Blind inverse problems with isolated spikes

Valentin Debarnot, Pierre Weiss

Blind inverse problems with isolated spikes

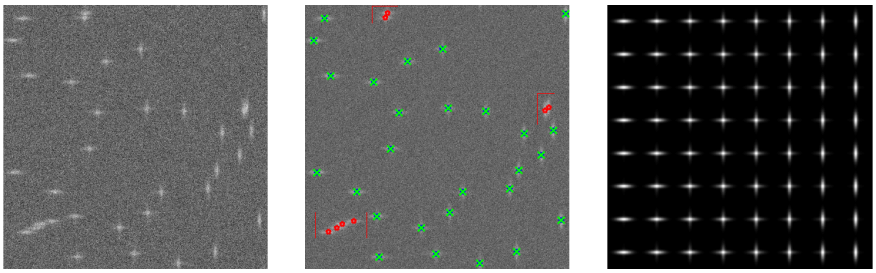
Valentin Debarnot & Pierre Weiss

March 30, 2021

Abstract

Assume that an unknown integral operator living in some known subspace is observed indirectly, by evaluating its action on a few Dirac masses at unknown locations. Is this information enough to recover the operator and the impulse responses locations stably? We study this question and answer positively under realistic technical assumptions. We illustrate the well-foundedness of this theory on two challenging optical imaging problems: blind super-resolution and deconvolution. This provides a simple, practical and theoretically grounded approach to solve these long resisting problems.

1 Introduction



(a) (b) (c)

Figure 1: A sketch of the contribution: (a) a noisy image of the action of an unknown operator on a few Dirac masses, (b) detection of isolated spikes, (c) an operator estimate applied to a Dirac comb. The results in (b) and (c) were obtained using the algorithms proposed in this paper, see Section 4.3 for the technical details.

To motivate this paper, let us start with a concrete problem in imaging. In Figure 1a, we simulated an image of fluorescent proteins observed with an optical microscope. Assume that an algorithm is able to recover the proteins locations at a sub-pixel accuracy from this image. By taking thousands of such images and stacking the protein locations, it is possible to break the diffraction limit and to construct an image with a resolution of the order of a nanometer. This principle was awarded the 2014 Nobel prize in chemistry [1, 2].

From a mathematical viewpoint, this problem can be modelled as follows. Let $\bar{\mu} = \sum_{n=1}^{N} \bar{w}_n \delta_{\bar{x}_n} \in \mathcal{M}(\mathbb{R}^d)$ denote a Radon measure that encodes the

protein locations ($\bar{x}_n$) and their intensity ($\bar{w}_n$). Assume that this measure is observed indirectly through a linear regularizing operator $\bar{A} : \mathcal{M}(\mathbb{R}^d) \to \mathcal{C}_0^0(\mathbb{R}^d)$:

$$y_m = (\bar{A}\bar{\mu})(z_m) + b, \tag{1}$$

where $y \in \mathbb{R}^M$ is the observed data, $(z_1, z_2, \ldots, z_M)$ denotes a set of sampling locations in $\mathbb{R}^D$ and $b \in \mathbb{R}^M$ is some additive noise.

Numerous approaches have been developped over the years to recover the positions ($\bar{x}_n$) from the measurements y. We refer the interested reader to the summaries of the super-resolution challenges [3, 4] for more insight on the possible approaches. The main hurdles to solve this problem are the following:

- The number of measurements M can be huge, making it essential to design computationally efficient methods.
- The weights ($\bar{w}_n$) are usually unknown.
- It is important to work off-the-grid to avoid biases in the location estimation.
- The proteins can sometimes be aggregated in clusters, resulting in a difficult disentanglement of their individual locations.
- Most importantly for this paper: the operator $\bar{A}$ is often only partially known, making it important to estimate both the positions and weights ($\bar{w}_n, \bar{x}_n$), but also the operator $\bar{A}$ itself.

The main objective of this work is to design certified methods, which are able to cope with the above difficulties. This work will strongly rely on the following unformal assumption:

Assumption 1.1. *A few Dirac masses are sufficiently separated from the others, so that the resulting impulse responses can be sensed* independently*.*

1.1 Related works

When the operator $\bar{A}$ is known, recovering $\bar{\mu}$ is a challenging problem, since the inverse problem is ill-posed and infinite dimensional. Specifying prior assumptions on the signal $\bar{\mu}$ to certify its approximate recovery is essential [5]. A few mathematical breakthroughs were achieved in the recent past.

Off-the-grid total variation minimization with a known operator In [6, 7], the authors proposed to recover the individual point sources by solving a generalization of the basis pursuit to an infinite dimensional setting. They showed that the recovery was stable given that the spikes were sufficiently separated. In [8], the authors showed that the separation was not needed, provided that the weights ($\bar{w}_n$) were positive. From a numerical perspective, the solution of this problem can be found rather efficiently using techniques of semi-infinite programming [9, 10, 11]. This type of approach is currently amongst the best competitors when a high density of proteins is used.

Gridded lifting for an unknown operator Assume that the operator $\bar{A}$ is unknown but lives in a *known finite dimensional subspace* $\mathcal{A}$. Also assume that the positions $(\bar{x}_n)$ are known, but that the weights $(\bar{w}_n)$ are unknown. Under these hypotheses, an elegant solution to recover $\mathcal{A}$ and $\bar{\mu}$ was proposed by Ahmed et al in [12] based on a trick called lifting. This approach allows to tackle bilinear problems of the form

$$\inf_{A\in\mathcal{A},u\in\mathcal{U}} \frac{1}{2}\|Au-y\|_2^2, \tag{2}$$

where $\mathcal{U}$ is a finite dimensional subspace of signals, by transforming the bilinear problem into a linear one restricted to rank-1 matrices. This nonconvex constraint can then be relaxed to a convex one by using the nuclear norm. This approach can be guaranteed to stably estimate $(\bar{w}_n)$ and $\bar{A}$ under rather stringent assumptions. The assumptions were relaxed in a series of works [12, 13, 14, 15, 16]. One important achievement was to allow to handle sparsity constraints over a fixed grid instead of subspace constraints. This is particularly relevant for the considered setting.

Off-the-grid lifting In [14], Y. Chi showed that the lifting trick could also be used when the unknown positions $(\bar{x}_n)$ live off-the-grid, $D = 1$ and the operators in $\mathcal{A}$ are convolution operators. The approach was then extended to the 2 dimensional setting for convolution operators in [17]. In [18], an alternative formulation was proposed based on the Hankel lifting for convolution operators in 1D. This approach is elegant but is currently restricted to convolution operators, while it is important in many applications to consider space variant systems. In addition, we will see that a convex relaxation may not be the most efficient approach from a practical viewpoint in the numerical experiments.

1.2 Our contribution

Our main contribution in this work is to propose a simple estimation method that strongly relies on Assumption (1.1). The proposed methodology offers many significant advantages:

- We can work with near arbitrary subspaces of operators $\mathcal{A}$.
- We work under a general linear sampling model with arbitray linear forms.
- The proposed theory doesn't require a grid.
- Our theory is rather simple and leads to recovery conditions that can be checked in advance (for some of them) or a posteriori (for some others).
- The proposed theory provides answers to alternative questions such as the ability to recover the position of Dirac masses from simple correlation algorithms.
- We propose an original study of the stability to noise under a white Gaussian noise assumption on b. This requires analyzing the suprema of continuous Gaussian processes and second order chaos. While most stability bounds in the sparse recovery literature are valid asymptotically when

the noise level goes to 0, our results hold for noise levels of the order of the norm of the measurements. In particular, we explain why correlation algorithms can perform extremely well even under large noise levels.

- The proposed framework - though strongly dependent on Assumption 1.1 - is still realistic for various applications. For instance, we recently proposed a set of algorithms to estimate the subspace of operators $\mathcal{A}$ in [19, 20].
- Most importantly, the proposed algorithms are simple to implement and efficient in practice.

2 Preliminaries

All the proofs in this paper are post-poned to the appendix.

2.1 Notation

Throughout the paper $\mathcal{M}(\mathbb{R}^D)$ will denote the set of Radon measures, i.e. the dual of the set $\mathcal{C}_0^0(\mathbb{R}^D)$ of continuous functions vanishing at infinity. For $\mu \in \mathcal{M}(\mathbb{R}^D)$ and $u \in \mathcal{C}_0^0(\mathbb{R}^D)$, we let $\langle \mu, u \rangle \in \mathbb{R}$ denote the value of the linear form μ on u. We also let $\mu \star u$ denote the convolution product between μ and u defined for all $x \in \mathbb{R}^D$ by $(\mu \star u)(x) = \langle \mu, u(x - \cdot) \rangle$.

In all the paper, the notation $\langle \cdot, \cdot \rangle$ will also refer to the usual scalar product on the vector space $\mathbb{R}^N$, where $N \in \mathbb{N}$ and for $u \in \mathbb{R}^N$, $\|u\|_2$ will denote the ℓ^2-norm of u defined by $\|u\|_2^2 = \langle u, u \rangle$. For two matrices M_1, M_2 in $\mathbb{R}^{M \times N}$, the notation M_1^T will stand for the transpose of M_1 and $\langle M_1, M_2 \rangle_F = \mathrm{Tr}(M_1^T M_2)$ will denote the Frobenius scalar product.

We let $\langle \cdot, \cdot \rangle_{L^2(\mathbb{R}^D)}$ denote the usual scalar product of $L^2(\mathbb{R}^D)$. For a compact and symmetric set $\Omega \subset \mathbb{R}^D$, we let $PW(\Omega)$ denote the Paley-Wiener set of bandlimited functions on Ω, i.e. the set of functions in $L^2(\mathbb{R}^D)$ that have a Fourier transform that vanishes outside Ω.

2.2 Assumptions

All our results will be established under the following two assumptions on the family of operators.

Assumption 2.1 (The operators' structure)**.** *We assume that the family of observation operators $\mathcal{A} = \mathrm{span}\{A_1, \dots, A_I\}$ is a subspace of linear operators from $\mathcal{M}(\mathbb{R}^D)$ to $C_0^0(\mathbb{R}^D)$.*

For any $A \in \mathcal{A}$, there exists a vector $\gamma = (\gamma_i) \in \mathbb{R}^I$ such that for any $\mu \in \mathcal{M}(\mathbb{R}^D)$:

$$A\mu = A(\gamma)\mu \stackrel{\text{def.}}{=} \sum_{i=1}^{I} \gamma_i A_i \mu \tag{3}$$

The next assumption describes the general sampling model considered in this work.

Assumption 2.2 (The observation model)**.** *Let $(\nu_m)_{1 \leq m \leq M}$ in $\mathcal{M}(\mathbb{R}^D)$ denote a collection of M linear forms on $C_0^0(\mathbb{R}^D)$. Let $(\bar{x}_n)_{1 \leq n \leq N}$ denote a collection of N points in $\mathbb{R}^D$ and $(\bar{w}_n)_{1 \leq n \leq N}$ denote N weights.*

We assume that we are given the $N \times M$ measurements

$$y_{n,m} \stackrel{\text{def.}}{=} \bar{w}_n \langle \nu_m, A(\bar{\gamma})\delta_{\bar{x}_n} \rangle + b_{n,m}. \tag{4}$$

In what follows, we will let $y_n = (y_{n,m})_m$ denote the measurement vector in $\mathbb{R}^M$ associated to the n-th Dirac mass $\delta_{\bar{x}_n}$. The observation model 2.2 allows to describe nearly any sampling device. For instance the traditional pointwise sampling would consist in choosing $\nu_m = \delta_{z_m}$, where $(z_m)_{1 \leq m \leq M}$ is a set of sampling locations. The critical element in this assumption is that the impulse responses are observed *independently from each other*.

Under Assumptions 2.1 and 2.2, the impulse response of an operator $A(\gamma)$ at a location $z \in \mathbb{R}^D$ is given by $A(\gamma)\delta_z = \sum_{i=1}^I \gamma_i A_i \delta_z$. Letting

$$(E(z))_{m,i} \stackrel{\text{def.}}{=} \langle \nu_m, A_i \delta_z \rangle, \tag{5}$$

we can rewrite equation (4) compactly as

$$y_{n,m} = (E(\bar{x}_n)\bar{\alpha}_n)_m + b_{n,m}, \tag{6}$$

with $\bar{\alpha}_n = \bar{w}_n \bar{\gamma}$. The matrix-valued function $E : \mathbb{R}^D \to \mathbb{R}^{M \times I}$ will play an essential role in our analysis. Some of our results will depend on two additional hypotheses.

Assumption 2.3 (Identifiability of the operator)**.** *For all $x \in \mathbb{R}^D$, the mapping $E(x) : \mathbb{R}^I \to \mathbb{R}^M$ is injective: we have*

$$\sigma_- \mathrm{Id} \preccurlyeq E^*(x)E(x) \preccurlyeq \sigma_+ \mathrm{Id} \tag{7}$$

with $0 < \sigma_- \leq \sigma_+ < +\infty$. In what follows, we let $\kappa \stackrel{\text{def.}}{=} \frac{\sigma_+}{\sigma_-}$.

This assumption will be useful to guarantee that an operator can be stably estimated once the location of a Dirac mass is known. Throughout the paper, we let

$$R(x) \stackrel{\text{def.}}{=} \mathrm{Ran}(E(x)) \tag{8}$$

denote the subspace of possible measurements for an impulse response located at $x \in \mathbb{R}^D$ and $\Pi_{R(x)}$ denote the orthogonal projector onto the range $R(x)$. Another important technical assumption is the following.

Assumption 2.4 (Identifiability of the Dirac masses location)**.** *The mapping E satisfies the following inequality for any pair $x, \bar{x} \in \mathbb{R}^D$*

$$\|\Pi_{R(x)}\Pi_{R(\bar{x})}\|_{2\to 2} \leq 1 - \phi(\|x - \bar{x}\|_2) \tag{9}$$

for some nondecreasing function $\phi : \mathbb{R}_+ \to [0, 1]$ with $\phi(0) = 0$ and $\phi(t) > 0$ for $t > 0$.

This assumption will allow to guarantee the stable recovery of the Dirac masses locations. This can be understood informally as follows. Take two locations $x \neq \bar{x}$ in $\mathbb{R}^D$. Then, the two ranges $R(x)$ and $R(\bar{x})$ do not contain two identical elements. Hence, the knowledge of a measurement of the form $E^*(\bar{x})\bar{\gamma}$ different from 0 should be enough to perfectly recover $\bar{x}$.

An additional implicit assumption is that the range $R(x)$ cannot be the singleton $\{0\}$ for any x. Indeed, taking $x = \bar{x}$ implies that $\|\Pi_{R(x)}\|_{2\to 2} = 1$.

2.3 Some intuition

Before stating our main results, we provide some intuition on the meaning of Assumption 2.3 and Assumption 2.4.

An injectivity condition

Proposition 2.1. *Under Assumptions 2.3 and 2.4, the mapping* $(x, \gamma) \mapsto E(x)\gamma$ *is injective on* $(\mathbb{R}^D \times \mathbb{R}^I \backslash \{0\})$.

The injectivity of the mapping is a necessary condition to guarantee the identifiability of a position and an operator from a *single* measurement. For instance, it implies that - for any x - the subspace $\mathrm{span}(A_i\delta_x, 1 \leq i \leq I)$ does not contain two elements that are shifted versions of each other. This hypothesis is essential to discard the standard ambiguity in blind deconvolution related to the fact that the signal and the convolution kernel can be shifted in opposite directions and still yield the same measurement vector, see e.g. [21].

A correlation condition Assumption 2.4 allows to control the correlation between measurements of an impulse response at x with an operator $A(\gamma)$ and another at $\bar{x}$ with an operator $A(\bar{\gamma})$. Indeed, we obtain using Cauchy-Schwarz inequality:

$$\begin{aligned}\langle E(x)\gamma, E(\bar{x})\bar{\gamma}\rangle &= \langle \Pi_{R(x)}E(x)\gamma, \Pi_{R(\bar{x})}E(\bar{x})\bar{\gamma}\rangle = \langle \Pi_{R(\bar{x})}\Pi_{R(x)}E(x)\gamma, E(\bar{x})\bar{\gamma}\rangle \\ &\leq \|\Pi_{R(\bar{x})}\Pi_{R(x)}\|_{2\to 2}\|E(x)\gamma\|_2\|E(\bar{x})\bar{\gamma}\|_2\end{aligned}$$

A geometric condition The quantity $\|\Pi_{R(x)}\Pi_{R(\bar{x})}\|_{2\to 2}$ is related to the principal angle between the subspaces $R(x)$ and $R(\bar{x})$. To realize this, let us recall that the principal angle between two subspaces U and V of a Hilbert space is defined by

$$\angle(U, V) = \arccos \max_{\substack{u\in U, v\in V \\ u\neq 0, v\neq 0}} \frac{\langle u, v\rangle}{\|u\|_2\|v\|_2}. \tag{10}$$

Proposition 2.2. *We have*

$$\|\Pi_{R(x)}\Pi_{R(\bar{x})}\|^2_{2\to 2} \leq \cos\left(\angle\left(R(x), R(\bar{x})\right)\right).$$

Convolution operators In this paragraph, we aim at providing some insights on Assumptions 2.3 and 2.4 for the particular case of convolution operators. We will work under the following assumption.

Assumption 2.5. *We assume that we are given an orthogonal* [1] *family* $(e_i)_{1\leq i\leq I}$ *of functions in* $PW(\Omega)$ *that vanish at infinity. The operators* A_i *are convolutions with the filters* e_i*, i.e.* $A_i\mu = e_i \star \mu$ *for* $\mu \in \mathcal{M}(\mathbb{R}^D)$.

The linear forms ν_m *describing the sampling device correspond to a Shannon sampler, i.e.* $\nu_m = \delta_{z_m}$*, where the positions* z_m *correspond to a Cartesian grid with a grid-size smaller than* $\frac{2\pi}{\mathrm{diam}(\Omega)}$.

[1] The orthogonality is not a strong assumption, since any family can be orthogonalized.

Under Assumption 2.5, any $(u, v) \in PW(\Omega)^2$ satisfy

$$\langle u, v \rangle_{L^2(\mathbb{R}^D)} = \sum_{m \in \mathbb{N}} u(z_m) v(z_m), \tag{11}$$

which is a variant of the Shannon-Nyquist theorem, see e.g. [22, Thm 3.5].

Proposition 2.3 (Operator identifiability for convolution operators)**.** *Under Assumption* (2.5)*, we have* $E^*(z)E(z) = \mathrm{Id}$*, hence Assumption 2.3 is satisfied with* $\sigma_- = \sigma_+ = 1$*.*

Proposition 2.4 (Location identifiability for convolution operators)**.** *Let* $\mathcal{C} : \mathbb{R}^D \to \mathbb{R}^{I \times I}$ *denote the following cross-correlation matrix-valued function:*

$$[\mathcal{C}(x - x')]_{i,i'} \stackrel{\text{def.}}{=} \langle e_i(\,\cdot - x), e_{i'}(\,\cdot - x') \rangle_{L^2(\mathbb{R}^D)} \tag{12}$$

Under Assumption (2.5)*, we have*

$$\|\Pi_{R(x)} \Pi_{R(x')}\|_{2 \to 2} = \|\mathcal{C}(x - x')\|_{2 \to 2}. \tag{13}$$

Proposition 2.4 shows that the condition (9) characterizes the speed of decay of a cross-correlation matrix. For instance, consider the simplest case $I = 1$, corresponding to a convolution with a known filter e_1. Then (9) simply measures how fast the auto-correlation function of e_1 decays away from 0. Intuitively, this information is central to derive stability results for algorithms that estimate the Dirac locations by finding correlation maxima. This statement will be made precise in Theorem 3.2.

Product-convolution operators To encode space varying operators, we now turn to product-convolution expansions [23]. These decompositions allow to represent compactly most linear integral operators arising in applications. For the sake of the current paper, we will work under the simplifying assumptions below.

Assumption 2.6 (Product-convolution expansion)**.** *We assume that we are given an orthogonal family* $(e_k)_{1 \leq k \leq K}$ *of bandlimited functions in* $PW(\Omega)$*, and another orthogonal family* $(f_l)_{1 \leq l \leq L}$ *of functions in* $L^2(\mathbb{R}^D) \cap C_0^0(\mathbb{R}^D)$*.*

The family of observation operators $\mathcal{A}$ *is a subspace of product-convolution expansions from* $\mathcal{M}(\mathbb{R}^D)$ *to* $C_0^0(\mathbb{R}^D)$ *defined as follows. For any* $A \in \mathcal{A}$*, there exists a matrix* $\gamma = (\gamma_{k,l}) \in \mathbb{R}^{K \times L}$ *such that for any* $\mu \in \mathcal{M}(\mathbb{R}^D)$*:*

$$A\mu = A(\gamma)\mu \stackrel{\text{def.}}{=} \sum_{k=1}^{K} \sum_{l=1}^{L} \gamma_{k,l} e_k \star (f_l \odot \mu). \tag{14}$$

Similarly to Assumption 2.5, we assume that a Shannon sampler is used. Letting $i = (k, l)$*, this implies that* $(E(z))_{i,m} = f_l(z_m) e_k(z_m - z)$*.*

Let us mention that the blurring operators appearing in optics can be represented very efficiently using this structure [24]. In addition, we recently showed how a subspace of product-convolution operators $\mathcal{A}$ could be constructed in practice in optical imaging [25, 19, 20].

Proposition 2.5. *Under Assumptions 2.1, 2.2 and 2.6, we have*

$$\|\Pi_{R(x)}\Pi_{R(x')}\|_{2\to 2} = \|\mathcal{C}(x-x')\|_{2\to 2}. \tag{15}$$

However, for $L \geq 2$, Assumption 2.3 is not valid: the mapping $E(z)$ is not injective for any z.

As a consequence of this proposition, we will see that the identification of a product-convolution operator with $L \geq 2$ is possible only under the condition $N \geq L$, i.e. by observing multiple impulse responses.

3 Main results

A significant difficulty in the problem of operator estimation comes from the fact that the weights $\bar{w}_n$ are possibly unknown. This issue has been carefully analyzed in a series of recent works, leading to a better understanding of the strengths and limitations of the idea of convex lifting and relaxation [12, 26, 15, 21]. The second difficulty comes from the fact that the positions $\bar{x}_n$ are unknown. This issue received less attention in the literature and we will first focus on this, by assuming that the weights $\bar{w}_n \neq 0$ are known. We will then turn to the case of unknown weights.

3.1 The case of known weights

For all $1 \leq n \leq N$, any measurement y_n of the form (6) can be written as

$$y_n = E(\bar{x}_n)\bar{\alpha}_n + b_n, \tag{16}$$

where $\bar{x}_n \in \mathbb{R}^D$ is an unknown location that we wish to recover and $\bar{\alpha}_n = \bar{w}_n\bar{\gamma}$ is a vector colinear to the unknown operator parameterization $\bar{\gamma}$. Our aim in this section is to design an estimate $\hat{X} = (\hat{x}_1, \dots, \hat{x}_N)$ of $\bar{X} = (\bar{x}_1, \dots, \bar{x}_N)$ and $\hat{\gamma}$ of $\bar{\gamma}$ and certify their proximity, despite the noise term $b_n \in \mathbb{R}^M$. Letting $X = (x_1, \dots, x_N) \in \mathbb{R}^{D\times N}$ and

$$F(\gamma, X) \overset{\text{def.}}{=} \frac{1}{2}\sum_{n=1}^{N} \|\bar{w}_n E(x_n)\gamma - y_n\|_2^2 = \frac{1}{2}\sum_{n=1}^{N} F_n(\gamma, x), \tag{17}$$

a natural approach to achieve this goal is to solve the following nonlinear least-square problem

$$(\hat{\gamma}, \hat{X}) \overset{\text{def.}}{=} \operatorname*{argmin}_{X\in\mathbb{R}^{D\times N}, \gamma\in\mathbb{R}^{I\times N}} F(\gamma, X). \tag{18}$$

3.1.1 Characterizing the solutions

Theorem 3.1. *The global minimizer $(\hat{\gamma}, \hat{X})$ of Problem (18) can be obtained in a two step procedure.*

First, estimating the positions in (18) boils down to solving the following maximum correlation problem

$$\hat{x}_n \in \operatorname*{argmax}_{x\in\mathbb{R}^D} \langle \Pi_{R(x)} y_n, y_n \rangle, \tag{19}$$

independently for all $1 \leq n \leq N$.

Second, the vector $\hat{\gamma}$ *is given by any solution of the following linear system system:*

$$\left(\sum_{n=1}^{N} \bar{w}_n^2 E^*(\hat{x}_n) E(\hat{x}_n) \right) \gamma = \sum_{n=1}^{N} \bar{w}_n E^*(\hat{x}_n) y_n. \tag{20}$$

Assumption 2.3 is sufficient to ensure that $\hat{\gamma}$ *is unique with a single observation.*

Theorem 3.1 shows that under realistic assumptions, the blind inverse problem (18) can be solved exactly in a two step procedure. The first step consists in finding independently N maxima of nonconvex functions in $\mathbb{R}^D$. This is not a trivial task, but it can be implemented with standard methods from nonlinear programming (gradient or Newton-like method) starting from a sufficiently good initialization. Let us mention that this procedure is one of the most popular approach to estimate the positions [3]. The second step consists in solving a small dimensional linear system of equations.

Remark 3.1. *In the proposed formulation, two implicit regularization terms are used: i) we look for a single Dirac mass and ii) the operator live in a known subspace. If the dimension* I *of the subspace is large, the proposed methodology might fail: as* I *increases, so does* $\dim(R(x))$. *For instance in the case of convolution operators (see Assumption 2.5), we have* $\dim(R(x)) = I$. *If* $I = M$, *the correlation function* (19) *is constant and there is no hope to recover* $\bar{x}$.

A possible solution for this problem is to add a weighted ℓ^2*-regularization on* γ *of the form* $\frac{1}{2}\sum_{i=1}^{I} \theta_i \gamma_i^2$, *where* θ_i *are weights adapted to the problem at hand. Most of the theory developed in this paper applies to this setting as well. The main difference is that this regularization introduces a bias in the operator estimate.*

3.1.2 Stability of the location estimates

As seen in Theorem 3.1, recovering the location of each Dirac location x_n can be achieved independently. The following theorem shows that this estimation is robust despite the noise term b_n.

Theorem 3.2 (Stability of the Dirac locations)**.** *Let* $y_{0,n} = E(\bar{x}_n)\bar{\gamma}$ *denote the noiseless measurements and assume that* $\|b_n\|_2 \leq \theta \|y_{0,n}\|_2$ *with* $\theta < \frac{\sqrt{6}}{2} - 1 \simeq 0.225$. *Then, under Assumption 2.4, the following inequality holds:*

$$\|\hat{x}_n - \bar{x}_n\|_2 \leq \phi_+^{-1}\left(2\theta^2 + 4\theta\right), \tag{21}$$

where $\phi_+^{-1}(t) = \inf\{s \text{ s.t. } \phi(s) \geq t\}$ *is the quantile function of* ϕ *(in particular* $\phi_+^{-1} = \phi^{-1}$ *if* ϕ *is bijective).*

The above bound can be quite pessimistic for large M, but - apart from the constant $\frac{\sqrt{6}}{2} - 1$ - it cannot be improved for an arbitrary noise term b_n. In the case of random noise however, we can obtain a much better probabilistic control using concentration inequalities. We focus on the case $N = 1$ observed impulse response, which allows us to discard the indices n.

Proposition 3.3. *Assume that $b \sim \mathcal{N}(0, \sigma^2 \mathrm{Id})$ is white Gaussian noise of variance σ^2. Define the following two random processes*

$$\Delta_1(x) \stackrel{\text{def.}}{=} \langle \Pi_{R(x)} y_0, \Pi_{R(x)} b \rangle \text{ and } \Delta_2(x) \stackrel{\text{def.}}{=} \frac{1}{2} \|\Pi_{R(x)}(b)\|_2^2. \tag{22}$$

Then under Assumption 2.4

$$\|\hat{x} - \bar{x}\|_2 \leq \phi_+^{-1}\left(\frac{2\mathrm{Ampl}(\Delta_1 + \Delta_2)}{\|y_0\|_2^2}\right), \tag{23}$$

where $\mathrm{Ampl}(f) \stackrel{\text{def.}}{=} \sup_{x \in \mathbb{R}^D} f(x) - \inf_{x \in \mathbb{R}^D} f(x)$.

This proposition reveals that the critical quantity to control, to evaluate the localization error is the ammplitude of the random process $\Delta_1 + \Delta_2$. Obtaining tight analytical bounds for this is a difficult problem in general. Hopefully the following proposition shows that it can be evaluated efficiently using numerical procedures.

Proposition 3.4. *We have* $\mathrm{Ampl}(\Delta_1 + \Delta_2) \leq \mathrm{Ampl}(\Delta_1) + \mathrm{Ampl}(\Delta_2)$. *In addition, the random variable* $Z_1 = \mathrm{Ampl}(\Delta_1)$ *is sub-Gaussian:*

$$\mathbb{P}(|Z_1 - \bar{Z}_1| \geq t) \leq 2\exp\left(-t^2/(8\sigma^2\|y_0\|_2^2)\right)$$

and the random variable $Z_2 = \mathrm{Ampl}(\Delta_2)$ *is sub-exponential:*

$$\mathbb{P}(|Z_2 - \bar{Z}_2| \geq t) \leq 2\exp\left(-Ct/\sigma^2\right),$$

where $\bar{Z}_1$ *and* $\bar{Z}_2$ *are the expectations of* Z_1 *and* Z_2 *and* C *is a universal constant.*

This proposition has two consequences. First, we see that the deviation around the mean of Z_1 scales as $\sigma\|y_0\|_2$ and the deviation around the mean of Z_2 scales as σ. Second, Hoeffding [27, Thm 2.6.1] and Bernstein [27, 2.8.1] inequalities imply that computing an empirical average of Z_1 and Z_2 will converge rapidly to the true means $\bar{Z}_1$ and $\bar{Z}_2$. Hence, it is possible to obtain a precise numerical estimate using an empirical average and we know that the probability that the variables deviate from the means by more than $\sigma(\|y_0\|_2+1)$ is negligible.

Unfortunately, the averages $\bar{Z}_1$ and $\bar{Z}_2$ are difficult to compute in general. Hence, the above proposition can only be used to estimate average deviations with a computer. The following proposition provides upper-bounds for $\bar{Z}_1$ and $\bar{Z}_2$ under additional regularity assumptions in the 1D setting.

Theorem 3.5. *Assume that* $D = 1$, *and that:*

- *The range* $R(x)$ *is constant for* $|x| > 1$.
- *the mapping* $x \mapsto \Pi_{R(x)}$ *is Lipschitz continuous:*

$$\|\Pi_{R(x)} - \Pi_{R(x')}\|_{2\to 2} \leq L\|x - x'\|_2 \text{ with } L > 2.^2 \tag{24}$$

- *The following inequality holds for* x, x' *in* $[-1, 1]$ *(this can be seen as a specification of Assumption 2.4):*

$$\|\Pi_{R(x)}\Pi_{R(x')}\|_{2\to 2} \leq \frac{1}{1 + \beta\|x - x'\|_2} \text{ with } \beta > 0. \tag{25}$$

[2]The technical condition $L > 2$ can be easily discarded, but simplifies the presentation.

Then, we have:

$$\bar{Z}_1 \leq C\sigma\|y_0\|_2\sqrt{\frac{L}{\beta}+1} \quad \textit{and} \quad \bar{Z}_2 \leq C'\cdot\sigma^2\left(\log(L)+\sqrt{\log(L)I}\right).$$

for some absolute constants C and C'.

Proposition 3.3 and Theorem 3.5 improve Theorem 3.2 massively. A sufficient condition for the bound (21) to be informative is that $\mathrm{Ampl}(\Delta_1) + \mathrm{Ampl}(\Delta_2) \lesssim \|y_0\|_2^2$, which holds true for $\sigma \lesssim \|y_0\|_2 \min\left(\sqrt{\frac{\beta}{L+\beta}}, \frac{1}{\sqrt{\log L+\sqrt{\log(L)I}}}\right)$. Here each noise component can have an amplitude of the order of the signal's ℓ^2-norm! Previously, it was the whole ℓ^2-norm of the noise which had to be less than the signal's norm. This surprising phenomenon is actually observed in practice, with a good localization despite a huge amount of noise.

Example 3.1. *To illustrate the interest and tightness of Theorem 3.5, let us consider a simple example. Assume that we observe a* discrete *Dirac mass y_0 of size $M \in 2\mathbb{N}$: $(y_0)_{M/2} = 1$ and $(y_0)_m = 0$ for all $m \neq M/2$. Let $y = y_0 + b$, where $b \sim \mathcal{N}(0, \sigma^2\mathrm{Id})$. What is the probability that the maximum of y is located at $M/2$ and not elsewhere? Standard concentration inequalities (e.g. Slepian's lemma) state that the maximum of the absolute value of a white Gaussian noise concentrates around $\sigma\sqrt{2\log M}$, i.e. that the probability that it is higher than $c\cdot\sigma\sqrt{2\log M}$ decays exponentially in c. Hence, there is a high probability that the maximum be located at $M/2$ if $\sigma\sqrt{2\log M} < 1/2$ and it should located elsewhere if $\sigma\sqrt{2\log M} > 2$.*

In the continuous setting, this example can be approached as follows. Let us consider the case of a single convolution operator ($I = 1$) with a kernel e defined as a hat function e of width r : $e(x) = \max(1 - |x|/r, 0)$. We assume that the sampling locations are regularly spaced in $[-1, 1]$, i.e. $\nu_m = \delta_{-1+m\cdot 2/M}$. In order to satisfy Assumption 2.4, we need the radius r to satisfy $r > 1/M$. For instance, for $r = 1.1/M$, the convolution operator is very localized, and the continuous setting approximately reproduces the discrete one just described. In that case, it is possible to show that $L \lesssim 1/r$ and $\beta \sim 1/r$. The condition $\sigma \lesssim \frac{\|y_0\|_2}{\sqrt{\log L}}$ obtained in Theorem 3.5 exactly corresponds to what we just described and we see that it cannot *be improved.*

3.1.3 Stability of the operator estimates

Finally, it is possible to guarantee the closeness between $\bar{\gamma}$ and $\hat{\gamma}$ under an additional Lipschitz regularity assumption on E. Here, we work with $N = 1$ observed impulse response. A similar result can be obtained for arbitrary N, but we want to emphasize that the estimation is possible and stable with *a single* observation.

Theorem 3.6 (Stability of the operator estimate with a single observation)**.** *Assume that $N = 1$ and that E is $\sqrt{\sigma_+}L_E$-Lipschitz continuous*[3]*:*

$$\|E(x) - E(x')\|_{2\to 2} \leq \sqrt{\sigma_+}L_E\|x - x'\|_2 \quad \textit{for all} \quad (x, x') \in \mathbb{R}^D \times \mathbb{R}^D. \quad (26)$$

[3]The scaling in $\sqrt{\sigma_+}$ is natural considering Assumption 2.3.

Then, under Assumption 2.3, we have

$$\frac{\|\hat{\gamma}-\bar{\gamma}\|_2}{\|\bar{\gamma}\|_2} \leq \kappa^{3/2}\frac{\|b_1\|_2}{\|y_{0,1}\|_2} + \epsilon_2(\hat{x}) \tag{27}$$

with

$$\epsilon_2(\hat{x}) = C\kappa^{5/2}L_E\|\hat{x}-\bar{x}\|_2\left(1+\frac{\|b_1\|_2}{\|y_{0,1}\|_2}\right) + o_{\hat{x}\to\bar{x}}\left(\|\hat{x}-\bar{x}\|_2^2\right)$$

for some absolute constant C.

Together with Theorem 3.2, this last result ensures that $\hat{\gamma} \to \bar{\gamma}$ when the noise level $\|b_1\|_2$ vanishes. This means that we can stably recover an operator when observing a single impulse response.

Unfortunately, Assumption 2.3 is not always met in practical situations of interest as outlined in Proposition 2.5. In that case, observing multiple impulse responses $N > 1$ can still make a stable estimation possible.

Theorem 3.7 (Stability of the operator estimate with multiple observations)**.** *Given $X = (x_1, \dots, x_n)$, let $\bar{w}_- = \min_n |\bar{w}_n|$, $\bar{w}_+ = \max_n |\bar{w}_n|$ and*

$$\mathcal{C}(X) \overset{\text{def.}}{=} \sum_{n=1}^{N} \bar{w}_n^2 E^*(x_n)E(x_n).$$

Let $\tilde{\sigma}_- = \bar{w}_-^2\hat{\sigma}_-$ and $\tilde{\sigma}_+ = \bar{w}_+^2\hat{\sigma}_+$ and assume that

$$\tilde{\sigma}_-\mathrm{Id} \preccurlyeq \mathcal{C}(\hat{X}) \preccurlyeq \tilde{\sigma}_+\mathrm{Id}. \tag{28}$$

Similarly to Theorem 3.6, assume that E is $\sqrt{\hat{\sigma}_+}L_E$-Lipschitz continuous and let $\tilde{\kappa} = \frac{\tilde{\sigma}_+}{\tilde{\sigma}_-}$.

Then we have

$$\frac{\|\hat{\gamma}-\bar{\gamma}\|_2}{\|\bar{\gamma}\|_2} \leq \tilde{\kappa}^{3/2}\frac{\|B\|_2}{\|Y_0\|_2} + \epsilon_2(\hat{X}) \tag{29}$$

with

$$\epsilon_2(\hat{X}) \leq C\tilde{\kappa}^{5/2}L_E\|\bar{X}-\hat{X}\|_2\left(1+\frac{\|Y_0\|_2+\|B\|_2}{\|Y_0+B\|_2}\right) + o_{\hat{X}\to\bar{X}}\left(\|\bar{X}-\hat{X}\|_2^2\right),$$

for some absolute constant C.

Assumption (28) is a geometrical condition intertwining the locations of the Dirac masses and the properties of the observation mapping E. It can be hard to verify in advance. However it only requires computing the $I \times I$ matrix $\mathcal{C}(\hat{X})$, which can be achieved once $\hat{X}$ has been evaluated. The stable estimation of $\hat{X}$ on its side only depends on Assumption 2.4, which can be verified in advance and can be satisfied independently of Assumption 2.3. Hence, Theorem 3.7 actually yields a constructive result to guarantee the stable recovery of an operator with the following approach:

- If Assumption 2.4 is satisfied and the noise level is low, estimate $\hat{X}$.
- Evaluate the condition number $\tilde{\kappa}$ of $\mathcal{C}(\hat{X})$.
- If $\tilde{\kappa}$ is sufficiently low, $\hat{\gamma}$ provides a good estimate of $\bar{\gamma}$.

3.2 The case of unknown weights

If the weights $\bar{w}_n$ are unknown, the previously described approaches cannot be used directly. It is then tempting to solve the following nonlinear least squares problem:

$$\inf_{\substack{w\in\mathbb{R}^N,\gamma\in\mathbb{R}^I\\ X\in\mathbb{R}^{N\times D}}} J(w,\gamma,X) \tag{30}$$

where

$$J(w,\gamma,X) \overset{\text{def.}}{=} \frac{1}{2}\sum_{n=1}^{N} J_n(w_n,\gamma,x_n) \text{ with } J_n(w_n,\gamma,x_n) \overset{\text{def.}}{=} \frac{1}{2}\|w_n E(x_n)\gamma - y_n\|_2^2.$$

In what follows, we let $(\hat{w},\hat{\gamma},\hat{X})$ denote any minimizer of (30).

3.2.1 A bilinear inverse problem

Theorem (3.1) shows that computing the locations estimate $\hat{X}$ can be achieved independently of the weights $\hat{w}$ and of the operator $\hat{\gamma}$. Minimizing J with respect to (w,γ) for a fixed $X=\hat{X}$ is a *bilinear* inverse problem. It received a considerable attention lately, with numerous progress both on the necessary and sufficient conditions to guarantee the recovery [12, 15, 21, 16, 28], on the optimal stability to noise [29], and on the numerical aspects through convex lifting [30] or local optimization [31, 32, 33, 34, 35, 26]. We briefly explain below how these results can be applied to the proposed setting.

Let $\hat{I}_n \overset{\text{def.}}{=} \dim(R(\hat{x}_n))$ and $\hat{I}=\sum_{n=1}^N \hat{I}_n$. Using a singular value decomposition, we can decompose $E(\hat{x}_n)$ as

$$E(\hat{x}_n) = \hat{U}_n\hat{V}_n^*, \tag{31}$$

where $\hat{U}_n\in\mathbb{R}^{M\times\hat{I}_n}$ is orthogonal in the sense that $\hat{U}_n^*\hat{U}_n=\mathrm{Id}$ and $\hat{V}_n\in\mathbb{R}^{I\times\hat{I}_n}$ contains orthogonal colums. Hence, letting $c_n \overset{\text{def.}}{=} \hat{U}_n^* y_n$, we obtain

$$\begin{aligned}\operatorname*{argmin}_{w\in\mathbb{R}^N,\gamma\in\mathbb{R}^I} J(w,\gamma,\hat{X}) &= \operatorname*{argmin}_{w\in\mathbb{R}^N,\gamma\in\mathbb{R}^I} \frac{1}{2}\sum_{n=1}^{N}\|\hat{U}_n\hat{V}_n^* w_n\gamma - y_n\|_2^2\\ &= \operatorname*{argmin}_{w\in\mathbb{R}^N,\gamma\in\mathbb{R}^I} \frac{1}{2}\sum_{n=1}^{N}\|\hat{V}_n^* w_n\gamma - c_n\|_2^2.\end{aligned}$$

Letting $\hat{\mathcal{B}}:\mathbb{R}^N\times\mathbb{R}^I\to\mathbb{R}^{\hat{I}}$ denote the following bilinear mapping:

$$\hat{\mathcal{B}}(w,\gamma) \overset{\text{def.}}{=} \begin{pmatrix}\hat{V}_1^* w_1\gamma\\ \vdots\\ \hat{V}_N^* w_N\gamma\end{pmatrix} \text{ and } c \overset{\text{def.}}{=} \begin{pmatrix}c_1\\ \vdots\\ c_N\end{pmatrix}, \tag{32}$$

we can rewrite J more compactly as $J(w,\gamma,\hat{X})=\frac{1}{2}\|\hat{\mathcal{B}}(w,\gamma)-c\|_2^2$, and hence:

$$\operatorname*{argmin}_{w\in\mathbb{R}^N,\gamma\in\mathbb{R}^I} J(w,\gamma,\hat{X}) = \operatorname*{argmin}_{w\in\mathbb{R}^N,\gamma\in\mathbb{R}^I} \frac{1}{2}\|\hat{\mathcal{B}}(w,\gamma)-c\|_2^2. \tag{33}$$

Notice that the dimension M, which might be huge in applications, completely disappeared from this formulation.

3.2.2 Global injectivity conditions

Recovering w and γ is possible only up to a multiplicative constant since

$$J(tw, \gamma/t, \hat{X}) = J(w, \gamma, \hat{X}) \text{ for all } t \neq 0.$$

Now, consider the noiseless setting $B = 0$ and assume that the locations are perfectly recovered: $\hat{X} = \bar{X}$. In that situation, a necessary condition to recover $(\bar{w}, \bar{\gamma})$ modulo the above scaling ambiguity is that there exists a unique pair (w, γ) with $\|w\|_2 = 1$ such that $\hat{\mathcal{B}}(w, \gamma) = c$. To the best of our knowledge, deriving conditions to ensure this local injectivity condition received little attention in the literature.

In [36, 21], the authors study a more stringent *global* injectivity condition of the form

$$\forall c \in \mathbb{R}^{\hat{I}}, \exists \text{ a unique } (w, \gamma) \text{ with } \|w\|_2 = 1 \text{ s.t. } \hat{\mathcal{B}}(w, \gamma) = c. \tag{34}$$

Their main result states that a *necessary* condition for $\hat{\mathcal{B}}$ to be globally injective is that

$$\hat{I} \geq 2(N + I) - 4, \tag{35}$$

which provides a rule on how to choose the number of measurements N. In addition, they prove that almost every bilinear mapping $\hat{\mathcal{B}}$ with respect to the Lebesgue measure is globally injective provided that the inequality (35) holds. In the proposed setting, there are three limitations to this result. First, the operator $\hat{\mathcal{B}}$ that appears in our formulation possesses a peculiar structure which may well fall in a set of 0 measure. Second, we observed that the condition (35) could be violated significantly in practice and that stable recovery still occurred for a particular c. Third, the result does not certify that a low complexity algorithm can actually recover the factors.

3.2.3 Optimization of the factors

Solving (33) can be achieved using local optimization over each factor w and γ [32, 35, 26]. A simple approach consists in using an alternate minimization between the factors as outlined in Algorithm 1. Notice that every step of the

Algorithm 1 Alternating minimization

Require: Initial guess: $w_1 \in \mathbb{R}^N$.
Require: Iteration number K.
  for all $k = 1 \to K - 1$ **do**
    $\gamma_{k+1} = \underset{\gamma \in \mathbb{R}^I}{\operatorname{argmin}} \frac{1}{2} \|\hat{\mathcal{B}}(w_k, \gamma) - c\|_2^2.$
    $w_{k+1} = \underset{w \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2} \|\hat{\mathcal{B}}(w, \gamma_{k+1}) - c\|_2^2.$
  end for
  return (w_K, γ_K).

algorithm can be performed efficiently since the dimensions of the problem are significantly reduced. This approach can be certified to recover a stable estimate $(\hat{w}, \hat{\gamma})$ of $(\bar{w}, \bar{\gamma})$ provided that a clever initialization is used [35, 26]. Sufficient

recovery guarantees are for instance provided when the bilinear mapping $\hat{\mathcal{B}}$ is chosen at random. This method also allows to easily incorporate constraints (e.g. nonnegativity) in the factors, which can sometimes allow a significantly improved reconstruction. In all our numerical experiments, we will use the spectral initialization from [26] as a starting point.

3.2.4 Optimization over rank-1 matrices

The bilinear mapping $\hat{\mathcal{B}}(w, \gamma)$ can be rewritten as a linear mapping $\hat{\mathcal{L}}$ on the rank-1 outer product $T = w\gamma^T : \hat{\mathcal{B}}(w, \gamma) = \hat{\mathcal{L}}(T)$. Hence, we have:

$$\inf_{w \in \mathbb{R}^N, \gamma \in \mathbb{R}^I} J(w, \gamma, \hat{X}) = \inf_{T \in \mathbb{R}^{N \times I}, \text{rank}(T)=1} \frac{1}{2} \|\hat{\mathcal{L}}(T) - c\|_2^2. \tag{36}$$

The interest of the right-hand side in equation (36) compared to the left hand side is that the scaling ambiguity is discarded. Letting $\mathcal{T}$ denote the set of rank-1 matrices, this alternative formulation can be solved using a projected gradient descent described in Algorithm 2. To the best of our knowledge, this formula-

Algorithm 2 Projected gradient descent

Require: Initial guess: $T \in \mathbb{R}^{N \times I}$.
Require: Iteration number K.
  Compute $\tau = \frac{1}{\|\hat{\mathcal{L}}\|_{2\to 2}^2}$ using a power iteration.
  for all $k = 1 \to K - 1$ **do**
    $T_{k+1} = \Pi_{\mathcal{T}} \Big(T_k - \tau \hat{\mathcal{L}}^* (\hat{\mathcal{L}}(T_k) - c) \Big)$.
  end for
  Decompose $T_K = w_K \gamma_K^*$.
  return (w_K, γ_K).

tion received no attention yet in the literature and we will provide numerical comparisons in the next section. Again, we will use the spectral initialization from [26] as a starting guess for this algorithm.

3.2.5 Convex relaxation using the nuclear norm

Finally, a popular method [12, 16, 30, 14] is a convex relaxation using the nuclear norm. The usual convex relaxation of the nonconvex problem (36) is the following:

$$\inf_{T \in \mathbb{R}^{N \times I}, \hat{\mathcal{L}}(T)=c} \|T\|_* \quad \text{or} \quad \inf_{T \in \mathbb{R}^{N \times I}} \frac{1}{2} \|\hat{\mathcal{L}}(T) - c\|_2^2 + \lambda \|T\|_*, \tag{37}$$

where $\lambda > 0$ is a regularization parameter and $\|\cdot\|_*$ is the nuclear norm, i.e. the sum of the singular values of T. This convex function over the space of matrices is well known to promote low-rank solutions since the extreme points of the associated unit ball are the rank-1 matrices [37]. The stable recovery of the tensor $\bar{w}\bar{\gamma}^T$ has been established under rather stringent conditions based on random subspace assumptions [12, 16]. Experimentally, the method seems to provide satisfactory results under much weaker conditions.

From a numerical perspective, Problem (37) can be solved using a diversity of proximal algorithms, such as an accelerated proximal gradient descent or a

Douglas-Rachford algorithm [38]. We do not further detail these algorithms, which are well documented in the literature.

4 Applications

The aim of this section is to illustrate the proposed theory using simple 1D examples and to explain the setting of the 2D experiment in Figure 1.

4.1 Convolution operators with known weights

We start with an illustration of Theorem 3.2 using convolution operators only. We focus on the case of pointwise sampling on $[0,1]$, by setting $\nu_m = \delta_{z_m}$, with $z_m = m/M$ for $m \in \{1,\dots,M\}$. Notice that this case also covers the case of product-convolution operators since the ranges $R(x)$ of convolution and product-convolution operators are identical.

4.1.1 The families of operators

We consider three families of convolution operators $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$ differing by the choice of the convolution filters.

Family $\mathcal{A}_1$ is defined through a set of convolution operators A_i with Gaussian filters (e_i) defined by:

$$e_i(x) = \exp(-x^2/(2\sigma_i^2)) \text{ with } \sigma_i = 1e^{-2} \cdot \frac{i-1}{I-1} + 3e^{-2} \cdot \left(1 - \frac{i-1}{I-1}\right).$$

Using this family in a blind deconvolution problem allows to identify the variance of a Gaussian convolution filter. Gaussian convolution filters are amongst the most popular simplified point spread function models in microscopy.

Family $\mathcal{A}_2$ is also defined using Gaussian convolution filters, but the standard deviation ranges in $[3e^{-2}, 9e^{-2}]$ instead of $[1e^{-2}, 3e^{-2}]$.

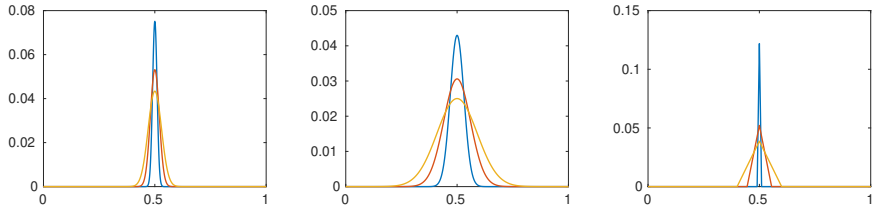
Family $\mathcal{A}_3$ is defined with less regular convolution filters. Let $\psi(x) = (1 - |x-1|)_+$ denote the hat function.

$$e_i = \psi(x \cdot \sigma_i) \text{ where } \sigma_i = 2e^{-2} \cdot \frac{i-1}{I-1} + 2e^{-1} \cdot \left(1 - \frac{i-1}{I-1}\right).$$

In all settings we set $I = 3$. The filters corresponding to each family are displayed in Figure 2. We then orthogonalize the filters using a singular value decomposition on a very fine grid. This leads to a new family of orthogonal filters $(e_i^\perp)$ which will be used in all experiments to satisfy Assumption 2.5.

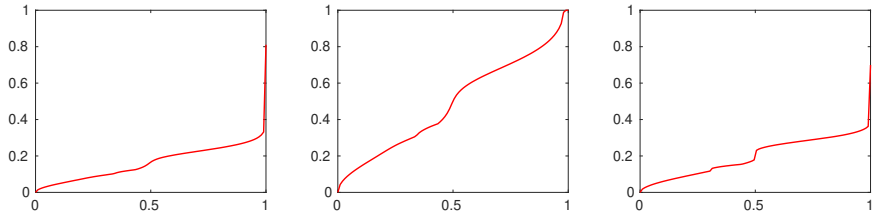
4.1.2 The inverse functions ϕ^{-1}

As stated in Theorem 3.2, the critical element to guarantee a stable recovery of the locations $\bar{x}_m$ is the function ϕ and its inverse, which characterizes the angle between the ranges $R(x)$ and $R(x')$. To evaluate this function, we first sample the function $\|\Pi_{R(0)}\Pi_{R(k\Delta x)}\|_{2\to 2}$ on a fine grid. We store the result in the vector



(a) Filters in family $\mathcal{A}_1$ (b) Filters in family $\mathcal{A}_2$ (c) Filters in family $\mathcal{A}_3$

Figure 2: The different families of convolution filters used in Section 4.1.



(a) $\phi_{1,+}^{-1}$ (b) $\phi_{2,+}^{-1}$ (c) $\phi_{3,+}^{-1}$

Figure 3: The corresponding inverse functions $\phi_{i,+}^{-1}$ (in red) for $i \in \{1, 2, 3\}$, for the different convolution systems.

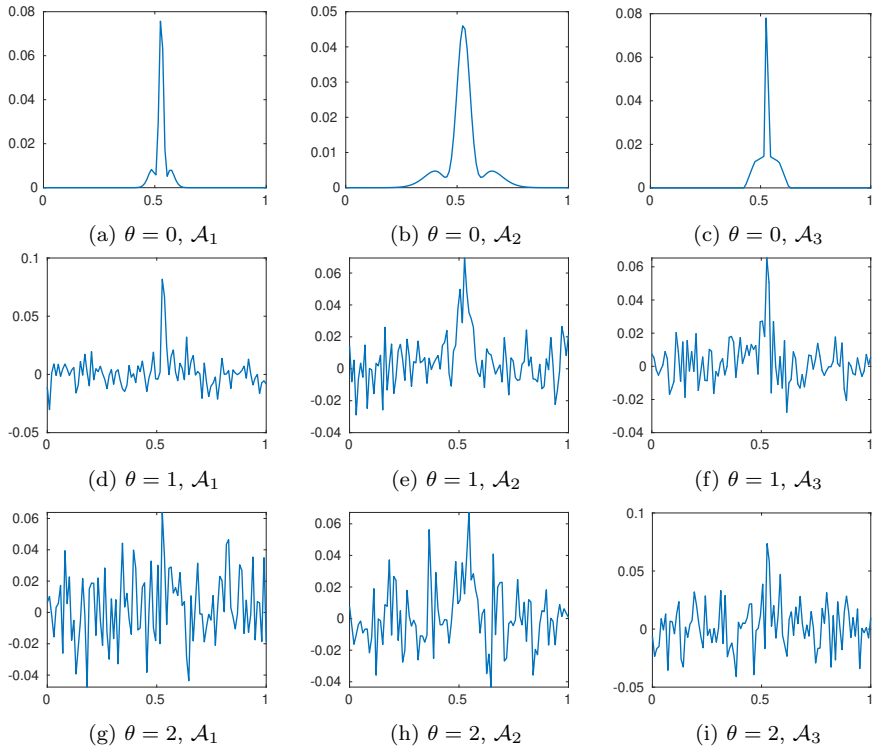
$\phi_0(k) \overset{\text{def.}}{=} 1 - \|\Pi_{R(0)}\Pi_{R(k\Delta x)}\|_{2\to 2}$ with a sampling step Δx. This function is not necessarily nondecreasing. Hence, we find the closest nondecreasing function by solving an isotonic regression problem of the form:

$$\inf_{\phi} \frac{1}{2}\|\phi - \phi_0\|_2^2 \text{ with } \phi_{k+1} - \phi_k \geq 0 \text{ and } \phi \geq \phi_0.$$

This problem is convex and can be solved using the CVX library [39] for instance. We use the solution $\hat{\phi}$ of this problem in place of ϕ in Assumption 2.4. The inverse filters are displayed in Figure 3. The stability to noise is dependent on the speed of ascent of ϕ_+^{-1}. As can be seen by comparing the two Gaussian families, the smallest the filter, the slower the ascent. Hence, very localized impulse responses should be easier to detect with a good accuracy than larger ones. Also notice that the regularity of the convolution kernels seem to have little importance since the inverses $\phi_{1,+}^{-1}$ and $\phi_{3,+}^{-1}$ behave roughly similarly in terms of speed of ascent.

4.1.3 Stability of the locations

Here, we study the robustness of the estimation to noise. To this end, we compute the empirical average of the error $\mathbb{E}(|\hat{x} - \bar{x}|)$ for various noise levels and realizations. The expectation is estimated by averaging 100 noise realizations. We fix $\bar{\gamma}$ once for all. We use white Gaussian noise, i.e. $b_n \sim \mathcal{N}(0, \sigma^2 \mathrm{Id})$, with $\sigma = \theta\|y_0\|_2/\sqrt{M}$ and $\theta \in [0, 2]$. Figure 4 shows the resuting signals with $M = 100$ for the noise levels $\theta \in \{0, 1, 2\}$ and each family. Notice that $\theta = 1$ corresponds to an expected norm of noise equal to the signal's norm. Hence, we consider rather extreme noise levels. We will see that the localization accuracy is surprisingly good in spite of this challenging setting.



(a) $\theta = 0$, $\mathcal{A}_1$ (b) $\theta = 0$, $\mathcal{A}_2$ (c) $\theta = 0$, $\mathcal{A}_3$

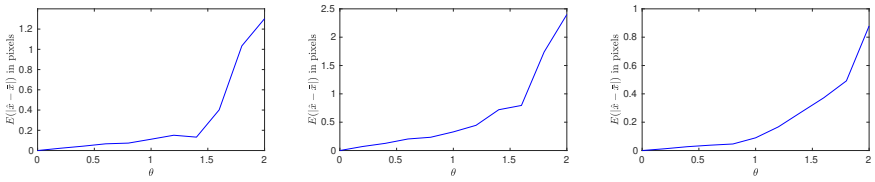
(d) $\theta = 1$, $\mathcal{A}_1$ (e) $\theta = 1$, $\mathcal{A}_2$ (f) $\theta = 1$, $\mathcal{A}_3$

(g) $\theta = 2$, $\mathcal{A}_1$ (h) $\theta = 2$, $\mathcal{A}_2$ (i) $\theta = 2$, $\mathcal{A}_3$

Figure 4: Examples of measurements vectors y for different noise levels and each family.

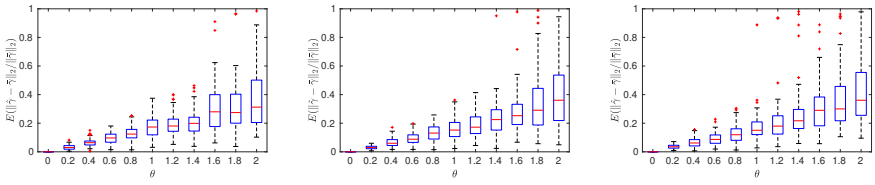
The empirical estimate of $\mathbb{E}(|\hat{x}-\bar{x}|)$ is displayed with respect to the noise level σ in Figure 5. The error is displayed as a proportion of a pixel. For instance, an error of 0.1 means that the localization was accurate at a tenth of a pixel. Hence, we can expect a super-resolution effect for a precision below 0.5. This accuracy is obtained for all families up to the noise level $\theta = 1$, corresponding to a measurement dominated by noise.

To end this experiment, we evaluate $\hat{\gamma}$ for all experiments and display the relative error $\frac{\|\hat{\gamma}-\bar{\gamma}\|_2}{\|\bar{\gamma}\|_2}$ for all families of operators. Letting $\bar{h} = \sum_{i=1}^{I} \bar{\gamma}_i e_i^{\perp}$ and $\hat{h} = \sum_{i=1}^{I} \hat{\gamma}_i e_i^{\perp}$ denote the true convolution filter and the estimated one, notice that we have $\frac{\|\bar{h}-\hat{h}\|_{L^2(\mathbb{R}^D)}}{\|\bar{h}\|_{L^2(\mathbb{R}^D)}} = \frac{\|\hat{\gamma}-\bar{\gamma}\|_2}{\|\bar{\gamma}\|_2}$, since the family $(e_i^{\perp})$ is orthogonal. In Figure 6, we see that the reconstruction errors for any family of convolution filters behave really similarly. In particular, the errors using the family $\mathcal{A}_1$ and the family $\mathcal{A}_2$ are nearly identical. This might come as a surprise since the localization errors were significantly higher for the family $\mathcal{A}_2$, which is a scaled version of $\mathcal{A}_1$. This fact can be explained by the fact that the Lipschitz constant L_E in Theorem 3.6 is inversely proportional to the scaling of the Gaussian, which compensates for the localization errors.



(a) Family $\mathcal{A}_1$ (b) Family $\mathcal{A}_2$ (c) Family $\mathcal{A}_3$

Figure 5: Average localization error $\mathbb{E}(|\hat{x} - \bar{x}|)$ as a fraction of a pixel for different noise levels $\theta \in [0, 1]$ and the three families $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$.



(a) $\mathcal{A}_1$ (b) $\mathcal{A}_2$ (c) $\mathcal{A}_3$

Figure 6: Boxplots of the relative errors $\frac{\|\hat{\gamma}-\bar{\gamma}\|_2}{\|\bar{\gamma}\|_2}$ using a single measurement for various noise levels.

4.2 Product-convolution operators and unknown weights

The objective of this section is to compare the different algorithms described in Section 3.2 for the specific case of 1D product-convolution operators described in Assumption 2.6. In this experiment, we work on a grid and set $\hat{X} = \bar{X}$ since the objective is not to assess the localization performance, but rather the ability to solve a bilinear inverse problem.

We use the pointwise sampling model $\nu_m = \delta_{z_m}$ with $z_m = 10 \cdot m/M$ and $M = 1000$. This corresponds to a uniform sampling of the interval $[0, 10]$. For the filters (e_k), we use the family of Gaussian convolution kernels $\mathcal{A}_1$ with $K = 3$. We set the vectors f_l as smooth random Gaussian processes by convolving a random vector with distribution $\mathcal{N}(0, \mathrm{Id}_M)$ with a Gaussian filter of large variance.

Once the family of operators is defined through the pairs of families (e_k) and (f_l), we can sample operators at random in this family by setting $\gamma \sim \mathcal{N}(0, \mathrm{Id}_I)$. In Figure 7, we visualize a set of operators indirectly by applying them to a Dirac comb with 4 spikes. As can be seen, the operators are space-varying. Their impulse responses belong to $\mathrm{span}(e_k, k \in \{1, \dots, K\})$.

To assess the performance of the different algorithms, we evaluate the percentage of perfect recovery results with various values of L and N. We run the algorithms 100 times with random locations for the N spikes $\bar{x}_n$, with random weights $\bar{w}_n$ and with a random family (f_l). The recovery results are displayed in Figure 8. For the considered families, the nuclear norm relaxation performs very poorly, suggesting that the relaxation approaches suggested both for discrete and gridless problems might not be the best competitors. In comparison, the alternate minimization (Algorithm 1) with the spectral initialization from [26] and the seemingly novel projected gradient descent (Algorithm 2) perform satisfactorily for a good range of values of L and N. Between those two, the

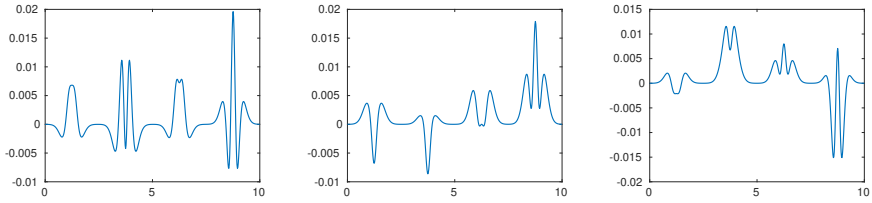
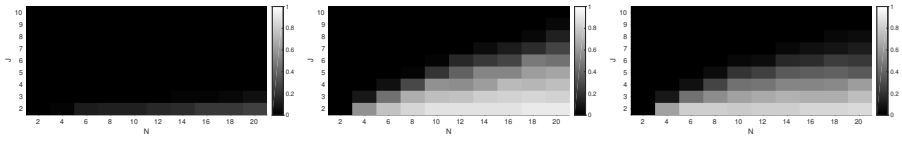


Figure 7: Examples of random product-convolution operators for a fixed family (e_k) and (f_l) and 3 random realizations of γ.



(a) Lifting & nuclear norm (b) Projected gradient (c) Alternate minimization

Figure 8: Percentages of perfect recovery results for the different algorithms.

projected gradient descent seems to provide better results for a wider range of parameters. A theoretical analysis of this idea might be worth an exploration. Unfortunately, no algorithm is able to succeed systematically. This might be related to the fact that the random positions $(\bar{x}_n)$ are badly located.

In this setting, the condition for global injectivity (35) reads:

$$N \geq \frac{2KL-4}{K-2},$$

i.e. $N \geq 6L-4$ for $K=3$. We see the shortcomings of this rule in this experiment: perfect recovery does not always occur when this condition is satisfied, because the condition does not certify the success of an algorithm. And the algorithms manage to recover some operators when this condition is not met. However, it is clear that a necessary condition for identifiability is $N \geq L$, since otherwise, even the problem with known weights cannot be identified.

4.3 A 2D experiment

To end this paper, we briefly describe how the experiment from Figure 1 was carried out. We generated a family of product-convolution operators with astigmatic impulse responses as follows. We set the family (e_k) as anisotropic Gaussian vectors with $K=8$. We also set the family (f_l) as the monomials 1, x and y, resulting in $L=3$ basis elements to describe the space variations.

We launched the maximum correlation algorithm to locate the beads positions in Figure 1b. The average localization error is 0.015 pixel, despite a significant amount of additive Gaussian noise. We then discarded by hand the locations that were too close from each other (red stars). Notice that this part can be easily automatized by thresholding the minimal distance between adjacent locations. We kept the other locations (blue stars) to estimate the operator, resulting in a total of $N=27$ impulse responses with a slightly inaccurate localization. We then used this information to recover the operator. Here we assumed that the weight $(\bar{w}_n)$ were known and all equal to 1. The

relative error between the operator estimated and the true one is 0.006. Here, we measured the distance between operators with the Hilbert-Schmidt norm. The whole process takes less than a second on a usual personal computer with Matlab.

Acknowledgments

The authors wish to thank Kristian Bredies for his careful analysis of the paper, which allowed to correct some flaws in the concentration results. The authors wish to thank Landry Duguet for his initial exploration of the problem during a one month internship. P. Weiss thanks T. Rezk for fruitful discussions.

This work was supported by Fondation pour la Recherche Médicale (FRM grant number ECO20170637521 to V. Debarnot), by the ANR project Optimization on Measures Spaces ANR-17-CE23-0013-01 and from ANR-3IA Artificial and Natural Intelligence Toulouse Institute.

5 Proofs

5.1 Proofs of the propositions from Section 2.3

Proof of Proposition 2.1

Proof. We show that for all $(x,\gamma),(x',\gamma') \in \mathbb{R}^D \times \mathbb{R}^I \backslash \{0\}$

$$E(x)\gamma = E(x')\gamma' \Rightarrow \gamma = \gamma'.$$

If $x = x'$: by Assumption 2.3 the mapping $\gamma \mapsto E(x)\gamma$ is injective, meaning that for all $\gamma,\gamma' \in \mathbb{R}^I \backslash \{0\}$ we have

$$E(x)\gamma = E(x)\gamma' \Rightarrow (x,\gamma) = (x,\gamma').$$

If $x \neq x'$: by Assumption 2.4 we have $R(x) \cap R(x') = \{0\}$, which implies that for any $\gamma,\gamma' \in \mathbb{R}^I \backslash \{0\}$, we have $E(x)\gamma \neq E(x')\gamma'$. We prove it by contradiction. Assume that $R(x) \cap R(x') \neq \{0\}$. Then, there exists $\gamma \in \mathbb{R}^I \backslash \{0\}$ such that $E(x)\gamma \in R(x')$. Let denote $v \overset{\text{def.}}{=} \frac{E(x)\gamma}{\|E(x)\gamma\|_2\|}$. Then $\|\Pi_{R(x')}\Pi_{R(x)}\|_{2\to 2} \geq \|\Pi_{R(x')}v\|_2 = \|v\|_2 = 1$. Nevertheless, by assumption 2.4 we have $\|\Pi_{R(x')}\Pi_{R(x)}\|_{2\to 2} < 1$ for $x \neq x'$. This concludes the proof. □

Proof of Proposition 2.2

Proof. We have:

$$\begin{aligned}
&\|\Pi_{R(x)}\Pi_{R(\bar{x})}\|_{2\to 2}^2 \\
&= \sup_{y\in\mathbb{R}^M, \|y\|_2=1} \|\Pi_{R(x)}\Pi_{R(\bar{x})}y\|_2^2 = \sup_{y\in\mathbb{R}^M, \|y\|_2=1} \langle \Pi_{R(x)}\Pi_{R(\bar{x})}y, \Pi_{R(x)}\Pi_{R(\bar{x})}y\rangle \\
&= \sup_{y\in\mathbb{R}^M, \|y\|_2=1} \langle \Pi_{R(x)}\Pi_{R(\bar{x})}y, y\rangle = \sup_{y\in\mathbb{R}^M, \|y\|_2=1} \langle \Pi_{R(\bar{x})}y, \Pi_{R(x)}y\rangle \\
&= \sup_{y\in R(x), \|y\|_2=1} \langle \Pi_{R(\bar{x})}y, y\rangle \leq \sup_{\substack{y\in R(x), \bar{y}\in R(\bar{x}) \\ \|y\|_2=1, \|\bar{y}\|_2=1}} \langle \bar{y}, y\rangle = \cos\left(\angle\left(R(x), R(\bar{x})\right)\right).
\end{aligned}$$

□

Proof of Proposition 2.3

Proof. We have $[E(z)]_{m,i} = \langle e_i(\cdot - z), \delta_{z_m} \rangle = e_i(z_m - z)$. Hence

$$[E^*(z)E(z)]_{i,i'} = \sum_{m \in \mathbb{N}} e_i(z_m - z) e_{i'}(z_m - z) = \begin{cases} 1 & \text{if } i = i' \\ 0 & \text{otherwise,} \end{cases}$$

where we used (11) to obtain the last equality. □

Proof of Proposition 2.4

Proof. By Assumption 2.5, we have $\Pi_{R(x)} = E(x)E^*(x)$. Then, by definition, we have

$$\begin{aligned} \|\Pi_{R(x)}\Pi_{R(x')}\|_{2\to 2} &= \|E(x)E^*(x)E(x')E^*(x')\|_{2\to 2} \\ &= \sup_{u \in \ell^2(\mathbb{Z}^D), \|u\|_{\ell^2}=1} \|E(x)E^*(x)E(x')E^*(x')u\|_2 \\ &= \sup_{v \in R^I, \|v\|_2=1} \|E(x)E^*(x)E(x')v\|_2 \\ &= \|E(x)E^*(x)E(x')\|_{2\to 2} = \|E^*(x)E(x')\|_{2\to 2} \\ &= \|\mathcal{M}(x - x')\|_{2\to 2}, \end{aligned}$$

using the fact that $E^*(x)E(x) = \mathrm{Id}$. □

Proof of Proposition 2.5

Proof. The first part of the proof is trivial: the range $R(x)$ of $E(x)$ is unchanged.

As for the second part, it suffices to realize that E contains columns which are colinear. □

5.2 Proof of Theorem 3.1

Proof. For a fixed $X \in \mathbb{R}^{D \times N}$, any optimal coefficient matrix $\gamma(X)$ is characterized by the linear $I \times I$ system

$$\left(\sum_{n=1}^{N} \bar{w}_n^2 E^*(x_n)E(x_n) \right) \gamma = \sum_{n=1}^{N} \bar{w}_n E^*(x_n) y_n. \tag{38}$$

Letting

$$\mathcal{E}(X) = \begin{pmatrix} \bar{w}_1 E(x_1) \\ \vdots \\ \bar{w}_N E(x_N) \end{pmatrix} \text{ and } Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix},$$

this system can be rewritten compactly as

$$\mathcal{E}^*(X)\mathcal{E}(X)\gamma = \mathcal{E}^*(X)Y.$$

Letting $\mathcal{E}^+(X)$ denote the Moore-Penrose pseudo-inverse of $\mathcal{E}(X)$, we can for instance choose $\gamma(X) = \mathcal{E}^+(X)Y$, leading to

$$\begin{aligned} G(X) &\overset{\text{def.}}{=} F(\gamma(X), X) = \frac{1}{2}\|[\mathcal{E}^+(X)\mathcal{E}^*(X) - \mathrm{Id}]Y\|_2^2 = \frac{1}{2}\|[\Pi_{\mathrm{Ran}(\mathcal{E}(X))} - \mathrm{Id}]Y\|_2^2 \\ &= \frac{1}{2}\|\Pi_{\mathrm{Ran}(\mathcal{E}(X))}Y\|_2^2 + \frac{1}{2}\|Y\|_2^2 - \langle \Pi_{\mathrm{Ran}(\mathcal{E}(X))}Y, Y\rangle \\ &= \frac{1}{2}\|Y\|_2^2 - \frac{1}{2}\|\Pi_{\mathrm{Ran}(\mathcal{E}(X))}Y\|_2^2. \end{aligned}$$

Hence, minimizing G amounts to maximizing the quadratic form

$$H(X) \overset{\text{def.}}{=} \frac{1}{2}\langle \Pi_{\mathrm{Ran}(\mathcal{E}(X))}Y, Y\rangle = \frac{1}{2}\sum_{n=1}^{N}\langle \Pi_{\mathrm{Ran}(E(x_n))}y_n, y_n\rangle,$$

which boils down to N independent subproblems. Under Assumption 2.3, the mapping $E(x)$ is injective for all $x \in \mathbb{R}^D$. Hence, $E^*(x)E(x)$ is a positive definite matrix and so is $\mathcal{E}^*(X)\mathcal{E}(X)$. This and the fact that $\bar{w}_n \neq 0$ justifies that $\gamma(X)$ is unique for any points configuration X. □

5.3 Proof of Theorem 3.2

We start with a simple lemma.

Lemma 5.1. *Let $D \in \mathbb{N}$, $f : \mathbb{R}^D \to \mathbb{R}$ and $\epsilon : \mathbb{R}^D \to \mathbb{R}$. Define $g \overset{\text{def.}}{=} f + \epsilon$ and assume that the following sets are non-empty*

$$\hat{X} = \operatorname*{argmax}_{x\in\mathbb{R}^D} g(x) \quad \text{and} \quad \bar{X} = \operatorname*{argmax}_{x\in\mathbb{R}^D} f(x).$$

Further assume that $\|\epsilon\|_\infty \leq \eta$ for some $\eta > 0$ and that there exists an increasing function $\varphi : \mathbb{R} \to \mathbb{R}_+$ such that

$$f(x) \leq f(\bar{x}) - \varphi(\|x - \bar{x}\|_2),\ \forall x \in \mathbb{R}^D. \tag{39}$$

Then $\bar{X} = \{\bar{x}\}$ is a singleton and any $\hat{x} \in \hat{X}$ satisfies $\|\hat{x} - \bar{x}\|_2 \leq \varphi_+^{-1}(2\eta)$.

Proof. By inequality (39) and strict monotonicity of ϕ, we have $f(\bar{x}) > f(x)$ for all $x \neq \bar{x}$. Hence $\bar{x}$ is the unique maximizer of f. We have

$$g(\hat{x}) \geq g(\bar{x}) = f(\bar{x}) - \epsilon(\bar{x}) \geq f(\bar{x}) - \eta. \tag{40}$$

In addition, for any $x \in \mathbb{R}^D$, we have

$$\begin{aligned} g(x) &= f(x) + \epsilon(x) \leq f(x) + \eta \\ &\leq f(\bar{x}) - \varphi(\|x - \bar{x}\|_2) + \eta. \end{aligned}$$

For any x such that $\|x - \bar{x}\|_2 > \varphi_+^{-1}(2\eta)$, we have $g(x) < f(\bar{x}) - \eta$, and thus $g(x) < g(\bar{x})$, which implies that $x \neq \hat{x}$. The contraposition is that any $\hat{x} \in \hat{X}$ satisfies $\|\hat{x} - \bar{x}\|_2 \leq \varphi_+^{-1}(2\eta)$. □

Proof. i) Let $y_{0,n} = \bar{w}_n E(\bar{x}_n)\bar{\gamma} = y_n - b_n$ denote the noiseless measurements and let $F_{0,n}(x,\gamma) = \frac{1}{2}\|\bar{w}_n E(x)\gamma - y_{0,n}\|_2^2$. We have $F_{0,n}(\bar{x}_n, \bar{\gamma}) = 0$. Let $\gamma_0(x) \in \operatorname{argmin}_{\gamma\in\mathbb{R}^I} F_{0,n}(x,\gamma)$ denote any minimizer (for instance the one given by the pseudo-inverse). Now, define $G_{0,n}(x) \overset{\text{def.}}{=} F_{0,n}(x, \gamma_0(x))$. We have $G_{0,n}(x) = \frac{1}{2}\left(\|y_{0,n}\|_2^2 - \|\Pi_{R(x)}y_{0,n}\|_2^2\right)$, by an argument similar to the one in the proof of Theorem 3.1. By Assumption 2.4, $R(x) \cap R(\bar{x}) = \{0\}$ for $x \neq \bar{x}$. Hence, $\|\Pi_{R(x)}y_{0,n}\|_2^2 < \|y_{0,n}\|_2^2$ for $x \neq \bar{x}$ and $\bar{x}_n$ is the unique minimizer of $G_{0,n}$. Therefore, the function

$$H_{0,n}(x) \overset{\text{def.}}{=} \frac{1}{2}\|\Pi_{R(x)}y_{0,n}\|_2^2 = \frac{1}{2}\left\langle \Pi_{R(x)}y_{0,n}, y_{0,n}\right\rangle,$$

also admits a unique maximizer in $\bar{x}$. Overall, we see that under Assumption 2.4, $G_{0,n}$ admits a unique minimizer equal to $\bar{x}$. Under the additional Assumption 2.3, $F_{0,n}$ admits a unique solution $(\bar{x}, \bar{\gamma})$.

ii) Now, let $F_n(x,\gamma) = \frac{1}{2}\|\bar{w}_n E(x)\gamma - y_{0,n}\|_2^2$, $\gamma(x)$ denote any minimizer of F_n w.r.t. γ, $G_n(x) = F_n(x,\gamma(x))$ and $H_n(x) \overset{\text{def.}}{=} \frac{1}{2}\langle \Pi_{R(x)}y_n, y_n\rangle$. Let $\hat{x}$ denote any maximizer of H_n and assume for now that we manage to obtain a bound of the form $|H_n(x) - H_{0,n}(x)| \leq \eta$ for some $\eta \geq 0$. We have

$$\begin{aligned}
H_{0,n}(x) &= \frac{1}{2}\left\langle \Pi_{R(x)}y_{0,n}, y_{0,n}\right\rangle = \frac{1}{2}\left\langle \Pi_{R(x)}y_{0,n}, \Pi_{R(\bar{x})}y_{0,n}\right\rangle \\
&= \frac{1}{2}\left\langle \Pi_{R(\bar{x})}\Pi_{R(x)}y_{0,n}, y_{0,n}\right\rangle \leq \frac{1}{2}\left\|\Pi_{R(\bar{x})}\Pi_{R(x)}\right\|_{2\to 2}\|y_{0,n}\|_2^2 \\
&\leq \frac{1}{2}\left(1 - \phi(\|x-\bar{x}\|_2)\right)\|y_{0,n}\|_2^2 = H_{0,n}(\bar{x}) - \frac{1}{2}\phi(\|x-\bar{x}\|_2)\|y_{0,n}\|_2^2.
\end{aligned}$$

by Assumption 2.4. Hence, we can use Lemma 5.1 with $f(x) = H_{0,n}(x)$, $g(x) = H_n(x)$ and $\varphi(r) = \frac{1}{2}\phi(r)\|y_{0,n}\|_2^2$. This allows us to conclude that

$$\|\hat{x} - \bar{x}\|_2 \leq \phi_+^{-1}\left(\frac{4\eta}{\|y_{0,n}\|_2^2}\right). \tag{41}$$

iii) The last remaining point is to control $\|H_{0,n} - H_n\|_\infty$. We have

$$\begin{aligned}
H_n(x) &= \frac{1}{2}\langle \Pi_{R(x)}\left(y_{0,n} + b_n\right), y_{0,n} + b_n\rangle \\
&= H_{0,n}(x) + \langle \Pi_{R(x)}y_{0,n}, b_n\rangle + \frac{1}{2}\|\Pi_{R(x)}(b_n)\|_2^2.
\end{aligned}$$

Hence, using Cauchy-Schwarz inequality, we obtain for all x

$$|H_n(x) - H_{0,n}(x)| \leq \|\Pi_{R(x)}y_{0,n}\|_2\|\Pi_{R(x)}b_n\|_2 + \frac{1}{2}\|\Pi_{R(x)}(b_n)\|_2^2.$$

Using the facts that $\|\Pi_{R(x)}b_n\|_2 \leq \|b_n\|_2$ and that $\|b_n\|_2 \leq \theta\|y_{0,n}\|_2$, we obtain

$$\|H_n - H_{0,n}\|_\infty \leq \|y_{0,n}\|_2^2\left(\theta + \frac{1}{2}\theta^2\right).$$

For the inequality (41) to make sense, we need $4(\theta + \frac{1}{2}\theta^2) \leq 1$, which is equivalent to $\theta < \frac{\sqrt{6}}{2} - 1$ and Theorem 3.2 is proven. □

5.4 Proof of Proposition 3.3

Proof. This proposition derives from point iii) in the previous proof. The inequality (41) can be improved as

$$\|\hat{x} - \bar{x}\|_2 \leq \phi_+^{-1}\left(\frac{4\inf_{c\in\mathbb{R}}\|H - H_0 - c\|_\infty}{\|y_0\|_2^2}\right), \tag{42}$$

since a constant term does not affect the location of a minimizer. We have

$$\Delta(x) \overset{\text{def.}}{=} H(x) - H_0(x) = \langle \Pi_{R(x)} y_0, \Pi_{R(x)} b\rangle + \frac{1}{2}\|\Pi_{R(x)}(b)\|_2^2.$$

It is therefore natural to set $c = \frac{1}{2}\left(\sup_{x\in\mathbb{R}^D}\Delta(x) - \inf_{x\in\mathbb{R}^D}\Delta(x)\right)$, to minimize in the infinite norm in Problem (42).

□

5.5 Proof of Proposition 3.4

Proof. First notice that

$$\begin{aligned}
\mathrm{Ampl}(\Delta_1 + \Delta_2) &= \sup_{x\in\mathbb{R}^D} \Delta_1(x) + \Delta_2(x) - \inf_{x\in\mathbb{R}^D} \Delta_1(x) + \Delta_2(x) \\
&\leq \sup_{x\in\mathbb{R}^D} \Delta_1(x) + \sup_{x\in\mathbb{R}^D} \Delta_2(x) - \left(\inf_{x\in\mathbb{R}^D} \Delta_1(x) + \inf_{x\in\mathbb{R}^D} \Delta_2(x)\right) \\
&= \mathrm{Ampl}(\Delta_1) + \mathrm{Ampl}(\Delta_2).
\end{aligned}$$

We will treat the two random processes Δ_1 and Δ_2 separately.

i) Consider the function $f_1(b) \overset{\text{def.}}{=} \sup_{x\in\mathbb{R}^D}\langle \Pi_{R(x)} y_0, b\rangle$ and define the random variable $V_1^+ = f_1(b)$ with mean $\bar{V}_1^+$. Similarly, define $V_1^- = \inf_{x\in\mathbb{R}^D}\langle \Pi_{R(x)} y_0, b\rangle$. In addition, notice that $Z_1 = \mathrm{Ampl}(\Delta_1) \leq V_1^+ - V_1^-$. We first show that f_1 is Lipschitz continuous. We have

$$\begin{aligned}
f_1(b+\epsilon) &= \sup_{x\in\mathbb{R}^D}\langle \Pi_{R(x)} y_0, b+\epsilon\rangle \leq \sup_{x\in\mathbb{R}^D}\langle \Pi_{R(x)} y_0, b\rangle + \|y_0\|_2\|\epsilon\|_2 \\
f_1(b+\epsilon) &= \sup_{x\in\mathbb{R}^D}\langle \Pi_{R(x)} y_0, b+\epsilon\rangle \geq \sup_{x\in\mathbb{R}^D}\langle \Pi_{R(x)} y_0, b\rangle - \|y_0\|_2\|\epsilon\|_2
\end{aligned}$$

Hence, $|f_1(b) - f_1(b+\epsilon)| \leq \|y_0\|_2\|\epsilon\|_2$ and f_1 is $\|y_0\|_2$-Lipschitz continuous. Using a Gaussian logarithmic Sobolev inequality [40, Thm 5.6], we obtain that V_1^+ is sub-Gaussian with

$$\mathbb{P}\left(|V_1^+ - \bar{V}_1^+| \geq t\right) \leq 2\exp(-t^2/(2\sigma^2\|y_0\|^2)).$$

The same result holds for V_1^-. Finally, the sum of two dependent sub-Gaussian variables with parameters σ_1 and σ_2 is sub-Gaussian with a sub-Gaussian parameter smaller than $\sigma_1 + \sigma_2$, so that

$$\mathbb{P}\left(|Z_1 - \bar{Z}_1| \geq t\right) \leq 2\exp(-t^2/(8\sigma^2\|y_0\|^2)).$$

ii) Now, define the random variable $Y_2^+ \overset{\text{def.}}{=} \sup_{x\in\mathbb{R}^D}\|\Pi_{R(x)} b\|_2$ and $V_2^+ \overset{\text{def.}}{=} \frac{1}{2}(Y_2^+)^2$. The function $b \mapsto \sup_{x\in\mathbb{R}^D}\|\Pi_{R(x)} b\|_2$ is 1-Lipschitz continuous. Hence

using a Gaussian logarithmic Sobolev inequality again, we obtain that Y_2^+ is sub-Gaussian with

$$\mathbb{P}\left(|Y_2^+ - \bar{Y}_2^+| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2\sigma^2}\right).$$

Using [27, Lemma 2.7.6], we conclude that V_2^+ is sub-exponential and satisfies

$$\mathbb{P}\left(|V_2^+ - \bar{V}_2^+| \geq t\right) \leq 2\exp(-Ct/\sigma^2),$$

for an absolute constant C. We can make a similar proof for the random variable $Y_2^- \overset{\text{def.}}{=} \inf_{x\in\mathbb{R}^D} \|\Pi_{R(x)}b\|_2$ and conclude as in the previous proof. □

5.6 Proof of Theorem 3.5

Controlling $\bar{Z}_1$. Here, we wish to control the supremum of the centered Gaussian process Δ_1. A traditional approach to bound it consists in computing Dudley's entropy integral, see e.g. [40, Corollary 13.2]. To this end, we introduce the pseudo-metric:

$$d(x,x') \overset{\text{def.}}{=} \sqrt{\mathbb{E}\left((\Delta_1(x) - \Delta_1(x'))^2\right)}. \tag{43}$$

Let $B(c,\delta) = \{x \in \mathbb{R}^D, d(c,x) \leq \delta\}$ denote a ball of radius δ centered at c with respect to d. Let $S \subset \mathbb{R}$ denote a set and define the covering number $N(\delta, S)$ as the minimum number of δ-balls needed to cover S. We then have

$$\mathbb{E}\left(\sup_{x\in\mathbb{R}} \Delta_1(x)\right) \leq 12\int_0^{\eta/2} \sqrt{\log(N(u,\mathbb{R}))}\,du \text{ with } \eta = \inf_{t\in\mathbb{R}}\sup_{t'\in\mathbb{R}} d(t,t'). \tag{44}$$

Using that $y_0 \in R(\bar{x})$, we have

$$\begin{aligned} d(x,x')^2 &= \mathbb{E}\left((\Delta_1(x) - \Delta_1(x'))^2\right) \\ &= \mathbb{E}\left((\langle\left(\Pi_{R(x)} - \Pi_{R(x')}\right)\Pi_{R(\bar{x})}y_0, b\rangle)^2\right) \\ &\leq \sigma^2\|y_0\|_2^2\|\left(\Pi_{R(x)} - \Pi_{R(x')}\right)\Pi_{R(\bar{x})}\|_{2\to 2}^2. \end{aligned}$$

Thus

$$\begin{aligned} \|\left(\Pi_{R(x)} - \Pi_{R(x')}\right)\Pi_{R(\bar{x})}\|_{2\to 2} &\leq \|\Pi_{R(x)}\Pi_{R(\bar{x})}\|_{2\to 2} + \|\Pi_{R(x')}\Pi_{R(\bar{x})}\|_{2\to 2} \\ &\leq \frac{2}{1+\beta\min\left(\|x-\bar{x}\|_2, \|x'-\bar{x}\|_2\right)} \end{aligned}$$

This leads to

$$d(x,x') \leq \frac{2\sigma\|y_0\|_2}{1+\beta\min\left(\|x-\bar{x}\|_2, \|x'-\bar{x}\|_2\right)} \tag{45}$$

and the Lipschitz continuity also implies that

$$d(x,x') \leq \sigma\|y_0\|_2 L\|x-x'\|_2. \tag{46}$$

In what follows, we let

$$\tilde{d}(x,x') \overset{\text{def.}}{=} \sigma\|y_0\|_2\min\left(L\|x-x'\|_2, \frac{2}{1+\beta\min\left(\|x-\bar{x}\|_2, \|x'-\bar{x}\|_2\right)}\right) \tag{47}$$

and $\tilde{N}$ denote the corresponding covering number. The inequality $d(x,x') \leq \tilde{d}(x,x')$ implies that $N(\delta,S) \leq \tilde{N}(\delta,S)$ for all S and δ.

Without loss of generality, we assume that $\bar{x}=0$. Now, let $c = \frac{2\sigma\|y_0\|_2}{\beta\delta}$ and remark that the decay condition (45) implies that all x with $|x| \geq c$ belong to the ball $B(c,\delta)$. Hence

$$\tilde{N}(\delta,\mathbb{R}) \leq 1 + \tilde{N}(\delta,[-c,c]).$$

The second inequality (46) implies that the δ-balls have a diameter no smaller than $\frac{2\delta}{L\sigma\|y_0\|_2}$. Hence, the interval $[-c,c]$ is covered by at most $\left\lceil 2c \cdot \frac{L\sigma\|y_0\|_2}{2\delta} \right\rceil$ δ-balls, leading to

$$N(\delta,\mathbb{R}) \leq \tilde{N}(\delta,\mathbb{R}) \leq 2\frac{L}{\beta}\left(\frac{\sigma\|y_0\|_2}{\delta}\right)^2 + 2.$$

We have $\eta = \inf_{t\in\mathbb{R}} \sup_{t'\in\mathbb{R}} d(t,t') \leq 2\sigma\|y_0\|_2$. Hence:

$$\begin{aligned}&\mathbb{E}\left(\sup_{x\in\mathbb{R}} \Delta_1(x)\right)\\ &\leq 12\int_0^{\sigma\|y_0\|_2} \sqrt{\log\left(2\frac{L}{\beta}\left(\frac{\sigma\|y_0\|_2}{u}\right)^2 + 2\right)}\,du.\end{aligned}$$

For all $u \in [0,\sigma\|y_0\|_2]$, we have

$$2\frac{L}{\beta}\left(\frac{\sigma\|y_0\|_2}{u}\right) + 2 \leq \left(1+\frac{\beta}{L}\right)2\frac{L}{\beta}\left(\frac{\sigma\|y_0\|_2}{u}\right).$$

Hence

$$\begin{aligned}\mathbb{E}\left(\sup_{x\in\mathbb{R}} \Delta_1(x)\right) &\leq 12\int_0^{\sigma\|y_0\|_2} \sqrt{\log\left[\left(1+\frac{\beta}{L}\right)2\frac{L}{\beta}\left(\frac{\sigma\|y_0\|_2}{u}\right)^2\right]}\,du\\ &= 12\sigma\|y_0\|_2\int_0^1 \sqrt{\log\left(\left(1+\frac{\beta}{L}\right)2\frac{L}{\beta}\right) - 2\log(u)}\,du\\ &\leq 12\sigma\|y_0\|_2\left[2\sqrt{\pi\left(\frac{L}{\beta}+1\right)} + \sqrt{\log\left(2\frac{L}{\beta}+2\right)}\right],\end{aligned}$$

where we skipped the elementary (but technical) details to obtain the last bound. To conclude, we use the fact that $\bar{Z}_1 \leq 2\mathbb{E}\left(\sup_{x\in\mathbb{R}} \Delta_1(x)\right)$. □

Controlling $\bar{Z}_2$. Controlling $\bar{Z}_2$ amounts to studying the supremum of a so-called Gaussian chaos of order 2. This problem arises in different fields and has been addressed using tools of generic chaining in [41, 42, 43, 40]. We will make use of the following theorem, which has been rewritten using our notation.

Theorem 5.2 (Theorem 3.1 in [43]). *Let $\mathcal{A} = \{\Pi_{R(x)}, x \in [-1,1]\}$ denote the family of projectors. Let $d_{2\to2}(\mathcal{A}) = \sup_{A\in\mathcal{A}} \|A\|_{2\to2}$ and $d_F(\mathcal{A}) = \sup_{A\in\mathcal{A}} \|A\|_F$. Let us define the following Dudley integral*

$$\gamma = c_1 \int_0^{d_{2\to2}(\mathcal{A})} \sqrt{\log N(u,\mathcal{A})}\,du,$$

where the covering number is evaluated w.r.t. the spectral distance $\|\cdot\|_{2\to 2}$. *Then*

$$\mathbb{P}\left(\sup_{x\in[-1,1]}|\Delta_2(x)-\mathbb{E}(\Delta_2(x))|\geq c_2\sigma^2E+t\right)\leq 2\exp\left(-c_3\min\left(\frac{t^2}{\sigma^4V^2},\frac{t}{\sigma^2U}\right)\right),$$

with

$$E=\gamma^2+\gamma d_F(\mathcal{A})+d_F(\mathcal{A})d_{2\to 2}(\mathcal{A}),\quad U=d_{2\to 2}^2(\mathcal{A}),\quad V=d_{2\to 2}(\mathcal{A})\left(\gamma+d_F(\mathcal{A})\right).$$

and where c_1,c_2,c_3 *are absolute constants.*

In what follows, the constants c_1,c_2,c_3 may change from line to line, but are always absolute. The fact that $R(x)$ is constant outside $[-1,1]$ allows to restrict our study to this interval. In addition, notice that

$$\mathrm{Ampl}(\Delta_2)\leq 2\sup_{x\in[-1,1]}|\Delta_2(x)-\mathbb{E}(\Delta_2(x))|.$$

Under the Assumptions of Theorem 3.5, we have $d_{2\to 2}(\mathcal{A})=1$ (since each matrix $\Pi_{R(x)}$ is a projector) and $d_F(\mathcal{A})=\sqrt{I}$ (since $\dim(R(x))=I$). In addition, the Lipschitz continuity assumption (24) yields $N(u,\mathcal{A})\leq\frac{L}{u}$ for $L\geq 1$ and $u\leq 1$. Hence, we get that

$$\begin{aligned}\gamma&=c_1\int_0^{d_{2\to 2}(\mathcal{A})}\sqrt{\log N(u,\mathcal{A})}\,du &&\leq c_1\int_0^1\sqrt{\log\frac{L}{u}}\,du\\&=c_1\left(\sqrt{\log(L)}+\frac{\sqrt{\pi}}{2}L\cdot\mathrm{erfc}\left(\sqrt{\log(L)}\right)\right) &&\leq c_1\left(\sqrt{\log(L)}+\frac{1}{2\sqrt{\log(L)}}\right),\end{aligned}$$

where we used the inequality $\mathrm{erfc}(z)<\frac{\exp(-z^2)}{\sqrt{\pi}z}$ to obtain the last result. Hence, we obtain:

$$E\leq c_1\cdot\left(\log(L)+\sqrt{\log(L)I}+\sqrt{I}\right).$$

$$V\leq c_1\cdot\left(\sqrt{\log(L)}+\sqrt{I}\right)\text{ and }U=1.$$

Overall, we see that

$$\mathbb{P}\left(\mathrm{Ampl}(\Delta_2)\geq c_1\sigma^2E+t\right)\leq 2\exp\left(-c_2\min\left(\frac{t^2}{\sigma^4(\log(L)+I)},\frac{t}{\sigma^2}\right)\right).$$

This yields

$$\mathbb{E}\left(\mathrm{Ampl}(\Delta_2)\right)=\bar{Z}_2\leq c\cdot\sigma^2\left(\log(L)+\sqrt{\log(L)I}+\sqrt{I}\right).$$

□

5.7 Proof of Theorem 3.6

Proof. Let $P(x)\overset{\text{def.}}{=}E^*(x)E(x)$. By definition, we have

$$\hat{\gamma}=P(\hat{x})^{-1}E^*(\hat{x})(y_{0,1}+b_1)\tag{48}$$

We have $E(\hat{x}) = E(\bar{x}) + \Delta$ with $\|\Delta\|_{2\to 2} \leq \sqrt{\sigma_+} L_E \|\hat{x} - \bar{x}\|_2$. Hence $P(\hat{x}) = P(\bar{x}) + \Delta'$ with

$$\|\Delta'\|_{2\to 2} \leq 2L_E \|\hat{x} - \bar{x}\|_2 \sqrt{\sigma_+} \|E(\bar{x})\|_{2\to 2} + \sigma_+ L_E^2 \|\hat{x} - \bar{x}\|_2^2. \tag{49}$$

The linear system to recover $\hat{\gamma}$ is then

$$(P(\bar{x}) + \Delta')\hat{\gamma} = (E^*(\bar{x}) + \Delta)(y_{0,1} + b_1) = E^*(\bar{x}) y_{0,1} + \delta \tag{50}$$

with $\delta = \Delta(y_{0,1} + b_1) + E^*(\bar{x}) b_1$. Under Assumption 2.3, the unique solution of $P(\bar{x})\gamma = E^*(\bar{x}) y_{0,1}$ is $\bar{\gamma}$. If $\hat{x}$ is sufficiently close to $\bar{x}$, we have $\|\Delta'\|_{2\to 2} < \sigma_-$ and $\|P(\bar{x})^{-1}\Delta'\|_{2\to 2} < 1$. We can now use standard results of linear algebra, see e.g. [44, 3.6], to obtain that

$$\begin{aligned}
&\frac{\|\hat{\gamma} - \bar{\gamma}\|_2}{\|\bar{\gamma}\|_2} \\
&\leq \frac{\|P(\bar{x})\|_{2\to 2} \|P(\bar{x})^{-1}\|_{2\to 2}}{1 - \|P(\bar{x})^{-1}\Delta'\|_{2\to 2}} \left(\frac{\|\Delta'\|_{2\to 2}}{\|P(\bar{x})\|_{2\to 2}} + \frac{\|\delta\|_2}{\|E^*(\bar{x}) y_{0,1}\|_2} \right) \\
&\leq \frac{\sigma_+}{\sigma_-} \frac{1}{\left(1 - \frac{\|\Delta'\|_{2\to 2}}{\sigma_-}\right)} \left(\frac{\|\Delta'\|_{2\to 2}}{\sigma_-} + \frac{\|\Delta\|_{2\to 2}(\|y_{0,1}\|_2 + \|b_1\|_2) + \sqrt{\sigma_+}\|b_1\|_2}{\sqrt{\sigma_-}\|y_{0,1}\|_2} \right)
\end{aligned}$$

Letting $\hat{x} \to \bar{x}$, we have $\frac{1}{\left(1 - \frac{\|\Delta'\|_{2\to 2}}{\sigma_-}\right)} \sim 1 + \frac{\|\Delta'\|_{2\to 2}}{\sigma_-}$.

Using that

$$\|\Delta'\|_{2\to 2} \leq 2L_E \sqrt{\sigma_+} \|E(\bar{x})\|_{2\to 2} \|\hat{x} - \bar{x}\|_2 + o_{\hat{x}\to\bar{x}}\left(\|\hat{x} - \bar{x}\|_2^2\right),$$

and

$$\|\Delta\|_{2\to 2} \leq L_E \sqrt{\sigma_+} \|\hat{x} - \bar{x}\|_2,$$

we have

$$\begin{aligned}
&\frac{\sigma_+}{\sigma_-} \frac{1}{\left(1 - \frac{\|\Delta'\|_{2\to 2}}{\sigma_-}\right)} \left(\frac{\|\Delta'\|_{2\to 2}}{\sigma_-} + \frac{\|\Delta\|_{2\to 2}(\|y_{0,1}\|_2 + \|b_1\|_2) + \sqrt{\sigma_+}\|b_1\|_2}{\sqrt{\sigma_-}\|y_{0,1}\|_2} \right) \\
&= \frac{\sigma_+}{\sigma_-} \left(1 + \frac{\|\Delta'\|_{2\to 2}}{\sigma_-} + o_{\hat{x}\to\bar{x}}\left(\|\hat{x} - \bar{x}\|_2^2\right)\right) \Big(\frac{\|\Delta'\|_{2\to 2}}{\sigma_-} \\
&+ \frac{\|\Delta\|_{2\to 2}(\|y_{0,1}\|_2 + \|b_1\|_2) + \sqrt{\sigma_+}\|b_1\|_2}{\sqrt{\sigma_-}\|y_{0,1}\|_2} \Big) \\
&\leq \kappa \|\hat{x} - \bar{x}\|_2 \Big(\frac{2L_E\sqrt{\sigma_+}\|E(\bar{x})\|_{2\to 2}}{\sigma_-} + \frac{L_E\sqrt{\sigma_+}}{\sqrt{\sigma_-}}(1 + \frac{\|b_{0,1}\|_2}{\|y_{0,1}\|_2}) \\
&+ \sqrt{\kappa}\frac{\|b_{0,1}\|_2}{\|y_{0,1}\|_2} 2L_E \frac{\sqrt{\sigma_+}\|E(\bar{x})\|_{2\to 2}}{\sigma_-} \Big) + \kappa^{3/2} \frac{\|b_1\|_2}{\|y_{0,1}\|_2} + o_{\hat{x}\to\bar{x}}\left(\|\hat{x} - \bar{x}\|_2^2\right) \\
&\leq C\kappa^{5/2} L_E \|\hat{x} - \bar{x}\|_2 \left(1 + \frac{\|b_1\|_2}{\|y_{0,1}\|_2}\right) + o_{\hat{x}\to\bar{x}}\left(\|\hat{x} - \bar{x}\|_2^2\right)
\end{aligned}$$

since $\kappa \geq 1$, and where C denote a universal constant. Let

$$\epsilon_2 \stackrel{\text{def.}}{=} C\kappa^{5/2} L_E \|\hat{x} - \bar{x}\|_2 \left(1 + \frac{\|b_1\|_2}{\|y_{0,1}\|_2}\right) + o_{\hat{x}\to\bar{x}}\left(\|\hat{x} - \bar{x}\|_2^2\right),$$

this concludes the proof. □

Proof. Let $P(x) \overset{\text{def.}}{=} E^*(x)E(x)$. By definition, we have

$$\hat{\gamma} = P(\hat{x})^{-1}E^*(\hat{x})(y_{0,1} + b_1) \tag{51}$$

We have $E(\hat{x}) = E(\bar{x}) + \Delta$ with $\|\Delta\|_{2\to 2} \leq \sqrt{\sigma_+}L_E\|\hat{x} - \bar{x}\|_2$. Hence $P(\hat{x}) = P(\bar{x}) + \Delta'$ with

$$\|\Delta'\|_{2\to 2} \leq 2L_E\|\hat{x} - \bar{x}\|_2\sqrt{\sigma_+}\|E(\bar{x})\|_{2\to 2} + \sigma_+ L_E^2\|\hat{x} - \bar{x}\|_2^2. \tag{52}$$

The linear system to recover $\hat{\gamma}$ is then

$$(P(\bar{x}) + \Delta')\hat{\gamma} = (E^*(\bar{x}) + \Delta)(y_{0,1} + b_1) = E^*(\bar{x})y_{0,1} + \delta \tag{53}$$

with $\delta = \Delta(y_{0,1} + b_1) + E^*(\bar{x})b_1$. Under Assumption 2.3, the unique solution of $P(\bar{x})\gamma = E^*(\bar{x})y_{0,1}$ is $\bar{\gamma}$. If $\hat{x}$ is sufficiently close to $\bar{x}$, we have $\|\Delta'\|_{2\to 2} < \sigma_-$ and $\|P(\bar{x})^{-1}\Delta'\|_{2\to 2} < 1$. We can now use standard results of linear algebra, see e.g. [44, 3.6], to obtain that

$$\begin{aligned}
&\frac{\|\hat{\gamma} - \bar{\gamma}\|_2}{\|\bar{\gamma}\|_2} \\
&\leq \frac{\|P(\bar{x})\|_{2\to 2}\|P(\bar{x})^{-1}\|_{2\to 2}}{1 - \|P(\bar{x})^{-1}\Delta'\|_{2\to 2}}\left(\frac{\|\Delta'\|_{2\to 2}}{\|P(\bar{x})\|_{2\to 2}} + \frac{\|\delta\|_2}{\|E^*(\bar{x})y_{0,1}\|_2}\right) \\
&\leq \frac{\sigma_+}{\sigma_-}\frac{1}{\left(1 - \frac{\|\Delta'\|_{2\to 2}}{\sigma_-}\right)}\left(\frac{\|\Delta'\|_{2\to 2}}{\sigma_-} + \frac{\|\Delta\|_{2\to 2}(\|y_{0,1}\|_2 + \|b_1\|_2) + \sqrt{\sigma_+}\|b_1\|_2}{\sqrt{\sigma_-}\|y_{0,1}\|_2}\right) \\
&\leq \epsilon_1 + \epsilon_2(\bar{x}).
\end{aligned}$$

Letting $\hat{x} \to \bar{x}$, we see that the noise term can be decomposed as an error term ϵ_1 which does not vanish as $\hat{x} \to \bar{x}$ and an additional term $\epsilon_2(\hat{x})$ which does vanish. The asymptotics for this term is

$$\epsilon_1 = \kappa^{3/2}\frac{\|b_1\|_2}{\|y_{0,1}\|_2},$$

and using $\frac{1}{\left(1 - \frac{\|\Delta'\|_{2\to 2}}{\sigma_-}\right)} \sim 1 + \frac{\|\Delta'\|_{2\to 2}}{\sigma_-}$ we obtain

$$\begin{aligned}
\epsilon_2(\bar{x}) &= C\kappa L_E\|\hat{x} - \bar{x}\|_2\left[\kappa + \sqrt{\kappa}\left(1 + \frac{\|b_1\|_2}{\|y_{0,1}\|_2}\right) + \kappa^{3/2}\frac{\|b_1\|_2}{\|y_{0,1}\|_2}\right] \\
&\quad + o_{\hat{x}\to\bar{x}}\left(\|\hat{x} - \bar{x}\|_2^2\right) \\
&\leq C\kappa^{5/2}L_E\|\hat{x} - \bar{x}\|_2\left(1 + \frac{\|b_1\|_2}{\|y_{0,1}\|_2}\right) + o_{\hat{x}\to\bar{x}}\left(\|\hat{x} - \bar{x}\|_2^2\right),
\end{aligned}$$

for some absolute constant C. □

5.8 Proof of Theorem 3.7

Proof. Let

$$\mathcal{E}(X) = \begin{pmatrix} \bar{w}_1 E(x_1) \\ \vdots \\ \bar{w}_N E(x_N) \end{pmatrix}, Y_0 = \begin{pmatrix} y_{0,1} \\ \vdots \\ y_{0,N} \end{pmatrix} \text{ and } B = \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix}.$$

By definition, we have

$$\hat{\gamma} = \mathcal{C}(\hat{X})^{-1}\left(\mathcal{E}^*(\hat{X})Y_0 + \mathcal{E}^*(\hat{X})B\right) \tag{54}$$

We have

$$\begin{aligned}\|\mathcal{E}(\hat{X}) - \mathcal{E}(\bar{X})\|_{2\to 2}^2 &\leq \sum_{n=1}^{N} \bar{w}_n^2 \|E(\hat{x}_n) - E(\bar{x}_n)\|_{2\to 2}^2 \\ &\leq \tilde{\sigma}_+ L_E^2 \|\hat{X} - \bar{X}\|_2^2.\end{aligned}$$

Hence, we have $\mathcal{E}(\hat{X}) = \mathcal{E}(\bar{X}) + \Delta$ with $\|\Delta\|_{2\to 2} \leq \sqrt{\tilde{\sigma}_+} L_E \|\hat{X} - \bar{X}\|_2$ and $\mathcal{C}(\bar{X}) = \mathcal{C}(\hat{X}) + \Delta'$ with

$$\|\Delta'\|_{2\to 2} \leq 2\|\Delta\|_{2\to 2}\|\mathcal{E}(\hat{X})\|_{2\to 2} + \|\Delta\|_{2\to 2}^2. \tag{55}$$

Under Assumption (28), the linear system $\mathcal{C}(\hat{X})\gamma = \mathcal{E}^*(\hat{X})(Y_0 + B)$ admits a unique solution $\hat{\gamma}$. Under Assumption (28) again, and given that $\hat{X}$ is sufficiently close to $\bar{X}$ so that $\|\Delta'\|_{2\to 2} < \tilde{\sigma}_-$, the linear system $\mathcal{C}(\bar{X})\gamma = \mathcal{E}^*(\bar{X})Y_0$ admits $\bar{\gamma}$ as a unique solution. In addition $\|\mathcal{C}(\hat{X})^{-1}\Delta'\|_{2\to 2} < 1$.

Let $\delta = \mathcal{E}^*(\hat{X})(Y_0 + B) - \mathcal{E}^*(\bar{X})Y_0$ denote the residual of the right hand-side term between the two previous linear systems. We have

$$\|\delta\|_2 \leq \|\Delta\|_{2\to 2}\|Y_0\|_2 + \sqrt{\tilde{\sigma}_+}\|B\|_2,$$

provided that $\|\mathcal{E}(\hat{X})\|_{2\to 2} \leq \sqrt{\tilde{\sigma}_+}$.

We can now use standard results of linear algebra, see e.g. [44, 3.6], to obtain that

$$\begin{aligned}&\frac{\|\hat{\gamma} - \bar{\gamma}\|_2}{\|\bar{\gamma}\|_2} \\ &\leq \frac{\|\mathcal{C}(\hat{X})\|_{2\to 2}\|\mathcal{C}(\hat{X})^{-1}\|_{2\to 2}}{1 - \|\mathcal{C}(\hat{X})^{-1}\Delta'\|_{2\to 2}}\left(\frac{\|\Delta'\|_{2\to 2}}{\|\mathcal{C}(\hat{X})\|_{2\to 2}} + \frac{\|\delta\|_2}{\|\mathcal{E}^*(\hat{X})\,(Y_0 + B)\,\|_2}\right) \\ &\leq \frac{\tilde{\sigma}_+}{\tilde{\sigma}_-}\frac{1}{\left(1 - \frac{\|\Delta'\|_{2\to 2}}{\tilde{\sigma}_-}\right)}\left(\frac{2\sqrt{\tilde{\sigma}_+}\|\Delta\|_{2\to 2} + \|\Delta\|_{2\to 2}^2}{\tilde{\sigma}_-} + \frac{\|\Delta\|_{2\to 2}\|Y_0\|_2 + \sqrt{\tilde{\sigma}_+}\|B\|_2}{\sqrt{\tilde{\sigma}_-}\|Y_0 + B\|_2}\right) \\ &= \frac{\tilde{\sigma}_+}{\tilde{\sigma}_-}\frac{1}{\left(1 - \frac{\|\Delta'\|_{2\to 2}}{\tilde{\sigma}_-}\right)}\left(\frac{2\sqrt{\tilde{\sigma}_+}\|\Delta\|_{2\to 2} + \|\Delta\|_{2\to 2}^2}{\tilde{\sigma}_-} + \frac{\|\Delta\|_{2\to 2}\|Y_0\|_2}{\sqrt{\tilde{\sigma}_-}\|Y_0 + B\|_2} + \frac{\sqrt{\tilde{\sigma}_+}}{\sqrt{\tilde{\sigma}_-}}\frac{\|B\|_2}{\|Y_0 + B\|_2}\right)\end{aligned}$$

Letting $\hat{X} \to \bar{X}$, we can decompose previous equation in two terms: an error term ϵ_1 which does not vanish as $\hat{X} \to \bar{X}$ and an additional term $\epsilon_2(\hat{X})$ which does vanish. We have

$$\epsilon_1 = \tilde{\kappa}^{3/2}\frac{\|B\|_2}{\|Y_0\|_2},$$

and the asymptotic for the second term is

$$\begin{aligned}\epsilon_2(\hat{X}) &= C\|\Delta\|_{2\to 2}\left[\frac{\sqrt{\tilde{\sigma}_+}}{\tilde{\sigma}_-} + \frac{1}{\sqrt{\tilde{\sigma}_-}}\left(\frac{\|Y_0\|_2 + \|B\|_2}{\|Y_0 + B\|_2}\right) + \frac{\tilde{\sigma}_+\|B\|_2}{\tilde{\sigma}_-^{3/2}\|Y_0 + B\|_2}\right] \\ &\quad + o_{\hat{X}\to\bar{X}}\left(\|\bar{X} - \hat{X}\|_2^2\right) \\ &\leq C\tilde{\kappa}^{5/2}L_E\|\bar{X} - \hat{X}\|_2\left(1 + \frac{\|Y_0\|_2 + \|B\|_2}{\|Y_0 + B\|_2}\right) + o_{\hat{X}\to\bar{X}}\left(\|\bar{X} - \hat{X}\|_2^2\right),\end{aligned}$$

for some absolute constant C. □

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] William E Moerner and Lothar Kador. Optical detection and spectroscopy of single molecules in a solid. *Physical review letters*, 62(21):2535, 1989.

[3] Daniel Sage, Hagai Kirshner, Thomas Pengo, Nico Stuurman, Junhong Min, Suliana Manley, and Michael Unser. Quantitative evaluation of software packages for single-molecule localization microscopy. *Nature methods*, 12(8):717–724, 2015.

[4] Daniel Sage, Thanh-An Pham, Hazen Babcock, Tomas Lukes, Thomas Pengo, Jerry Chao, Ramraj Velmurugan, Alex Herbert, Anurag Agrawal, Silvia Colabrese, et al. Super-resolution fight club: assessment of 2d and 3d single-molecule localization microscopy software. *Nature methods*, 16(5):387–395, 2019.

[5] Otmar Scherzer, Markus Grasmair, Harald Grossauer, Markus Haltmeier, and Frank Lenzen. *Variational methods in imaging*. Springer, 2009.

[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] Vincent Duval and Gabriel Peyré. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355, 2015.

[8] 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.

[9] Kristian Bredies and Hanna Katriina Pikkarainen. Inverse problems in spaces of measures. *ESAIM: Control, Optimisation and Calculus of Variations*, 19(1):190–218, 2013.

[10] Quentin Denoyelle, Vincent Duval, Gabriel Peyré, and Emmanuel Soubies. The sliding frank–wolfe algorithm and its application to super-resolution microscopy. *Inverse Problems*, 36(1):014001, 2019.

[11] Axel Flinth, Frédéric de Gournay, and Pierre Weiss. On the linear convergence rates of exchange and continuous methods for total variation minimization. *Mathematical Programming*, pages 1–37, 2020.

[12] Ali Ahmed, Benjamin Recht, and Justin Romberg. Blind deconvolution using convex programming. *IEEE Transactions on Information Theory*, 60(3):1711–1732, 2013.

[13] Shuyang Ling and Thomas Strohmer. Self-calibration and biconvex compressive sensing. *Inverse Problems*, 31(11):115002, 2015.

[14] Yuejie Chi. Guaranteed blind sparse spikes deconvolution via lifting and convex optimization. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):782–794, 2016.

[15] Peter Jung, Felix Krahmer, and Dominik Stöger. Blind demixing and deconvolution at near-optimal rate. *IEEE Transactions on Information Theory*, 64(2):704–727, 2017.

[16] Ali Ahmed and Laurent Demanet. Leveraging diversity and sparsity in blind deconvolution. *IEEE Transactions on Information Theory*, 64(6):3975–4000, 2018.

[17] Mohamed A Suliman and Wei Dai. Blind two-dimensional super-resolution and its performance guarantee. *arXiv preprint arXiv:1811.02070*, 2018.

[18] Jinchi Chen, Weiguo Gao, Sihan Mao, and Ke Wei. Vectorized hankel lift: A convex approach for blind super-resolution of point sources. *arXiv preprint arXiv:2008.05092*, 2020.

[19] Valentin Debarnot, Paul Escande, and Pierre Weiss. A scalable estimator of sets of integral operators. *Inverse Problems*, 2019.

[20] Valentin Debarnot, Paul Escande, Thomas Mangeat, and Pierre Weiss. Learning low-dimensional models of microscopes. *IEEE Transactions on Computational Imaging*, 2020.

[21] Yanjun Li, Kiryung Lee, and Yoram Bresler. Identifiability and stability in blind deconvolution under minimal assumptions. *IEEE Transactions on Information Theory*, 63(7):4619–4633, 2017.

[22] Stéphane Mallat. *A wavelet tour of signal processing*. Elsevier, 1999.

[23] Paul Escande and Pierre Weiss. Approximation of integral operators using product-convolution expansions. *Journal of Mathematical Imaging and Vision*, 58(3):333–348, 2017.

[24] Ralf C. Flicker and Francois J. Rigaut. Anisoplanatic deconvolution of adaptive optics images. *J. Opt. Soc. Am. A*, 22(3):504–513, Mar 2005.

[25] Jérémie Bigot, Paul Escande, and Pierre Weiss. Estimation of linear operators from scattered impulse responses. *Applied and Computational Harmonic Analysis*, 47(3):730–758, 2019.

[26] Xiaodong Li, Shuyang Ling, Thomas Strohmer, and Ke Wei. Rapid, robust, and reliable blind deconvolution via nonconvex optimization. *Applied and computational harmonic analysis*, 47(3):893–934, 2019.

[27] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[28] Felix Krahmer and Dominik Stöger. On the convex geometry of blind deconvolution and matrix completion. *arXiv preprint arXiv:1902.11156*, 2019.

[29] Yuxin Chen, Jianqing Fan, Bingyan Wang, and Yuling Yan. Convex and nonconvex optimization are both minimax-optimal for noisy blind deconvolution. *arXiv preprint arXiv:2008.01724*, 2020.

[30] Robert Beinert and Kristian Bredies. Tensor-free proximal methods for lifted bilinear/quadratic inverse problems with applications to phase retrieval. *arXiv preprint arXiv:1907.04875*, 2019.

[31] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[32] Jérôme Bolte, Shoham Sabach, and Marc Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1-2):459–494, 2014.

[33] Valerio Cambareri and Laurent Jacques. Through the haze: a non-convex approach to blind gain calibration for linear random sensing models. *Information and Inference: A Journal of the IMA*, 8(2):205–271, 2019.

[34] Yann Traonmilin and Jean-Francois Aujol. The basins of attraction of the global minimizers of the non-convex sparse spike estimation problem. *Inverse Problems*, 36(4):045003, 2020.

[35] Z. Zhu, Q. Li, G. Tang, and M. B. Wakin. Global optimality in low-rank matrix optimization. *IEEE Transactions on Signal Processing*, 66(13):3614–3628, 2018.

[36] Michael Kech and Felix Krahmer. Optimal injectivity conditions for bilinear inverse problems with applications to identifiability of deconvolution problems. *SIAM Journal on Applied Algebra and Geometry*, 1(1):20–37, 2017.

[37] Claire Boyer, Antonin Chambolle, Yohann De Castro, Vincent Duval, Frédéric De Gournay, and Pierre Weiss. On representer theorems and convex regularization. *SIAM Journal on Optimization*, 29(2):1260–1281, 2019.

[38] Patrick L Combettes and Jean-Christophe Pesquet. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer, 2011.

[39] Michael Grant and Stephen Boyd. Cvx: Matlab software for disciplined convex programming, version 2.1, 2014.

[40] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[41] Rafał Latała et al. Estimates of moments and tails of gaussian chaoses. *The Annals of Probability*, 34(6):2315–2331, 2006.

[42] Michel Talagrand. *The generic chaining: upper and lower bounds of stochastic processes*. Springer Science & Business Media, 2006.

[43] Felix Krahmer, Shahar Mendelson, and Holger Rauhut. Suprema of chaos processes and the restricted isometry property. *Communications on Pure and Applied Mathematics*, 67(11):1877–1904, 2014.

[44] Eugene E Tyrtyshnikov. *A brief introduction to numerical analysis*. Springer Science & Business Media, 2012.