calculations, but we obtain:

$$\det(C_\pi^* C_\pi) \leqslant 64 \left(\bar{C}(f_0, f_1) + 2k\right)^3 \tag{176}$$

with a condition on η that is weaker than (173).

K.2.2 Bounding $\det(D^*D)$

Then, since D^*D is Hermitian positive definite, we can apply Hadamard's inequality (Theorem 7.8.1 in [45]) to bound its determinant by the product of its main diagonal entries (i.e. the squared 2-norms of the columns of D), so we obtain, after we use (143) with $M_1 = M_2 = 2$:

$$\begin{aligned} \det(D^*D) \leqslant & \left(8\eta^6 + \sum_{i=1}^k g(t_i - s_3)^2 + \sum_{i=1}^k g'(t_i - s_3)^2\right) \\ & \cdot \prod_{u=2}^k \left(8\eta^6 + \sum_{i=1}^k g(t_i - s_{2u})^2 + \sum_{i=1}^k g'(t_i - s_{2u})^2\right) \\ & \cdot \prod_{u=2}^k \left(32\eta^4 + \sum_{i=1}^k g'(\xi_{i,u})^2 + \sum_{i=1}^k g''(\xi'_{i,u})^2\right), \end{aligned} \tag{177}$$

where $\xi_{i,u}, \xi'_{i,u} \in [t_i - s_{2u} - \eta, t_i - s_{tu}]$.

For fixed $u \in \{1, \ldots, k\}$, we may write that

$$\sum_{i=1}^k g(t_i - s_{2u})^2 \leqslant 2 \sum_{i=0}^\infty e^{-\frac{i^2\Delta^2}{\sigma^2}} = \frac{2}{1 - e^{-\frac{\Delta^2}{\sigma^2}}}. \tag{178}$$

To see the inequality, if we note that $s_{2i} \leqslant t_i \leqslant s_{2i+1}$, we have that

$$\begin{aligned} g(t_i - s_{2i}) &\leqslant g(0), & g(t_i - s_{2\cdot(i+1)}) &\leqslant g(0), \\ g(t_i - s_{2\cdot(i-1)}) &\leqslant g(\Delta), & g(t_i - s_{2\cdot(i+2)}) &\leqslant g(\Delta), \\ &\vdots & &\vdots \end{aligned}$$

and by adding the inequalities we obtain (178). Likewise, it holds that

$$\sum_{i=1}^k g'(t_i - s_{2u})^2 \leqslant \frac{4}{\sigma^4} \cdot \frac{2}{1 - e^{-\frac{\Delta^2}{\sigma^2}}},$$

$$\sum_{i=1}^{k} g''(t_i - s_{2u})^2 \leqslant \left(\frac{2}{\sigma^2} + \frac{4}{\sigma^4}\right)^2 \cdot \frac{2}{1 - e^{-\frac{\Delta^2}{\sigma^2}}}. \tag{179}$$

and, for $\xi_{i,u}, \xi'_{u,u} \in [t_i - s_{2u} - \eta, t_i - s_{2u}]$:

$$\sum_{i=1}^{k} g'(\xi_{i,u})^2 \leqslant \frac{4}{\sigma^4} \cdot \frac{3}{1 - e^{-\frac{\Delta^2}{\sigma^2}}},$$

$$\sum_{i=1}^{k} g''(\xi'_{i,u})^2 \leqslant \left(\frac{2}{\sigma^2} + \frac{4}{\sigma^4}\right)^2 \cdot \frac{3}{1 - e^{-\frac{\Delta^2}{\sigma^2}}}. \tag{180}$$

Note that we obtained (180) in the same way as (178), except that we obtain a factor of 3 instead of 2:

$$\begin{array}{lll} g(\xi_{i,i}) \leqslant g(0), & g(\xi_{i,i-1}) \leqslant g(0) & g(\xi_{i,i+1}) \leqslant g(0) \\ & g(\xi_{i,i-2}) \leqslant g(\Delta) & g(\xi_{i,i+2}) \leqslant g(\Delta) \\ & g(\xi_{i,i-3}) \leqslant g(2\Delta) & g(\xi_{i,i+3}) \leqslant g(2\Delta) \\ & \vdots & \vdots \end{array}$$

Lastly, the above bounds also hold if we have $g(t_i - s_{2u+1})$ instead of $g(t_i - s_{2u})$. Substituting these bounds back into (177) and using (143) with $M_1 = M_2 = 2$, we obtain:

$$\begin{aligned} \det(D^*D) &\leqslant \left[8\eta^6 + \left(1 + \frac{4}{\sigma^4}\right) \frac{2}{1 - e^{-\frac{\Delta^2}{\sigma^2}}}\right] \\ &\quad \cdot \prod_{u=2}^{k} \left[8\eta^6 + \left(1 + \frac{4}{\sigma^4}\right) \frac{2}{1 - e^{-\frac{\Delta^2}{\sigma^2}}}\right] \\ &\quad \cdot \prod_{u=2}^{k} \left[32\eta^4 + \left(\frac{4}{\sigma^4} + \left(\frac{2}{\sigma^2} + \frac{4}{\sigma^4}\right)^2\right) \frac{3}{1 - e^{-\frac{\Delta^2}{\sigma^2}}}\right] \\ &\leqslant \left[8 + \left(1 + \frac{4}{\sigma^4}\right) \frac{2}{1 - e^{-\frac{\Delta^2}{\sigma^2}}}\right]^k \left[32 + \left(\frac{4}{\sigma^4} + \left(\frac{2}{\sigma^2} + \frac{4}{\sigma^4}\right)^2\right) \frac{3}{1 - e^{-\frac{\Delta^2}{\sigma^2}}}\right]^{k-1} \\ &= F_1\left(\Delta, \frac{1}{\sigma}\right)^k \cdot F_2\left(\Delta, \frac{1}{\sigma}\right)^{k-1}, \end{aligned} \tag{181}$$

where

$$\begin{aligned} F_1\left(\Delta, \frac{1}{\sigma}\right) &= 8 + \left(1 + \frac{4}{\sigma^4}\right) \frac{2}{1 - e^{-\frac{\Delta^2}{\sigma^2}}}, \\ F_2\left(\Delta, \frac{1}{\sigma}\right) &= 32 + \left(\frac{1}{\sigma^4} + \frac{2}{\sigma^6} + \frac{2}{\sigma^8}\right) \frac{24}{1 - e^{-\frac{\Delta^2}{\sigma^2}}}. \end{aligned} \tag{182}$$

Note that the bound (181) on $\det(D^*D)$ is the same for all $j = 3, \ldots, m$. Combining (174) with (181) in (159), we obtain:

$$N_{\underline{l},j} \leqslant \eta^{k+\frac{1}{2}} C(f_0, f_1)^{\frac{1}{2}} \left(4k + 4 + \frac{4k}{\sigma^2}\right)^{\frac{1}{2}} F_1\left(\Delta, \frac{1}{\sigma}\right)^{\frac{k}{2}} F_2\left(\Delta, \frac{1}{\sigma}\right)^{\frac{k-1}{2}} \tag{183}$$

for $j = 3, \ldots, m$ if (173) holds.

Finally, we need to upper bound $N_{\underline{l},j}$ for $j = 2$ and $j = m+1$. For simplicity, consider $j = 2$. Applying the same assumptions and operations as in (159), we have

$$N_{\underline{l},2} = \eta^k \begin{vmatrix} f_0 & O(\eta^3) & \cdots & O(\eta^3) & O(2\eta^2) & \cdots & O(\eta^3) \\ \vdots & & & & & & \\ 0 & g(t_i - s_2) & \cdots & g(t_i - s_{2u}) & -g'(\xi_{i,u}) & \cdots & O(\eta^3) \\ 0 & g'(t_i - s_2) & \cdots & g'(t_i - s_{2u}) & -g''(\xi'_{i,u}) & \cdots & O(\eta^3/\sigma^2) \\ \vdots & & & & & & \\ f_1 & O(\eta^3) & \cdots & O(\eta^3) & O(2\eta^2) & \cdots & 1 \end{vmatrix}, \tag{184}$$

where $\xi_{i,u}, \xi'_{i,u} \in [t_i - s_{2u} - \eta, t_i - s_{2u}]$. We bound $N_{\underline{l},2}$ by using Hadamard's inequality (the more general version, see [46]) and, after we use (143) with $M_1 = M_2 = 2$, we obtain:

$$\begin{aligned} N_{\underline{l},2} \leqslant{}& \eta^k \sqrt{f_0^2 + f_1^2} \left[1 + 4(k+1)\eta^6 + 4k\frac{\eta^6}{\sigma^4} \right]^{\frac{1}{2}} \\ &\cdot \prod_{u=1}^{k} \left[8\eta^6 + \sum_{i=1}^{k} g(t_i - s_{2u})^2 + \sum_{i=1}^{k} g'(t_i - s_{2u})^2 \right]^{\frac{1}{2}} \\ &\cdot \prod_{u=1}^{k} \left[32\eta^4 + \sum_{i=1}^{k} g'(\xi_{i,u})^2 + \sum_{i=1}^{k} g''(\xi'_{i,u})^2 \right]^{\frac{1}{2}}, \end{aligned} \tag{185}$$

and, by applying the bounds on the sums and $\eta \leqslant 1$, we obtain:

$$N_{\underline{l},2} \leqslant \eta^k \sqrt{f_0^2 + f_1^2} \left(4k + 5 + \frac{4k}{\sigma^4} \right)^{\frac{1}{2}} F_1\left(\Delta, \frac{1}{\sigma}\right)^{\frac{k}{2}} F_2\left(\Delta, \frac{1}{\sigma}\right)^{\frac{k}{2}}, \tag{186}$$

for F_1 and F_2 defined as in (182), and note that the same bound also holds for $N_{\underline{l},m+1}$.

To conclude, from (183) and (186), we can derive a general bound valid for all j:

$$N_{\underline{l},j} \leqslant \eta^k C(f_0, f_1)^{\frac{1}{2}} \left(4k + 5 + \frac{4k}{\sigma^4} \right)^{\frac{1}{2}} F_{\max}\left(\Delta, \frac{1}{\sigma}\right)^k \tag{187}$$

for all $j = 2, \ldots, m+1$ if (173) holds, where

$$F_{\max}\left(\Delta, \frac{1}{\sigma}\right) = \left(8 + \left(1 + \frac{4}{\sigma^4}\right) \frac{2}{1 - e^{-\frac{\Delta^2}{\sigma^2}}} \right)^{\frac{1}{2}} \left(32 + \left(\frac{1}{\sigma^4} + \frac{2}{\sigma^6} + \frac{2}{\sigma^8} \right) \frac{24}{1 - e^{-\frac{\Delta^2}{\sigma^2}}} \right)^{\frac{1}{2}}. \tag{188}$$

K.3 Condition on η

We now return to the condition (153) that η must satisfy so that our application of Lemma 29 is valid. Since A is positive definite, we have $\left\|A^{-1}\right\|_2 \leqslant 1/\lambda_{\min}(A)$. Using this, we obtain:

$$\rho(A^{-1}(A' + C')) \leqslant \|A^{-1}(A' + C')\|_2 \leqslant \|A^{-1}\|_2 \|A' + C'\|_2 \leqslant \frac{1}{\lambda_{\min}(A)} \cdot \|A' + C'\|_F, \tag{189}$$

and

$$A' + C' = \begin{bmatrix} 0 & \cdots & O(\eta^2) & O(2\eta) & \cdots & O(\eta^2) \\ \vdots & & \vdots & \vdots & & \vdots \\ O(\eta^2) & \cdots & \frac{t_u - s_{2u}}{\eta} g'(\xi^*_{i,u}) & -\frac{t_u - s_{2u}}{\eta} g''(\xi^{*\prime}_{i,u}) + \frac{1}{2} g''(\xi_{i,u}) & \cdots & O(\eta^2) \\ O(\eta^2/\sigma^2) & \cdots & \frac{t_u - s_{2u}}{\eta} g''(\xi^{*\prime}_{i,u}) & -\frac{t_u - s_{2u}}{\eta} g'''(\xi^{*\prime\prime}_{i,u}) + \frac{1}{2} g'''(\xi'_{i,u}) & \cdots & O(\eta^2/\sigma^2) \\ \vdots & & \vdots & \vdots & & \vdots \\ O(\eta^2) & \cdots & O(\eta^2) & O(2\eta) & \cdots & 0 \end{bmatrix}, \tag{190}$$

where

$$
\begin{aligned}
&\xi_{i,u}, \xi'_{i,u} \in [t_i - s_{2u} - \eta, t_i - s_{2u}],\\
&\xi^*_{i,u}, \xi^{*\prime}_{i,u}, \xi^{*\prime\prime}_{i,u} \in [t_i - t_u - |t_u - s_{2u}|, t_i - t_u + |t_u - s_{2u}|].
\end{aligned} \tag{191}
$$

for all $i, u = 1, \dots, k$.

Because the eigenvalues of A are $1, 1$ and the eigenvalues of B and $\lambda_{\min}(B) \geqslant F_{\min}(\Delta, \frac{1}{\sigma})$ with $0 < F_{\min}(\Delta, \frac{1}{\sigma}) < 1$, then we have that that $\lambda_{\min}(A) \geqslant F_{\min}\left(\Delta, \frac{1}{\sigma}\right)$. Moreover, after applying (143) with $M_1 = M_2 = 2$, we have that:

$$
\begin{aligned}
\|A' + C'\|_F^2 \leqslant\ & 8(k+1)\eta^4 + 8k\frac{\eta^4}{\sigma^4} \qquad \text{(from the first and last columns)}\\
&+ 8k\eta^4 + 32k\eta^2 + \qquad \text{(from the first and last rows)}\\
&+ \sum_{i,u=1}^{k} g'(\xi^*_{i,u})^2 + \sum_{i,u=1}^{k} g''(\xi^{*\prime}_{i,u})^2\\
&+ \sum_{i,u=1}^{k} \left(-\frac{t_u - s_{2u}}{\eta} g''(\xi^{*\prime}_{i,u}) + \frac{1}{2} g''(\xi_{i,u}) \right)^2 + \sum_{i,u=1}^{k} \left(-\frac{t_u - s_{2u}}{\eta} g'''(\xi^{*\prime\prime}_{i,u}) + \frac{1}{2} g'''(\xi'_{i,u}) \right)^2.
\end{aligned} \tag{192}
$$

We upper bound this using the inequalities in (180) and, similarly, for g''':

$$
\sum_{i=1}^{k} g'''(\xi_{i,u})^2 \leqslant \left(\frac{12}{\sigma^4} + \frac{8}{\sigma^6} \right)^2 \cdot \frac{3}{1 - e^{-\frac{\Delta^2}{\sigma^2}}}, \tag{193}
$$

and note that these inequalities hold for all numbers ξ in (191). By expanding the parentheses and applying Cauchy-Schwartz to the products, and using that $|t_u - s_{2u}| \leqslant \eta$, we obtain:

$$
\begin{aligned}
\|A' + C'\|_F^2 \leqslant\ & 8(2k+1)\eta^4 + 32k\eta^2 + 8k\frac{\eta^4}{\sigma^4}\\
&+ k \cdot \frac{4}{\sigma^4} \cdot \frac{3}{1 - e^{-\frac{\Delta^2}{\sigma^2}}} + k \cdot \left(\frac{2}{\sigma^2} + \frac{4}{\sigma^4} \right)^2 \cdot \frac{3}{1 - e^{-\frac{\Delta^2}{\sigma^2}}}\\
&+ \frac{9k}{4} \cdot \left(\frac{2}{\sigma^2} + \frac{4}{\sigma^4} \right)^2 \cdot \frac{3}{1 - e^{-\frac{\Delta^2}{\sigma^2}}}\\
&+ \frac{9k}{4} \cdot \left(\frac{12}{\sigma^4} + \frac{8}{\sigma^6} \right)^2 \cdot \frac{3}{1 - e^{-\frac{\Delta^2}{\sigma^2}}}
\end{aligned} \tag{194}
$$

Using the assumption that $\eta \leqslant \sigma^2$ to write $\frac{\eta^4}{\sigma^4} \leqslant \sigma^4$, and then by applying $\eta \leqslant 1$ and $\sigma \leqslant \sqrt{2}$, we can write

$$
\|A' + C'\|_F^2 \leqslant 80k + 8 + kP\left(\frac{1}{\sigma}\right) \frac{3}{1 - e^{-\frac{\Delta^2}{\sigma^2}}}, \tag{195}
$$

where $P\left(\frac{1}{\sigma}\right)$ is a polynomial in $\frac{1}{\sigma}$ defined as follows:

$$
P\left(\frac{1}{\sigma}\right) = \frac{4}{\sigma^4} + \frac{13}{4}\left(\frac{2}{\sigma^2} + \frac{4}{\sigma^4} \right)^2 + \frac{9}{4}\left(\frac{12}{\sigma^4} + \frac{8}{\sigma^6} \right)^2. \tag{196}
$$

Inserting the above observations in the condition (153), we finally obtain

$$
\eta \leqslant \frac{8F_{\min}(\Delta, \frac{1}{\sigma})}{34(2k+2)\left(80k + 8 + kP\left(\frac{1}{\sigma}\right) \frac{3}{1 - e^{-\frac{\Delta^2}{\sigma^2}}} \right)^{\frac{1}{2}}}. \tag{197}
$$

With this choice of η, the condition (153) is satisfied.

K.4 Bound for (136)

Combining the results from K.1 and K.2, we arrive at

$$
\begin{aligned}
|b_j| &= \frac{N_{\underline{l},j+1}}{N_{\underline{l},1}} \qquad \text{(see (138))} \\
&\leqslant \frac{\bar{C}(f_0,f_1)^{\frac{5}{4}}\left(4k+5+\frac{4k}{\sigma^4}\right)^{\frac{1}{2}}}{1-\frac{\sqrt{e}}{2}}\left[\frac{F_{\max}\left(\Delta,\frac{1}{\sigma}\right)}{F_{\min}\left(\Delta,\frac{1}{\sigma}\right)^2}\right]^k, \qquad \text{(see (187) and (157))} \qquad (198)
\end{aligned}
$$

$$
\text{and} \quad |b_j^\pi| \leqslant \frac{\left(\bar{C}(f_0,f_1)+2k\right)^{\frac{3}{2}}}{\eta\left(1-\frac{\sqrt{e}}{2}\right)}\left[\frac{F_{\max}\left(\Delta,\frac{1}{\sigma}\right)}{F_{\min}\left(\Delta,\frac{1}{\sigma}\right)^2}\right]^k, \qquad \text{(see (176))} \qquad (199)
$$

for every $j \in [m]$, provided that the conditions (173) and (197) hold. Consequently,

$$
\|b\|_2 = \sqrt{\sum_{j=1}^m b_j^2} \leqslant \frac{\sqrt{(2k+2)\left(4k+5+\frac{4k}{\sigma^4}\right)}}{1-\frac{\sqrt{e}}{2}}\bar{C}(f_0,f_1)^{\frac{5}{4}}F\left(\Delta,\frac{1}{\sigma}\right)^k, \qquad (200)
$$

$$
\|b^\pi\|_2 \leqslant \frac{\sqrt{2k+2}}{\eta\left(1-\frac{\sqrt{e}}{2}\right)}\left(\bar{C}(f_0,f_1)+2k\right)^{\frac{3}{2}}F\left(\Delta,\frac{1}{\sigma}\right)^k, \qquad (201)
$$

where we write $F\left(\Delta,\frac{1}{\sigma}\right) = \frac{F_{\max}\left(\Delta,\frac{1}{\sigma}\right)}{F_{\min}\left(\Delta,\frac{1}{\sigma}\right)^2}$ if, $\sigma \leqslant \sqrt{2}$ and $\Delta > \sigma\sqrt{\log\left(3+\frac{4}{\sigma^2}\right)}$ and the conditions (173) and (197) hold. This completes the proof of Lemma 24 since $m = 2k+2$ by assumption.

L Proof of Lemma 29

To begin with, let us note that

$$
\det(A+\epsilon B) = \det(A)\det(I+\epsilon A^{-1}B)
$$

We will now upper and lower bound the second term in this equality. Denoting $C = A^{-1}B$, consider

$$
\log\det(I+\epsilon C) = \sum_{i\in R}\log(1+\epsilon\lambda_i(C)) + \sum_{i\in I}\log(1+2\epsilon\,\mathrm{Re}(\lambda_i(C))+\epsilon^2|\lambda_i(C)|^2),
$$

where

$$
\begin{aligned}
&R = \{i \in [m] : \mathrm{Im}(\lambda_i(C)) = 0\} \text{ and} \\
&I = \{i \in [m] : \lambda_{i_1}(C), \lambda_{i_2}(C) \text{ are complex conjugate and } \mathrm{Im}(\lambda_{i_1}(C)) \neq 0\}.
\end{aligned}
$$

1. For $i \in R$, if $\epsilon < \frac{1}{|\lambda_i(C)|}$, use apply Taylor expansion for $\log(1+x)$ and obtain:

$$
\log(1+\epsilon\lambda_i(C)) = \epsilon\lambda_i(C) - \frac{\epsilon^2\lambda_i^2(C)}{2\xi_{i,\epsilon}^2}, \qquad \text{where } \xi_{i,\epsilon} \in \left[1-\epsilon|\lambda_i(C)|, 1+\epsilon|\lambda_i(C)|\right]. \qquad (202)
$$

2. For $i \in C$, we apply the same Taylor expansion and, writing $y = 2\epsilon\,\mathrm{Re}(\lambda_i(C)) + \epsilon^2|\lambda_i(C)|^2$, for $|y| < 1$ we obtain:

$$
\log(1+y) = \frac{y}{\xi_{i,\epsilon}}, \qquad \text{where } \xi_{i,\epsilon} \in \left[1-|y|, 1+|y|\right].
$$

   Then, we have that

$$
2\epsilon|\lambda_i(C)| + \epsilon^2|\lambda_i(C)|^2 \leqslant 4\epsilon|\lambda_i(C)| < 1,
$$

where the first inequality is true if $\epsilon \leqslant \frac{2}{|\lambda_i(C)|}$ and the second inequality is true if $\epsilon < \frac{1}{4|\lambda_i(C)|}$. From the condition on $\xi_{i,\epsilon}$ and noting that $|\operatorname{Re}(\lambda_i(C))| \leqslant |\lambda_i(C)|$, we have that

$$\xi_{i,\epsilon} \leqslant 1 + |y| \leqslant 1 + 4\epsilon|\lambda_i(C)| \qquad \text{and} \qquad \xi_{i,\epsilon} \geqslant 1 - |y| \geqslant 1 - 4\epsilon|\lambda_i(C)| \geqslant \frac{1}{2}, \tag{203}$$

where the last inequality holds if $\epsilon \leqslant \frac{1}{8|\lambda_i(C)|}$. Therefore, $\frac{1}{\xi_{i,\epsilon}} \leqslant 2$ if $\epsilon \leqslant \frac{1}{8|\lambda_i(C)|}$.

We now use (202), (203) and $\epsilon \leqslant \frac{1}{8|\lambda_i(C)|}$. Let $\epsilon \leqslant \frac{1}{8|\lambda_i(C)|}$ for all eigenvalues $\lambda_i(C)$, then:

$$\log\det(I + \epsilon C) = \sum_{i \in R} \epsilon\lambda_i(C) - \frac{\epsilon^2}{2}\sum_{i \in R} \frac{\lambda_i^2(C)}{\xi_{i,\epsilon}^2} + \sum_{i \in I} \frac{1}{\xi_{i,\epsilon}}\left(2\epsilon\operatorname{Re}(\lambda_i(C)) + \epsilon^2|\lambda_i(C)|^2\right), \tag{204}$$

and, by using that $\xi_{i,\epsilon} \geqslant 1 - \epsilon|\lambda_i(C)| \geqslant \frac{7}{8}$ for $i \in R$ and $\frac{1}{\xi_{i,\epsilon}} \leqslant 2$ for $i \in C$ and the fact that the index $i \in I$ accounts for two eigenvalues, we obtain:

$$\begin{aligned}
|\log\det(I + \epsilon C)| &\leqslant \epsilon\sum_{i \in R}|\lambda_i(C)| + \frac{32\epsilon^2}{49}\sum_{i \in R}|\lambda_i(C)|^2 + 4\epsilon\sum_{i \in I}|\lambda_i(C)| + 2\epsilon^2\sum_{i \in I}|\lambda_i(C)|^2 \\
&\leqslant 2\epsilon\sum_{i=1}^{m}|\lambda_i(C)| + \epsilon^2\sum_{i=1}^{m}|\lambda_i(C)|^2 \\
&\leqslant m\epsilon\rho(C)(2 + \epsilon\rho(C)) \\
&\leqslant m\epsilon\rho(C)(2 + \frac{1}{8}) = \frac{17}{8}m\epsilon\rho(C) \\
&\leqslant \frac{1}{2}
\end{aligned} \tag{205}$$

where the second last inequality holds if $\epsilon \leqslant \frac{1}{8\rho(C)}$ and the last inequality holds if $\epsilon \leqslant \frac{8}{34m\rho(C)}$, which is also the dominating condition for ϵ in the proof if $m \geqslant 2$.

Now, note that for $|x| \leqslant \frac{1}{2}$, we have that $e^x = 1 + xe^{\xi}$ for some $\xi \in [-|x|, |x|] \subseteq [-\frac{1}{2}, \frac{1}{2}]$ by Taylor expansion, and by taking $x = \log\det(I + \epsilon C)$, we obtain:

$$\det(I + \epsilon C) = e^x \leqslant 1 + |x|e^{\frac{1}{2}} \leqslant 1 + \frac{17\sqrt{e}}{8}m\epsilon\rho(C) \leqslant 1 + \frac{\sqrt{e}}{2},$$

and similarly

$$\det(I + \epsilon C) = e^x \geqslant 1 - |x|e^{\frac{1}{2}} \geqslant 1 - \frac{17\sqrt{e}}{8}m\epsilon\rho(C) \geqslant 1 - \frac{\sqrt{e}}{2}.$$

From the last two inequalities, by multiplying by $\det(A)$, we obtain the result of our lemma.

M Proof of Lemma 30

Recalling the definition of B in (156), we apply Gershgorin disc theorem to find the discs $D(a_{ii}, \sum_{i \neq j} a_{ij})$ which contain the eigenvalues of B. Due to the structure of H, we consider two cases:

1. On odd rows, the centre of the disc is $a_{\text{odd}_{ii}} = g(0) = 1$ and the radius is

$$R_{\text{odd}_i} = \sum_{\substack{j=1 \\ j \neq i}}^{k} |g(t_i - t_j)| + |-g'(t_i - t_j)| \tag{206a}$$

$$= \sum_{\substack{j=1 \\ j \neq i}}^{k} g(t_i - t_j)\left(1 + \frac{2|t_i - t_j|}{\sigma^2}\right), \qquad \text{for} \quad i = 1, \ldots, k. \tag{206b}$$

2. On even rows, the centre of the disc is $a_{\text{even}_{ii}} = -g''(0) = \frac{2}{\sigma^2}$ and the radius is

$$R_{\text{even}_i} = \sum_{\substack{j=1\\ j\neq i}}^{k} |g'(t_i - t_j)| + |-g''(t_i - t_j)| \tag{207a}$$

$$= \sum_{\substack{j=1\\ j\neq i}}^{k} g(t_i - t_j)\left(\frac{2|t_i - t_j|}{\sigma^2} + \left|-\frac{2}{\sigma^2} + \frac{4(t_i - t_j)^2}{\sigma^4}\right|\right), \qquad \text{for} \quad i = 1, \dots, k. \tag{207b}$$

Since the eigenvalues of B are real, they are lower bounded by

$$\min_{i=1,\dots,k} 1 - R_{\text{odd}_i} \quad \text{or} \quad \min_{i=1,\dots,k} \frac{2}{\sigma^2} - R_{\text{even}_i}. \tag{208}$$

Because $|t_i - t_j| \leqslant 1$, we have that

$$1 - R_{\text{odd}_i} \geqslant 1 - \left(1 + \frac{2}{\sigma^2}\right)\sum_{\substack{j=1\\ j\neq i}}^{k} g(t_i - t_j) \tag{209}$$

Since $\frac{4(t_i-t_j)^2}{\sigma^4} - \frac{2}{\sigma^2} > \frac{4\Delta^2}{\sigma^4} - \frac{2}{\sigma^2} > 0$ due to our assumptions on σ (see (135)), we obtain

$$\frac{2}{\sigma^2} - R_{\text{even}_i} \geqslant \frac{2}{\sigma^2} - \frac{4}{\sigma^4}\sum_{\substack{j=1\\ j\neq i}}^{k} g(t_i - t_j), \tag{210}$$

for all $i = 1, \dots, k$. Using the fact that g is decreasing and $|t_i - t_j| \geqslant |i - j|\Delta$, we obtain

$$\sum_{\substack{j=1\\ j\neq i}}^{k} g(t_i - t_j) \leqslant \sum_{\substack{j=1\\ j\neq i}}^{k} g(|i - j|\Delta) \tag{211a}$$

$$\leqslant 2\sum_{j=1}^{\infty} g(j\Delta) = 2\sum_{j=1}^{\infty} e^{-\frac{j^2\Delta^2}{\sigma^2}} \leqslant 2\sum_{j=1}^{\infty}\left(e^{-\frac{\Delta^2}{\sigma^2}}\right)^j \tag{211b}$$

$$= \frac{2e^{-\frac{\Delta^2}{\sigma^2}}}{1 - e^{-\frac{\Delta^2}{\sigma^2}}}, \qquad \text{for } i = 1, \dots, k. \tag{211c}$$

The sum of of the series is valid because $e^{-\frac{\Delta^2}{\sigma^2}} < 1$. Combining this with (209) and (210), we obtain:

$$1 - R_{\text{odd}_i} \geqslant 1 - \left(1 + \frac{2}{\sigma^2}\right)\frac{2e^{-\frac{\Delta^2}{\sigma^2}}}{1 - e^{-\frac{\Delta^2}{\sigma^2}}}, \tag{212a}$$

$$\frac{2}{\sigma^2} - R_{\text{even}_i} \geqslant \frac{2}{\sigma^2} - \frac{4}{\sigma^4}\frac{2e^{-\frac{\Delta^2}{\sigma^2}}}{1 - e^{-\frac{\Delta^2}{\sigma^2}}}, \tag{212b}$$

for all $i = 1, \dots, k$. By using the assumption that $\Delta > \sigma\sqrt{\log\left(3 + \frac{4}{\sigma^2}\right)}$, we can check that the lower bound in (212a) is greater than zero, and by also using that $\sigma \leqslant \sqrt{2}$, we can check that it is smaller than than the lower bound in (212b).

To conclude, since all the eigenvalues of the matrix B are real and in the union of the discs $D(1, R_{\text{odd}_i})$ and $D(\frac{2}{\sigma^2}, R_{\text{even}_i})$ for $i = 1, \dots, k$, then, by using the above observation and the lower bound in (212a), we obtain a lower bound of all the eigenvalues of B:

$$\lambda_j(B) \geqslant 1 - \left(1 + \frac{2}{\sigma^2}\right)\frac{2e^{-\frac{\Delta^2}{\sigma^2}}}{1 - e^{-\frac{\Delta^2}{\sigma^2}}}, \qquad \forall j = 1, \dots, 2k. \tag{213}$$

Note that we may be able to obtain better bounds given by (212b) for k eigenvalues if we scale the Gershgorin discs so that $D(1, R_{\text{odd}_i})$ and $D(1, R_{\text{even}_i})$ become disjoint.

N Proof of Lemma 25

Because $\Delta > \sigma\sqrt{\log\frac{5}{\sigma^2}}$, we have that

$$e^{-\frac{\Delta^2}{\sigma^2}} < \frac{\sigma^2}{5}, \tag{214}$$

which implies that

$$\frac{1}{1-e^{-\frac{\Delta^2}{\sigma^2}}} < \frac{5}{5-\sigma^2}. \tag{215}$$

Then, $-\frac{2e^{-\frac{\Delta^2}{\sigma^2}}}{1-e^{-\frac{\Delta^2}{\sigma^2}}} > -\frac{2\sigma^2}{5-\sigma^2}$, so

$$1-(1+\frac{2}{\sigma^2})\frac{2e^{-\frac{\Delta^2}{\sigma^2}}}{1-e^{-\frac{\Delta^2}{\sigma^2}}} > \frac{1-3\sigma^2}{5-\sigma^2}, \quad \text{i.e.} \quad F_{\min}\left(\Delta, \frac{1}{\sigma}\right) > \frac{1-3\sigma^2}{5-\sigma^2}. \tag{216}$$

Similarly, we have that:

$$8+\left(1+\frac{4}{\sigma^4}\right)\frac{2}{1-e^{-\frac{\Delta^2}{\sigma^2}}} < \frac{-8\sigma^6+50\sigma^4+40}{\sigma^4(5-\sigma^2)} < \frac{98}{\sigma^4(5-\sigma^2)} \tag{217}$$

and

$$32+\left(\frac{1}{\sigma^4}+\frac{2}{\sigma^6}+\frac{2}{\sigma^8}\right)\frac{24}{1-e^{-\frac{\Delta^2}{\sigma^2}}} < \frac{-32\sigma^{10}+160\sigma^8+120\sigma^4+240\sigma^2+240}{\sigma^8(5-\sigma^2)} < \frac{792}{\sigma^8(5-\sigma^2)}, \tag{218}$$

where in the last two inequalities we used $\sigma < 1$. Combining (216),(217) and (218), we obtain (56):

$$\frac{F_{\max}\left(\Delta, \frac{1}{\sigma}\right)}{F_{\min}\left(\Delta, \frac{1}{\sigma}\right)^2} < \frac{c_1}{\sigma^6(1-3\sigma^2)^2}. \tag{219}$$

Then, using that $\bar{f} < 1$ and that $\frac{\bar{C}^{\frac{5}{4}}}{\bar{f}} \leqslant \frac{(\bar{C}+2k)^{\frac{3}{2}}}{\bar{f}}, (\bar{C}+2k)^{\frac{3}{2}} \leqslant \frac{(\bar{C}+2k)^{\frac{3}{2}}}{\bar{f}}$, we have that:

$$\begin{aligned}
\left((6+\frac{2}{\bar{f}})\sqrt{4k+5+\frac{4k}{\sigma^4}}\bar{C}^{\frac{5}{4}}+\frac{6}{\eta}(\bar{C}+2k)^{\frac{3}{2}}\right)\frac{\sqrt{2k+2}}{1-\frac{\sqrt{e}}{2}} &< c_2\frac{(\bar{C}+2k)^{\frac{3}{2}}}{\bar{f}}\sqrt{2k+2}\frac{\eta\sqrt{4k\sigma^4+5\sigma^4+4k}+\sigma^2}{\eta\sigma^2} \\
&< c_2\frac{(\bar{C}+2k)^{\frac{3}{2}}}{\bar{f}}\sqrt{2k+2}\cdot\frac{\sqrt{8k+5}+1}{\eta\sigma^2} \qquad (\sigma<1) \\
&< c_2\frac{(\bar{C}+2k)^{\frac{3}{2}}}{\bar{f}}\frac{k}{\eta\sigma^2},
\end{aligned} \tag{220}$$

for a large enough constant c_2 and, from (219) and (220), we obtain (57). Similarly we show that (58) holds:

$$\begin{aligned}
\frac{\sqrt{(2k+2)\left(4k+5+\frac{4k}{\sigma^4}\right)}}{1-\frac{\sqrt{e}}{2}}\frac{\bar{C}(f_0,f_1)^{\frac{5}{4}}}{\bar{f}} &< c_5\frac{\sqrt{(k+1)(4k\sigma^4+5\sigma^4+4k)}}{\sigma^2}\frac{\bar{C}^{\frac{5}{4}}}{\bar{f}} \\
&< c_5\cdot\frac{\sqrt{(k+1)(8k+5)}}{\sigma^2}\frac{\bar{C}^{\frac{5}{4}}}{\bar{f}} \qquad (\sigma<1) \\
&< c_5\frac{k}{\sigma^2}\frac{\bar{C}^{\frac{5}{4}}}{\bar{f}}.
\end{aligned} \tag{221}$$

To show that (49) is satisfied if (13) holds, from (215) and (216), we have that:

$$
\begin{aligned}
\frac{8F_{\min}(\Delta,\frac{1}{\sigma})}{34(2k+2)\left(80k+8+kP\left(\frac{1}{\sigma}\right)\frac{3}{1-e^{-\frac{\Delta^2}{\sigma^2}}}\right)^{\frac{1}{2}}} &> \frac{8(1-3\sigma^2)}{34(5-\sigma^2)(2k+2)\left(80k+8+\frac{15kP\left(\frac{1}{\sigma}\right)}{5-\sigma^2}\right)^{\frac{1}{2}}} \\
&= \frac{8(1-3\sigma^2)}{34(2k+2)\left((80k+8)(5-\sigma^2)^2+15(5-\sigma^2)kP\left(\frac{1}{\sigma}\right)\right)^{\frac{1}{2}}} \\
&> \frac{8(1-3\sigma^2)}{34(2k+2)\left(2000k+200+75kP\left(\frac{1}{\sigma}\right)\right)^{\frac{1}{2}}} \qquad (5-\sigma^2<5) \\
&> \frac{8(1-3\sigma^2)\sigma^6}{34(2k+2)\left(2000k\sigma^{12}+200\sigma^{12}+c_3k\right)^{\frac{1}{2}}} \qquad \left(P\left(\frac{1}{\sigma}\right)<\frac{c_3}{\sigma^{12}}\right) \\
&> \frac{8(1-3\sigma^2)\sigma^6}{34(2k+2)\left(c_3k+200\right)^{\frac{1}{2}}} \qquad (\sigma<1) \\
&> c_3\cdot\frac{\sigma^6(1-3\sigma^2)}{(k+1)^{\frac{3}{2}}},
\end{aligned} \tag{222}
$$

and similarly:

$$
\frac{\bar{C}^{\frac{1}{6}}}{\left(4k+4+\frac{4k}{\sigma^2}\right)^{\frac{1}{3}}} = \frac{\bar{C}^{\frac{1}{6}}\sigma^{\frac{2}{3}}}{(4k\sigma^2+4\sigma^2+4k)^{\frac{1}{3}}} > \frac{\bar{C}^{\frac{1}{6}}\sigma^{\frac{2}{3}}}{(8k+4)^{\frac{1}{3}}} > c_4\cdot\frac{\bar{C}^{\frac{1}{6}}\sigma^{\frac{2}{3}}}{(k+1)^{\frac{1}{3}}}, \tag{223}
$$

for some constants c_3, c_4. Finally, from (214) and (215), we also obtain:

$$
\frac{e^{-\frac{\Delta^2}{\sigma^2}}+e^{-\frac{2\Delta^2}{\sigma^2}}}{1-e^{-\frac{\Delta^2}{\sigma^2}}} < \left(\sigma^2+\frac{\sigma^4}{5}\right)\frac{1}{5-\sigma^2} < \frac{16}{210}, \tag{224}
$$

where, in the last inequality, we used $\sigma < \frac{1}{\sqrt{3}}$. Furthermore, if $\lambda < \frac{2}{5}$, then $1-2\lambda > \frac{1}{5}$ and

$$
e^{-\frac{\Delta^2(1-2\lambda)}{\sigma^2}} < \left(\frac{\sigma^2}{5}\right)^{1-2\lambda} < \frac{1}{15^{1-2\lambda}} < \frac{1}{15^{1/5}}, \tag{225}
$$

so we can combine the last two inequalities to obtain:

$$
1-\frac{e^{-\frac{\Delta^2}{\sigma^2}}+e^{-\frac{2\Delta^2}{\sigma^2}}}{1-e^{-\frac{\Delta^2}{\sigma^2}}}-e^{-\frac{\Delta^2(1-2\lambda)}{\sigma^2}} > 1-\frac{16}{210}-\frac{1}{15^{1/5}} > \frac{1}{3}. \tag{226}
$$

Finally, to bound $e^{\frac{\Delta^2\lambda^2}{\sigma^2}}$, note from the definition of λ and η that $\lambda\Delta < \frac{\eta}{2}$, and from the assumption that $\eta \leqslant \sigma^2$, we obtain that:

$$
e^{\frac{\Delta^2\lambda^2}{\sigma^2}} < e^{\frac{\eta^2}{4\sigma^2}} \leqslant e^{\sigma^2/4} < c_5, \tag{227}
$$

where we used that $\sigma < 1$. Combining the last two inequalities, we obtain that (61) holds for some constant $c_5 > 0$.

O Proof of Lemma 27

In the proof of Lemma 23 in Appendix J we require that $f_0 \gg \bar{f}$, $f_0 \gg f_1$ and $f_0 \gg 1$. These conditions come from equations (126), (130) and (132) respectively. We can, therefore, fix $\bar{f}$ and f_1 such that $\bar{f} < 1$ and

$1 < f_1 < f_0$, and give an expression of f_0 as a function of $\bar{f}$ as $\epsilon \to 0$. From equation (126), we have that

$$\frac{f_0}{\bar{f}} > \frac{N_{\underline{l},1}}{\min_{\tau_{\underline{l}} \in T_\epsilon^C} N_{1,1}(\tau_{\underline{l}})}. \tag{228}$$

which is required by the condition that $\det(M_N)$ in (123) is positive when $\tau_{\underline{l}} \in T_\epsilon^C$. While $N_{\underline{l},1}$ does not depend on ϵ, we can argue that, for $\epsilon \to 0$, the minimum in the denominator is lower bounded by $N_{1,1}(\tau_{\underline{l}})$, where

$$\tau_{\underline{l}} \in \bar{T}_\epsilon = \{t_1 - \epsilon, t_1 + \epsilon, \ldots, t_k - \epsilon, t_k + \epsilon\}. \tag{229}$$

Therefore, a sufficient condition to ensure (228) is:

$$\frac{f_0}{\bar{f}} > \frac{N_{\underline{l},1}}{\min_{\tau_{\underline{l}} \in \bar{T}_\epsilon} N_{1,1}(\tau_{\underline{l}})}. \tag{230}$$

Note that $\min_{\tau_{\underline{l}} \in \bar{T}_\epsilon} N_{1,1}(\tau_{\underline{l}}) \to 0$ as $\epsilon \to 0$, since two rows of the determinant become equal. More explicitly, let us assume, for simplicity, that the minimum over the finite set of points $\bar{T}_\epsilon$ is attained $t_1 + \epsilon$. Then, we subtract the row with $\tau_{\underline{l}}$ from the first row of $N_{1,1}(\tau_{\underline{l}})$, and then expand the determinant along this row. By taking $\tau_{\underline{l}} = t_1 + \epsilon$, we obtain

$$N_{1,1}(t_1 + \epsilon) = \sum_{j=1}^{m} (-1)^{j+1} N_{1,1,j} \left[g(t_1 - s_j) - g(t_1 - s_j + \epsilon) \right],$$

where the minors $N_{1,1,j}$ are fixed (i.e. independent of ϵ). Therefore, as $\epsilon \to 0$, we have that $N_{1,1}(t_1+\epsilon) \to 0$, with $N_{1,1}(t_1 + \epsilon) > 0, \forall \epsilon$. Then, everything else in $N_{1,1}(t_1 + \epsilon)$ being fixed, there exists $\epsilon_0 > 0$ such that [6]

$$\min_{\tau_{\underline{l}} \in T_\epsilon^C} N_{1,1}(\tau_{\underline{l}}) = N_{1,1}(t_1 + \epsilon), \quad \forall \epsilon < \epsilon_0. \tag{231}$$

We will now find the exact rate at which $N_{1,1}(t_1 + \epsilon) \to 0$ for $\epsilon < \epsilon_0$. In the row with $\tau_{\underline{l}}$ in $N_{1,1}(t_1 + \epsilon)$, we Taylor expand the entries in the columns $j = 1, \ldots, m$ as follows:

$$g(\tau_{\underline{l}} - s_j) = g(t_1 - s_j + \epsilon) = g(t_1 - s_j) + \epsilon g'(t_1 - s_j) + \frac{\epsilon^2}{2} g''(\xi_j), \tag{232}$$

for some $\xi_j \in [t_1 - s_j, t_1 - s_j + \epsilon]$, and note that $\xi_j \to t_1 - s_j$ as $\epsilon \to 0$. Then

$$N_{1,1}(t_1+\epsilon) = \epsilon \begin{vmatrix} g\left(t_1 - s_1\right) & \cdots & g\left(t_1 - s_m\right) \\ g'\left(t_1 - s_1\right) & \cdots & g'\left(t_1 - s_m\right) \\ \vdots & & \vdots \\ g'(t_1 - s_1) + \frac{\epsilon}{2} g''(\xi_1) & \cdots & g'(t_1 - s_m) + \frac{\epsilon}{2} g''(\xi_m) \\ \vdots & & \vdots \\ g\left(t_k - s_1\right) & \cdots & g\left(\tau - s_m\right) \\ g'\left(t_k - s_1\right) & \cdots & g'\left(t_k - s_m\right) \\ g\left(1 - s_1\right) & \cdots & g\left(1 - s_m\right) \end{vmatrix} \qquad \text{(subtract 1st row from } \tau_{\underline{l}} \text{ row)}$$

$$= \frac{\epsilon^2}{2} \begin{vmatrix} g\left(t_1 - s_1\right) & \cdots & g\left(t_1 - s_m\right) \\ g'\left(t_1 - s_1\right) & \cdots & g'\left(t_1 - s_m\right) \\ \vdots & & \vdots \\ g''(\xi_1) & \cdots & g''(\xi_m) \\ \vdots & & \vdots \\ g\left(t_k - s_1\right) & \cdots & g\left(\tau - s_m\right) \\ g'\left(t_k - s_1\right) & \cdots & g'\left(t_k - s_m\right) \\ g\left(1 - s_1\right) & \cdots & g\left(1 - s_m\right) \end{vmatrix} =: \frac{\epsilon^2}{2} N_{1,1}^\epsilon, \qquad \text{(subtract 2nd row from } \tau_{\underline{l}} \text{ row)} \tag{233}$$

[6] To see this, take $f(\epsilon) = N_{1,1}(t_1 + \epsilon)$ and we know that f is continuous, $f(0) = 0$ and $f(\epsilon) > 0, \forall \epsilon$. If $f(\epsilon) = C\epsilon^2$ on $[0, \epsilon']$ for some $\epsilon' > 0$ and $C > 0$, which we show later in the proof (see (233)), then take $B = \min_{\epsilon \geqslant \epsilon'} f(\epsilon)$, and then there exists $\epsilon_0 \leqslant \epsilon'$ such that $f(\epsilon) \leqslant B, \forall \epsilon < \epsilon_0$. So we have that $f(\epsilon) \leqslant \min_{\tau \geqslant \epsilon} f(\tau), \forall \epsilon < \epsilon_0$, which implies that $\min_{\tau \geqslant \epsilon} f(\tau) = f(\epsilon), \forall \epsilon < \epsilon_0$.

for $\epsilon < \epsilon_0$ and note that swapping the $\tau_{\underline{l}}$ row with the third row involves an even number of adjacent row swaps, so the sign of the determinant remains the same. Also, for $\epsilon < \epsilon_0$, we have that:

$$N_{1,1}^{\epsilon} \to \begin{vmatrix} g(t_1 - s_1) & \cdots & g(t_1 - s_m) \\ g'(t_1 - s_1) & \cdots & g'(t_1 - s_m) \\ g''(t_1 - s_1) & \cdots & g''(t_1 - s_m) \\ \vdots & & \vdots \\ g(t_k - s_1) & \cdots & g(\tau - s_m) \\ g'(t_k - s_1) & \cdots & g'(t_k - s_m) \\ g(1 - s_1) & \cdots & g(1 - s_m) \end{vmatrix} =: N_{1,1}' > 0, \tag{234}$$

where the last inequality is true because Gaussians form an extended T-system (see [27]) and the determinant in the limit does not depend on ϵ. [7]

Substituting (233) and (231) into (228), and noting the the minimum in (231) can be attained at any $\tau_{\underline{l}} \in \bar{T}_\epsilon$ defined in (229) (not necessarily at $t_1 + \epsilon$, which we assumed above for simplicity), we obtain:

$$\frac{f_0}{\bar{f}} > \frac{2N_{\underline{l},1}}{\epsilon^2 \min_{\tau_{\underline{l}} \in \bar{T}_\epsilon} N_{1,1}^{\epsilon}}, \quad \forall \epsilon < \epsilon_0, \tag{235}$$

which is the condition we must impose on $f_0/\bar{f}$ so that $\det(M_N) > 0$ for $\epsilon < \epsilon_0$ instead of (127). Therefore, for $\epsilon < \epsilon_0$, we set

$$f_0 = C_\epsilon \cdot \frac{\bar{f}}{\epsilon^2}, \quad \text{where} \quad C_\epsilon = \frac{3N_{\underline{l},1}}{\min_{\tau_{\underline{l}} \in \bar{T}_\epsilon} N_{1,1}^{\epsilon}} \quad \text{and} \quad \lim_{\epsilon \to 0} C_\epsilon \in (0, +\infty). \tag{236}$$

Using that $f_1 < f_0$ and $1 < f_0$, we bound $\bar{C}(f_0, f_1)$ in (52):

$$\bar{C}(f_0, f_1) < \bar{c}_1 f_0^2, \tag{237}$$

where $\bar{c}_1$ is a universal constant. Finally, we insert (236) and (237) into (59) and obtain:

$$\begin{aligned} C_1\left(\frac{1}{\epsilon}\right) &< \frac{(\bar{c}_1 f_0^2 + 2k)^{\frac{3}{2}}}{\bar{f}} = \frac{(\bar{c}_1 C_\epsilon^2 \bar{f}^2/\epsilon^4 + 2k)^{\frac{3}{2}}}{\bar{f}} \\ &< \frac{(\bar{c}_1 C_\epsilon^2 + 2k\epsilon^4)^{\frac{3}{2}}}{\bar{f}} \cdot \frac{1}{\epsilon^6}, \quad \forall \epsilon < \epsilon_0, \end{aligned} \tag{238}$$

where we have used that $\bar{f} < 1$. Similarly, from (60) we obtain

$$\begin{aligned} C_2\left(\frac{1}{\epsilon}\right) &< \frac{\bar{c}_2 f_0^{\frac{5}{2}}}{\bar{f}} = \bar{c}_2 C_\epsilon^{\frac{5}{2}} \bar{f}^{\frac{3}{2}} \cdot \frac{1}{\epsilon^5} \\ &< \bar{c}_2 C_\epsilon^{\frac{5}{2}} \cdot \frac{1}{\epsilon^5}, \quad \forall \epsilon < \epsilon_0, \end{aligned} \tag{239}$$

where c_2 is a universal constant, and this concludes the proof.

---

[7] Note that in the previous footnote we assumed that $f(\epsilon) = C\epsilon^2$, where C is independent of ϵ, but actually from (233) we have that $f(\epsilon) = N_{1,1}(t_1 + \epsilon) = \frac{\epsilon^2}{2} N_{1,1}^{\epsilon}$, where $N_{1,1}^{\epsilon} \to N_{1,1}' > 0$ as $\epsilon \to 0$. Our conclusion that the $N_{1,1}$ goes to zero at the rate ϵ^2 does not change because, for small enough $E > 0$, there exists $\epsilon' > 0$ such that:

$$0 < N_{1,1}' - E < N_{1,1}^{\epsilon} < N_{1,1}' + E, \quad \forall \epsilon < \epsilon',$$

so $f(\epsilon) \geqslant C_2 \epsilon^2$ for all $\epsilon < \epsilon'$, where $C_2 = (N_{1,1}' - E)/2 > 0$ is independent of ϵ. Then, using the argument in the previous footnote, there exists $\epsilon_0 > 0$ such that $\min_{\tau \geqslant \epsilon} f(\tau) \geqslant C_2 \epsilon^2$ for all $\epsilon < \epsilon_0$, and therefore the obtain a version of (235) where the factor in front of $1/\epsilon^2$ is independent of ϵ.