The dual approach to non-negative super-resolution: perturbation analysis

Stéphane Chrétien[1,2,3], Andrew Thompson[2], and Bogdan Toader[4]

[1]Laboratoire ERIC, Université Lyon 2, Bron, France
[2]National Physical Laboratory, Teddington, UK
[3]Alan Turing Institute, London, UK
[4]Department of Statistics and Data Science, Yale University, US

July 6, 2023

Abstract

We study the problem of super-resolution, where we recover the locations and weights of non-negative point sources from a few samples of their convolution with a Gaussian kernel. It has been shown that exact recovery is possible by minimising the total variation norm of the measure, and a practical way of achieve this is by solving the dual problem. In this paper, we study the stability of solutions with respect to the solutions dual problem, both in the case of exact measurements and in the case of measurements with additive noise. In particular, we establish a relationship between perturbations in the dual variable and perturbations in the primal variable around the optimiser and a similar relationship between perturbations in the dual variable around the optimiser and the magnitude of the additive noise in the measurements. Our analysis is based on a quantitative version of the implicit function theorem.

1 Problem setup

In the study of non-negative super-resolution, the aim is to estimate a signal x which consists of a number of point sources with unknown locations and non-negative magnitudes, from only a few measurements of the convolution of x with a known convolution kernel ϕ. This is a problem that arises in a number of applications, for example fluorescence microscopy [1], astronomy [2] or ultrasound imaging [3]. In such applications, the measurement device has a limited resolution and cannot distinguish between distinct point sources that are close to each other in the input signal x. This is often modelled as a deconvolution problem with a Gaussian kernel.

Specifically, let x be a non-negative measure on $I = [0, 1]$ consisting of k unknown non-negative point sources:

$$x = \sum_{i=1}^{k} a_i \delta_{t_i},$$

with $a_i > 0$, for all $i = 1, \ldots, k$, and let y_j be the possibly noisy measurements obtained by sampling the convolution of x with a known kernel ϕ at locations s_j:

$$y_j = \int_I \phi(t - s_j) x(\mathrm{d}t) + w_j = \sum_{i=1}^{k} a_i \phi(t_i - s_j) + w_j, \tag{1}$$

for all $j = 1, \ldots, m$ or, in vector notation:

$$y = \sum_{i=1}^{k} a_i \Phi(t_i) + w, \tag{2}$$

where

$$y = [y_1, \ldots, y_m]^T, \tag{3}$$

$$\Phi(t) = [\phi(t - s_1), \ldots, \phi(t - s_m)]^T, \tag{4}$$

$$w = [w_1, \ldots, w_m]^T. \tag{5}$$

Of particular interest is the case of the Gaussian kernel:

$$\phi(t) = e^{-t^2/\sigma^2}, \tag{6}$$

where σ is assumed to be known to the practitioner.

In the setting where the measurements y are exact, namely when $w = 0$, the signal x can be recovered by solving the following problem:

$$\min_{x \geqslant 0} \|x\|_{TV} \quad \text{subject to} \quad y = \int_I \Phi(t) x(\mathrm{d}t), \tag{7}$$

where $\|\cdot\|_{TV}$ is the Total Variation (TV) norm for Radon measures defined as

$$\|x\|_{\mathrm{TV}} = \sup \left\{ \int \psi \mathrm{d}x; \quad \psi \in C(I), \|\psi\|_\infty \leqslant 1 \right\}. \tag{8}$$

When additive measurement noise is present, the signal x can be recovered as the solution to

$$\min_{x \geqslant 0} \left\| y - \int_I \Phi(t) x(\mathrm{d}t) \right\|_1 \quad \text{such that} \quad \|x\|_{TV} \leqslant \Pi, \tag{9}$$

where Π plays the role of the regularisation parameter. Opting for an ℓ_1-type fidelity term is a reasonable choice in a robust estimation framework, as discussed in e.g. [4].

In the context of problems (7) and (9), in this manuscript we give bounds on the errors in the source locations $\{t_i\}_{i=1}^k$ and weights $\{a_i\}_{i=1}^k$ as a function of the errors in the dual variable when solving the dual problem, which we then extend to the case when the measurements are corrupted by additive noise, where we give an exact dependence of the error in the dual variable on the level of noise. A subset of the results in this paper have been presented in the conference article [5].

The problem of super-resolution has been studied extensively in the literature since the seminal paper [6], which addressed the case of complex amplitudes. Since the original contributions of Candès and Fernandez-Granda, there have been numerous follow-up results such as the ones by Schiebinger et al. [7], Duval and Peyré [8], Denoyelle et al. [9], Bendory et al. [10], Azaïs et al. [11], Eftekhari et al. [12, 13] and many others. For instance, the authors of [7] consider the noiseless setting by taking real-valued samples of y with a more general choice of ϕ (such as a Gaussian) and also assume x to be non-negative as in the present work. Their proposed approach again involves TV norm minimization with linear constraints. Bendory et al. [10] consider ϕ to be Gaussian or Cauchy, do not place sign assumptions on x, and also analyze the TV norm minimization with linear fidelity constraints for estimating x from noiseless samples of y.

1.1 Main goals of our study

A standard way to approach problem (7) is by considering its dual:

$$\max_{\lambda \in \mathbb{R}^m} \; y^T \lambda \quad \text{subject to} \quad \lambda^T \Phi(t) \leqslant 1 \quad \forall t \in I, \tag{10}$$

which is a finite-dimensional problem with infinitely many constraints, known as a semi-infinite program (SIP). One of the main motivations for the study of the dual problem stems from the fact that this dual problem is *finite (and even sometimes low) dimensional* and as such, is amenable to efficient optimisation algorithms such as exchange methods [14] or sequential quadratic programming [15]. Moreover, the constraints $\lambda^T \Phi(t) \leqslant 1, \forall t \in I$ can be handled using an exact penalty approach, i.e. can be reformulated as

$$\min_{\lambda \in \mathbb{R}^m} -y^T \lambda + C \cdot \max \left\{ \sup_s \left(\lambda^T \Phi(s) - 1 \right), 0 \right\}, \tag{11}$$

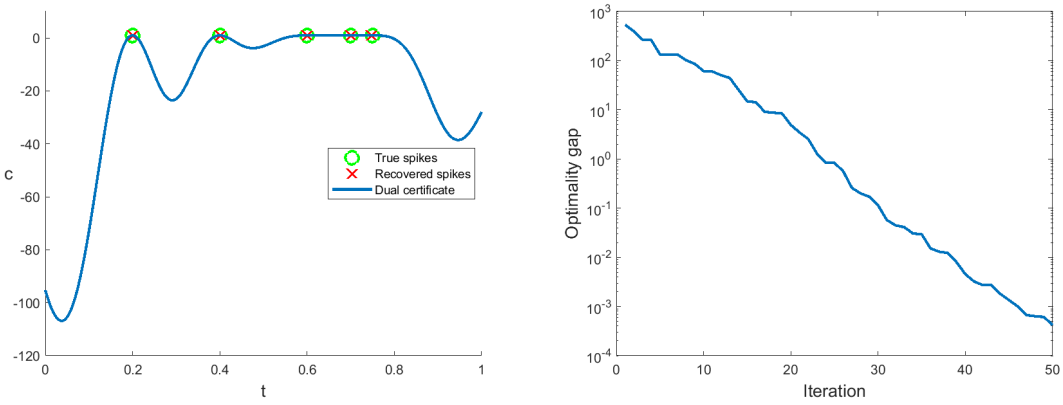

Figure 1: (a) Solution and dual certificate obtained using the level bundle method. (b) Decrease in optimality gap with iteration number.

for a penalty parameter C, thus making the problem amenable to non-smooth optimisation algorithms such as bundle methods [16, 17]. To illustrate the use of such methods for solving the dual problem, we present the result of an experiment in which we use the level bundle method [16] to solve a continuous sparse inverse problem of the kind introduced in this section. Here a signal (consisting of five spikes with locations $\{0.2, 0.4, 0.6, 0.7, 0.75\}$ each with amplitude 1) is convolved with a Gaussian kernel $\phi(t) = e^{-t^2/\sigma^2}$ with $\sigma = 0.1$ and sampled at 15 equispaced points on $[0, 1]$. The dual problem is solved using the level bundle method, and the spike locations are identified from the global maximisers of the dual certificate obtained. Figure 1(a) displays the recovered solution using the level bundle method along with the corresponding dual certificate, showing that the method is able to recover the signal to high accuracy even though the minimum separation is somewhat small (0.05). Figure 1(b) shows the speed of convergence in terms of the decrease in the optimality gap (the model gap - see Appendix B). We observe linear convergence in practice.

Solving the dual problem for λ leads to the *dual certificate*, a function of the form $q(s) = \sum_{j=1}^{m} \lambda_j \phi(s - s_j)$ (defined in Section 2), whose global maximisers are the source locations $\{t_i\}_{i=1}^{k}$. The weights $\{a_i\}_{i=1}^{k}$ are then found by solving a least squares problem using the measurements and the source locations. Using the idea of dual certificate, our perturbation results are quite intuitive: the locations of the global maximisers of the dual certificate are perturbed when λ is perturbed, which leads to perturbed source locations t_i. Providing a quantitative analysis of the recovery error as a function of the error in the dual solution is the main goal of the present work. In addition, we extend the analysis to the noisy setting, where we give the explicit dependence of the error of the dual solution on potential additive noise in the measurements.

1.2 Our contributions

In this paper, we restrict our study to the case of Gaussian kernels. Our main results are the following

- In the setting of exact measurements, we provide bounds on how far the estimated locations t_k and magnitudes a_k are from their true values as the dual variable λ is perturbed from its optimal value λ^* when x is recovered by solving the dual problem (10). These bounds are given in Theorems 2 and 4. These give us an insight into the size of the error in the locations and magnitudes when we apply an optimisation algorithm to the dual of the super-resolution problem.

- In the setting of measurements corrupted by additive noise, we leverage the perturbation bounds obtained for the noiseless case in order to study the impact of additive noise in the observations, when the signal is recovered by solving the alternative problem (9). For this purpose, we make precise links between the dual solutions to (9) and (10). Our main result for this noisy setup is Theorem 8, where we give an explicit bound on the impact of noise on the estimation of the dual solution to (10). This makes again the case for the study of (10) under perturbation.

While the bounds given in these theorems apply only to the case when the convolution kernel is Gaussian, the same techniques can be applied to obtain perturbation bounds for other kernels, with a few differences in the way some sums in the proofs are bounded, which would would be specific to the kernel used.

1.3 Comparison with previous work

1.3.1 Alternative formulations for the noiseless setting

For the particular case of non-negative x, Boyd et al. [18] proposed an improved Frank-Wolfe algorithm in the primal. In certain instances, for e.g., with Fourier samples (such as in [6, 19]), the dual, which is a SIP, can also be reformulated as a semi-definite program (SDP). From a practical point of view, SDP is notoriously slow for even moderately large number of variables. The algorithm of [18] is a first order scheme with potential local correction steps, and is practically more viable.

As already mentioned, the main reason we advocate for using the dual problem (10) is that exact penalty can be used in order to reformulate the dual problem as a non-smooth minimisation problem for which methods such as bundle methods [16], [20] are efficient in practice. To the best of our knowledge, there is no analysis of the impact of obtaining approximate solutions of the dual on the quality of the recovered locations.

1.3.2 The penalised least squares approach

The approach adopted in [8, 9] is to solve a least-squares-type minimization procedure with a TV norm based penalty term (also referred to as the Beurling LASSO (for example [21])) for recovering x from samples of y. The approach in [22] considers a natural finite approximation on the grid to the continuous problem, and studies the limiting behaviour as the grid becomes finer; see also [23]. These works develop a perturbation analysis which is different from ours since it applies to specific types of perturbations of a different problem (ℓ_2 vs. ℓ_1 type fidelity terms), and do not provide precise quantitative dependencies with respect to all the parameters of the problem.

1.3.3 The Prony/Matrix Pencil approach

Another efficient approach is the one of [24] based on the original work of Hua and Sarkar [25] using a Matrix Pencil approach, and recently extended to the multi-kernel setting in [26]. Perturbation analysis of the Matrix Pencil approach is provided in [24]; see also [26] for a more detailed exposition of these results with the correct order of dependencies. The reason we develop an analysis of the dual problem (10) here is that it easily extends to the multidimensional setting as well, at least for small dimensions. In contrast, the Matrix Pencil method, although very efficient in one dimension, becomes much more involved in several dimensions [27].

1.4 Plan of the paper

We start by presenting the noise-free perturbation results related to problem (10) in Section 2, followed by the perturbation results in the setting when the measurements are corrupted by noise in Section 3. The proofs of our results are given in Section 4 and we show numerical experiments to verify the validity of our results in practice in Section 5. Lastly, we conclude the paper in Section 6.

2 Bound on the error as λ is perturbed – the noise-free case

In this section we present our first main results, namely two theorems that give bounds on the perturbations around the source locations t_i and the magnitudes a_i respectively, as the dual variable is perturbed away from the optimiser λ^*, when the convolution kernel is a Gaussian with known width σ as defined in (6).

First, let us briefly give an informal statement of the main results in this section.

Informal Theorem. ***(Stability of primal recovery)*** *Let $\lambda^* \in \mathbb{R}^m$ be a solution of the dual program* (10) *with ϕ Gaussian and λ a perturbation of λ^* in a ball of radius δ_λ, given in Theorem 2 and let $\mathbf{t}^*, \mathbf{a}^*$ be the*

vectors of source locations and weights in the true signal x and $\tilde{\mathbf{t}}, \tilde{\mathbf{a}}$ respectively their perturbations due to λ. Then, the error between $\mathbf{t}^*$ and $\tilde{\mathbf{t}}$ is bounded by:

$$\|\tilde{\mathbf{t}} - \mathbf{t}^*\|_2 \leqslant \sqrt{k} C_{t*} \|\lambda - \lambda^*\|_2. \tag{12}$$

Moreover, if the above error is bounded by δ_t, given in Theorem 4, then the error between $\mathbf{a}^*$ and $\tilde{\mathbf{a}^*}$ is bounded by:

$$\|\tilde{\mathbf{a}} - \mathbf{a}^*\|_2 \leqslant C_{a*} \|\tilde{\mathbf{t}} - \mathbf{t}^*\|_2 + O(\|\tilde{\mathbf{t}} - \mathbf{t}^*\|_2^2). \tag{13}$$

As the two error bounds above are derived independently using different ideas, we will discuss them individually. Before giving the exact statement of each theorem, we define the concept of dual certificate, which plays an important role throughout this paper.

Definition 1. ***(Dual certificate)*** *Consider a solution λ^* of the dual problem (10) or (27). Then a dual certificate is a function of the form*

$$q(t) = \sum_{j=1}^{m} \lambda_j^* \phi(t - s_j) = {\lambda^*}^T \Phi(t), \tag{14}$$

which satisfies the conditions:

$$q(t_i) = 1, \quad \forall i = 1, \ldots, k, \tag{15}$$

$$q(t) < 1, \quad \forall t \neq t_i, \forall i = 1, \ldots, k. \tag{16}$$

The idea of dual certificate is common in the super-resolution literature, and we know that the global maximisers of $q(t)$ correspond to the source locations $\{t_i\}_{i=1}^k$ (see, for example [6, 7, 13], Once these are found, amplitudes $\{a_i\}_{i=1}^k$ are obtained by solving a linear system.

We are now ready to discuss the perturbation results in the noise-free setting. In the following theorem, we consider the dual (10) of (7) and quantify how the source locations given by the global maximisers of the dual certificate formed by the dual solution λ^* are affected by perturbations of λ^*.

Theorem 2. ***(Dependence of $|t - t^*|$ on $\|\lambda - \lambda^*\|_2$)*** *Let $\lambda^* \in \mathbb{R}^m$ be a solution of the dual program (10) with ϕ Gaussian as given in (6) such that the the dual certificate $q(s)$ defined in (14) satisfies conditions (15) and (16), λ a perturbation of λ^* in a ball of radius δ_λ and t an arbitrary local maximiser of $q_\lambda(s) = \sum_{j=1}^m \lambda_j \phi(s - s_j)$. Note that, for $\lambda = \lambda^*$, the corresponding local (and global) maximiser t^* of $q_{\lambda*} = q$ is a true source location in $\{t_i\}_{i=1}^k$. Let $R = \frac{\|\lambda^*\|_2}{\sigma}$ and $c \approx 3.9036$ a universal constant. If the radius δ_λ is bounded by*

$$\delta_\lambda \leqslant \frac{|q''(t^*)|^2 \sigma^3 \sqrt{e}}{4\sqrt{2}\,(2 + cR)\, m}, \tag{17}$$

then

$$|t - t^*| \leqslant C_{t*} \|\lambda - \lambda^*\|_2, \tag{18}$$

where

$$C_{t*} = \frac{1}{4 + cR} \left[1 + \frac{2\sqrt{2m}(2 + cR)}{|q''(t^*)|\sqrt{e}} \right]. \tag{19}$$

$$\tag{20}$$

Proposition 3. ***(Simplified C_{t*})*** *Under the conditions of Theorem 2, the constant C_{t*} can be further bounded by:*

$$C_{t*} < \frac{1}{4} + \frac{2\sqrt{2}}{\sqrt{e}} \cdot \frac{\sqrt{m}}{|q''(t^*)|}. \tag{21}$$

The proofs of Theorem 2 and Proposition 3 are given in Section 4.1. As a brief summary, Theorem 2 is proved by applying the implicit function theorem to the function $F(t, \lambda) = q'(t)$, where $q(t)$ is the dual certificate given in Definition 1, since we know that $F(t^*, \lambda^*) = 0$. This allows us to express t as a function $t(\lambda)$ in a neighbourhood of (t^*, λ^*), and a quantitative version of the theorem [28] gives an explicit expression for $\partial_\lambda t(\lambda)$ and the neighbourhood in terms of the derivatives of F. By bounding this derivative and the neighbourhood and then applying a truncated Taylor expansion to $t(\lambda)$, we obtain the result of Theorem 2.

One of the main conclusions which can be drawn from this result is that the primal spike location error is controlled in l_∞, but degrades as a function of the number of measurements in the order of $\sqrt{m}$. Alternatively, we can write (18) in terms of the ℓ_2 norm of the error between the vector of true source locations $\mathbf{t}^*$ and the perturbed source locations $\tilde{\mathbf{t}}$:

$$\|\tilde{\mathbf{t}} - \mathbf{t}^*\|_2 \leqslant \sqrt{k} C_{t*} \|\lambda - \lambda^*\|_2.$$

Of crucial importance is the curvature of the dual certificate at the true solution: the flatter the certificate, the worse the estimation error. Our theorem also gives important information about the accuracy in the dual variable required to guarantee our upper bound on the error of recovery. This accuracy is of the inverse order of the number of measurements, which is quite a stringent constraint. Both the m and the $\sqrt{m}$ factors are a consequence of the way we bound sums of shifted copies of the kernel, namely $\sum_{j=1}^{m} \phi(t - s_j) \leqslant m \max_{t \in \mathbb{R}} \phi(t)$. Given the fast decay of the Gaussian, it is clear that this is not a tight bound. However, any bound would reflect the density of samples close to each source location.

We will now give a result regarding the perturbation of the magnitudes a_i when λ^* is perturbed. Let Φ be the matrix whose entries are defined as

$$\Phi_{ij} = \phi(t_j - s_i), \tag{22}$$

and $\mathbf{t}^*$ and $\mathbf{a}^*$ the vectors of source locations and weights:

$$\mathbf{t}^* = [t_1, \ldots, t_k]^T, \qquad \mathbf{a}^* = [a_1, \ldots, a_k]^T.$$

When we solve (10) exactly, we obtain the source locations by finding the global maximisers of $q(s)$. Then, the vector of weights $\mathbf{a}^*$ is found by solving the system

$$\Phi \mathbf{a} = \mathbf{y}.$$

When the source locations are perturbed, we denote the resulting perturbed data matrix by:

$$\tilde{\Phi} = \Phi + E, \tag{23}$$

and we calculate the vector of perturbed weights $\tilde{\mathbf{a}}$ as the solution of the least squares problem

$$\min_{\mathbf{a}} \|\tilde{\Phi}\mathbf{a} - \mathbf{y}\|_2. \tag{24}$$

The following theorem, proved in Section 4.2, gives a bound on the error $\|\mathbf{a}^* - \tilde{\mathbf{a}}\|_2$ between the vector of true weights $\mathbf{a}^*$ and the vector of weights $\tilde{\mathbf{a}}$ obtained by solving the least squares problem (24) with the perturbed matrix $\tilde{\Phi}$, as a function of the error $\|\tilde{\mathbf{t}} - \mathbf{t}^*\|_2$ between the perturbed source locations $\tilde{\mathbf{t}}$ and the true source locations $\mathbf{t}^*$.

Theorem 4. ***(Dependence of** $\|\tilde{\mathbf{a}} - \mathbf{a}^*\|_2$ **on** $\|\tilde{\mathbf{t}} - \mathbf{t}^*\|_2$**)*** *Let $\mathbf{t}^* \in [0, 1]^k$ be the vector of true source locations, $\tilde{\mathbf{t}} \in [0, 1]^k$ the perturbed source locations in a ball of radius δ_t, $\mathbf{a}^*$ the vector of true weights and $\tilde{\mathbf{a}}$ the vector of perturbed weights obtained by solving problem* (24)*. If the radius δ_t is bounded by:*

$$\delta_t < \frac{\sigma^2 \sigma_{\max}(\Phi)}{4 e^{4/\sigma^2} \sqrt{m}} \left(\sqrt{1 + \frac{\sigma_{\min}^2(\Phi)}{\sigma_{\max}^2(\Phi)}} - 1 \right), \tag{25}$$

where $\sigma_{\max}(\Phi)$, $\sigma_{\min}(\Phi)$ are the largest and respectively smallest singular values of the matrix Φ defined in (22)*, then:*

$$\|\tilde{\mathbf{a}} - \mathbf{a}^*\|_2 \leqslant C_{a*} e^{\frac{4}{\sigma^2} \max_j |\tilde{t}_j - t_j|} \|\tilde{\mathbf{t}} - \mathbf{t}^*\|_2 + O(\|\tilde{\mathbf{t}} - \mathbf{t}^*\|_2^2), \tag{26}$$

where

$$C_{a*} = \frac{4\sqrt{m} \|\mathbf{a}^*\|_2}{\sigma^2 \sigma_{\min}(\Phi)}.$$

Note that we write the $O(\|\tilde{\mathbf{t}} - \mathbf{t}^*\|_2^2)$ term in the bound above in order to simplify the presentation of the result. We can, however, calculate the constants corresponding to the higher order terms in the bound by using the inequality (123) in the proof of Theorem 4 in Section 4.2. For example, the constant in the second order term is equal to $C_{a*}^2/\|\mathbf{a}^*\|_2 \left[1 + 2\sigma_{\max}^2(\Phi)/\sigma_{\min}^2(\Phi)\right]$.

3 Bound on $\|\lambda - \lambda^*\|_2$ in terms of the noise w

In this section we assume that the measurements are corrupted by additive noise and we give a result where we bound the perturbation in the dual variable λ around the minimiser λ^* as a function of the noise w in the measurements. Specifically, the noisy measurements are defined as in (1):

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

for $w_j \neq 0$ and $j = 1, \ldots, m$.

The aim is to estimate how the source locations $\{t_i\}_{i=1}^k$ and weights $\{a_i\}_{i=1}^k$ are affected by the additive noise w in the measurements around the solution of the problem. In the previous section we have established how the source locations and weights are perturbed around their true values as the dual variable λ is perturbed around its optimal value λ^*. In the noisy setting, we want to establish a precise quantitative relationship between the perturbations of λ around λ^* and the magnitude of the noise.

Before we state the main result, which gives a relationship of this kind, first we need to describe the exact mathematical setting under which the result holds. Then we introduce the function $\bar{F}$ in (39) to which we apply the implicit function theorem, whose Jacobian is crucial for this result.

In order to account for noise in the measurements, we consider a slightly modified version of the dual problem (10). To be specific, we use an additional box constraint on the dual variable λ and obtain the dual problem:

$$\begin{aligned} \max_{\lambda \in \mathbb{R}^m} y^T \lambda \quad \text{such that} \quad & \lambda^T \Phi(t) \leqslant 1, \quad \forall t \in I, \\ \text{and} \quad & \|\lambda\|_\infty \leqslant \tau, \end{aligned} \tag{27}$$

which is the dual of (9) and whose derivation is given in Appendix A. The parameter τ is the inverse of the Lagrange multiplier corresponding to the constraint in (9), and therefore it plays the same regularisation role as Π. Looking at the specific formulation of the primal problem (9), we can see that it takes measurement noise into account by doing ℓ_1 minimisation of the error instead of requiring the measurements to be satisfied exactly.

To motivate the exact form of the function $\bar{F}$ in (39) to which we apply the implicit function theorem to obtain the perturbation result from Theorem 8, consider the exact penalty formulation of (27):

$$\min_{\lambda \in \mathbb{R}^m} \Psi_\Pi(\lambda) \quad \text{such that} \quad \|\lambda\|_\infty \leqslant \tau, \tag{28}$$

where

$$\Psi_\Pi(\lambda) = -y^T \lambda + \Pi \cdot \max\left\{ \sup_s \left(\sum_{j=1}^{m} \lambda_j \phi(s - s_j) - 1 \right), 0 \right\}. \tag{29}$$

For a large enough value of Π, a solution to (28) which satisfies the constraints in (27) is also a solution of (27) (see, for example, Section 1.2 in [20]). This is a non-smooth optimisation problem and its solution can be found by using any method that relies on calculating subgradients, for example the level method [16].

A subgradient of $\Psi_\Pi(\lambda)$ has the form:

$$\partial \Psi_\Pi = \begin{cases} -y + \Pi \sum_{i=1}^{k'} \nu_i g(s_i^*), & (\nu_1 + \ldots + \nu_{k'} = 1) \qquad \text{if} \quad \sup_s \sum_{j=1}^m \lambda_j \phi(s - s_j) > 1, \\ -y + \Pi \sum_{i=1}^{k'} \nu_i g(s_i^*), & (\nu_1 + \ldots + \nu_{k'} \leqslant 1) \qquad \text{if} \quad \sup_s \sum_{j=1}^m \lambda_j \phi(s - s_j) = 1, \\ -y, & \qquad\qquad\qquad\quad \text{if} \quad \sup_s \sum_{j=1}^m \lambda_j \phi(s - s_j) < 1, \end{cases} \tag{30}$$

where $\{s_i^*\}_{i=1}^{k'}$ are the global maximisers of the function $\sum_{j=1}^m \lambda_j \phi(s-s_j)$, the vectors $g(s)$ are of the form $g(s) = [\phi(s-s_1), \ldots, \phi(s-s_m)]^T$ and $\nu_i \geqslant 0$ for all $i = 1, \ldots, k'$. Note that here we apply the formula for the subgradient of the max function and for the sup function (see for example [20]). The coefficients in the convex combination from the formula for the subgradient of the max function with zero account for the case when $\nu_1 + \ldots + \nu_{k'} < 1$[1].

As in the noise-free setting, we assume here that the dual solution λ^* forms a dual certificate, namely the function $q(s)$ as defined in (14) satisfies conditions (15) and (16). Then, the subdifferential at λ^* has the form:

$$\partial \Psi_\Pi(\lambda^*) = -y + \Pi \sum_{i=1}^k \nu_i g(t_i), \tag{31}$$

where $\{t_i\}_{i=1}^k$ are the source locations, so the optimality condition for (28):

$$0 \in \partial \Psi_\Pi(\lambda^*), \tag{32}$$

is equivalent to:

$$y = \Pi \sum_{i=1}^k \nu_i g(t_i), \tag{33}$$

for some $\nu_1, \ldots, \nu_k \geqslant 0$ with $\nu_1 + \ldots + \nu_k \leqslant 1$ and for $w = 0$. Note that, given the definition of y from (1), the optimality condition (33) is satisfied for

$$\nu_i = \frac{a_i}{\Pi}, \quad \forall i = 1, \ldots, k, \tag{34}$$

$$w_j = 0, \quad \forall j = 1, \ldots, m, \tag{35}$$

where in order to satisfy $\nu_1 + \ldots + \nu_k \leqslant 1$, we need Π such that:

$$\Pi \geqslant a_1 + \ldots + a_i, \tag{36}$$

which is the same as the constraint in (9).

Motivated by the above reasoning, we now want to apply the quantitative implicit function theorem, as given in [28], to a function F of the form:

$$F([\lambda, \nu]^T, w) = \sum_{i=1}^k a_i \Phi(t_i^*) - \sum_{i=1}^k \nu_i \Phi(t_i(\lambda)) + w, \tag{37}$$

where we know that $F([\lambda^*, a]^T, 0) = 0$. For the sake of simplicity, we include the parameter Π in the coefficients ν_i, so in the second sum in F each ν_i actually corresponds to $\Pi\nu_i$, and $\nu_1 + \ldots + \nu_k \leqslant \Pi$ rather than $\nu_1 + \ldots + \nu_k \leqslant 1$.

However, note that $F : \mathbb{R}^{m+k} \times \mathbb{R}^m \to \mathbb{R}^m$ and in order to apply the implicit function theorem to obtain the dependence of the first argument of F as a function of the second argument, it is required that spaces of the first argument, the second argument and the codomain of F have the same dimension. To overcome this issue, we assume that the solution we work with has a few particular properties, since the dual certificate, given in Definition 1, is not unique in general. As before, we will assume that the solution λ^* to the dual problem (27) satisfies the dual certificate condition. In addition, we assume the existence of a solution $\bar{\lambda}^*$ of (27) as follows:

Definition 5. *Let $\lambda^* \in \mathbb{R}^m$ be a solution to the dual problem (27) with $m-k$ entries on the boundary of the box constraint of (27) i.e. there exist indices $\gamma_1, \ldots, \gamma_{m-k} \in \{1, \ldots, m\}$ such that $\lambda^*_{\gamma_j} = \pm\tau$, $j = 1, \ldots, m-k$. Then we define $\bar{\lambda}^* \in \mathbb{R}^k$ to be the vector that consists of the non-fixed entries of λ^*, in the same order, and $\bar{\lambda} \in \mathbb{R}^k$ a perturbation of $\bar{\lambda}^*$.*

[1] More specifically, both functions in the max attain their maximum, so we have that $\partial\Psi_\pi = -y + \Pi\left[\alpha_1 \partial \sup_s \left(\sum_{j=1}^m \lambda_j \phi(s-s_j) - 1\right) + \alpha_2 \partial 0\right]$, with $\alpha_1, \alpha_2 > 0$ and $\alpha_1 + \alpha_2 = 1$, and therefore $\partial\Psi_\pi = -y + \Pi \sum_{i=1}^{k'} \alpha_1 \nu_i' g(s_i^*)$, with $\nu_1' + \ldots + \nu_{k'}' = 1$ and $0 \leqslant \alpha_1 \leqslant 1$.

In practice, such a solution $\bar{\lambda}^*$ would be achieved due to the complementarity conditions at optimality corresponding to the box constraint $\|\lambda\| \leqslant \tau$. Similarly, we define a vector consisting of $2k$ entries of $\Phi(t)$ in (4).

Definition 6. *Let $\theta := \{\theta_1, \ldots, \theta_{2k}\} \subset \{1, \ldots, m\}$. Then we define $\bar{\Phi}_\theta(t)$ to be the vector consisting of the entries of $\Phi(t)$ in (4) corresponding to the indices in θ:*

$$\bar{\Phi}_\theta(t) := [\phi(t - s_{\theta_1}), \phi(t - s_{\theta_2}), \ldots, \phi(t - s_{\theta_{2k}})]^T. \tag{38}$$

We will also use $\bar{\Phi}(t)$ to denote $\bar{\Phi}_\theta(t)$ when the specific choice of θ is not relevant in the context.

Lastly, given the definitions of $\bar{\lambda}$ and $\bar{\Phi}_\theta(t)$ above, we define the following function to which we will be able to apply the implicit function theorem:

Definition 7. *Let $\lambda^* \in \mathbb{R}^m$ be a solution of (27) with $m - k$ fixed entries and $\bar{\lambda}^* \in \mathbb{R}^k$ consisting of the non-fixed entries of λ^*, as given in Definition 5, and let $\bar{\Phi}_\theta(t)$ be given as in Definition 6 for an index set θ of $2k$ indices between 1 and m. Then, for the perturbation $\bar{\lambda}$ of $\bar{\lambda}^*$, we define the function $\bar{F} : \mathbb{R}^{2k} \times \mathbb{R}^{2k} \to \mathbb{R}^{2k}$ as:*

$$\bar{F}([\bar{\lambda}, \nu]^T, w_\theta) = \sum_{i=1}^{k} a_i \bar{\Phi}_\theta(t_i^*) - \sum_{i=1}^{k} \nu_i \bar{\Phi}_\theta(t_i(\bar{\lambda})) + w_\theta, \tag{39}$$

where $t_i(\bar{\lambda}^*) = t_i^*$, for $i = 1, \ldots, k$, are the source locations corresponding to λ^* and $w_\theta \in \mathbb{R}^{2k}$ contains the entries of the noise vector $w \in \mathbb{R}^m$ corresponding to the indices in θ.

We can now state the main result of this section, namely a bound on the perturbation of λ^* (or more specifically $\bar{\lambda}^*$) as a function of the measurement noise. The proof is given in Section 4.3.

Theorem 8. ***(Dependence of $\|\lambda - \lambda^*\|_2$ on the noise w)*** *Let $\lambda^* \in \mathbb{R}^m$ be a solution to the dual problem (27) with $w = 0$, namely the optimal solution of (27) with noiseless measurements, which satisfies the conditions in Definition 5, and the vector $\bar{\lambda}^* \in \mathbb{R}^k$ of non-fixed entries of λ^*. For the function $\bar{F}$ in Definition 7, let J^* be its Jacobian with respect to the first variable, evaluated at $([\bar{\lambda}^*, a]^T, 0)$ and $\sigma_{\min}(J^*)$ its smallest singular value. We also assume that the solution λ^* forms a dual certificate, namely the function $q(t)$ defined in (14) satisfies conditions (15) and (16). If J^* is invertible, $\|w\|_2 \leqslant \delta_w$ and*

$$|q''(t^*)| \leqslant 2\left(1 + \frac{4m\tau}{\sigma^2}\right), \tag{40}$$

then, for a perturbation λ of λ^* with the same fixed entries to the boundary of the box constraint, we have that:

$$\left\|\lambda - \lambda^*\right\|_2 \leqslant C_{\lambda^*} \cdot \|w\|_2, \tag{41}$$

where

$$C_{\lambda^*} = \frac{2}{\sigma_{\min}(J^*)}, \tag{42}$$

$$\delta_w = \frac{\sigma_{\min}(J^*)^2}{4P(m, k, \sigma, \Pi, \tau, C_{t^*})}, \tag{43}$$

and

$$\begin{aligned} P(m, k, \sigma, \Pi, \tau, C_{t^*}) = \sqrt{2}k\Bigg[& \frac{1}{\sigma^2}\left(2\sqrt{k}C_{t^*}^2\Pi + 4kC_{t^*}\bar{\Delta}_2\tau\Pi\right) \\ & + \frac{1}{\sigma}\left(\frac{\sqrt{2k}C_{t^*}}{\sqrt{e}} + 4\sqrt{k}C_{t^*}^2\Pi + \frac{2\sqrt{2}\bar{\Delta}_2\Pi}{\sqrt{e}} + 8kC_{t^*}\bar{\Delta}_2\tau\Pi + \frac{\sqrt{2k}\bar{\Delta}_2\Pi}{\sqrt{e}} + \sqrt{\frac{2}{e}}C_{t^*}\right)\Bigg], \end{aligned} \tag{44}$$

where C_{t^*} is given in (19) in Theorem 2, $c \approx 3.9036$, $c_2 = 4 + \frac{c\sqrt{2}}{\sqrt{e}} \approx 7.3484$ are universal constants and

$$\bar{\Delta}_2 = \frac{\sqrt{k}}{\sigma^4}\left(c_2 C_{t^*} m\tau + \frac{2\sqrt{2}}{\sqrt{e}}\sigma\right). \tag{45}$$

The theorem above makes explicit the dependence of the perturbation in the dual variable λ around the solution λ^* on the additive noise w in the measurement vector y, with the assumption given in Definition 5. This is a linear relation where the constant depends on the specific configuration of the problem we are solving, namely the locations and weights of the sources, and width of the Gaussian and the sampling locations. The theorem also gives an upper bound on the magnitude of the noise where this result holds as a function of the same parameters.

As an additional interpretation of Theorem 8 regarding the assumption on the fixed entries in λ and λ^*, it states that, for a solution λ to the dual problem (27) with noisy measurements that has $m-k$ entries equal to the boundary of the box constraint, there is a solution λ^* to the noise-free dual problem $w=0$ with the same entries fixed to the boundary of the box constraint and the error for the remaining k entries bounded by (41).

Moreover, under a few additional assumptions, we give a simplified approximation of the constant P in (44) for clarity:

Proposition 9. ***(Simplified P)*** *Under the conditions of Theorem 8 and, in addition, if* $\Pi, \tau \leqslant 1$*, then:*

$$P(m,k,\sigma,C_{t^*}) = \mathcal{O}\left(\frac{mk^{5/2}C_{t^*}^2}{\sigma^6}\right). \tag{46}$$

One important observation is that, while the above result only applies to a subset of the entries in λ and w, which entries are selected is not arbitrary. The choice of the entries in λ and w reflects which samples s_j contain the most information, and therefore which noise entries in w affect the solution to the optimisation problem the most. More specifically, in order for the Jacobian J^* to be invertible, we are led to select the samples (and therefore λ and w entries) that satisfy this condition the best, namely the ones that are the closest to the source locations. We discuss this aspect in more detail in Section 3.1.

Lastly, note that the results in Section 2 and Section 3 refer to different optimisation problems: the duals (10) and (27) of problems (7) and (9) respectively. However, the proofs of our perturbation results rely on the property that the dual solution λ forms a dual certificate, the global maximisers of which give the locations of the point sources in the input signal x, with the additional bound on λ from (27) being used in the proof of Theorem 8. Moreover, since our analysis is independent of the exact formulation of the primal problems, we can conclude that the results from both Section 2 and Section 3 apply to the problem of super-resolution in the noisy setting, namely they give bounds of the perturbations of the source locations and weights as a consequence of noise in the measurements.

3.1 Discussion

One of the conditions in Theorem 8 is that the Jacobian J^* is invertible. While we do not provide a rigorous analysis of the conditions in which this is satisfied, in this section we discuss in more detail what the condition requires and give further motivation for why it is true in a reasonable scenario. Specifically, we assume that the samples that are used for calculating the Jacobian are the closest samples to the sources, i.e. the set θ for which we define $\bar{F}$ in Definition 7 contains the two indices corresponding to the closest two samples to each source location, for each of the k sources. Therefore, the rows in the system given by $\bar{F}$ in (39), as well as the entries in $\bar{\lambda}$ and the entries in the noise vector w_θ, correspond to these samples.

Recall that J^* is the Jacobian of the function $\bar{F}$ from (39) with respect to the first argument. The entries in J^* are:

$$\partial_{\bar{\lambda}_l}\bar{F}([\bar{\lambda},\nu]^T,w)\Big|_{\substack{\bar{\lambda}=\bar{\lambda}^*\\ \nu=a\\ w_\theta=0}} = -\sum_{i=1}^{k} a_i\phi'(t_i^*-s_{\theta_j})\partial_{\bar{\lambda}_l}t_i(\bar{\lambda}^*) \tag{47}$$

$$= \sum_{i=1}^{k}\frac{a_i\phi'(t_i^*-s_{\theta_j})\phi'(t_i^*-s_l)}{q''(t_i^*)}, \tag{48}$$

for $l=1,\dots,k$, $j=1,\dots,2k$, where $\{s_l\}_{l=1}^k$ correspond to the non-fixed entries of λ (i.e. $\bar{\lambda}$) and

$$\partial_{\nu_l}\bar{F}([\bar{\lambda},\nu]^T,w_\theta)\Big|_{\substack{\bar{\lambda}=\bar{\lambda}^*\\ \nu=a\\ w_\theta=0}} = -\phi(t_l^*-s_{\theta_j}), \tag{49}$$

for $l = 1, \dots, k$, $j = 1, \dots, 2k$, where in the first equality we used (59) with (63) and (64) plugged in, so the result holds under the conditions in Theorem 2, namely for λ with $\|\lambda - \lambda^*\|_2 \leqslant \delta_\lambda$, where δ_λ is given in (17).

Writing J^* as

$$J^* = [J_\lambda J_\nu], \tag{50}$$

where the entries in the blocks J_λ and J_ν are given by (48) and (49) respectively, we have that:

$$J_\lambda = \sum_{i=1}^{k} \frac{a_i}{q''(t_i^*)} \bar{\Phi}'(t_i^*) \bar{\Phi}'(t_i^*)^T, \tag{51}$$

and

$$J_\nu = -[\bar{\Phi}(t_1^*) \dots \bar{\Phi}(t_k^*)], \tag{52}$$

where

$$\bar{\Phi}(t) = [\phi(t - s_{\theta_1}), \dots, \phi(t - s_{\theta_{2k}})]^T, \tag{53}$$

$$\bar{\Phi}'(t) = [\phi'(t - s_{\theta_1}), \dots, \phi'(t - s_{\theta_{2k}})]^T. \tag{54}$$

Note that $rank(J_\nu) = k$ by the T-systems property of the Gaussian (assuming that the $t_1 \leqslant \dots \leqslant t_k$ and $s_{\theta_1} \leqslant \dots \leqslant s_{\theta_{2k}}$) and in order for the matrix J^* to be invertible we need $rank(J^*) = 2k$, as it is a square matrix with $2k$ columns. By rewriting the columns of J_λ, we have that:

$$J^* = \left[\sum_{i=1}^{k} \frac{a_i \phi'(t_i^* - s_1)}{q''(t_i^*)} \bar{\Phi}'(t_i^*) \quad \dots \quad \sum_{i=1}^{k} \frac{a_i \phi'(t_i^* - s_k)}{q''(t_i^*)} \bar{\Phi}'(t_i^*) \quad -\bar{\Phi}(t_1^*) \quad \dots \quad -\bar{\Phi}(t_k^*) \right], \tag{55}$$

and by taking its determinant and using the multi-linearity property of the determinant with respect to its columns, we have that:

$$\begin{aligned} \det(J^*) = (-1)^k & \frac{a_1 \dots a_k}{q''(t_1^*) \dots q''(t_k^*)} \\ & \cdot \sum_{l=1}^{k!} \left(\prod_{i=1}^{k} \phi'(P_l(t_i^*) - s_i) \right) \left| P_l(\bar{\Phi}'(t_1^*)) \quad \dots \quad P_l(\bar{\Phi}'(t_k^*)) \quad \bar{\Phi}(t_1^*) \quad \dots \quad \bar{\Phi}(t_k^*) \right|, \end{aligned} \tag{56}$$

where P_l for $l = 1, \dots, k!$ are the permutations of k elements. Note that when we expand the determinant, the terms in the final sum are determinants with all the possible combinations of the vectors in each sum, which results in most determinants having repeated columns, so they are equal to zero. The only non-zero determinants in the resulting sum are the ones where the first k columns are the vectors $\{\Phi'(t_i^*)\}_{i=1}^k$ and their permutations, multiplied by the corresponding constants. We now order the columns of the determinant:

$$\begin{aligned} \det(J^*) &= (-1)^k \frac{a_1 \dots a_k}{q''(t_1^*) \dots q''(t_k^*)} \\ &\qquad \sum_{l=1}^{k!} sign(P_i) \left(\prod_{i=1}^{k} \phi'(P_l(t_i^*) - s_i) \right) \left| \bar{\Phi}(t_1^*) \quad \bar{\Phi}'(t_1^*) \quad \dots \quad \bar{\Phi}(t_k^*) \quad \bar{\Phi}'(t_k^*) \right|, \\ &= (-1)^k \frac{a_1 \dots a_k}{q''(t_1^*) \dots q''(t_k^*)} \left| \bar{\Phi}(t_1^*) \quad \bar{\Phi}'(t_1^*) \quad \dots \quad \bar{\Phi}(t_k^*) \quad \bar{\Phi}'(t_k^*) \right| \\ &\qquad \sum_{l=1}^{k!} sign(P_l) \left(\prod_{i=1}^{k} \phi'(P_l(t_i^*) - s_i) \right), \end{aligned} \tag{57}$$

where by $sign(P_i)$ we denote the sign of the determinant corresponding to the permutation P_i after reordering the columns as above. Because of the extended T-system property of the Gaussian function [29], the determinant above is strictly positive. The dominant term in the sum is the one corresponding to the identity permutation, where for each $i = 1, \dots, k$, the sample s_i is the closest sample to the source location t_i^*. As

the samples get further, the terms of the sum approach zero. This can be expressed more quantitatively by imposing explicit conditions on the distances between the closest samples and the sources, the separation of sources and the separation of samples, as done, for example, in [13].

As a last remark to motivate the choice of the dimension of $\bar{\lambda}^*$ and $\bar{\lambda}$ in Definition 5, note the expansion in (56) of $\det(J^*)$. If the vector $\bar{\lambda}$ had more than k entries, then the columns consisting of the permutations of $\bar{\Phi}'(t_i^*)$ would inevitably be repeated, since there are k sources t_i^* and more than k such columns. This implies that all the determinants in the sum would be zero and, therefore, J^* would not be invertible, implying that Theorem 8 would not be true in this case. This explains why choosing $\bar{\lambda}$ to contain more than k entries of λ would be incompatible with our analysis in the proof of Theorem 8.

4 Proofs

In this section we present the proofs of the theorems from Sections 2 and 3.

4.1 Proof of Theorem 2 (Dependence of $|t - t^*|$ on $\|\lambda - \lambda^*\|_2$)

Let t^* be an arbitrary local maximiser of the function $q(t)$ in (14), so t^* is also a source location, and λ^* the solution to (10). The key step in this proof is applying a quantitative version of the implicit function theorem [28] to the function:

$$F(t, \lambda) = \sum_{j=1}^{m} \lambda_j \phi'(t - s_j), \tag{58}$$

where $F(t^*, \lambda^*) = 0$ because t^* is a maximizer of $q(s)$ in (14). The theorem allows us to express t as a function $t(\lambda)$ of λ with:

$$\partial_\lambda t(\lambda) = -\left[\partial_t F(t(\lambda), \lambda)\right]^{-1} \partial_\lambda F(t(\lambda), \lambda), \tag{59}$$

for t in a ball of radius δ_0 around t^* and for λ in a ball of radius $\delta_1 \leqslant \delta_0$ around λ^*, where δ_0 is chosen such that

$$\sup_{(t,\lambda)\in V_\delta} \left\| I - \left[\partial_t F(t^*, \lambda^*)\right]^{-1} \partial_t F(t, \lambda) \right\| \leqslant \frac{1}{2}, \tag{60}$$

where $V_\delta = \left\{(t, \lambda) \in \mathbb{R}^{m+1} : |t - t^*| \leqslant \delta_0, \|\lambda - \lambda^*\| \leqslant \delta_0\right\}$ and δ_1 is given by

$$\delta_1 = (2M_t B_\lambda)^{-1} \delta_0, \tag{61}$$

where

$$B_\lambda = \sup_{(t,\lambda)\in V_\delta} \|\partial_\lambda F(t, \lambda)\|_2,$$
$$M_t = \left\| \partial_t F(t^*, \lambda^*)^{-1} \right\|_2.$$

The following two lemmas, proved in Sections 4.1.1 and 4.1.2 respectively, give us values of δ_0 and δ_1 that define balls around t^* and λ^* respectively which are included in the balls required by the quantitative implicit function theorem with radii defined in (60) and (61).

Lemma 10. ***(Radius of ball around t^*)*** *The condition* (60) *is satisfied if*

$$\delta_0 = \frac{\sigma^2 |q''(t^*)|}{\sqrt{m}\left(4 + 2c \cdot \frac{\|\lambda^*\|_2}{\sigma}\right)}. \tag{62}$$

Lemma 11. ***(Radius of ball around λ^*)*** *For δ_0 from Lemma 10 and δ_1 from condition* (61)*, the following choice of δ_λ:*

$$\delta_\lambda = \frac{\sigma\sqrt{e}|q''(t^*)|}{2\sqrt{2m}} \cdot \delta_0$$

satisfies $\delta_\lambda < \delta_1$.

Given the definition of the function F in (58), we have that

$$\partial_t F(t,\lambda) = \sum_{j=1}^{m} \lambda_j \phi''(t - s_j), \tag{63}$$

$$\partial_\lambda F(t,\lambda) = [\phi'(t - s_1), \quad \dots \quad , \phi'(t - s_m)]^T. \tag{64}$$

By applying Taylor expansion to $t(\lambda)$ around λ^* in the region defined by δ_0 and δ_λ, we have that

$$t(\lambda) = t(\lambda^*) + \left\langle \lambda - \lambda^*, \partial_\lambda t(\lambda_\delta) \right\rangle,$$

for some λ_δ on the line segment determined by λ^* and λ, so

$$\begin{aligned} \left|t(\lambda) - t(\lambda^*)\right| &\leqslant \left\|\lambda - \lambda^*\right\|_2 \cdot \left\|\partial_\lambda t(\lambda_\delta)\right\|_2 \\ &\leqslant \frac{\delta_0}{\sum_{j=1}^m \lambda_{\delta j} \phi''(t(\lambda_\delta) - s_j)} \cdot \left\| \left[\phi'(t(\lambda_\delta) - s_1), \quad \dots \quad , \phi'(t(\lambda_\delta) - s_m)\right] \right\|_2, \end{aligned} \tag{65}$$

where in the last inequality we used that $\|\lambda - \lambda^*\| \leqslant \delta_0$ and (59). We now need to bound the terms in (65) for the Gaussian kernel $\phi(t) = e^{-t^2/\sigma^2}$. First, we rewrite the last inequality as

$$\begin{aligned} |t(\lambda) - t(\lambda^*)| \sum_{j=1}^{m} &(\lambda_{\delta j} + \lambda_j^* - \lambda_j^*) \phi''(t(\lambda_\delta) - s_j) \\ &\leqslant \delta_0 \cdot \left\| \left[\phi'(t(\lambda_\delta) - s_1), \quad \dots \quad , \phi'(t(\lambda_\delta) - s_m)\right] \right\|_2, \end{aligned} \tag{66}$$

we apply the reverse triangle inequality in the sum on the left hand side:

$$\begin{aligned} |t(\lambda) - t(\lambda^*)| &\left[- \left| \sum_{j=1}^{m} (\lambda_{\delta j} - \lambda_j^*) \phi''(t(\lambda_\delta) - s_j) \right| + \left| \sum_{j=1}^{m} \lambda_j^* \phi''(t(\lambda_\delta) - s_j) \right| \right] \\ &\leqslant \delta_0 \cdot \left\| \left[\phi'(t(\lambda_\delta) - s_j)\right]_{j=1}^m \right\|_2 \end{aligned} \tag{67}$$

and then we apply the Cauchy-Schwartz inequality to the first sum on the left hand side above to obtain:

$$\begin{aligned} |t(\lambda) - t(\lambda^*)| &\left[- \left\|\lambda_\delta - \lambda^*\right\|_2 \cdot \left\| \left[\phi''(t(\lambda_\delta) - s_j)\right]_{j=1}^m \right\|_2 + \left| \sum_{j=1}^{m} \lambda_j^* \phi''(t(\lambda_\delta) - s_j) \right| \right] \\ &\leqslant \delta_0 \cdot \left\| \left[\phi'(t(\lambda_\delta) - s_j)\right]_{j=1}^m \right\|_2. \end{aligned} \tag{68}$$

To simplify the notation, we write $\delta_t = |t(\lambda) - t(\lambda^*)|$ and

$$A = \left\| \left[\phi''(t(\lambda_\delta) - s_j)\right]_{j=1}^m \right\|_2, \tag{69}$$

$$B = \left| \sum_{j=1}^{m} \lambda_j^* \phi''(t(\lambda_\delta) - s_j) \right|, \tag{70}$$

$$C = \left\| \left[\phi'(t(\lambda_\delta) - s_j)\right]_{j=1}^m \right\|_2, \tag{71}$$

and by using[2] $\|\lambda_\delta - \lambda^*\|_2 \leqslant \delta_0$, we have that:

$$\delta_t(-\delta_0 A + B) \leqslant \delta_0 C, \tag{72}$$

which can be further re-written as:

$$\delta_t \leqslant \frac{C + \delta_t A}{B} \cdot \delta_0. \tag{73}$$

The aim now is to obtain a bound on δ_t as a function of δ_0 and the parameters of the problem. Therefore, we need to lower bound B and upper bound $C + \delta_t A$.

[2]Since $\|\lambda - \lambda^*\| \leqslant \delta_0$ and λ_δ is on the line segment between λ^* and λ, then λ_δ is in the ball centred at λ^* with radius δ_0.

Bounding A,B,C

We start with B, for which we want to calculate a lower bound. First, we Taylor expand each term of the sum around $t(\lambda^*) - s_j$ as follows:

$$B = \left| \sum_{j=1}^{m} \lambda_j^* \phi''(t(\lambda^*) - s_j + t(\lambda_\delta) - t(\lambda^*)) \right|$$

$$= \left| \sum_{j=1}^{m} \lambda_j^* \phi''(t(\lambda^*) - s_j) + \big(t(\lambda_\delta) - t(\lambda^*)\big) \sum_{j=1}^{m} \lambda_j^* \phi'''(\xi_j) \right| \tag{74}$$

$$\geqslant \left| \sum_{j=1}^{m} \lambda_j^* \phi''(t(\lambda^*) - s_j) \right| - \big|t(\lambda_\delta) - t(\lambda^*)\big| \left| \sum_{j=1}^{m} \lambda_j^* \phi'''(\xi_j) \right|, \tag{75}$$

where $\xi_j \in \big[t(\lambda^*) - s_j - |t(\lambda_\delta) - t(\lambda^*)|, \quad t(\lambda^*) - s_j + |t(\lambda_\delta) - t(\lambda^*)|\big]$ for $j = 1, \ldots, m$, and on the last line we used the reverse triangle inequality. We calculate an upper bound of the last sum in the previous equation as follows:

$$\left| \sum_{j=1}^{m} \lambda_j^* \phi'''(\xi_j) \right| \leqslant \|\lambda^*\|_2 \cdot \left\| \big[\phi'''(\xi_j)\big]_{j=1}^{m} \right\|_2, \qquad \text{by Cauchy-Schwartz,} \tag{76}$$

$$\leqslant \frac{c\|\lambda^*\|_2 \sqrt{m}}{\sigma^3}, \tag{77}$$

where in the last line we used the maximum value of $\phi'''(t)$ and c is a constant.[3]

Finally, by using the δ_0 from Lemma 10 as a bound on $|t(\lambda_\delta) - t(\lambda^*)|$ and (77), we obtain:

$$B \geqslant |q''(t^*)| \left[1 - \frac{c\|\lambda^*\|_2}{4\sigma + 2c\|\lambda^*\|_2} \right]. \tag{78}$$

Note that the last fraction above is subunitary, so the bound is indeed positive.

Lastly, we upper bound $C + \delta_t A$. We bound both A and C using the upper bounds on ϕ' and ϕ'' given in footnote 3 and obtain:

$$A \leqslant \frac{2\sqrt{m}}{\sigma^2}, \tag{79}$$

$$C \leqslant \frac{\sqrt{2m}}{\sigma\sqrt{e}}, \tag{80}$$

and for δ_t we use the bound (62). Putting (62), (78), (79) and (80) together, we obtain:

$$\big|t(\lambda) - t(\lambda^*)\big| \leqslant C_{t^*} \cdot \big\|\lambda - \lambda^*\big\|_2, \tag{81}$$

where

$$C_{t^*} = \frac{2\sqrt{2m}\,\big(2\sigma + c\|\lambda^*\|_2\big)}{|q''(t^*)|\sigma\sqrt{e}\,\big(4\sigma + c\|\lambda^*\|_2\big)} + \frac{2\sigma}{4\sigma + F\|\lambda^*\|_2}, \tag{82}$$

which can also be written in the form in (19) in Theorem 2.

4.1.1 Proof of Lemma 10 (Radius δ_0 of the ball around t^*)

Let us now find the radius δ_0 which satisfies (60). Using (63), the expression inside the sup in (60) is

$$E = \left| 1 - \frac{\sum_{j=1}^{m} \lambda_j \phi''(t - s_j)}{\sum_{j=1}^{m} \lambda_j^* \phi''(t^* - s_j)} \right| = \frac{\left| \sum_{j=1}^{m} \lambda_j^* \phi''(t^* - s_j) - \lambda_j \phi''(t - s_j) \right|}{\left| \sum_{j=1}^{m} \lambda_j^* \phi''(t^* - s_j) \right|}. \tag{83}$$

[3] $\max_{t\in\mathbb{R}} \phi'(t) = \frac{\sqrt{2}}{\sigma\sqrt{e}}, \max_{t\in\mathbb{R}} \phi''(t) = \frac{2}{\sigma^2}, \max_{t\in\mathbb{R}} \phi'''(t) = \frac{c}{\sigma^3},$ where $c = \frac{4\sqrt{9-3\sqrt{6}}}{e^{\frac{3-\sqrt{6}}{2}}} \approx 3.9036.$

By denoting each term in the sum in the numerator in the last equation above by T_j and then adding and subtracting λ_j^* and t^*, we obtain:

$$\begin{aligned} T_j &= \lambda_j^* \phi''(t^* - s_j) - (\lambda_j - \lambda_j^*)\phi''(t - s_j) - \lambda_j^* \phi''(t^* - s_j + t - t^*) \\ &= -(\lambda_j - \lambda_j^*)\phi''(t - s_j) - \lambda_j^*(t - t^*)\phi'''(\xi_j), \end{aligned} \tag{84}$$

for some $\xi_j \in \left[t^* - s_j - |t - t^*|, t^* - s_j + |t - t^*|\right]$. Then:

$$\begin{aligned} E &\leqslant \frac{\left|\sum_{j=1}^m (\lambda_j - \lambda_j^*)\phi''(t - s_j)\right| + \left|\sum_{j=1}^m \lambda_j^*(t - t^*)\phi'''(\xi_j)\right|}{\left|\sum_{j=1}^m \lambda_j^* \phi''(t^* - s_j)\right|} \\ &\leqslant \frac{\|\lambda - \lambda^*\|_2 \left\|\left[\phi''(t - s_j)\right]_{j=1}^m\right\|_2 + |t - t^*| \left|\sum_{j=1}^m \lambda_j^* \phi'''(\xi_j)\right|}{\left|\sum_{j=1}^m \lambda_j^* \phi''(t^* - s_j)\right|} = E' \end{aligned} \tag{85}$$

We now have that

$$\sup_{(t,\lambda)\in V_{\delta_0}} E \leqslant \sup_{\substack{|t-t^*|\leqslant\delta_0,\\ \|\lambda-\lambda^*\|\leqslant\delta_0}} E' \tag{86}$$

$$\leqslant \delta_0 \cdot \frac{\left\|\left[\phi''(t - s_j)\right]_{j=1}^m\right\|_2 + \left|\sum_{j=1}^m \lambda_j^* \phi'''(\xi_j)\right|}{\left|\sum_{j=1}^m \lambda_j^* \phi''(t^* - s_j)\right|}. \tag{87}$$

We now further upper bound the fraction on the last line of the previous equation. The terms in the numerator are bounded by taking the maxima of the functions ϕ'' and ϕ''' from footnote 3 respectively:

$$\left\|\left[\phi''(t - s_j)\right]_{j=1}^m\right\|_2 = \sqrt{\sum_{j=1}^m \phi''(t - s_j)^2} \leqslant \sqrt{m \cdot \max_j |\phi''(t - s_j)|^2} \leqslant \frac{2\sqrt{m}}{\sigma^2} \tag{88}$$

and

$$\left|\sum_{j=1}^m \lambda_j^* \phi'''(\xi_j)\right| \leqslant \|\lambda^*\|_2 \left\|\left[\phi'''(\xi_j)\right]_{j=1}^m\right\|_2 \qquad \text{by Cauchy-Schwartz} \tag{89}$$

$$= \|\lambda^*\|_2 \sqrt{\sum_{j=1}^m \phi'''(\xi_j)^2} \tag{90}$$

$$\leqslant \|\lambda^*\|_2 \max_j |\phi'''(\xi_j)| \sqrt{m} \tag{91}$$

$$= c \cdot \frac{\|\lambda^*\|_2 \sqrt{m}}{\sigma^3}, \tag{92}$$

where $c = \frac{4\sqrt{9-3\sqrt{6}}}{e^{\frac{3-\sqrt{6}}{2}}} \approx 3.9036$. By writing

$$q(t) = \sum_{j=1}^m \lambda_j^* \phi(t - s_j) \tag{93}$$

and using the above bounds, we have that

$$\sup_{(t,\lambda)\in V_{\delta_0}} E \leqslant \delta_0 \cdot \frac{\frac{2\sqrt{m}}{\sigma^2} + c \cdot \frac{\|\lambda^*\|_2 \sqrt{m}}{\sigma^3}}{\left|q''(t^*)\right|} \tag{94}$$

Finally, in order to satisfy condition (60), we need to impose the condition that the right hand side of (94) is less than or equal to $\frac{1}{2}$. We select δ_0 to be the largest value that satisfies this, so:

$$|t - t^*| \leqslant \delta_0 = \frac{\left|q''(t^*)\right|}{\frac{4\sqrt{m}}{\sigma^2} + 2c \cdot \frac{\|\lambda^*\|_2 \sqrt{m}}{\sigma^3}} = \frac{\sigma^2 |q''(t^*)|}{\sqrt{m}\left(4 + 2c \cdot \frac{\|\lambda^*\|_2}{\sigma}\right)}. \tag{95}$$

4.1.2 Proof of Lemma 11 (Radius δ_λ of the ball around λ^*)

The radius δ_λ of the perturbation of λ^* is given by:

$$\delta_\lambda = (2M_t B_\lambda)^{-1} \delta_0, \tag{96}$$

where

$$B_\lambda = \sup_{(t,\lambda) \in V_\delta} \|\partial_\lambda F(t, \lambda)\|_2, \tag{97}$$

$$M_t = \left\| \partial_t F(t^*, \lambda^*)^{-1} \right\|_2. \tag{98}$$

For B_λ, we have:

$$\|\partial_\lambda F(t, \lambda)\|_2 = \sqrt{\sum_{j=1}^{m} \phi'(t - s_j)^2} \leqslant \frac{\sqrt{2m}}{\sigma\sqrt{e}}, \tag{99}$$

where we have used the global maximum of the first derivative of the Gaussian from footnote 3, so by taking sup on both sides in the last equation, we obtain:

$$B_\lambda \leqslant \frac{\sqrt{2m}}{\sigma\sqrt{e}}. \tag{100}$$

Note that here we do not use any assumptions on the locations of the sources t_i and the samples s_j. If we did, we would be able to obtain a tighter bound than by only using the absolute maximum of the function.

For M_t, note that we have

$$M_t = |q''(t^*)|^{-1}, \tag{101}$$

where $q(t)$ is defined in (93), so

$$(2M_t B_\lambda)^{-1} \delta_0 \geqslant \frac{\sigma\sqrt{e}|q''(t^*)|}{2\sqrt{2m}} \cdot \delta_0, \tag{102}$$

We then take δ_λ to be equal to the lower bound in the equation above:

$$\delta_\lambda = \frac{\sigma\sqrt{e}|q''(t^*)|}{2\sqrt{2m}} \cdot \delta_0, \tag{103}$$

and, after substituting our choice of δ_0 from (62), we obtain the radius (17) in Theorem 2.

4.1.3 Proof of Proposition 3

Starting from the definition of C_{t^*} in (19), we have that:

$$\begin{aligned} C_{t^*} &= \frac{1}{4 + cR}\left[1 + \frac{2\sqrt{2m}(2 + cR)}{|q''(t^*)|\sqrt{e}}\right] \\ &< \frac{1}{4 + c\|\lambda^*\|_2/\sigma} + \frac{2\sqrt{2m}}{|q''(t^*)|\sqrt{e}} \\ &< \frac{1}{4} + \frac{2\sqrt{2}}{\sqrt{e}} \cdot \frac{\sqrt{m}}{|q''(t^*)|}, \end{aligned} \tag{104}$$

where in the first inequality we used the definition of $R = \frac{\|\lambda^*\|_2}{\sigma}$ and $\frac{2+cR}{4+cR} < 1$ and in the second inequality we used $c\|\lambda^*\|_2/\sigma > 0$, where $c \approx 3.9036$ is a universal constant.

4.2 Proof of Theorem 4 (Dependence of $\|\tilde{\mathbf{a}} - \mathbf{a}^*\|_2$ on $\|\tilde{\mathbf{t}} - \mathbf{t}^*\|_2$)

We apply equation (4.2) in [30], with $e = 0$ (the noise in the observations), and obtain

$$\tilde{\mathbf{a}} = \mathbf{a}^* - \Phi^\dagger E\mathbf{a}^* - F^T E\mathbf{a}^*, \tag{105}$$

where $\Phi^\dagger = (\Phi^T\Phi)^{-1}\Phi^T$ is the pseudo-inverse of Φ and $F = O(E)$ is the perturbation of the $\Phi^\dagger$ due to the perturbation E of Φ, namely

$$\tilde{\Phi}^\dagger = \Phi^\dagger + F^T.$$

In order to obtain an explicit expression for F, we write $\tilde{\Phi}^\dagger$:

$$\begin{aligned}
\tilde{\Phi}^\dagger &= (\tilde{\Phi}^T\tilde{\Phi})^{-1}\tilde{\Phi}^T \\
&= \left[(\Phi + E)^T(\Phi + E)\right]^{-1}(\Phi + E)^T \qquad \text{by (23)} \\
&= (\Phi^T\Phi + \Delta)^{-1}(\Phi^T + E^T),
\end{aligned} \tag{106}$$

where

$$\Delta = E^T\Phi + \Phi^T E + E^T E \in \mathbb{R}^{k\times k}. \tag{107}$$

In order to compute the first factor in (106), consider the QR decomposition of Φ:

$$\Phi = QR, \quad \text{where} \quad Q \in \mathbb{R}^{m\times k} \quad \text{and} \quad R \in \mathbb{R}^{k\times k}, \tag{108}$$

with $Q^TQ = I_k$ R upper triangular. We have that:

$$\Phi^\dagger = R^{-1}Q^T, \tag{109}$$

$$\Phi^T\Phi = R^TR. \tag{110}$$

We then write the first factor in (106) as

$$\begin{aligned}
(\Phi^T\Phi + \Delta)^{-1} &= (R^TR + \Delta)^{-1} \\
&= \left[R^T\left(I + R^{-T}\Delta R^{-1}\right)R\right]^{-1} \\
&= R^{-1}\left[I + \sum_{l=1}^{\infty}(-1)^l\left(R^{-T}\Delta R^{-1}\right)^l\right]R^{-T} \\
&= (R^TR)^{-1} + S_\Phi \\
&= (\Phi^T\Phi)^{-1} + S_\Phi,
\end{aligned} \tag{111}$$

where

$$S_\Phi = R^{-1}\left[\sum_{l=1}^{\infty}(-1)^l\left(R^{-T}\Delta R^{-1}\right)^l\right]R^{-T} \in \mathbb{R}^{k\times k}, \tag{112}$$

and in the second inequality in (111) we applied the Neumann series expansion to the matrix $I - R^{-T}\Delta R^{-1}$, which converges if

$$\| - R^{-T}\Delta R^{-1}\|_2 < 1. \tag{113}$$

We will return to condition (113) at the end of this section. We now substitute (111) in (106), giving

$$\begin{aligned}
\tilde{\Phi}^\dagger &= \left[(\Phi^T\Phi)^{-1} + S_\Phi\right](\Phi^T + E^T) \\
&= \Phi^\dagger + (\Phi^T\Phi)^{-1}E^T + S_\Phi\Phi^T + S_\Phi E^T,
\end{aligned}$$

so we have that

$$F^T = (\Phi^T\Phi)^{-1}E^T + S_\Phi\Phi^T + S_\Phi E^T, \tag{114}$$

which is indeed $O(E)$, since $S_\Phi = O(\Delta)$ and $\Delta = O(E)$. We next upper bound $\|S_\Phi\|_2$. Firstly, note that, because $\Phi^\dagger = \left(QR^{-1}\right)^T$, we have that:[4]

$$\|\Phi^\dagger\|_2 = \|QR^{-1}\|_2 = \|R^{-1}\|_2. \tag{116}$$

Then, by using (116), norm submultiplicativity and triangle inequality, from (112) we have

$$\|S_\Phi\|_2 \leqslant \|\Phi^\dagger\|_2^2 \sum_{l=1}^{\infty} \|\Phi^\dagger\|_2^{2l} \|\Delta\|_2^l. \tag{117}$$

Now let D be an upper bound on $\|\Delta\|_2$, obtained by applying the triangle inequality in (107), so that

$$\|\Delta\|_2 \leqslant D = 2\|E\|_2\|\Phi\|_2 + \|E\|_2^2. \tag{118}$$

Then, from (117) we have

$$\begin{aligned} \|S_\Phi\|_2 &\leqslant \|\Phi^\dagger\|_2^2 \sum_{l=1}^{\infty} \|\Phi^\dagger\|_2^{2l} D^l \\ &= \|\Phi^\dagger\|_2^2 \left(\frac{1}{1 - D\|\Phi^\dagger\|_2^2} - 1 \right) = \frac{D\|\Phi^\dagger\|_2^4}{1 - D\|\Phi^\dagger\|_2^2}, \end{aligned} \tag{119}$$

where the series converges if $D\|\Phi^\dagger\|_2^2 < 1$, in which case the denominator in the last fraction above is positive. We return to this condition at the end of the section. We also know that[5]

$$\|\Phi^\dagger\|_2 = \frac{1}{\sigma_{\min}(\Phi)}. \tag{120}$$

By applying triangle inequality in (114) and then using (119) and the fact that $\|(\Phi^T\Phi)^{-1}\|_2 = 1/\sigma_{\min}^2(\Phi) = \|\Phi^\dagger\|_2^2$ (from (120)), we obtain

$$\|F\|_2 \leqslant \|E\|_2\|\Phi^\dagger\|_2^2 + \frac{D\|\Phi^\dagger\|_2^4}{1 - D\|\Phi^\dagger\|_2^2} \left(\|\Phi\|_2 + \|E\|_2 \right), \tag{121}$$

where D is given in (118). It remains to establish an upper bound on $\|E\|_F$, and consequently on $\|E\|_2$. The following lemma, proved in Section 4.2.1 gives us such a bound.

Lemma 12. ***(Upper bound on*** $\|E\|_F$***)*** *Let* $E = \tilde{\Phi} - \Phi$ *for* Φ *and* $\tilde{\Phi}$ *as defined in* (22) *and* (23) *respectively for* $t_j, \tilde{t}_j \in [0,1]$ *for* $j = 1, \ldots, k$*. Then:*

$$\|E\|_F \leqslant \frac{4e^{\frac{4}{\sigma^2} \max_j |\tilde{t}_j - t_j|} \sqrt{m}}{\sigma^2} \|\tilde{\mathbf{t}} - \mathbf{t}^*\|_2. \tag{122}$$

By using triangle inequality and norm submultiplicativity in (105), and then substituting (121) and (122), we obtain

$$\begin{aligned} \|\mathbf{a}^* - \tilde{\mathbf{a}}\|_2 &\leqslant \|E\|_2\|\Phi^\dagger\|_2\|\mathbf{a}^*\|_2 + \|E\|_2^2\|\Phi^\dagger\|_2^2\|\mathbf{a}^*\|_2 \\ &\quad + \frac{\|E\|_2 D\|\Phi^\dagger\|_2^4}{1 - D\|\Phi^\dagger\|_2^2} (\|\Phi\|_2 + \|E\|_2)\|\mathbf{a}^*\|_2 \\ &\leqslant \frac{4e^{\frac{4}{\sigma^2} \max_j |\tilde{t}_j - t_j|} \sqrt{m} \|\mathbf{a}^*\|_2}{\sigma^2 \sigma_{\min}(\Phi)} \|\tilde{\mathbf{t}} - \mathbf{t}^*\|_2 + O(\|\tilde{\mathbf{t}} - \mathbf{t}^*\|_2^2), \end{aligned} \tag{123}$$

[4]To see the second equality in (116), for a matrix $Q \in \mathbb{R}^{m \times k}$ with $Q^TQ = I$ and any matrix $A \in \mathbb{R}^{k \times k}$ we have that

$$\|QA\|_2 = \sup_{\|v\|_2 = 1} \|QAv\|_2 = \sup_{\|v\|_2 = 1} \|Av\|_2 = \|A\|_2,$$

since

$$\|QAv\|_2^2 = v^TA^TQ^TQAv = v^TA^TAv = \|Av\|_2^2. \tag{115}$$

[5]Using the SVD $\Phi = U\Sigma V^T$, we have $\Phi^\dagger = (\Phi^T\Phi)^{-1}\Phi^T = (V\Sigma^2V^T)^{-1}V\Sigma U^T = V\Sigma^{-1}U^T$, so the conclusion follows.

which is the bound given in Theorem 4. Note that because $\|E\|_2 = O(\|\tilde{\mathbf{t}} - \mathbf{t}^*\|_2)$ (see (122)), the first term is the only term that is $O(\|\tilde{\mathbf{t}} - \mathbf{t}^*\|_2)$ in the first inequality above, so the other terms are included in the $O(\|\tilde{\mathbf{t}} - \mathbf{t}^*\|_2^2)$ term at the end.[6]

Lastly, we return to condition (113), which must be satisfied in order for the bound above to hold. By using norm submultiplicativity and the bound on $\|\Delta\|_2$ from (118), we obtain

$$\|\Phi^{\dagger^T} \Delta \Phi^\dagger\|_2 \leqslant \|\Phi^\dagger\|_2^2 \|E\|_2^2 + 2\|\Phi\|_2 \|\Phi^\dagger\|_2^2 \|E\|_2 \tag{124}$$

and by requiring that the right hand side above is less than one, we obtain a quadratic constraint on $\|E\|_2$, satisfied if

$$\|E\|_2 < \sigma_{\max}(\Phi) \left(\sqrt{1 + \frac{\sigma_{\min}^2(\Phi)}{\sigma_{\max}^2(\Phi)}} - 1 \right).$$

By using the bound on $\|E\|_2$ from (122) with $\max_j |\tilde{t}_j - t_j| \leqslant 1$, the above holds if

$$\delta_t < \frac{\sigma^2 \sigma_{\max}(\Phi)}{4e^{4/\sigma^2}\sqrt{m}} \left(\sqrt{1 + \frac{\sigma_{\min}^2(\Phi)}{\sigma_{\max}^2(\Phi)}} - 1 \right),$$

which is the condition (25) in the statement of the theorem. Note that by imposing this, we also ensure that the condition for the series in (119) to converge holds, since $D\|\Phi^\dagger\|_2^2$ is equal to the right hand side of (124).

4.2.1 Proof of Lemma 12 (Bound of $\|E\|_F$)

Since $E = \tilde{\Phi} - \Phi$, for $\tilde{t}_j$ being a perturbation of t_j, we have that

$$|E_{ij}| = \left| e^{-\frac{(s_i - \tilde{t}_j)^2}{\sigma^2}} - e^{-\frac{(s_i - t_j)^2}{\sigma^2}} \right| = e^{-\frac{(s_i - t_j)^2}{\sigma^2}} \left| e^{\frac{1}{\sigma^2}\left[(s_i - t_j)^2 - (s_i - \tilde{t}_j)^2\right]} - 1 \right|.$$

Then the exponent can be written as

$$\left| \frac{1}{\sigma^2} \left[(s_i - t_j)^2 - (s_i - \tilde{t}_j)^2 \right] \right| = \left| \frac{1}{\sigma^2} \left[2s_i(\tilde{t}_j - t_j) + (t_j + \tilde{t}_j)(t_j - \tilde{t}_j) \right] \right| \leqslant \frac{4|\tilde{t}_j - t_j|}{\sigma^2},$$

where we used that $s_i, \tilde{t}_j, t_j \in [0, 1]$, so

$$e^{-\frac{4}{\sigma^2}|\tilde{t}_j - t_j|} \leqslant e^{\frac{1}{\sigma^2}\left[(s_i - t_j)^2 - (s_i - \tilde{t}_j)^2\right]} \leqslant e^{\frac{4}{\sigma^2}|\tilde{t}_j - t_j|},$$

which implies that

$$\begin{aligned}
|E_{ij}| &\leqslant \left| e^{\frac{1}{\sigma^2}\left[(s_i - t_j)^2 - (s_i - \tilde{t}_j)^2\right]} - 1 \right| \\
&\leqslant \max\left\{ 1 - e^{-\frac{4}{\sigma^2}|\tilde{t}_j - t_j|}, e^{\frac{4}{\sigma^2}|\tilde{t}_j - t_j|} - 1 \right\} \\
&= e^{\frac{4}{\sigma^2}|\tilde{t}_j - t_j|} - 1 \\
&= \frac{4}{\sigma^2}|\tilde{t}_j - t_j| \cdot e^{\xi},
\end{aligned}$$

for some $\xi \in \left[-\frac{4}{\sigma^2}|\tilde{t}_j - t_j|, \frac{4}{\sigma^2}|\tilde{t}_j - t_j| \right]$ and where in the first inequality we have used that $e^{-\frac{(s_i - t_j)^2}{\sigma^2}} \leqslant 1$. Then

$$|E_{ij}| \leqslant \frac{4}{\sigma^2}|\tilde{t}_j - t_j| \cdot e^{\frac{4}{\sigma^2}|\tilde{t}_j - t_j|},$$

[6] In these terms, note that $\max_j |\tilde{t}_j - t_j| \leqslant 1$ and therefore the notation $O(\|\tilde{\mathbf{t}} - \mathbf{t}^*\|_2^2)$ is correct.

and we conclude that

$$\|E\|_2 \leqslant \|E\|_F \leqslant \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{k}\left(\frac{4e^{\frac{4}{\sigma^2}|\tilde{t}_j - t_j|}}{\sigma^2}\right)^2 |\tilde{t}_j - t_j|^2}$$
$$\leqslant \frac{4e^{\frac{4}{\sigma^2}\max_j |\tilde{t}_j - t_j|}\sqrt{m}}{\sigma^2}\|\tilde{\mathbf{t}} - \mathbf{t}^*\|_2, \tag{125}$$

provided that $\tilde{t}_j, t_j \in [0,1]$ for all $j = 1, \ldots, k$.

4.3 Proof of Theorem 8 (Dependence of $\|\lambda - \lambda^*\|_2$ on the noise w)

We apply the quantitative implicit function theorem to the function $\bar{F}$ defined in (39). First, note that in the bound (41), which we want to prove, we only need to consider the k non-fixed entries of λ and λ^*, as the error in the other entries zero. Therefore, in this proof we will only work with the vectors of non-fixed entries $\bar{\lambda}, \bar{\lambda}^* \in \mathbb{R}^k$, but we will abuse the notation for simplicity and write λ and λ^* respectively. Similarly, we will write w to denote the vector $w_\theta \in \mathbb{R}^{2k}$ corresponding to the $2k$ entries of the noise vector $w \in \mathbb{R}^m$ and $s_1, \ldots, s_{2k}$ to denote $s_{\theta_1}, \ldots, s_{\theta_{2k}}$. The partial derivatives of $\bar{F}$ from Definition 7 are:

$$\partial_{\lambda_l}\bar{F}_j = -\sum_{i=1}^{k}\nu_i\phi'(t_i(\lambda) - s_j)\partial_{\lambda_l}t_i(\lambda) \quad l = 1, \ldots, k, \quad j = 1, \ldots, 2k, \tag{126}$$

$$\partial_{\nu_l}\bar{F}_j = -\phi(t_l(\lambda) - s_j), \quad l = 1, \ldots, k, \quad j = 1, \ldots, 2k, \tag{127}$$

$$\partial_{w_l}\bar{F}_j = \begin{cases} 1, & \text{if} \quad l = j, \\ 0, & \text{otherwise,} \end{cases} \qquad l, j = 1, \ldots, 2k. \tag{128}$$

Let $\gamma = [\lambda, \nu]^T$ and $\gamma^* = [\lambda^*, a]^T$, so that we can write $\bar{F}([\lambda, \nu]^T, w)$ as $\bar{F}(\gamma, w)$ and $\bar{F}(\gamma^*, 0) = 0$. In order to apply the implicit function theorem, the following conditions must be satisfied:

1. $\partial_\gamma \bar{F}(\gamma^*, 0)$ is invertible,
2. We choose the radius δ_γ of the ball V_{δ_γ} around γ where the result of the quantitative implicit function theorem holds:
$$\sup_{(\gamma,w)\in V_{\delta_\gamma}} \left\| I - \left[\partial_\gamma \bar{F}(\gamma^*, 0)\right]^{-1}\partial_\gamma \bar{F}(\gamma, w)\right\|_2 \leqslant \frac{1}{2}, \tag{129}$$
3. The radius δ_w of the ball around $w^* = 0$ that contains w is:
$$\delta_w = (2M_w B_{\delta_\gamma})^{-1}\delta_\gamma, \tag{130}$$
where
$$B_{\delta_\gamma} = \sup_{(\gamma,w)\in V_{\delta_\gamma}} \|\partial_w \bar{F}(\gamma, w)\|_2, \tag{131}$$
$$M_w = \|\partial_\gamma \bar{F}(\gamma^*, 0)^{-1}\|_2. \tag{132}$$

The first condition is also one of the conditions in the theorem, and it has been discussed in Section 3.1. We now need to establish the two radii for the balls of the perturbations.

Perturbation radii

Before proceeding to calculating the radii of the balls where the implicit function theorem holds, we state the following lemma, which allows us to write the Jacobian of $\bar{F}$ with respect to the first variable as a sum of the Jacobian evaluated at $(\gamma^*, w^*) = ([\lambda^*, a]^T, 0)$ and a perturbation matrix, whose norm is bounded explicitly. The proof of Lemma 13 is given in Section 4.3.2.

Lemma 13. ***(Bound on the perturbation of the Jacobian of $\bar{F}$)*** *Let $J(\lambda, \nu, w)$ be the Jacobian of $\bar{F}(\gamma, w)$ with respect to $\gamma = [\lambda, \nu]^T$ and δ_γ an upper bound on the perturbation of $\gamma^* = [\lambda^*, a]^T$, namely:*

$$\left\| \begin{bmatrix} \lambda - \lambda^* \\ \nu - a \end{bmatrix} \right\|_2 \leqslant \delta_\gamma.$$

Then:

$$J(\lambda, \nu, w) = J(\lambda^*, a, 0) + E, \tag{133}$$

with

$$\|E\|_F \leqslant P(k, \sigma, \Pi, \tau, C_{t*}) \cdot \delta_\gamma, \tag{134}$$

where:

$$\begin{aligned} P(k, \sigma, \Pi, \tau, C_{t*}) = \sqrt{2}k \Bigg[& \frac{1}{\sigma^2} \left(2\sqrt{k} C_{t*}^2 \Pi + 4k C_{t*} \bar{\Delta}_2 \tau \Pi + 2 C_{t*} \right) \\ & + \frac{1}{\sigma} \left(\frac{\sqrt{2k} C_{t*}}{\sqrt{e}} + 4\sqrt{k} C_{t*}^2 \Pi + \frac{2\sqrt{2} \bar{\Delta}_2 \Pi}{\sqrt{e}} + 8k C_{t*} \bar{\Delta}_2 \tau \Pi + \frac{\sqrt{2k} \bar{\Delta}_2 \Pi}{\sqrt{e}} \right) \Bigg], \end{aligned} \tag{135}$$

for $\|\lambda - \lambda^*\| \leqslant \delta_\lambda$, where δ_λ and C_{t*} are given in (17) *and* (19) *respectively in Theorem 2 and $\bar{\Delta}_2$ is given in* (176) *in the proof.*

We can now use Lemma 13 to write

$$\partial_\gamma \bar{F}(\gamma, w) = \partial_\gamma \bar{F}(\gamma^*, 0) + E, \tag{136}$$

then

$$\begin{aligned} I - \left[\partial_\gamma \bar{F}(\gamma^*, 0)\right]^{-1} \partial_\gamma \bar{F}(\gamma, w) &= I - \left[\partial_\gamma \bar{F}(\gamma^*, 0)\right]^{-1} \left[\partial_\gamma \bar{F}(\gamma^*, 0) + E\right] \\ &= - \left[\partial_\gamma \bar{F}(\gamma^*, 0)\right]^{-1} E, \end{aligned} \tag{137}$$

so

$$\begin{aligned} \left\| I - \left[\partial_\gamma \bar{F}(\gamma^*, 0)\right]^{-1} \partial_\gamma \bar{F}(\gamma, w) \right\|_2 &\leqslant \left\| \left[\partial_\gamma \bar{F}(\gamma^*, 0)\right]^{-1} \right\|_2 \cdot \|E\|_F \\ &\leqslant \frac{\|E\|_F}{\sigma_{\min}(\partial_\gamma \bar{F}(\gamma^*, 0))} \\ &\leqslant \frac{P(k, \sigma, \Pi, \tau, C_{t*})}{\sigma_{\min}(\partial_\gamma \bar{F}(\gamma^*, 0))} \cdot \delta_\gamma, \end{aligned} \tag{138}$$

where $P \cdot \delta_\gamma$ is the upper bound on $\|E\|_F$ given in (135).

Therefore, from the condition that the right-hand side of the last inequality is less than or equal to $\frac{1}{2}$, we choose the radius δ_γ to be:

$$\delta_\gamma = \frac{\sigma_{\min}(\partial_\gamma \bar{F}(\gamma^*, 0))}{2P(k, \sigma, \Pi, \tau, C_{t*})}. \tag{139}$$

Using (128), we have that

$$B_{\delta_\gamma} = 1. \tag{140}$$

Then

$$M_w = \frac{1}{\sigma_{\min}(\partial_\gamma \bar{F}(\gamma^*, 0))}, \tag{141}$$

so, using (139), we obtain

$$\delta_w = \frac{\sigma_{\min}(\partial_\gamma \bar{F}(\gamma^*, 0))^2}{4P(k, \sigma, \Pi, \tau, C_{t*})}. \tag{142}$$

Applying the quantitative implicit function theorem

Having calculated the radii where the quantitative implicit function theorem holds, we apply it to obtain:

$$\partial_w g(w) = -\left[\partial_1 \bar{F}(g(w), w)\right]^{-1}, \tag{143}$$

where ∂_1 is the partial derivative with respect with the first argument and $g(w)$ gives the dependence of $[\lambda, \nu]^T$ on w. Specifically, we write:

$$\lambda_i(w) = g_i(w) \quad \text{for} \quad i = 1, \ldots, k, \tag{144}$$

$$\nu_i(w) = g_{k+i}(w) \quad \text{for} \quad i = 1, \ldots, k. \tag{145}$$

Let $J(\lambda, \nu, w) = \partial_1 \bar{F}([\lambda, \nu]^T, w)$, where $\lambda = \lambda(w)$ and $\nu = \nu(w)$ by (144) and (145). Lemma 13 gives

$$J(\lambda, \nu, w) = J(\lambda^*, a, 0) + E, \tag{146}$$

so E is the perturbation of $J(\lambda^*, a, 0)$ due to perturbed λ, ν, w and a bound on $\|E\|_F$ is given in the lemma.

We will now use the following result, proved in Section 4.3.1, which enables us to make use of the upper bound on the norm of the perturbation given by Lemma 13 in order to lower bound the smallest singular value of J.

Lemma 14. *Let $J \in \mathbb{R}^{m \times n}$. If $J = A + E$, then*

$$\sigma_{\min}(J) \geqslant \sigma_{\min}(A) - \|E\|_F.$$

By applying Lemma 14, we have that:

$$\begin{aligned} \frac{1}{\sigma_{\min}(J(\lambda, \nu, w))} &\leqslant \frac{1}{\sigma_{\min}(J(\lambda^*, a, 0)) - \|E\|_F} = \frac{1}{\sigma_{\min}(J(\lambda^*, a, 0)) \left(1 - \frac{\|E\|_F}{\sigma_{\min}(J(\lambda^*, a, 0))}\right)} \\ &\leqslant \frac{1}{\sigma_{\min}(J(\lambda^*, a, 0))} \cdot \left(1 + \frac{2\|E\|_F}{\sigma_{\min}(J(\lambda^*, a, 0))}\right), \end{aligned} \tag{147}$$

for

$$\frac{\|E\|_F}{\sigma_{\min}(J(\lambda^*, a, 0))} \leqslant \frac{1}{2}, \tag{148}$$

where we used the fact that $(1 - x)^{-1} \leqslant 1 + 2x$ for $x \in [0, \frac{1}{2}]$. Note that the condition above is the same as the condition that the right hand side of (138) is less than or equal to $\frac{1}{2}$, which is satisfied for our choice of δ_γ and δ_w.

From (143) and (147), we have that:

$$\begin{aligned} \|\partial_w g(w)\|_2 &= \frac{1}{\sigma_{\min}(J(\lambda^*, a, 0) + E)} \\ &\leqslant \frac{1}{\sigma_{\min}(J(\lambda^*, a, 0))} \left(1 + \frac{2}{\sigma_{\min}(J(\lambda^*, a, 0))} \cdot \|E\|_F\right) \\ &\leqslant \frac{2}{\sigma_{\min}(J(\lambda^*, a, 0))}, \end{aligned} \tag{149}$$

where $\|E\|_F$ is upper bounded in (135) and w, λ and ν satisfy

$$\|w\|_2 \leqslant \delta_w, \quad \|\lambda - \lambda^*\|_2 \leqslant \delta_\gamma, \quad \|\nu - a\|_2 \leqslant \delta_\gamma.$$

The first-order Taylor expansion of $g(w)$ around $w = 0$ is:

$$g(w) = g(0) + \partial_w g(w_\delta)^T w, \tag{150}$$

for some w_δ on the segment between the zero vector and w. Noting that $g(w)$ is our notation for the vector:

$$g(w) = \begin{bmatrix} \lambda(w) \\ \nu(w) \end{bmatrix}, \tag{151}$$

with $\lambda(0) = \lambda^*$ and $\nu(0) = a$, from (150) we have that:

$$\left\| \begin{bmatrix} \lambda(w) - \lambda^* \\ \nu(w) - a \end{bmatrix} \right\|_2 = \left\| \partial_w g(w_\delta)^T w \right\|_2 \leqslant \|\partial_w g(w_\delta)\|_2 \cdot \|w\|_2, \tag{152}$$

for w, λ and ν such that

$$\|w\|_2 \leqslant \delta_w, \quad \|\lambda - \lambda^*\|_2 \leqslant \delta_\gamma, \quad \|\nu - a\|_2 \leqslant \delta_\gamma,$$

where we use the bound from (149).

4.3.1 Proof of Lemma 14

We have that

$$\begin{aligned} \sigma(J) &= \min_{\|v\|_2=1} \max_{\|u\|_2=1} u^T(A+E)v \\ &\geqslant \min_{\|v\|_2=1} \max_{\|u\|_2=1} u^T A v - \max_{\|v\|_2=1} \max_{\|u\|_2=1} u^T E v \\ &\geqslant \sigma_{\min}(A) - \|E\|_F. \end{aligned}$$

4.3.2 Proof of Lemma 13 (Bound on the perturbation of the Jacobian of $\bar{F}$)

Let $J(\lambda, \nu, w) = \partial_1 \bar{F}([\lambda, \nu]^T, w)$, where $\lambda = \lambda(w)$ and $\nu = \nu(w)$ by (144) and (145), and we want to write J in the form

$$J(\lambda, \nu, w) = J(\lambda^*, a, 0) + E \tag{153}$$

i.e. E is the perturbation of $J(\lambda^*, a, 0)$ due to perturbed λ, ν, w. In order to apply Lemma 14, we need an upper bound on $\|E\|_F$, so we need to upper bound each entry of E. Let

$$J = [J_1 J_2], \tag{154}$$

where J_1 corresponds to the terms (126) and J_2 to the terms (127) and

$$E = [E_1 E_2] \tag{155}$$

the corresponding perturbation terms.

Entries in J_1

For $i = 1, \ldots, k$ and $j = 1, \ldots, 2k$:

$$\begin{aligned} J_{1_{j,i}} &= -\sum_{p=1}^{k} (\nu_p - a_p + a_p)\phi'(t_p(\lambda) - t_p^* + t_p^* - s_j)\partial_{\lambda_i} t_p(\lambda) \\ &= -\sum_{p=1}^{k} \partial_{\lambda_i} t_p(\lambda) \left[a_p \phi'(t_p^* - s_j + t_p(\lambda) - t_p^*) + (\nu_p - a_p)\phi'(t_p^* - s_j + t_p(\lambda) - t_p^*) \right] \\ &= -\sum_{p=1}^{k} \partial_{\lambda_i} t_p(\lambda) \big[a_p \phi'(t_p^* - s_j) + a_p(t_p(\lambda) - t_p^*)\phi''(\xi_{j,p}) + (\nu_p - a_p)\phi'(t_p^* - s_j) \\ &\quad + (\nu_p - a_p)(t_p(\lambda) - t_p^*)\phi''(\xi_{j,p}) \big] \\ &= -\sum_{p=1}^{k} \partial_{\lambda_i} t_p(\lambda) \left(a_p \phi'(t_p^* - s_j) + \Delta_{1_{j,p}} \right), \end{aligned} \tag{156}$$

where

$$\Delta_{1_{j,p}} = a_p(t_p(\lambda) - t_p^*)\phi''(\xi_{j,p}) + (\nu_p - a_p)\phi'(t_p^* - s_j) + (\nu_p - a_p)(t_p(\lambda) - t_p^*)\phi''(\xi_{j,p}), \tag{157}$$

for some $\xi_{j,p} \in [t_p^* - s_j - |t_p - t_p^*|, t_p^* - s_j + |t_p - t_p^*|]$. The factor involving the partial derivative in (156) has the same form as (59) so in order to bound it we write the Taylor expansion of (59) around λ^*:

$$\partial_\lambda t_p(\lambda) = \partial_\lambda t_p(\lambda^*) + \partial^2_{\lambda\lambda} t_p(\lambda_\delta)\left(\lambda - \lambda^*\right), \tag{158}$$

for some λ_δ on the segment between λ and λ^*. By using (59) with (63) and (64), the entry i, l in the Hessian matrix $H = \partial^2_{\lambda\lambda} t_p(\lambda_\delta)$ is

$$\left(H^T\right)_{i,l} = \frac{F_{i,l}(\lambda)}{\left[\sum_{j=1}^m \lambda_j \phi''(t_p(\lambda) - s_j)\right]^2}, \tag{159}$$

for $i, l = 1, \ldots, k$, where

$$\begin{aligned} F_{i,l}(\lambda) = &- \phi''(t_p(\lambda) - s_i)\partial_{\lambda_l} t_p(\lambda) \sum_{j=1}^m \lambda_j \phi''(t_p(\lambda) - s_j) \\ &+ \phi'(t_p(\lambda) - s_i)\left(\sum_{j=1}^m \lambda_j \phi'''(t_p(\lambda) - s_j)\partial_{\lambda_l} t_p(\lambda) + \phi''(t_p(\lambda) - s_l)\right). \end{aligned} \tag{160}$$

Note that in the denominator (159) we use all m entries of λ and samples due to how we defined the function from (59), and the same is true for the sums in (160). From (158) and (160), we then write:

$$\partial_{\lambda_i} t_p(\lambda) = \partial_{\lambda_i} t_p(\lambda^*) + \Delta_{2_{i,p}}, \tag{161}$$

where

$$\Delta_{2_{i,p}} = \sum_{l=1}^k \frac{(\lambda_l - \lambda_l^*) F_{i,l}(\lambda_\delta)}{\left[\sum_{j=1}^m \lambda_j \phi''(t_p(\lambda_\delta) - s_j)\right]^2}. \tag{162}$$

Note that l goes up to k because we only work with k entries in λ. Therefore, we have that:

$$J_{1_{j,i}} = -\sum_{p=1}^k \left(\partial_{\lambda_i} t_p(\lambda^*) + \Delta_{2_{i,p}}\right)\left(a_p \phi'(t_p^* - s_j) + \Delta_{1_{j,p}}\right), \tag{163}$$

where

$$\Delta_{1_{j,p}} = O(|t_p - t_p^*| + |\nu_p - a_p|), \tag{164}$$

$$\Delta_{2_{i,p}} = O(\|\lambda - \lambda^*\|_2), \tag{165}$$

for $i = 1, \ldots, k$, $j = 1, \ldots, 2k$ and $p = 1, \ldots, k$. The next step now is to upper bound $|\Delta_{1_{j,p}}|$ and $|\Delta_{2_{i,p}}|$.

Bounding $\Delta_{1_{j,p}}$

By the triangle inequality, we have that:

$$\begin{aligned} |\Delta_{1_{j,p}}| &\leqslant |a_p||t_p(\lambda) - t_p^*||\phi''(\xi_{j,p})| + |\nu_p - a_p||\phi'(t_p^* - s_j)| + |\nu_p - a_p||t_p(\lambda) - t_p^*||\phi''(\xi_{j,p})| \\ &\leqslant |a_p||t_p(\lambda) - t_p^*|\frac{2}{\sigma^2} + |\nu_p - a_p|\frac{\sqrt{2}}{\sqrt{e}\sigma} + |\nu_p - a_p||t_p(\lambda) - t_p^*|\frac{2}{\sigma^2} =: \bar{\Delta}_{1_p}, \end{aligned} \tag{166}$$

for $j = 1, \ldots, 2k$ and $p = 1, \ldots, k$, where we have used the maxima of the Gaussian and its derivatives given in footnote 3.

Bounding $\Delta_{2_{i,p}}$

By applying the Cauchy-Schwartz inequality, we have that:

$$|\Delta_{2_{i,p}}| \leqslant \frac{1}{\left|\sum_{j=1}^{m} \lambda_j \phi''(t_p(\lambda_\delta) - s_j)\right|^2} \cdot \left\| \begin{bmatrix} F_{i,1}(\lambda_\delta) \\ \vdots \\ F_{i,k}(\lambda_\delta) \end{bmatrix} \right\|_2 \cdot \|\lambda - \lambda^*\|_2 \tag{167}$$

We now bound $|F_{i,l}|$ for $i, l = 1, \ldots, k$:

$$\begin{aligned}
|F_{i,l}(\lambda_\delta)| &\leqslant |\phi''(t_p(\lambda_\delta) - s_i)| \left|\partial_{\lambda_l} t_p(\lambda_\delta)\right| \left| \sum_{j=1}^{m} \lambda_j \phi''(t_p(\lambda_\delta) - s_j) \right| \\
&\quad + |\phi'(t_p(\lambda_\delta) - s_i)| \left(\left|\partial_{\lambda_l} t_p(\lambda_\delta)\right| \left| \sum_{j=1}^{m} \lambda_j \phi'''(t_p(\lambda_\delta) - s_j) \right| + |\phi''(t_p(\lambda_\delta) - s_l)| \right) \\
&\leqslant \frac{2C_{t*}}{\sigma^2} \|\lambda_\delta\|_2 \cdot \left\| \left[\phi''(t_p(\lambda_\delta) - s_j)\right]_{j=1}^{m} \right\|_2 \\
&\quad + \frac{\sqrt{2}}{\sqrt{e}\sigma} \left(C_{t*} \|\lambda_\delta\|_2 \cdot \left\| \left[\phi'''(t_p(\lambda_\delta) - s_j)\right]_{j=1}^{m} \right\|_2 + \frac{2}{\sigma^2} \right) \\
&\leqslant \frac{2C_{t*}}{\sigma^2} \|\lambda_\delta\|_2 \cdot \frac{2\sqrt{m}}{\sigma^2} + \frac{\sqrt{2}}{\sqrt{e}\sigma} \left(C_{t*} \|\lambda_\delta\|_2 \cdot \frac{c\sqrt{m}}{\sigma^3} + \frac{2}{\sigma^2} \right),
\end{aligned} \tag{168}$$

where we used the Cauchy-Schwartz inequality, the bounds in footnote 3 and C_{t*} from (19). Therefore, the above inequality holds for $\lambda_\delta \in \mathcal{B}(\lambda^*, \delta_\lambda)$ with δ_λ from (17).

The final bound on $|F_{i,l}|$ is

$$|F_{i,l}| \leqslant \frac{c_2 C_{t*} m\tau}{\sigma^4} + \frac{2\sqrt{2}}{\sqrt{e}\sigma^3}, \tag{169}$$

where $c_2 = 4 + \frac{c\sqrt{2}}{\sqrt{e}} \approx 7.3484$, for $i, l = 1, \ldots, k$ and we used $\|\lambda_\delta\|_2 \leqslant \tau\sqrt{m}$.

The next step is to obtain a lower bound on the denominator in (167). By adding and subtracting λ_j^* to λ_j and applying the reverse triangle inequality, we obtain:

$$\left| \sum_{j=1}^{m} \lambda_j \phi''(t_p(\lambda) - s_j) \right| \geqslant \left| \sum_{j=1}^{m} \lambda_j^* \phi''(t_p(\lambda) - s_j) \right| - \left| \sum_{j=1}^{m} (\lambda_j - \lambda_j^*) \phi''(t_p(\lambda) - s_j) \right| \tag{170}$$

$$\geqslant |q''(t^*)| \left[1 - \frac{c\|\lambda^*\|_2}{4\sigma + 2c\|\lambda^*\|_2} \right] - \frac{2\sqrt{m}\|\lambda - \lambda^*\|_2}{\sigma^2}, \tag{171}$$

where the first term on the right hand side on (170) has the same form as B in (70), and therefore on the next line we use the bound in (78). For the second term, we apply the Cauchy-Schwartz inequality and the bound in footnote 3, where the constant $c \approx 3.9036$ is obtained. The last inequality above holds under the condition that the right hand side is positive. We set the stronger condition that the right hand side of (171) is greater than or equal to one:

$$\left| \sum_{j=1}^{m} \lambda_j \phi''(t_p(\lambda) - s_j) \right| \geqslant 1, \tag{172}$$

which is satisfied if:

$$|q''(t^*)| \geqslant \left(1 + \frac{2\sqrt{m}\|\lambda - \lambda^*\|_2}{\sigma^2} \right) \cdot \frac{4\sigma + 2c\|\lambda^*\|_2}{4\sigma + c\|\lambda^*\|_2}. \tag{173}$$

Then, using the box constraint $\|\lambda\|_\infty \leqslant \tau$ and the fact that $\|\lambda - \lambda^*\|_2 \leqslant 2\tau\sqrt{m}$ from the triangle inequality, and the fact that $\frac{4\sigma+2c\|\lambda^*\|_2}{4\sigma+c\|\lambda^*\|_2} \leqslant 2$ the condition (173) is satisfied if:

$$|q''(t^*)| \geqslant 2\left(1 + \frac{4m\tau}{\sigma^2}\right). \tag{174}$$

By combining (167), (169) and (172), we obtain the final bound on $|\Delta_{2_{i,p}}|$:

$$|\Delta_{2_{i,p}}| \leqslant \bar{\Delta}_2 \cdot \|\lambda - \lambda^*\|_2, \tag{175}$$

for $i = 1, \ldots, k$ and $p = 1, \ldots, k$, where

$$\bar{\Delta}_2 = \frac{\sqrt{k}}{\sigma^4}\left(c_2 C_{t^*} m\tau + \frac{2\sqrt{2}}{\sqrt{e}}\sigma\right), \tag{176}$$

and $c \approx 3.9036$, $c_2 = 4 + \frac{c\sqrt{2}}{\sqrt{e}} \approx 7.3484$.

Therefore, from (163) and by using the definitions of $\bar{\Delta}_{1_p}$ and $\bar{\Delta}_2$ from (166) and (176) respectively, we have that:

$$\begin{aligned}
|E_{1_{j,i}}| &= \left|\sum_{p=1}^{k} \partial_{\lambda_i} t_p(\lambda^*)\Delta_{1_{j,p}} + a_p\phi'(t_p^* - s_j)\Delta_{2_{i,p}} + \Delta_{1_{j,p}}\Delta_{2_{i,p}}\right| \\
&\leqslant C_t \sum_{p=1}^{k} \bar{\Delta}_{1_p} + \|\lambda - \lambda^*\|_2\bar{\Delta}_2\|a\|_2 \cdot \left\|\left[\phi'(t_p^* - s_j)\right]_{p=1}^{k}\right\|_2 + \|\lambda - \lambda^*\|_2\bar{\Delta}_2 \sum_{p=1}^{k} \bar{\Delta}_{1_p} \\
&\leqslant (C_t + \|\lambda - \lambda^*\|_2\bar{\Delta}_2)\Big(\frac{2}{\sigma^2}\|a\|_2\|t(\lambda) - t^*\|_2 + \frac{\sqrt{2}}{\sqrt{e}\sigma}\|\nu - a\|_1 \\
&\quad + \frac{2}{\sigma}\|\nu - a\|_2\|t(\lambda) - t^*\|_2\Big) + \frac{\sqrt{2k}}{\sqrt{e}\sigma}\|a\|_2\bar{\Delta}_2 \\
&\leqslant (C_t + \|\lambda - \lambda^*\|_2\bar{\Delta}_2)\Big(\frac{2\sqrt{k}C_{t^*}}{\sigma^2}\|a\|_2\|\lambda - \lambda^*\|_2 + \frac{\sqrt{2}}{\sqrt{e}\sigma}\|\nu - a\|_1 \\
&\quad + \frac{2\sqrt{k}C_{t^*}}{\sigma}\|\nu - a\|_2\|\lambda - \lambda^*\|_2\Big) + \frac{\sqrt{2k}}{\sqrt{e}\sigma}\|a\|_2\|\lambda - \lambda^*\|_2\bar{\Delta}_2 =: \bar{E}_1
\end{aligned} \tag{177}$$

for $i = 1, \ldots, k$ and $j = 1, \ldots, 2k$.

Entries in J_2

By adding and subtracting t_j^* then taking a Taylor expansion like before, we obtain:

$$\begin{aligned}
J_{2_{j,i}} &= -\phi(t_i^* - s_j + t_i(\lambda) - t_i^*) \\
&= -\phi(t_i^* - s_j) - (t_i(\lambda) - t_i^*)\phi'(\xi_j) \\
&= -\phi(t_i^* - s_j) - E_{2_{i,j}},
\end{aligned} \tag{178}$$

for some $\xi_j \in [t_i^* - s_j - |t_i(\lambda) - t_i^*|, t_i^* - s_j + |t_i(\lambda) - t_i^*|]$ and $E_{2_{j,i}}$ is the perturbation term. Then:

$$|E_{2_{j,i}}| \leqslant |t_i(\lambda) - t_i^*| \cdot \frac{\sqrt{2}}{\sigma\sqrt{e}}, \tag{179}$$

for $i = 1, \ldots, k$ and $j = 1, \ldots, 2k$.

Putting everything together

We have that

$$\begin{aligned}\|E\|_F &= \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{2k} E_{1_{j,i}}^2 + \sum_{i=1}^{k}\sum_{j=1}^{2k} E_{2_{j,i}}^2} \\ &\leqslant \sqrt{2k^2\bar{E}_1^2 + \frac{4k}{\sigma^2 e}\sum_{i=1}^{k}|t_i(\lambda) - t_i^*|^2} \\ &\leqslant k\bar{E}_1\sqrt{2} + \frac{2\sqrt{k}}{\sigma\sqrt{e}}\|t(\lambda) - t^*\|_2 \\ &\leqslant k\bar{E}_1\sqrt{2} + \frac{2C_{t^*}k}{\sigma\sqrt{e}}\|\lambda - \lambda^*\|_2, \end{aligned} \tag{180}$$

where we have used the bounds on the entries of E_1 and E_2 from (177) and (179) and Theorem 2, so this result holds for $\lambda \in \mathcal{B}(\lambda^*, \delta_\lambda)$ for δ_λ defined in the theorem. Finally, by substituting the expression of $\bar{E}_1$ from (177), we obtain:

$$\begin{aligned}\|E\|_F \leqslant \sqrt{2}k\Bigg[&\left(C_{t^*} + \|\lambda - \lambda^*\|_2\bar{\Delta}_2\right)\left(\frac{2C_{t^*}\sqrt{k}}{\sigma^2}\|a\|_2\|\lambda - \lambda^*\|_2\right. \\ &\left. + \frac{\sqrt{2}}{\sqrt{e}\sigma}\|\nu - a\|_1 + \frac{2C_{t^*}\sqrt{k}}{\sigma}\|\nu - a\|_2\|\lambda - \lambda^*\|_2\right) + \frac{\sqrt{2k}}{\sqrt{e}\sigma}\|a\|_2\|\lambda - \lambda^*\|_2\bar{\Delta}_2\Bigg] \\ &+ \frac{2C_{t^*}k}{\sigma\sqrt{e}}\|\lambda - \lambda^*\|_2. \end{aligned} \tag{181}$$

Let δ_γ be a bound on the perturbation:

$$\left\|\begin{bmatrix}\lambda - \lambda^* \\ \nu - a\end{bmatrix}\right\|_2 \leqslant \delta_\gamma, \tag{182}$$

and therefore:

$$\|\lambda - \lambda^*\|_2 \leqslant \delta_\gamma \qquad \text{and} \qquad \|\nu - a\|_2 \leqslant \delta_\gamma. \tag{183}$$

We also have that:

$$\|\nu - a\|_2 \leqslant \|\nu - a\|_1 \leqslant \|\nu\|_1 + \|a\|_1 \leqslant 2\Pi, \tag{184}$$

where we used that $\nu_1 + \ldots + \nu_k \leqslant \Pi$ and the fact that $x = \sum_{p=1}^{k} a_p\delta_{t_p}$ is the solution to (9), so it satisfies $\|x\|_{TV} = \|a\|_1 \leqslant \Pi$.

Similarly, we have that:

$$\begin{aligned}\|\lambda - \lambda^*\|_2 &\leqslant \|\lambda\|_2 + \|\lambda^*\|_2 \leqslant \sqrt{k}\|\lambda\|_\infty + \sqrt{k}\|\lambda^*\|_\infty \\ &\leqslant 2\sqrt{k}\tau, \end{aligned} \tag{185}$$

since both λ and λ^* satisfy the constraint in (27). In order to write the bound (181) as $P \cdot \delta_\gamma$, we expand the parentheses and use the following bounds:

$$\|\lambda - \lambda^*\|_2\|\nu - a\|_1 \leqslant 2\Pi \cdot \delta_\gamma \tag{186}$$

$$\|\lambda - \lambda^*\|_2^2\|\nu - a\|_2 \leqslant 4\sqrt{k}\tau\Pi \cdot \delta_\gamma \tag{187}$$

$$\|\lambda - \lambda^*\|_2\|\nu - a\|_2 \leqslant 2\Pi \cdot \delta_\gamma \tag{188}$$

to obtain:

$$\|E\|_F \leqslant \sqrt{2}k\Bigg(\frac{2\sqrt{k}C_{t*}^2\Pi}{\sigma^2}+\frac{\sqrt{2k}C_{t*}}{\sqrt{e}\sigma}+\frac{4\sqrt{k}C_{t*}^2\Pi}{\sigma} + \frac{4kC_{t*}\bar{\Delta}_2\tau\Pi}{\sigma^2}+\frac{2\sqrt{2}\bar{\Delta}_2\Pi}{\sqrt{e}\sigma}+\frac{8kC_{t*}\bar{\Delta}_2\tau\Pi}{\sigma}+\frac{\sqrt{2k}\bar{\Delta}_2\Pi}{\sqrt{e}\sigma} + \frac{\sqrt{2}C_{t*}}{\sigma\sqrt{e}}\Bigg)\cdot\delta_\gamma, \tag{189}$$

which we rearrange based on σ to obtain $\|E\|_F \leqslant P(k,\sigma,\Pi,\tau,C_{t*})\cdot\delta_\gamma$, where:

$$P(k,\sigma,\Pi,\tau,C_{t*}) = \sqrt{2}k\Bigg[\frac{1}{\sigma^2}\left(2\sqrt{k}C_{t*}^2\Pi+4kC_{t*}\bar{\Delta}_2\tau\Pi\right) + \frac{1}{\sigma}\left(\frac{\sqrt{2k}C_{t*}}{\sqrt{e}}+4\sqrt{k}C_{t*}^2\Pi+\frac{2\sqrt{2}\bar{\Delta}_2\Pi}{\sqrt{e}}+8kC_{t*}\bar{\Delta}_2\tau\Pi+\frac{\sqrt{2k}\bar{\Delta}_2\Pi}{\sqrt{e}}+\sqrt{\frac{2}{e}}C_{t*}\right)\Bigg],$$

which is the final bound in (135).

5 Numerical experiments

In this section, we present numerical experiments which verify the bounds given by our main results, Theorem 2, Theorem 4 and Theorem 8. To do this, we take an example of a source and sample configuration and a Gaussian kernel for a given σ and solve the exact penalty formulation (28) of the dual problem (27) using the level method [16], given in Appendix B. We introduce inaccuracies in λ by stopping the algorithm early and show how these perturbations affect the source locations and weights. Next, we add noise to the measurements to show how λ is affected. We are, therefore, able to compare the ratios of the perturbations obtained numerically with the constants in the theorems to show the validity of our results in practice. The specific details are discussed in the next subsections.

Setup

We place three sources at locations $t_i^* \in T = \{0.25, 0.63, 0.889\}$ with weights $a_i^* \in \{0.8, 0.5, 0.9\}$ and $m = 21$ equispaced samples in $[0,1]$, with a Gaussian kernel $\phi(t) = e^{-t^2/\sigma^2}$ with $\sigma = 0.07$. We show this configuration in Figure 2.

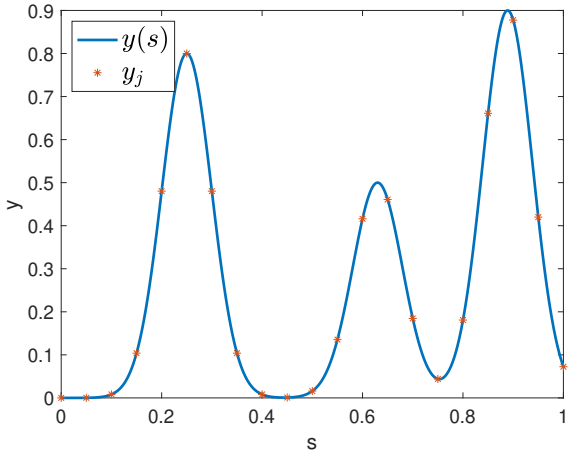

Figure 2: The source-sample configuration used for numerical experiments in the current section.

Effect of λ^* perturbations on t^*

We then solve the dual problem (27) in the exact penalty formulation (28) with box constraint parameter $\tau = 10^5$ and penalty parameter $\Pi = 100$ and run it for $P = 500$ iterations. This gives an accuracy in the source locations of $|t_i - t_i^*| \leqslant 10^{-8}$ for $t_i^* \in T$.

While it is possible to optimise the parameters τ, Π and P in order to obtain better accuracy in the source locations t_i and weights a_i, it is not the aim of this section. Note that Theorem 2 gives the result (18) in the form

$$|t_i - t_i^*| \leqslant C_{t*} \|\lambda - \lambda^*\|_2,$$

where $t_i^* \in T$ is an arbitrary true source location, λ^* is the solution to the dual problem (27)[7] and t is obtained by perturbing t^* as a consequence of the perturbation λ^* in λ.

One way of showing that a relationship of the type of (18) holds in practice is to plot the ratio $\frac{|t_i^{(p)} - t_i^*|}{\|\lambda^{(p)} - \lambda^*\|_2}$, for $p = p_0, \ldots, P$ and $i = 1, \ldots, k$, where P is the number of iterations the level method is run for, p is the index of each iteration and $t_i^{(p)}$ and $\lambda^{(p)}$ are the values of t_i and λ obtained at iteration p, where $p_0 \geqslant 1$ is large enough so that $\|\lambda^{(p)} - \lambda^*\|_2$ satisfies the condition in Theorem 2. The level method computes the value $\lambda^{(p)}$ after p iterations and $\{t_i^{(p)}\}_{i=1}^k$ are obtained by calculating the global maxima of the dual certificate $q^{(p)}(s) = \sum_{j=1}^m \lambda_j^{(p)} \phi(s - s_j)$. Since we know the true value of t_i^*, we can find $t_i^{(p)}$ by running a local optimisation algorithm with t_i^* as the initial condition. For a large enough value of p, this will give an accurate value of $t_i^{(p)}$ and we can, therefore, calculate $|t_i^{(p)} - t_i^*|$ for each $p = p_0, \ldots, P$ and $t_i^* \in T$. Then we check that:

$$\frac{|t_i^{(p)} - t_i^*|}{\|\lambda^{(p)} - \lambda^*\|_2} \leqslant C_{t*}, \tag{190}$$

for $p = p_0, \ldots, P$ and $i = 1, \ldots, k$. One issue is that the true value of λ^* is not known. The best estimate we have is $\lambda_{best}^* = \lambda^{(P)}$, namely the value of λ^* given by the level method after P iterations. Therefore, the result of Theorem 2 cannot be verified directly in practice, but must be adapted to take into account this inaccuracy. For $i = 1, \ldots, k$, we have that:

$$\begin{aligned} |t_i^{(p)} - t_i^*| &\leqslant C_{t*} \|\lambda^{(p)} - \lambda^*\|_2 \\ &\leqslant C_{t*} \left(\|\lambda^{(p)} - \lambda_{best}^*\|_2 + \|\lambda_{best}^* - \lambda^*\|_2 \right), \end{aligned} \tag{191}$$

and so

$$\frac{|t_i^{(p)} - t_i^*|}{\|\lambda^{(p)} - \lambda_{best}^*\|_2} \leqslant C_{t*} \left(1 + \frac{\|\lambda_{best}^* - \lambda^*\|_2}{\|\lambda^{(p)} - \lambda_{best}^*\|_2} \right). \tag{192}$$

For fixed P, which in the experiments in this section is $P = 500$, $\|\lambda_{best}^* - \lambda^*\|_2$ above is fixed and as p approaches P, we have that $\|\lambda^{(p)} - \lambda_{best}^*\|_2 \to 0$, and therefore the right hand side above goes to infinity. This is not a problem for our results, as it is not relevant how the ratio $\frac{|t_i^{(p)} - t_i^*|}{\|\lambda^{(p)} - \lambda_{best}^*\|_2}$ behaves for $\|\lambda^{(p)} - \lambda_{best}^*\|_2 \leqslant \|\lambda_{best}^* - \lambda^*\|_2$.

We can then find a range for p where $\frac{\|\lambda_{best}^* - \lambda^*\|_2}{\|\lambda^{(p)} - \lambda_{best}^*\|_2} \leqslant 1$ and where we can see that

$$\frac{|t_i^{(p)} - t_i^*|}{\|\lambda^{(p)} - \lambda_{best}^*\|_2} \leqslant 2C_{t*}. \tag{193}$$

In Figure 3, we plot $\frac{|t_i^{(p)} - t_i^*|}{\|\lambda^{(p)} - \lambda_{best}^*\|_2}$ for $p = 20, \ldots, 270$, where we see that the ratio is less than C_{t*}. Specifically, we show the ratio $\frac{\|t_i^{(p)} - t_i^*\|}{\|\lambda^{(p)} - \lambda_{best}^*\|_2}$ and the constant C_{t*} from Theorem 2 for each $i \in \{1, 2, 3\}$.

[7]Note that the analysis of the dual problem (10) from Section 2 applies to the dual problem (27) considered in Section 3 as well, as the only difference difference between (10) and (27) is a box constraint on λ.

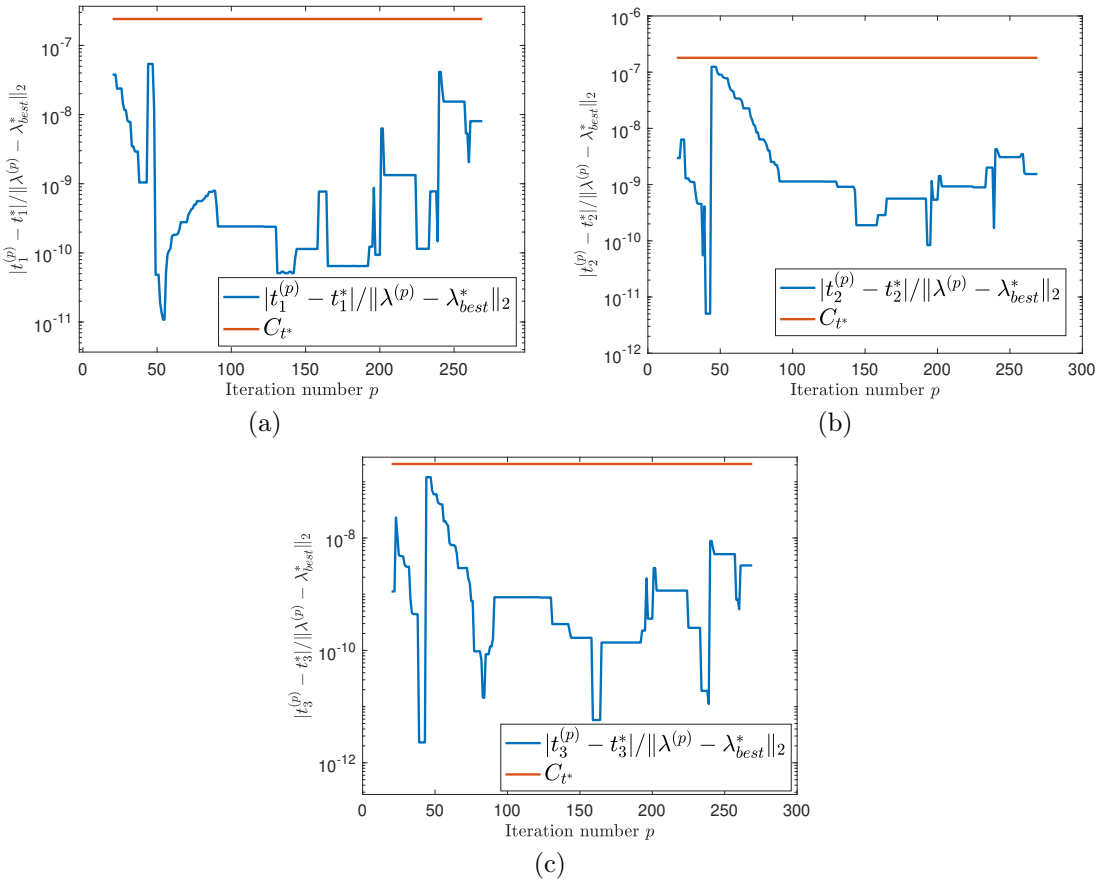

Figure 3: The result of Theorem 2 for $T = \{0.25, 0.63, 0.888\}$, $\sigma = 0.07$ and $m = 21$. For each $i \in \{1, 2, 3\}$, we show the ratio of the error in t_i and the error in λ compared to the constant C_{t*} given by Theorem 2.

Effect of t^* perturbations on a^*

In the case of Theorem 4, it is more straightforward to check the ratio of the errors, since we know the true values of the source locations and weights, which we denote by $t^* = [t_1^*, \dots, t_k^*]^T$ and $a^* = [a_1^*, \dots, a_k^*]^T$ respectively. The error bound (26) given by the theorem is of the form:

$$\|a - a^*\|_2 \leqslant C_{a^*} e^{\frac{4\|t-t^*\|_\infty}{\sigma^2}} \|t - t^*\|_2 + \mathcal{O}(\|t - t^*\|_2^2),$$

where t is the perturbed vector t^* and a is the perturbed vector a^* as a consequence of perturbing t^*. For the values $t_i^{(p)}, i \in \{1, 2, 3\}$, obtained after p iterations of the level method, we now solve the least squares problem $argmin_{\hat{a}} \|\Phi^{(p)}\hat{a} - y\|_2$ with the entries in the data matrix $\Phi^{(p)}$ given by $\Phi_{j,i}^{(p)} = \phi(t_i^{(p)} - s_j)$ to find the corresponding perturbed weights $a_i^{(p)}$ for $i \in \{1, 2, 3\}$. Then, according to Theorem 4, we have that:

$$\frac{\|a^{(p)} - a^*\|_2}{\|t^{(p)} - t^*\|_2} \leqslant C_{a^*}^t + \mathcal{O}(\|t^{(p)} - t^*\|_2), \tag{194}$$

where we write

$$C_{a^*}^t = C_{a^*} e^{\frac{4\|t^{(p)}-t^*\|_\infty}{\sigma^2}}.$$

In Figure 4, we show the ratio $\frac{\|a^{(p)}-a^*\|_2}{\|t^{(p)}-t^*\|_2}$ and $C_{a^*}^t$ in the same setting as in Figure 3, for iterations $p = 20, \dots, 270$.

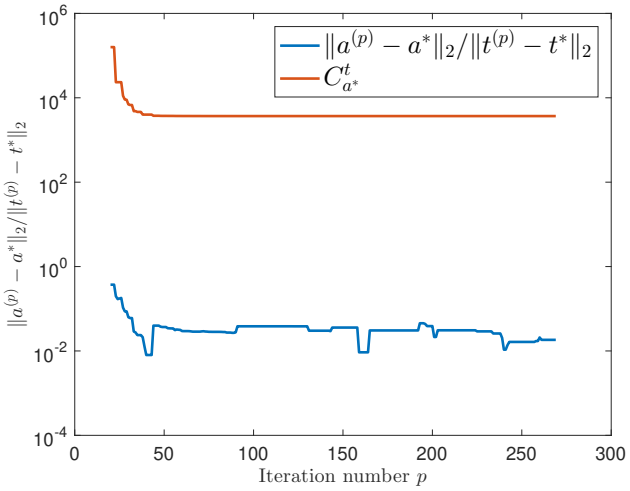

Figure 4: Plot of the ratio between $\|a^{(p)} - a^*\|_2$ and $\|t^{(p)} - t^*\|_2$ for $p = 20, \dots, 270$, and the bound $C_{a^*}^t$ from (194) in the setup described at the beginning of this section.

Effect of the noise w on λ^* and t^*

As in the case of Theorem 2, where we rely on a best approximation λ_{best}^* of λ^* for the numerical experiments, a similar approach is required to check the validity of the results of Theorem 8 in practice. Theorem 8 gives the bound (41) in the form:

$$\|\lambda_w^* - \lambda^*\|_2 \leqslant C_{\lambda^*} \cdot \|w\|_2,$$

where λ^* is the true solution of the dual problem (27) and λ_w^* is the solution to the same problem with y perturbed by the noise w.

As it is not possible to know exactly the values of λ^* and λ_w^*, let $\lambda_{best}^* = \lambda^{(P)}$ be the value of λ given by the level method after P iterations when y is exact and λ_{best} be the value of λ returned by the level method

after P iterations when y is corrupted by the additive noise w. Then we can reformulate the bound (41) in terms of λ^*_{best} and λ_{best}:

$$\begin{aligned}\|\lambda_{best} - \lambda^*_{best}\|_2 &= \|\lambda_{best} - \lambda^*_{best} + \lambda^* - \lambda^* + \lambda^*_w - \lambda^*_w\|_2 \\ &\leqslant \|\lambda_{best} - \lambda^*_w\|_2 + \|\lambda^* - \lambda^*_w\|_2 + \|\lambda^* - \lambda^*_{best}\|_2 \\ &\leqslant \|\lambda_{best} - \lambda^*_w\|_2 + C_{\lambda *}\|w\|_2 + \|\lambda^* - \lambda^*_{best}\|_2, \end{aligned} \tag{195}$$

so

$$\frac{\|\lambda_{best} - \lambda^*_{best}\|_2}{\|w\|_2} \leqslant C_{\lambda^*} + \frac{\|\lambda_{best} - \lambda^*_w\|_2 + \|\lambda^* - \lambda^*_{best}\|_2}{\|w\|_2}. \tag{196}$$

As before, we plot $\frac{\|\lambda_{best} - \lambda^*_{best}\|_2}{\|w\|_2}$, where λ^*_{best} is the solution we obtain by solving the dual problem (27) in its exact penalty formulation using the level method with $P = 100$ iterations and λ_{best} is the 'noisy' solution, which is obtained by solving the problem with $P = 100$ iterations when y is corrupted by additive noise w. We repeat this for different magnitudes of the noise w, which we increase gradually as follows. For each component y_j of y, we add a sample X_j from the standard uniform distribution $U(0, 1)$, multiplied by a coefficient w_c:

$$y_{noisy_j} = y_j + w_c \cdot X_j. \tag{197}$$

We repeat this for different values of the coefficient w_c from the set:

$$\begin{aligned} w_c \in \{&0.000002, 0.000004, \ldots, 0.00001, \\ &0.00002, 0.00004, \ldots, 0.0001, \\ &0.0002, 0.0004, \ldots, 0.001, \\ &0.002, 0.003, \ldots, 0.01, \\ &0.02, 0.03, \ldots, 0.1\}. \end{aligned} \tag{198}$$

Therefore, in Figure 5 we show the basic setup described at the beginning of this section. Panel (a) shows $\|\lambda_{best} - \lambda^*_{best}\|_2$ against the norm of the noise $\|w\|_2$, and in order to check that the algorithm actually converges to a useful λ^*_{best}, we also plot plot $\|t_{best} - t^*\|_2$ against $\|w\|_2$ in panel (b), since we know the true value t^*. Then, in panel (c) we plot the ratio $\frac{\|\lambda_{best} - \lambda^*_{best}\|_2}{\|w\|_2}$ and C_{λ^*} as given by Theorem 8, where we see that the ratio is smaller than the constant, as the theorem states. In the same plot, we also show the ratio $\frac{\|t_{best} - t^*\|_2}{\|w\|_2}$ and we see that it does not grow as the magnitude of the noise increases. In these experiments we only take into account $2k$ entries of λ and w, corresponding to the $2k$ samples that are the closest to the k sources, as described in Section 3, for which Theorem 8 holds.

6 Conclusion

In this paper, we proved primal stability in the non-negative super-resolution problem, when addressed via convex duality. The main ingredient in our analysis is a quantitative version of the implicit function theorem, a folklore result in the theory of dynamical systems community.

In the noise-free setting, our results provide quantitative bounds in terms of the number of measurements for the accuracy of the primal solution with respect to the convex dual problem solution in an ℓ_∞ error bound on the primal spike locations and an ℓ_2 error bound on the spike weights. In the case when the measurements are corrupted by additive noise, we have proved a similar result for how the dual variable is perturbed as a function of the magnitude of the noise.

Acknowledgements

This work was done while BT was affiliated to the Mathematical Institute, University of Oxford, UK. 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.

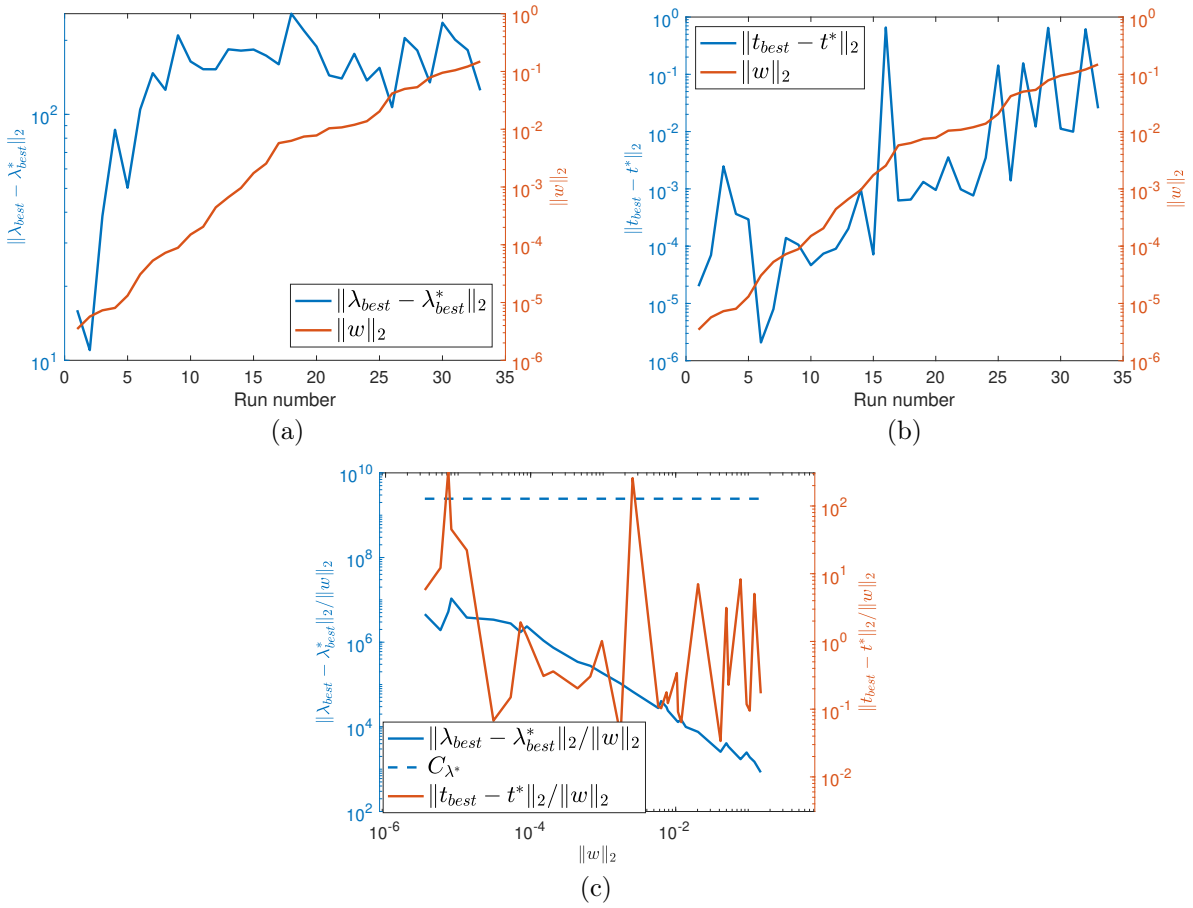

Figure 5: Plots of $\|\lambda_{best} - \lambda^*_{best}\|_2$ (panel(a)), $\|t_{best} - t\|_2$ (panel (b)) and their ratio to the noise $\|w\|_2$ (panel(c)) for $\|w\|_2$ in a range as given in (197) and (198), in the setting described at the beginning of this section.

References

[1] 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.

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

[3] 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.

[4] Federico Pierucci, Zaid Harchaoui, and Jérôme Malick. A smoothing approach for composite conditional gradient with nonsmooth loss. *Research report, RR-8662, INRIA Grenoble*, 2014.

[5] Stéphane Chrétien, Andrew Thompson, and Bogdan Toader. The dual approach to non-negative super-resolution: impact on primal reconstruction accuracy. In *2019 13th International conference on Sampling Theory and Applications (SampTA)*, pages 1–4, 2019.

[6] 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.

[7] Geoffrey Schiebinger, Elina Robeva, and Benjamin Recht. Superresolution without separation. *Information and Inference: A Journal of the IMA*, 7(1):1–30, 2018.

[8] Vincent Duval and Gabriel Peyré. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, 15:1315–1355, 2015.

[9] 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.

[10] 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.

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

[12] Armin Eftekhari, Jared Tanner, Andrew Thompson, Bogdan Toader, and Hemant Tyagi. Non-negative super-resolution is stable. In *2018 IEEE Data Science Workshop (DSW)*, pages 1–5, 2018.

[13] Armin Eftekhari, Jared Tanner, Andrew Thompson, Bogdan Toader, and Hemant Tyagi. Sparse non-negative super-resolution – simplified and stabilised. *Applied and Computational Harmonic Analysis*, 50:216–280, 2021.

[14] Armin Eftekhari and Andrew Thompson. Sparse inverse problems over measures: Equivalence of the conditional gradient and exchange methods. *SIAM Journal on Optimization*, 29(2):1329–1349, 2019.

[15] Marco A. López and Georg Still. Semi-infinite programming. *European Journal of Operational Research*, 180(2):491–518, 2007.

[16] Yurii Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Springer Publishing Company, 2014.

[17] Zhenan Fan, Yifan Sun, and Michael P. Friedlander. Bundle methods for dual atomic pursuit. In *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, pages 264–270, 2019.

[18] Nicholas Boyd, Geoffrey Schiebinger, and Benjamin Recht. The alternating descent conditional gradient method for sparse inverse problems. *SIAM Journal on Optimization*, 27(2):616–639, 2017.

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

[20] Jean-Baptiste Hiriart-Urruty and Claude Lemarechal. *Convex Analysis and Minimization Algorithms, Volume I: Algorithms*. Springer-Verlag, Berlin, 2nd edition, 1996.

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

[22] Vincent Duval and Gabriel Peyré. Sparse regularization on thin grids I: the Lasso. *Inverse Problems*, 33(5), 2017.

[23] Vincent Duval and Gabriel Peyré. Sparse spikes super-resolution on thin grids II: the continuous basis pursuit. *Inverse Problems*, 33(9), 2017.

[24] Ankur Moitra. Super-resolution, extremal functions and the condition number of Vandermonde matrices. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing (STOC 2015)*, 2015.

[25] Yingbo Hua and Tapan K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38(5):814–824, 1990.

[26] Stéphane Chrétien and Hemant Tyagi. Multi-kernel unmixing and super-resolution using the Modified Matrix Pencil method. *Journal of Fourier Analysis and Applications*, 26(1), 2020.

[27] Cédric Josz, Jean Bernard Lasserre, and Bernard Mourrain. Sparse polynomial interpolation: sparse recovery, super-resolution, or Prony? *Advances in Computational Mathematics*, 45(3):1401–1437, 2019.

[28] Carlangelo Liverani. Implicit function theorem (a quantitative version). Retrieved January 13, 2019, from https://www.mat.uniroma2.it/~liverani/SysDyn15/app1.pdf.

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

[30] G. W. Stewart. Perturbation theory and least squares with errors in the variables. In *Contemporary Mathematics 112: Statistical Analysis of Measurement Error Models and Applications*, pages 171–181. American Mathematical Society, 1990.

A Duality in the noisy case

In this section, we show the duality of the following problems:

$$\min_{x \geqslant 0} \left\| y - \int \Phi(t) x(\mathrm{d}t) \right\|_1 \quad \text{subject to} \quad \|x\|_{TV} \leqslant \Pi,$$

which is given in (9), and

$$\max_{\substack{\beta > 0 \\ \lambda \in \mathbb{R}^m}} \beta \left(\lambda^T y - \Pi \right) \quad \text{subject to} \quad \lambda^T \Phi(t) \leqslant 1, \quad \forall t \in [0,1] \quad \text{and} \quad \|\lambda\|_\infty \leqslant 1/\beta, \tag{199}$$

which is a more general version of the dual problem (27). We start from the primal problem (9) by introducing a new variable $z = \int \Phi(t) x(\mathrm{d}t)$:

$$\begin{aligned} \min_{\substack{x \geqslant 0 \\ z \in \mathbb{R}^m}} \|z - y\|_1 \quad \text{subject to} \quad & z = \int \Phi(t) x(\mathrm{d}t), \\ & \|x\|_{TV} \leqslant \Pi, \end{aligned} \tag{200}$$

and then we write the Lagrangian:

$$L(x, z, \beta, \lambda) = \|z - y\|_1 + \lambda^T \left(z - \int \Phi(t) x(\mathrm{d}t) \right) + \beta \left(\|x\|_{TV} - \Pi \right), \tag{201}$$

so the Lagrangian dual problem is:

$$\begin{aligned}
&\max_{\substack{\beta \geqslant 0 \\ \lambda \in \mathbb{R}^m}} \min_{\substack{x \geqslant 0 \\ z \in \mathbb{R}^m}} L(x, z, \beta, \lambda) = \\
&\qquad = \max_{\substack{\beta \geqslant 0 \\ \lambda \in \mathbb{R}^m}} \min_{\substack{x \geqslant 0 \\ z \in \mathbb{R}^m}} \left[\|z - y\|_1 + \lambda^T z + \int \left(\beta - \lambda^T \Phi(t) \right) x(\mathrm{d}t) \right] - \beta \Pi \\
&\qquad = \max_{\substack{\beta \geqslant 0 \\ \lambda \in \mathbb{R}^m}} \min_{\substack{x \geqslant 0 \\ w \in \mathbb{R}^m}} \left[\|w\|_1 + \lambda^T w + \int \left(\beta - \lambda^T \Phi(t) \right) x(\mathrm{d}t) \right] + \lambda^T y - \beta \Pi, \qquad (202)
\end{aligned}$$

where in the last equality we make the substitution $w = z - y$.

The integral on the right hand side is equal to $-\infty$ if there exists $t_0 \in [0, 1]$ such that $\lambda^T \Phi(t_0) > \beta$, as we can set $x = \infty \cdot \delta_{t_0}$. Therefore, we impose the condition that $\lambda^T \Phi(t) \leqslant \beta$ for all $t \in [0, 1]$, in which case the integral is equal to zero by taking x to be zero wherever the integrand is non-zero, and the dual becomes:

$$\max_{\substack{\beta \geqslant 0 \\ \lambda \in \mathbb{R}^m}} \min_{w \in \mathbb{R}^m} \left(\|w\|_1 + \lambda^T w \right) + \lambda^T y - \beta \Pi \quad \text{subject to} \quad \lambda^T \Phi(t) \leqslant \beta, \quad \forall t \in [0, 1]. \qquad (203)$$

which can be rewritten as:

$$\max_{\substack{\beta \geqslant 0 \\ \lambda \in \mathbb{R}^m}} - \max_{w \in \mathbb{R}^m} \left\{ -\lambda^T w - \|w\|_1 \right\} + \lambda^T y - \beta \Pi \quad \text{subject to} \quad \lambda^T \Phi(t) \leqslant \beta, \quad \forall t \in [0, 1]. \qquad (204)$$

and note that for $f(w) = \|w\|_1$:

$$f^*(\lambda) = \max_w \left\{ \lambda^T (-w) - \| - w\|_1 \right\} = \begin{cases} 0, & \text{if} \quad \|\lambda\|_\infty \leqslant 1, \\ \infty, & \text{otherwise,} \end{cases} \qquad (205)$$

is its conjugate [20]. Therefore, we impose the condition that $\|\lambda\|_\infty \leqslant 1$ and the dual becomes:

$$\max_{\substack{\beta \geqslant 0 \\ \lambda \in \mathbb{R}^m}} \lambda^T y - \beta \Pi \quad \text{subject to} \quad \lambda^T \Phi(t) \leqslant \beta, \quad \forall t \in [0, 1] \quad \text{and} \quad \|\lambda\|_\infty \leqslant 1. \qquad (206)$$

We then make the substitution $\lambda' = \lambda / \beta$ (for $\beta > 0$) to obtain:

$$\max_{\substack{\beta > 0 \\ \lambda' \in \mathbb{R}^m}} \beta \left(\lambda'^T y - \Pi \right) \quad \text{subject to} \quad \lambda'^T \Phi(t) \leqslant 1, \quad \forall t \in [0, 1] \quad \text{and} \quad \|\lambda'\|_\infty \leqslant 1/\beta, \qquad (207)$$

which is the problem (199).

Note that if we fix β and solve for λ', given that we are interested in the value of λ' rather than the value of the objective function, the problem above becomes:

$$\operatorname*{argmax}_{\lambda' \in \mathbb{R}^m} \lambda'^T y \quad \text{subject to} \quad \lambda'^T \Phi(t) \leqslant 1, \quad \forall t \in [0, 1] \quad \text{and} \quad \|\lambda'\|_\infty \leqslant 1/\beta, \qquad (208)$$

which is the problem (27) that we consider in Section 3.

B The level bundle method

In this section, we describe the level bundle method [16] applied to (28) for which experiments were presented in Section 1.1 and Section 5. The algorithm progressively builds up a polyhedral model of the objective function from a 'bundle' of subgradients at each iteration. The algorithm proceeds by projecting iterates onto a level set of the model, an approach which is known to improve robustness in comparison with the standard cutting planes subgradient method (Kelley's method). A statement of the algorithm is given in Algorithm 1.

In the experiments shown in Section 1.1, Π was chosen to be $2\|\mathbf{a}^*\|_\mathbf{1}$ and the level set parameter α was taken to be $1/4$.

Algorithm 1 Level bundle method for solving Program (28).

Input: Kernel function $\Phi : I \to \mathbb{R}^m$, measurements $y \in \mathbb{R}^m$, sample locations $\{s_j\}_{j\in\{1,\ldots,m\}} \in I$, penalty parameter $\Pi > 0$, level set parameter $\alpha \in (0,1)$ and number of iterations L.

Initialize: $l = 1$.

While $l \leqslant L$, **do**

1. Compute a subgradient as

$$t^l \in \arg\sup_{s\in I} (\lambda^{l-1})^T \Phi(s),$$

$$g^l = \begin{cases} -y + \Pi\left[(\lambda^{l-1})^T \Phi(t^l) - 1\right], & (\lambda^{l-1})^T \Phi(t^l) \geqslant 1 \\ -y, & (\lambda^{l-1})^T \Phi(t^l) < 1 \end{cases}$$

2. Build the polyhedral model

$$\widehat{\Psi}_\Pi^l(\lambda) = \max_{r=1,\ldots,l} \Psi_\Pi(\lambda^{r-1}) + (g^r)^T(\lambda - \lambda^{r-1}).$$

3. Compute $\nu^l = \inf_\lambda \widehat{\Psi}_\Pi^l(\lambda)$ and $\mu^l = \min_{r=1,\ldots,l} \Psi_\Pi(\lambda^r)$.

4. Project onto the level set as $\lambda^l = \mathcal{P}_{\mathcal{L}_\alpha^l}(\lambda^{l-1})$ where $\mathcal{L}_\alpha^l = \{\lambda : \widehat{\Psi}_\Pi(\lambda) \leqslant \alpha\mu^l + (1-\alpha)\nu^l\}$.

5. $l = l + 1$.

Output: $\lambda^L \in \mathbb{R}^m$.