XXX manuscript No.
(will be inserted by the editor)

Optimal Multivariate Gaussian Fitting with Applications to PSF Modeling in Two-Photon Microscopy Imaging

Emilie Chouzenoux[1,2] · **Tim Tsz-Kit Lau**[3] · **Claire Lefort**[4] · **Jean-Christophe Pesquet**[1]

Received: date / Accepted: date

Abstract Fitting Gaussian functions to empirical data is a crucial task in a variety of scientific applications, especially in image processing. However, most of the existing approaches for performing such fitting are restricted to two dimensions and they cannot be easily extended to higher dimensions. Moreover, they are usually based on alternating minimization schemes which benefit from few theoretical guarantees in the underlying nonconvex setting. In this paper, we provide a novel variational formulation of the multivariate Gaussian fitting problem, which is applicable to any dimension and accounts for possible non-zero background and noise in the input data. The block multiconvexity of our objective function leads us to propose a proximal alternating method to minimize it in order to estimate the Gaussian shape parameters. The resulting **FIGARO** algorithm is shown to converge to a critical point under mild assumptions. The algorithm shows a good robustness when tested on synthetic datasets. To demonstrate the versatility of **FIGARO**, we also illustrate its excellent performance in the fitting of the Point Spread Functions of experimental raw data from a two-photon fluorescence microscope.

Keywords Gaussian fitting · Kullback-Leibler divergence · Alternating minimization · Proximal methods · PSF identification · Two-photon Fluorescence microscopy

✉ Emilie Chouzenoux · emilie.chouzenoux@univ-mlv.fr
1 Center for Visual Computing, CentraleSupélec, INRIA Saclay, Université Paris-Saclay, 91190 Gif-sur-Yvette, France
2 Laboratoire d'Informatique Gaspard Monge, UMR CNRS 8049, Université Paris-Est Marne-la-Vallée, 77454 Marne-la-Vallée Cedex 2, France
3 Department of Statistics, Northwestern University, Evanston, IL 60208, United States of America
4 XLIM Research Institute, UMR CNRS 7252, Université de Limoges, 87032 Limoges, France

1 Introduction

Fitting Gaussian shapes from noisy observed data points is an essential task in various science and engineering applications. In the one-dimensional (1D) case, it lies for instance at the core of spectroscopy signal analysis techniques in physical science [21,31]. In the two-dimensional (2D) case, where Gaussian profile parameters are estimated from images, some worth mentioning applications include Gaussian beam characterization, particle tracking, and sensor calibration [28,37, 15]. In the domain of image recovery, a particularly important application of Gaussian shape fitting is the modeling of Point Spread Functions (PSF) from raw data of optical systems (e.g., microscopes, telescopes). The success of image restoration strategies strongly depends on the accuracy of the PSF estimation [13]. This estimation is often performed through a preliminary step of image acquisition of normalized and calibrated objects, associated with a model fitting strategy. The PSF model is chosen as a trade-off between accuracy and simplicity. Gaussian models often lead to both tractable and good quality approximations [35,32,1,42,41].

Let $\mathrm{L}^1(\mathbb{R}^Q)$ denote the space of real-valued summable functions defined on $\mathbb{R}^Q$. In this paper, we address the problem of fitting a Gaussian model to an observed function $y \in \mathrm{L}^1(\mathbb{R}^Q)$. We assume that the observed function y can be modeled as

$$(\forall \boldsymbol{u} \in \mathbb{R}^Q) \quad y(\boldsymbol{u}) = \overline{a} + \overline{b}\,\overline{p}(\boldsymbol{u}) + v(\boldsymbol{u}), \tag{1.1}$$

where $\overline{a} \in \mathbb{R}$ is a background term, $\overline{b} \in (0, +\infty)$ is a scaling parameter, $\overline{p} \in \mathrm{L}^1(\mathbb{R}^Q)$ represents a noiseless version of the observed field, and v is a function accounting for acquisition errors. The main assumption is that $\overline{p}$ is close, in a sense to be made precise, to the probability density function $\boldsymbol{u} \mapsto g(\boldsymbol{u}, \overline{\boldsymbol{\mu}}, \overline{\boldsymbol{C}})$, of a Q-dimensional normal distribution with mean $\overline{\boldsymbol{\mu}} \in \mathbb{R}^Q$ and precision (i.e., inverse covariance) matrix

$\overline{\boldsymbol{C}} \in \mathscr{S}_Q^{++}$ [1]. This distribution is expressed as

$$(\forall \boldsymbol{u} \in \mathbb{R}^Q)(\forall \boldsymbol{\mu} \in \mathbb{R}^Q)(\forall \boldsymbol{C} \in \mathscr{S}_Q^{++})$$
$$g(\boldsymbol{u},\boldsymbol{\mu},\boldsymbol{C}) = \sqrt{\frac{|\boldsymbol{C}|}{(2\pi)^Q}} \exp\left(-\frac{1}{2}(\boldsymbol{u}-\boldsymbol{\mu})^\top \boldsymbol{C}(\boldsymbol{u}-\boldsymbol{\mu})\right), \tag{1.2}$$

where $|\boldsymbol{C}|$ denotes the determinant of matrix $\boldsymbol{C}$. The fitting problem thus consists of finding an estimate $(\widehat{a},\widehat{b},\widehat{p},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{C}})$ of $(\overline{a},\overline{b},\overline{p},\overline{\boldsymbol{\mu}},\overline{\boldsymbol{C}})$ in accordance with model (1.1)

Because of its prominent importance in applications, there has been a significant amount of works on this subject [12, 25,24,23,34,42]. To the best of our knowledge, all existing works consider that $p = g(\cdot,\boldsymbol{\mu},\boldsymbol{C})$ and they are focused on fitting parameters $(\widehat{a},\widehat{b},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{C}})$ from y. Two main classes of methods can be distinguished. The first set of approaches [25,24,34] is based on the search for the best fitting parameters minimizing a least-squares cost between the observations and the sought model. The minimization process is based on the famous Levenberg-Marquardt alternating minimization strategy. However, it is worth mentioning that few established convergence guarantees are available for this method, which may be detrimental to its reliable use in practice. The second class of methods uses the so-called Caruana's formulation [12]. The idea here is to assume that the background term $\overline{a}$ is zero and to search for $(\widehat{b},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{C}})$ which minimize the difference of logarithms between the data and the model [23,1]. The advantage of such a strategy is that it gives rise to a convex formulation, for which efficient and reliable optimization techniques can be applied. It is however worth emphasizing that all the aforementioned works are focused on the resolution of the fitting problem in low dimensions, that is when $Q = 1$ [12,25,23,34] or $Q = 2$ [24, 1,42]. Moreover, except in [34] where a polynomial background is accounted for, the background term $\overline{a}$ is considered as zero. These assumptions however usually do not correspond to constraints inherent to an experimental setup or environment.

The aim of this paper is to propose a new multivariate Gaussian fitting strategy which avoids the aforementioned limitations. Our method relies on the minimization of a hybrid cost function combining a least-squares data fidelity term, a Kullback-Leibler divergence regularizer for improved robustness, and range constraints on the parameters. This original variational formulation results in a nonconvex minimization problem for which we propose a theoretically sound and efficient proximal alternating iterative resolution scheme. When applied to the analysis of 3D raw data acquired with a two-photon fluorescence microscope, our new computational strategy shows an unprecedented accuracy and reliability.

In Section 2, the data fitting problem is formulated in a variational manner. A proximal alternating optimization method called **FIGARO** is then proposed in Section 3 for finding a minimizer of the proposed nonconvex cost function. The implementation of the algorithm steps is discussed. The convergence of the sequence of iterates resulting from **FIGARO** is established in Section 4. Section 5 illustrates the high robustness of our approach to a model mismatch, when compared to a standard nonlinear least squares fitting strategy on 3D synthetic data. In Section 6, the scope of our approach is demonstrated through the analysis of the Point Spread Function of a 3D two-photon fluorescence microscope. Finally, Section 7 concludes the paper.

2 Proposed Variational Formulation

The key ingredient of our method relies on measuring the closeness of $\overline{p}$ to the Gaussian probability density functions by using the Kullback-Leibler (KL) divergence [5]. Let us first recall the definition of KL divergence. Let $\mathscr{P}$ denote the set of probability density functions supported on $\mathbb{R}^Q$:

$$\mathscr{P} = \Big\{ q \in \mathrm{L}^1(\mathbb{R}^Q) \mid (\forall \boldsymbol{u} \in \mathbb{R}^Q)\ q(\boldsymbol{u}) \geq 0 \int_\Omega q(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = 1 \Big\}. \tag{2.1}$$

Suppose that $(p,q) \in \mathscr{P}^2$ and q takes (strictly) positive values, the KL divergence from q to p reads

$$\mathscr{KL}(p \,\|\, q) = \int_{\mathbb{R}^Q} p(\boldsymbol{u}) \log\left(\frac{p(\boldsymbol{u})}{q(\boldsymbol{u})}\right) \mathrm{d}\boldsymbol{u}, \tag{2.2}$$

with the convention $0\log 0 = 0$.

In order to avoid singularity issues, we will assume that the Gaussian variances in each direction are bounded above by some maximal values. The spectrum of the precision matrix $\overline{\boldsymbol{C}}$ is thus bounded from below, in the sense that there exists some $\varepsilon > 0$ such that $\overline{\boldsymbol{C}} = \overline{\boldsymbol{D}} + \varepsilon \boldsymbol{I}_Q$ where $\overline{\boldsymbol{D}}$ belongs to $\mathscr{S}_Q^+$ and $\boldsymbol{I}_Q \in \mathbb{R}^{Q\times Q}$ denotes the identity matrix of $\mathbb{R}^Q$. We then propose to define $(\widehat{a},\widehat{b},\widehat{p},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{D}})$ as a minimizer of a hybrid cost function, gathering information regarding the observation model (1.1) and the Gaussian shape prior (1.2). The minimization problem reads

$$\underset{\substack{a\in\mathscr{A},b\in\mathscr{B}\\ \boldsymbol{\mu}\in\mathbb{R}^Q,p\in\mathscr{P},\boldsymbol{D}\in\mathscr{S}_Q^+}}{\text{minimize}} \quad \frac{1}{2}\int_{\mathbb{R}^Q} \big(y(\boldsymbol{u}) - a - bp(\boldsymbol{u})\big)^2 \mathrm{d}\boldsymbol{u} + \lambda \mathscr{KL}\big(p \,\|\, g(\cdot,\boldsymbol{\mu},\boldsymbol{D}+\varepsilon\boldsymbol{I}_Q)\big). \tag{2.3}$$

[1] Throughout the paper, $\mathscr{S}_Q^{++}$ will denote the set of symmetric positive definite matrices of $\mathbb{R}^{Q\times Q}$, $\mathscr{S}_Q^+$ the set of symmetric positive semidefinite matrices of $\mathbb{R}^{Q\times Q}$ and $\mathscr{S}_Q$ the set of symmetric matrices of $\mathbb{R}^{Q\times Q}$

Hereabove, $\mathscr{A}$ and $\mathscr{B}$ are some nonempty closed bounded real intervals corresponding to known bounds on $\overline{a}$ and $\overline{b}$ respectively, and $\lambda > 0$ is a regularization parameter weighting the KL penalty term favoring the proximity between p and the Gaussian model (1.2) parametrized by $(\boldsymbol{\mu}, \boldsymbol{D})$.

In practice, however, one generally has access only to a sampling of y, which is performed on a bounded Borel set Ω of $\mathbb{R}^Q$. The set Ω is supposed here chosen large enough so that it captures most of the probability mass of the sought Gaussian disribution. More precisely, we will assume that Ω is paved into $N \in \mathbb{N}$ voxels of volume $\Delta \in (0, +\infty)$ and mass centers $(\boldsymbol{x}_n)_{1 \leq n \leq N}$. The available vector of observations is then $\boldsymbol{y} = (y_n)_{1 \leq n \leq N}$ where, for every $n \in \{1, \dots, N\}$, $y_n = y(\boldsymbol{x}_n)$. After this discretization, by assuming that y and p are continuous functions in (2.3) and that Δ is small enough, the following more tractable optimization problem is thus substituted for the original variational formulation:

$$\begin{aligned} \underset{\substack{a \in \mathscr{A}, b \in \mathscr{B} \\ \boldsymbol{\mu} \in \mathbb{R}^Q, \boldsymbol{p} \in \mathscr{P}_{\mathrm{d}}, \boldsymbol{D} \in \mathscr{S}_Q^+}}{\text{minimize}} \quad & \frac{1}{2} \|\boldsymbol{y} - a\mathbf{1}_N - b\boldsymbol{p}\|^2 \\ & + \lambda \sum_{n=1}^N p_n \log\left(\frac{p_n}{g(\boldsymbol{x}_n, \boldsymbol{\mu}, \boldsymbol{D} + \varepsilon \boldsymbol{I}_Q)}\right), \end{aligned} \tag{2.4}$$

where $\|\cdot\|$ denotes the standard Euclidean norm. The probability density function p has been replaced by the vector $\boldsymbol{p} = (p_n)_{1 \leq n \leq N}$ which belongs $\mathscr{P}_{\mathrm{d}} = [0, +\infty)^N \cap \mathscr{C}$, where $\mathscr{C}$ is the affine hyperplane

$$\mathscr{C} = \left\{ \boldsymbol{p} \in \mathbb{R}^N \;\middle|\; \sum_{n=1}^N p_n = \Delta^{-1} \right\}. \tag{2.5}$$

The discrete KL term in (2.4) can be rewritten as

$$\begin{aligned} & \sum_{n=1}^N p_n \log\left(\frac{p_n}{g(\boldsymbol{x}_n, \boldsymbol{\mu}, \boldsymbol{D} + \varepsilon \boldsymbol{I}_Q)}\right) \\ &= \sum_{n=1}^N \mathrm{ent}(p_n) + p_n\Big(\frac{Q}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{D} + \varepsilon \boldsymbol{I}_Q|) \\ & \quad + \frac{1}{2}(\boldsymbol{x}_n - \boldsymbol{\mu})^\top (\boldsymbol{D} + \varepsilon \boldsymbol{I}_Q)(\boldsymbol{x}_n - \boldsymbol{\mu})\Big), \end{aligned} \tag{2.6}$$

where

$$(\forall \upsilon \in \mathbb{R}) \quad \mathrm{ent}(\upsilon) = \begin{cases} \upsilon \log \upsilon, & \upsilon > 0, \\ 0, & \upsilon = 0, \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.7}$$

Note that the above definition of the function ent allows us to impose directly the nonnegativity of the components of $\boldsymbol{p}$. For technical reasons which will appear later, we will also need to perform a twice continuously differentiable extension of the function $\boldsymbol{D} \mapsto -\log(|\boldsymbol{D} + \varepsilon \boldsymbol{I}_Q|)$ on the whole domain $\mathscr{S}_Q$. This extension φ is defined as follows. For every $\boldsymbol{D} \in \mathscr{S}_Q$ decomposed as $\boldsymbol{U}\,\mathrm{Diag}(\boldsymbol{\sigma})\boldsymbol{U}^\top$ with $\boldsymbol{U} \in \mathbb{R}^{Q \times Q}$ an orthogonal matrix and $\boldsymbol{\sigma} = (\sigma_q)_{1 \leq q \leq Q}$ the associated vector of eigenvalues of $\boldsymbol{D}$,

$$\begin{aligned} \varphi(\boldsymbol{D}) &= \widetilde{\varphi}(\boldsymbol{\sigma}) \\ &= \begin{cases} -\log(|\boldsymbol{D} + \varepsilon \boldsymbol{I}_Q|) = -\displaystyle\sum_{q=1}^Q \log(\sigma_q + \varepsilon), \\ \qquad \text{if} \quad \boldsymbol{D} \in \mathscr{S}_Q^+, \\ \widetilde{\varphi}(\mathbf{0}_Q) + \boldsymbol{\sigma}^\top \nabla \widetilde{\varphi}(\mathbf{0}_Q) + \frac{1}{2}\boldsymbol{\sigma}^\top \nabla^2 \widetilde{\varphi}(\mathbf{0}_Q)\boldsymbol{\sigma}, \\ \qquad \text{otherwise,} \end{cases} \end{aligned} \tag{2.8}$$

where $\mathbf{0}_Q$ is the Q-dimensional null vector, $\mathbf{1}_Q$ the Q-dimensional vector of all ones, and

$$\widetilde{\varphi}(\mathbf{0}_Q) = -Q\log\varepsilon, \nabla\widetilde{\varphi}(\mathbf{0}_Q) = -\varepsilon^{-1}\mathbf{1}_Q, \nabla^2\widetilde{\varphi}(\mathbf{0}_Q) = \varepsilon^{-2}\boldsymbol{I}_Q. \tag{2.9}$$

Let us denote by $\iota_{\mathscr{S}}$ the indicator function of a set $\mathscr{S}$, which is equal to 0 on this set and $+\infty$ otherwise. We are now ready to define the cost function which is minimized in our Gaussian fitting approach:

$$\begin{aligned} & (\forall a \in \mathbb{R})(\forall b \in \mathbb{R})(\forall \boldsymbol{p} \in \mathbb{R}^N)(\forall \boldsymbol{\mu} \in \mathbb{R}^Q)(\forall \boldsymbol{D} \in \mathscr{S}_Q) \\ & F(a, b, \boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{D}) = \frac{1}{2}\|\boldsymbol{y} - a\mathbf{1}_N - b\boldsymbol{p}\|^2 + \iota_{\mathscr{A}}(a) \\ & \qquad + \iota_{\mathscr{B}}(b) + \lambda \Psi(\boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{D}), \end{aligned} \tag{2.10}$$

where

$$\begin{aligned} & (\forall \boldsymbol{p} \in \mathbb{R}^N)(\forall \boldsymbol{\mu} \in \mathbb{R}^Q)(\forall \boldsymbol{D} \in \mathscr{S}_Q) \\ & \Psi(\boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{D}) = \sum_{n=1}^N \Big(\mathrm{ent}(p_n) + \frac{p_n}{2}\big(Q\log(2\pi) + \varphi(\boldsymbol{D}) \\ & \quad + (\boldsymbol{x}_n - \boldsymbol{\mu})^\top (\boldsymbol{D} + \varepsilon \boldsymbol{I}_Q)(\boldsymbol{x}_n - \boldsymbol{\mu})\big)\Big) + \iota_{\mathscr{C}}(\boldsymbol{p}) + \iota_{\mathscr{S}_Q^+}(\boldsymbol{D}). \end{aligned} \tag{2.11}$$

Remark 1 The proposed formulation deals with a regular grid but it can be easily extended to the case of irregular sampling by changing the definition of $\mathscr{C}$ into

$$\mathscr{C} = \left\{ \boldsymbol{p} \in \mathbb{R}^N \;\middle|\; \sum_{n=1}^N \Delta_n p_n = 1 \right\} \tag{2.12}$$

where, for every $n \in \{1, \dots, N\}$, $\Delta_n \in (0, +\infty)^N$ is the volume of the n-th voxel.

3 FIGARO Minimization Algorithm

3.1 Proposed Algorithm

The objective function (4.1) is nonconvex, yet convex with respect to each variable. A standard resolution approach is thus to adopt an alternating minimization strategy, where, at

each iteration, F is minimized with respect to one variable while the others remain fixed. This approach, sometimes referred to as Block Coordinate Descent or nonlinear Gauss-Seidel method, has been widely used in the context of PSF model fitting [42,30,32]. However, its convergence is only guaranteed under restrictive assumptions [38]. In order to get sounder convergence results, we propose to use an alternative strategy based on proximal tools which consists of replacing, at each iteration the direct minimization step by a proximal one ([33, Def. 1.22], [6, Def. 12.23], [18, Def. 10.1] [11]).

Definition 1 (Domain) Let f be a function from $\mathbb{R}^n$ to $(-\infty,+\infty]$. The *domain* of f is defined by

$$\mathrm{dom}\, f := \{x \in \mathbb{R}^n : f(x) < +\infty\}.$$

The function f is *proper* if and only if $\mathrm{dom}\, f$ is nonempty.

Definition 2 (Proximity operator) Let $f : \mathbb{R}^n \to (-\infty,+\infty]$ be a convex, proper, lower semi-continuous function. The *proximity operator* of f at $x \in \mathbb{R}^n$ is defined as

$$\mathrm{prox}_f(x) = \underset{y\in\mathbb{R}^n}{\mathrm{argmin}}\, f(y) + \frac{1}{2}\|y-x\|^2.$$

Let $\mathscr{S}$ be a nonempty closed convex subset of $\mathbb{R}^n$. Then $\mathrm{prox}_{\iota_{\mathscr{S}}}$ is equal to the projection $\mathsf{P}_{\mathscr{S}}$ onto $\mathscr{S}$.

The application of the proximal alternating method [4,2, 8] to the minimization of (4.1) yields Algorithm 1, called **FIGARO** (Fitting Gaussians with Proximal Optimization).

Algorithm 1 FIGARO method

$a_0 \in \mathscr{A}, b_0 \in \mathscr{B}, \boldsymbol{p}_0 \in \mathscr{C}, \boldsymbol{\mu}_0 \in \mathbb{R}^Q, \boldsymbol{D}_0 \in \mathscr{S}_Q^+$,
$(\gamma_a,\gamma_b,\gamma_p,\gamma_\mu,\gamma_D) \in (0,+\infty)^5$.
for $i = 1,2,\ldots$ **do**
$a^{(i+1)} = \mathrm{prox}_{\gamma_a F(\cdot,b^{(i)},\boldsymbol{p}^{(i)},\boldsymbol{\mu}^{(i)},\boldsymbol{D}^{(i)})}(a^{(i)})$
$b^{(i+1)} = \mathrm{prox}_{\gamma_b F(a^{(i+1)},\cdot,\boldsymbol{p}^{(i)},\boldsymbol{\mu}^{(i)},\boldsymbol{D}^{(i)})}(b^{(i)})$
$\boldsymbol{p}^{(i+1)} = \mathrm{prox}_{\gamma_p F(a^{(i+1)},b^{(i+1)},\cdot,\boldsymbol{\mu}^{(i)},\boldsymbol{D}^{(i)})}(\boldsymbol{p}^{(i)})$
$\boldsymbol{\mu}^{(i+1)} = \mathrm{prox}_{\gamma_\mu F(a^{(i+1)},b^{(i+1)},\boldsymbol{p}^{(i+1)},\cdot,\boldsymbol{D}^{(i)})}(\boldsymbol{\mu}^{(i)})$
$\boldsymbol{D}^{(i+1)} = \mathrm{prox}_{\gamma_D F(a^{(i+1)},b^{(i+1)},\boldsymbol{p}^{(i+1)},\boldsymbol{\mu}^{(i+1)},\cdot)}(\boldsymbol{D}^{(i)})$
end for

Remark that other methods such as those proposed in [40,17,10] are also applicable to our problem, but the considered alternating proximal point algorithm may appear preferable because of its simplicity.

3.2 Expressions of the Proximity Operators

In this part, we show that the proximity operators required in Algorithm 1 have closed form expressions.

Proposition 1 *Let $(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D}) \in \mathbb{R}\times\mathbb{R}\times\mathbb{R}^N\times\mathbb{R}^Q\times\mathscr{S}_Q$ and $(\gamma_a,\gamma_b) \in (0,+\infty)^2$. The proximity operator of $\gamma_a F(\cdot,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D})$ at a is given by*

$$\mathrm{prox}_{\gamma_a F(\cdot,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D})}(a) = \mathsf{P}_{\mathscr{A}}\left(\frac{a+\gamma_a \mathbf{1}_N^\top(\boldsymbol{y}-b\boldsymbol{p})}{1+\gamma_a N}\right) \tag{3.1}$$

and the proximity operator of $\gamma_b F(a,\cdot,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D})$ at b is given by

$$\mathrm{prox}_{\gamma_b F(a,\cdot,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D})}(b) = \mathsf{P}_{\mathscr{B}}\left(\frac{b+\gamma_b(\boldsymbol{y}-a\mathbf{1}_N)^\top\boldsymbol{p}}{1+\gamma_b\|\boldsymbol{p}\|^2}\right). \tag{3.2}$$

Proof Calculating the proximity operator of $\gamma_a F(\cdot,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D})$ is equivalent to calculating the proximity operator of the one-variable function $\vartheta + \iota_{\mathscr{A}}$ where

$$(\forall a \in \mathbb{R}) \quad \vartheta(a) = \frac{\gamma_a}{2}\sum_{n=1}^N (y_n - a - bp_n)^2. \tag{3.3}$$

It follows from [14] that

$$\mathrm{prox}_{\gamma_a F(\cdot,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D})} = \mathsf{P}_{\mathscr{A}} \circ \mathrm{prox}_{\vartheta}. \tag{3.4}$$

On the other hand, it follows from [18] that

$$\mathrm{prox}_{\vartheta}(a) = \frac{a+\gamma_a\mathbf{1}_N^\top(\boldsymbol{y}-b\boldsymbol{p})}{1+\gamma_a N}. \tag{3.5}$$

Expression (3.2) is obtained by similar arguments. □

Proposition 2 *Let $(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D}) \in \mathbb{R}\times\mathbb{R}\times\mathbb{R}^N\times\mathbb{R}^Q\times\mathscr{S}_Q$ and $\gamma_p > 0$. The proximity operator of $\gamma_p F(a,b,\cdot,\boldsymbol{\mu},\boldsymbol{D})$ at $\boldsymbol{p}$ is given by*

$$\mathrm{prox}_{\gamma_p F(a,b,\cdot,\boldsymbol{\mu},\boldsymbol{D})}(\boldsymbol{p}) = \left(\rho^{-1}\mathrm{W}\left(\rho\exp\left(w_n(\widehat{\nu})\right)\right)\right)_{1\le n\le N}, \tag{3.6}$$

where W *denotes the Lambert-*W *function [19],*

$$\rho = \frac{\gamma_p b^2 + 1}{\gamma_p \lambda}, \tag{3.7}$$

and, for every $n \in \{1,\ldots,N\}$, w_n is the function defined as

$$(\forall \nu \in \mathbb{R})\ w_n(\nu) = -1 - c_n + (\gamma_p\lambda)^{-1}(p_n + \gamma_p b(y_n - a) - \nu), \tag{3.8}$$

with

$$c_n = \frac{Q}{2}\log(2\pi) + \frac{1}{2}\varphi(\boldsymbol{D}) + \frac{1}{2}(\boldsymbol{x}_n-\boldsymbol{\mu})^\top(\boldsymbol{D}+\varepsilon\boldsymbol{I}_Q)(\boldsymbol{x}_n-\boldsymbol{\mu}). \tag{3.9}$$

Moreover, $\widehat{\nu} \in \mathbb{R}$ is the the unique zero of the function

$$(\forall \nu \in \mathbb{R}) \quad \Phi(\nu) = \rho^{-1}\sum_{n=1}^N \mathrm{W}\left(\rho\exp(w_n(\nu))\right) - \Delta^{-1}. \tag{3.10}$$

Proof Let $\widetilde{\boldsymbol{p}} \in \mathbb{R}^N$. Then,

$$\begin{aligned}\widehat{\boldsymbol{p}} &= \operatorname{prox}_{\gamma_p F(a,b,\cdot,\boldsymbol{\mu},\boldsymbol{D})}(\widetilde{\boldsymbol{p}}) \\ &= \operatorname*{argmin}_{\boldsymbol{p}\in\mathbb{R}^N} \gamma_p F(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D}) + \frac{1}{2}\|\boldsymbol{p}-\widetilde{\boldsymbol{p}}\|^2 \\ &= \operatorname*{argmin}_{\boldsymbol{p}\in\mathscr{C}} \gamma_p \sum_{n=1}^{N} \frac{1}{2}(y_n - a - bp_n)^2 + \gamma_p\lambda \sum_{n=1}^{N}(p_n\log p_n + p_n c_n) \\ &\qquad + \frac{1}{2}\sum_{n=1}^{N}(p_n-\tilde{p}_n)^2. \end{aligned} \tag{3.11}$$

The Lagrangian function associated with the above constrained problem reads

$$\begin{aligned}&(\forall \boldsymbol{p}\in[0,+\infty)^N)(\forall \nu\in\mathbb{R})\ \ \mathscr{L}(\boldsymbol{p},\nu) = \gamma_p \sum_{n=1}^{N}\frac{1}{2}(y_n-a-bp_n)^2 \\ &+ \sum_{n=1}^{N} \gamma_p\lambda\,(p_n\log p_n + p_n c_n) + \frac{1}{2}(p_n-\tilde{p}_n)^2 \\ &+ \nu\Big(\sum_{n=1}^{N} p_n - \Delta^{-1}\Big). \end{aligned} \tag{3.12}$$

Since Slater's condition obviously holds, there exists $\widehat{\nu}\in\mathbb{R}$ such that $(\widehat{\boldsymbol{p}},\widehat{\nu})$ is a saddle point of the $\mathscr{L}$ [7]. By Fermat's rule [6], $\widehat{\boldsymbol{p}} = (\widehat{p}_n)_{1\le n\le N}$ is thus obtained by finding a zero of the partial subdifferential of $\mathscr{L}$ with respect to variable $\boldsymbol{p}$. By using (3.7), this yields, for every $n\in\{1,\dots,N\}$,

$$\begin{aligned}&\gamma_p(b^2p_n - by_n + ab) + \gamma_p\lambda(1+\log p_n + c_n) \\ &\qquad + p_n - \tilde{p}_n + \widehat{\nu} = 0 \\ \Leftrightarrow\quad &\rho\, p_n + \log p_n = w_n(\widehat{\nu}) \\ \Leftrightarrow\quad &\rho\, p_n \exp(\rho\, p_n) = \rho\exp(w_n(\widehat{\nu})). \end{aligned} \tag{3.13}$$

By recalling that the Lambert-W function is such that $(\forall z\in\mathbb{R})\ \mathrm{W}(z)\exp(\mathrm{W}(z)) = z$, we deduce (3.6).

In addition, canceling the derivative of $\mathscr{L}$ with respect to ν amounts to finding a zero of function Φ defined in (3.10). This existence of a zero is guaranteed by the existence of $\widehat{\boldsymbol{p}}$. Let us now establish its uniqueness by evaluating the derivative Φ' using the following property of the Lambert W-function:

$$(\forall z\in\mathbb{R}^+)\quad \mathrm{W}'(z) = \frac{1}{(\mathrm{W}(z)+1)\mathrm{e}^{\mathrm{W}(z)}} = \frac{\mathrm{W}(z)}{(\mathrm{W}(z)+1)z}. \tag{3.14}$$

We have then

$$\begin{aligned}&(\forall \nu\in\mathbb{R}) \\ &\Phi'(\nu) = \sum_{n=1}^{N} \mathrm{W}'\big(\rho\exp(w_n(\nu))\big)\exp\big(w_n(\nu)\big)w_n'(\nu) \\ &= -\frac{1}{\gamma_p\lambda\rho}\sum_{n=1}^{N}\left(1 - \frac{1}{\mathrm{W}\big(\rho\exp(w_n(\nu))\big)+1}\right). \end{aligned} \tag{3.15}$$

Therefore, since W takes positive values on $(0,+\infty)$, $\Phi'(\nu) < 0$ for every $\nu\in\mathbb{R}$, i.e., Φ is strictly decreasing. We thus conclude that it has a unique zero $\widehat{\nu}$. □

The computation of the above proximity operator requires to determine the zero of the scalar function Φ. The following lemma shows that this can be achieved with high precision using Newton algorithm, the convergence of which is guaranteed for any initialization.

Lemma 1 *The Newton iteration*

$$(\forall t\in\mathbb{N})\quad \nu^{(t+1)} = \nu^{(t)} - \frac{\Phi(\nu^{(t)})}{\Phi'(\nu^{(t)})} \tag{3.16}$$

converges to the unique zero of Φ from any starting point $\nu^{(0)}\in\mathbb{R}$.

Proof We have already shown that Φ is strictly decreasing on $\mathbb{R}$ and has a unique zero. Let us now establish the convexity of Φ by calculating its second-order derivative

$$\begin{aligned}&(\forall \nu\in\mathbb{R}) \\ &\Phi''(\nu) = -\frac{1}{\gamma_p\lambda}\sum_{n=1}^{N}\frac{\mathrm{W}'(\rho\exp(w_n(\nu)))\exp(w_n(\nu))w_n'(\nu)}{(\mathrm{W}(\rho\exp(w_n(\nu)))+1)^2} \\ &= \frac{1}{\gamma_p^2\lambda^2\rho}\sum_{n=1}^{N}\frac{\mathrm{W}(\rho\exp(w_n(\nu)))}{(\mathrm{W}(\rho\exp(w_n(\nu)))+1)^3}. \end{aligned} \tag{3.17}$$

Since $\mathrm{W}(\rho\exp(w_n(\nu))) > 0$ for every $n\in\{1,\dots,N\}$ and $\nu\in\mathbb{R}$, we have $\Phi''(\nu) > 0$ for all $\nu\in\mathbb{R}$, i.e., Φ is strictly convex. Now, let us ascertain the convergence of Newton's method for finding the unique root of Φ. The remaining of our proof follows similar arguments as the one of [29, Chapter 3, Theorem 2]. For every $t\in\mathbb{N}$, let $e^{(t)}$ is the error defined as

$$(\forall t\in\mathbb{N})\quad e^{(t)} = \nu^{(t)} - \widehat{\nu},$$

where $\widehat{\nu}$ is the zero of Φ. From the definition of the Newton iteration, we have

$$\begin{aligned}(\forall t\in\mathbb{N})\quad e^{(t+1)} &= \nu^{(t+1)} - \widehat{\nu} = \nu^{(t)} - \frac{\Phi(\nu^{(t)})}{\Phi'(\nu^{(t)})} - \widehat{\nu} \\ &= e^{(t)} - \frac{\Phi(\nu^{(t)})}{\Phi'(\nu^{(t)})} = \frac{e^{(t)}\Phi'(\nu^{(t)}) - \Phi(\nu^{(t)})}{\Phi'(\nu^{(t)})}. \end{aligned} \tag{3.18}$$

By performing a second-order Taylor expansion, we get

$$\begin{aligned}(\forall t\in\mathbb{N})\quad 0 = \Phi(\widehat{\nu}) &= \Phi(\nu^{(t)} - e^{(t)}) \\ &= \Phi(\nu^{(t)}) - e^{(t)}\Phi'(\nu^{(t)}) + \frac{1}{2}(e^{(t)})^2\Phi''(\xi^{(t)}), \end{aligned} \tag{3.19}$$

where, for all $t\in\mathbb{N}$, $\xi^{(t)}\in[\min(\widehat{\nu},\nu^{(t)}),\max(\widehat{\nu},\nu^{(t)})]$. Combining the latter equality with (3.18) yields

$$(\forall t\in\mathbb{N})\quad e_{t+1} = \frac{1}{2}\frac{\Phi''(\xi^{(t)})}{\Phi'(\nu^{(t)})}(e^{(t)})^2. \tag{3.20}$$

Recall that $\Phi'(v) < 0$ and $\Phi''(v) > 0$ for all $v \in \mathbb{R}$. According to (3.20), for every $t \in \mathbb{N}$, $e^{(t+1)} < 0$, which implies that $v^{(t)} < \widehat{v}$ for all $t \geq 1$. Thus, since Φ is strictly decreasing, $(\forall t \geq 1)\ \Phi(v^{(t)}) > \Phi(\widehat{v}) = 0$. By (3.18), $(\forall t \geq 1)\ e^{(t+1)} > e^{(t)}$, and thus $(e^{(t)})_{t\geq 1}$ is increasing and upper bounded by 0. Hence, $(v^{(t)})_{t\geq 1}$ is also increasing and upper bounded by $\widehat{v}$. Therefore, the limits $e^* = \lim_{t\to+\infty} e^{(t)}$ and $v^* = \lim_{t\to+\infty} v^{(t)}$ exist. We deduce from (3.18) that $e^* = e^* - \Phi(v^*)/\Phi'(v^*)$, which implies that $\Phi(v^*) = 0$ and $v^* = \widehat{v}$. □

Proposition 3 *Let $(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D}) \in \mathbb{R}\times\mathbb{R}\times\mathbb{R}^N\times\mathbb{R}^Q\times\mathscr{S}_Q$ and $\gamma_\mu > 0$. The proximity operator of $\gamma_\mu F(a,b,\boldsymbol{p},\cdot,\boldsymbol{D})$ at $\boldsymbol{\mu}$ is given by*

$$\operatorname{prox}_{\gamma_\mu F(a,b,\boldsymbol{p},\cdot,\boldsymbol{D})}(\boldsymbol{\mu}) = \Big(\boldsymbol{I}_Q + \gamma_\mu\lambda(\mathbf{1}_N^\top\boldsymbol{p})(\boldsymbol{D}+\varepsilon\boldsymbol{I}_Q)\Big)^{-1} \times\left(\boldsymbol{\mu} + \gamma_\mu\lambda\sum_{n=1}^N p_n(\boldsymbol{D}+\varepsilon\boldsymbol{I}_Q)\boldsymbol{x}_n\right). \quad (3.21)$$

Proof Calculating the proximity of operator of $\gamma_\mu F(a,b,\boldsymbol{p},\cdot,\boldsymbol{D})$ is equivalent to calculating the proximity operator of the quadratic function

$$\boldsymbol{\mu} \mapsto \gamma_\mu\lambda\sum_{n=1}^N \frac{p_n}{2}(\boldsymbol{x}_n-\boldsymbol{\mu})^\top(\boldsymbol{D}+\varepsilon\boldsymbol{I}_Q)(\boldsymbol{x}_n-\boldsymbol{\mu}). \quad (3.22)$$

The result then follows from [18]. □

Proposition 4 *Let $(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D}) \in \mathbb{R}\times\mathbb{R}\times\mathbb{R}^N\times\mathbb{R}^Q\times\mathscr{S}(Q)$ and $\gamma_D > 0$. The proximity operator of $\gamma_D F(a,b,\boldsymbol{p},\boldsymbol{\mu},\cdot)$ at $\boldsymbol{D}$ is given by*

$$\operatorname{prox}_{\gamma_D F(a,b,\boldsymbol{p},\boldsymbol{\mu},\cdot)}(\boldsymbol{D}) = \frac{1}{2}\boldsymbol{V}\operatorname{Diag}\left(\left(\max\left(\omega_q - \varepsilon + \sqrt{(\omega_q+\varepsilon)^2+4m},0\right)\right)_{1\leq q\leq Q}\right)\times\boldsymbol{V}^\top, \quad (3.23)$$

where $\boldsymbol{\omega} = (\omega_q)_{1\leq q\leq Q}$ is a vector of eigenvalues of $\boldsymbol{D}-\boldsymbol{S}$ and $\boldsymbol{V}$ is a $Q\times Q$ orthogonal matrix such that $\boldsymbol{D}-\boldsymbol{S} = \boldsymbol{V}\operatorname{Diag}(\boldsymbol{\omega})\boldsymbol{V}^\top$ with $\boldsymbol{S} = \frac{1}{2}\gamma_D\lambda\sum_{n=1}^N p_n(\boldsymbol{x}_n-\boldsymbol{\mu})(\boldsymbol{x}_n-\boldsymbol{\mu})^\top$ and $m = \frac{1}{2}\gamma_D\lambda(\mathbf{1}_N^\top\boldsymbol{p})$.

Proof Let $\|\cdot\|_{\mathrm{F}}$ denote the Frobenius norm and let $\widetilde{\boldsymbol{D}} \in \mathscr{S}_Q$. We have

$$\begin{aligned}
&\operatorname{prox}_{\gamma_D F(a,b,\boldsymbol{p},\boldsymbol{\mu},\cdot)}(\widetilde{\boldsymbol{D}})\\
&= \operatorname*{argmin}_{\boldsymbol{D}\in\mathscr{S}_Q} \gamma_D F(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D}) + \frac{1}{2}\|\boldsymbol{D}-\widetilde{\boldsymbol{D}}\|_{\mathrm{F}}^2\\
&= \operatorname*{argmin}_{\boldsymbol{D}\in\mathscr{S}_Q} \frac{1}{2}\|\boldsymbol{D}-\widetilde{\boldsymbol{D}}\|_{\mathrm{F}}^2 + \operatorname{tr}(\boldsymbol{D}\boldsymbol{S}) + m\varphi(\boldsymbol{D}) + \iota_{\mathscr{S}_Q^+}(\boldsymbol{D})\\
&= \operatorname{prox}_{m\varphi+\iota_{\mathscr{S}_Q^+}}(\widetilde{\boldsymbol{D}}-\boldsymbol{S}). \quad (3.24)
\end{aligned}$$

Since φ and $\iota_{\mathscr{S}_Q^+}$ are spectral functions on $\mathscr{S}_Q$ associated with the functions $\widetilde{\varphi}$ and $\iota_{[0,+\infty)^Q}$, respectively, it follows from [6, Corollary 24.65] that

$$\operatorname{prox}_{\gamma_D F(a,b,\boldsymbol{p},\boldsymbol{\mu},\cdot)}(\widetilde{\boldsymbol{D}}) = \boldsymbol{V}\operatorname{Diag}\left(\operatorname{prox}_{m\widetilde{\varphi}+\iota_{[0,+\infty)^Q}}(\boldsymbol{\omega})\right)\boldsymbol{V}^\top, \quad (3.25)$$

where $\boldsymbol{\omega} = (\omega_q)_{1\leq q\leq Q}$ is a vector of eigenvalues of $\widetilde{\boldsymbol{D}}-\boldsymbol{S}$ and $\boldsymbol{V}$ is a $Q\times Q$ orthogonal matrix such that $\widetilde{\boldsymbol{D}}-\boldsymbol{S} = \boldsymbol{V}\operatorname{Diag}(\boldsymbol{\omega})\boldsymbol{V}^\top$. Since $m\widetilde{\varphi}+\iota_{[0,+\infty)^Q}$ is a separable function,

$$\operatorname{prox}_{m\widetilde{\varphi}+\iota_{[0,+\infty)^Q}}(\boldsymbol{\omega}) = (\widehat{\sigma}_q)_{1\leq q\leq Q}, \quad (3.26)$$

where, for every $q \in \{1,\ldots,Q\}$,

$$\begin{aligned}
\widehat{\sigma}_q &= \operatorname*{argmin}_{\sigma_q\in[0,+\infty)} -m\log(\sigma_q+\varepsilon) + \frac{1}{2}(\sigma_q-\omega_q)^2\\
&= \frac{1}{2}\max\left(\omega_q - \varepsilon + \sqrt{(\omega_q+\varepsilon)^2+4m},0\right). \quad (3.27)
\end{aligned}$$

4 Convergence Analysis

Let us now establish the convergence of the iterates generated by Algorithm 1. Our analysis will rely on the observation that **FIGARO** can be viewed as a special instance of the regularized Gauss-Seidel method from [4].

4.1 Preliminaries

Let us first recall some useful definitions concerning variational analysis and the fundamental Kurdyka-Łojasiewicz property that will be at the core of the convergence analysis of our algorithm.

Definition 3 (Subdifferential) [33, Def. 8.3] Let $f : \mathbb{R}^n \to (-\infty,+\infty]$ be a proper function.

(a) For a given $x \in \operatorname{dom} f$, the *Fréchet subdifferential* of f at x, written $\widehat{\partial} f(x)$, is the set of all vectors $u \in \mathbb{R}^n$ which satisfy

$$\liminf_{y\neq x\ y\to x} \frac{f(y)-f(x)-\langle u, y-x\rangle}{\|y-x\|} \geq 0.$$

When $x \notin \operatorname{dom} f$, we set $\widehat{\partial} f(x) = \varnothing$.

(b) The *limiting-subdifferential*, or simply the *subdifferential*, of f at $x \in \operatorname{dom} f$, written $\partial f(x)$, is defined as

$$\partial f(x) = \big\{v \in \mathbb{R}^n \mid \exists x^{(t)} \to x, f(x^{(t)}) \to f(x), v^{(t)} \in \widehat{\partial} f(x^{(t)}) \to v\big\}.$$

Definition 4 (Kurdyka-Łojasiewicz property) [10] The function $f:\mathbb{R}^n\to(-\infty,+\infty]$ is said to satisfy the *Kurdyka-Łojasiewicz (KL) property* at $x^*\in\operatorname{dom}\partial f$ if there exist $\eta\in(0,+\infty]$, a neighbourhood U of x^*, and a continuous concave function $\varphi:[0,\eta)\to\mathbb{R}^+$ such that

(a) $\varphi(0)=0$,
(b) φ is $\mathscr{C}^1$ on $(0,\eta)$,
(c) for all $s\in(0,\eta)$, $\varphi'(s)>0$,
(d) for all $x\in U\cap[f(x^*)<f<f(x^*)+\eta]$, the Kurdyka-Łojasiewicz inequality holds:

$$\varphi'(f(x)-f(x^*))\operatorname{dist}(0,\partial f(x))\geq 1.$$

Moreover, f is called a *KL function* if it satisfies the Kurdyka-Łojasiewicz inequality at every point in $\operatorname{dom}\partial f$.

4.2 Convergence Theorem

In order to establish convergence results, we will show that the objective function is KL, and that it can be split into the sum of a locally Lipschitz differentiable part involving all the variables, and non differentiable separable terms.

Lemma 2 *Function* (4.1) *is a KL function.*

Proof Let us recall that there exists an o-minimal structure, denoted by $\mathfrak{S}(\mathbb{R}_{\text{an,exp}})$ with $\mathbb{R}_{\text{an,exp}}:=(\mathbb{R},+,\cdot,(f),\exp)$, that contains the exponential functions and every restricted analytic functions (see [20, Example (6), pp. 505]). Note that $\mathfrak{S}(\mathbb{R}_{\text{an,exp}})$ also contains the logarithm function $\log\colon(0,+\infty)\to\mathbb{R}$ and $(\cdot)^r:\mathbb{R}\to\mathbb{R}$ defined by

$$a\mapsto\begin{cases}a^r, & a>0\\ 0, & a\leq 0,\end{cases}$$

where $r\in\mathbb{R}$. Then, by using [20, Section 5], we conclude that F is definable in an o-minimal structure. As a consequence, the results of [9] and Theorem 4.1 of [3] apply and hence F is a KL function.

Lemma 3 *Function* (4.1) *can be rewritten as*

$$\begin{aligned}&(\forall a\in\mathbb{R})(\forall b\in\mathbb{R})(\forall\boldsymbol{p}\in\mathbb{R}^N)(\forall\boldsymbol{\mu}\in\mathbb{R}^Q)(\forall\boldsymbol{D}\in\mathscr{S}_Q)\\ &F(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D})=G(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D})\\ &\qquad+f_1(a)+f_2(b)+f_3(\boldsymbol{p})+f_4(\boldsymbol{D}),\end{aligned}\tag{4.1}$$

where

$$\begin{aligned}&(\forall a\in\mathbb{R})(\forall b\in\mathbb{R})(\forall\boldsymbol{p}\in\mathbb{R}^N)(\forall\boldsymbol{\mu}\in\mathbb{R}^Q)(\forall\boldsymbol{D}\in\mathscr{S}(Q))\\ &G(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D})=\frac{1}{2}\|\boldsymbol{y}-a\mathbf{1}_N-b\boldsymbol{p}\|^2\\ &\quad+\lambda\sum_{n=1}^N\frac{p_n}{2}\Big((\boldsymbol{x}_n-\boldsymbol{\mu})^\top(\boldsymbol{D}+\varepsilon\boldsymbol{I}_Q)(\boldsymbol{x}_n-\boldsymbol{\mu})+\varphi(\boldsymbol{D})\Big),\end{aligned}\tag{4.2}$$

$$(\forall a\in\mathbb{R})\quad f_1(a)=\imath_{\mathscr{A}}(a),\tag{4.3}$$

$$(\forall b\in\mathbb{R})\quad f_2(b)=\imath_{\mathscr{B}}(b),\tag{4.4}$$

$$\begin{aligned}&(\forall\boldsymbol{p}\in\mathbb{R}^N)\quad f_3(\boldsymbol{p})=\imath_{\mathscr{C}}(\boldsymbol{p})\\ &\qquad+\lambda\sum_{n=1}^N\Big(\operatorname{ent}(p_n)+p_n\frac{Q}{2}\log(2\pi)\Big),\end{aligned}\tag{4.5}$$

$$(\forall\boldsymbol{D}\in\mathscr{S}(Q))\quad f_4(\boldsymbol{D})=\imath_{\mathscr{S}^+(Q)}(\boldsymbol{D}).\tag{4.6}$$

Moreover, G is C^2 on $\mathbb{R}\times\mathbb{R}\times\mathbb{R}^N\times\mathbb{R}^Q\times\mathscr{S}(Q)$.

Proof We first calculate the gradients $\nabla_aG,\nabla_bG,\nabla_{\boldsymbol{p}}G,\nabla_{\boldsymbol{\mu}}G$ and $\nabla_{\boldsymbol{D}}G$ of G with respect to the different variables. Let us denote by $(\boldsymbol{e}_{n,N})_{1\leq n\leq N}$ the canonical basis of $\mathbb{R}^N$. For every $(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D})\in\mathbb{R}\times\mathbb{R}\times\mathbb{R}^N\times\mathbb{R}^Q\times\mathscr{S}(Q)$,

$$\nabla_aG(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D})=b\mathbf{1}_N^\top\boldsymbol{p}+Na-\mathbf{1}_N^\top\boldsymbol{y},$$

$$\nabla_bG(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D})=b\|\boldsymbol{p}\|^2-\boldsymbol{y}^\top\boldsymbol{p}+a\mathbf{1}_N^\top\boldsymbol{p}=(b\boldsymbol{p}-\boldsymbol{y}+a\mathbf{1}_N)^\top\boldsymbol{p},$$

$$\begin{aligned}\nabla_{\boldsymbol{p}}G(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D})&=b^2\boldsymbol{p}-b\boldsymbol{y}+ab\mathbf{1}_N\\ &\quad+\frac{\lambda}{2}\sum_{n=1}^N\Big((\boldsymbol{x}_n-\boldsymbol{\mu})^\top(\boldsymbol{D}+\varepsilon\boldsymbol{I}_Q)(\boldsymbol{x}_n-\boldsymbol{\mu})+\varphi(\boldsymbol{D})\Big)\boldsymbol{e}_{n,N},\end{aligned}$$

$$\nabla_{\boldsymbol{\mu}}G(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D})=\lambda\sum_{n=1}^Np_n(\boldsymbol{D}+\varepsilon\boldsymbol{I}_Q)(\boldsymbol{\mu}-\boldsymbol{x}_n),$$

$$\nabla_{\boldsymbol{D}}G(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D})=\begin{cases}\dfrac{\lambda}{2}\displaystyle\sum_{n=1}^Np_n\Big((\boldsymbol{x}_n-\boldsymbol{\mu})(\boldsymbol{x}_n-\boldsymbol{\mu})^\top-(\boldsymbol{D}+\varepsilon\boldsymbol{I}_Q)^{-1}\Big)\\ \qquad\text{if}\quad\boldsymbol{D}\in\mathscr{S}^+(Q),\\ \dfrac{\lambda}{2}\displaystyle\sum_{n=1}^Np_n\Big((\boldsymbol{x}_n-\boldsymbol{\mu})(\boldsymbol{x}_n-\boldsymbol{\mu})^\top-\varepsilon^{-1}\boldsymbol{I}_Q+\varepsilon^{-2}\boldsymbol{D}\Big)\\ \qquad\text{otherwise.}\end{cases}\tag{4.7}$$

Let us now calculate the partial second-order derivatives of G. In the following, $\otimes$ denotes the matrix Kronecker product and $\operatorname{vec}(\boldsymbol{M})$ the columnwise ordering of a matrix $\boldsymbol{M}$. For every $(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D})\in\mathbb{R}\times\mathbb{R}\times\mathbb{R}^N\times\mathbb{R}^Q\times\mathscr{S}(Q)$, by setting

$\boldsymbol{d} = \mathrm{vec}(\boldsymbol{D})$, we have

$$
\begin{aligned}
&\nabla^2_a G(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D}) = N,\\
&\nabla^2_b G(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D}) = \|\boldsymbol{p}\|^2,\\
&\nabla^2_{\boldsymbol{p}} G(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D}) = b^2\boldsymbol{I}_N,\\
&\nabla^2_{\boldsymbol{\mu}} G(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D}) = \lambda\sum_{n=1}^{N} p_n(\boldsymbol{D}+\varepsilon\boldsymbol{I}_Q),\\
&\nabla^2_{a,b} G(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D}) = \mathbf{1}_N^\top\boldsymbol{p},\\
&\nabla^2_{\boldsymbol{p},a} G(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D}) = b\mathbf{1}_N,\\
&\nabla^2_{\boldsymbol{\mu},a} G(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D}) = \mathbf{0}_Q,\\
&\nabla^2_{\boldsymbol{d},a} G(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D}) = \mathbf{0}_{Q^2},\\
&\nabla^2_{\boldsymbol{p},a} G(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D}) = 2b\boldsymbol{p}-\boldsymbol{y}+a\mathbf{1}_N,\\
&\nabla^2_{\boldsymbol{\mu},b} G(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D}) = \mathbf{0}_Q,\\
&\nabla^2_{\boldsymbol{d},b} G(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D}) = \mathbf{0}_{Q^2},\\
&\nabla^2_{\boldsymbol{\mu},\boldsymbol{p}} G(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D}) = \lambda(\boldsymbol{D}+\varepsilon\boldsymbol{I}_Q)\sum_{n=1}^{N}(\boldsymbol{\mu}-\boldsymbol{x}_n)\boldsymbol{e}_{n,N}^\top,
\end{aligned}
$$

$$
\begin{aligned}
&\nabla^2_{\boldsymbol{p},\boldsymbol{d}} G(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D})\\
&= \begin{cases}
\dfrac{\lambda}{2}\displaystyle\sum_{n=1}^{N}\Big((\boldsymbol{x}_n-\boldsymbol{\mu})^\top\otimes\big(\boldsymbol{e}_{n,N}(\boldsymbol{x}_n-\boldsymbol{\mu})^\top\big)\\
\qquad -\boldsymbol{e}_{n,N}\mathrm{vec}\big((\boldsymbol{D}+\varepsilon\boldsymbol{I}_Q)^{-1}\big)^\top\Big) & \text{if } \boldsymbol{D}\in\mathscr{S}^+(Q),\\
\dfrac{\lambda}{2}\displaystyle\sum_{n=1}^{N}\Big((\boldsymbol{x}_n-\boldsymbol{\mu})^\top\otimes\big(\boldsymbol{e}_{n,N}(\boldsymbol{x}_n-\boldsymbol{\mu})^\top\big)\\
\qquad -\varepsilon^{-1}\boldsymbol{e}_{n,N}(\mathbf{1}_{Q^2}-\varepsilon^{-1}\boldsymbol{d})^\top\Big) & \text{otherwise,}
\end{cases}
\end{aligned}
$$

$$
\nabla^2_{\boldsymbol{\mu},\boldsymbol{d}} G(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D}) = \lambda\sum_{n=1}^{N} p_n(\boldsymbol{\mu}-\boldsymbol{x}_n)^\top\otimes\boldsymbol{I}_Q,
$$

$$
\nabla^2_{\boldsymbol{d}} G(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D}) = \begin{cases}
\dfrac{\lambda}{2}(\mathbf{1}_N^\top\boldsymbol{p})(\boldsymbol{D}+\varepsilon\boldsymbol{I}_Q)^{-1}\otimes(\boldsymbol{D}+\varepsilon\boldsymbol{I}_Q)^{-1}\\
\qquad \text{if} \quad \boldsymbol{D}\in\mathscr{S}^+(Q),\\
\dfrac{\lambda}{2}(\mathbf{1}_N^\top\boldsymbol{p})\varepsilon^{-2}\boldsymbol{I}_{Q^2}\\
\qquad \text{otherwise.}
\end{cases} \tag{4.8}
$$

Thanks to the definition of φ, the Hessian of G is thus defined and continuous on $\mathbb{R}\times\mathbb{R}\times\mathbb{R}^N\times\mathbb{R}^Q\times\mathscr{S}(Q)$. Hence the result.

We are now ready to prove the convergence of **FIGARO**.

Theorem 4.1 *Let $(\boldsymbol{t}^{(i)})_{i\in\mathbb{N}} = (a^{(i)},b^{(i)},\boldsymbol{p}^{(i)},\boldsymbol{\mu}^{(i)},\boldsymbol{D}^{(i)})_{i\in\mathbb{N}}$ be a sequence generated by Algorithm 1. If $(\boldsymbol{D}^{(i)})_{i\in\mathbb{N}}$ is upper bounded, then $(\boldsymbol{t}^{(i)})_{i\in\mathbb{N}}$ converges to $\widehat{\boldsymbol{t}} = (\widehat{a},\widehat{b},\widehat{\boldsymbol{p}},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{D}})$ satisfying the following equilibrium:*

$$
\begin{aligned}
&(\forall a\in\mathbb{R})(\forall b\in\mathbb{R})(\forall\boldsymbol{p}\in\mathbb{R}^N)(\forall\boldsymbol{\mu}\in\mathbb{R}^Q)(\forall\boldsymbol{D}\in\mathscr{S}(Q))\\
&\quad F(a,\widehat{b},\widehat{\boldsymbol{p}},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{D}}) \geq F(\widehat{a},\widehat{b},\widehat{\boldsymbol{p}},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{D}})\\
&\quad F(\widehat{a},b,\widehat{\boldsymbol{p}},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{D}}) \geq F(\widehat{a},\widehat{b},\widehat{\boldsymbol{p}},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{D}})\\
&\quad F(\widehat{a},\widehat{b},\boldsymbol{p},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{D}}) \geq F(\widehat{a},\widehat{b},\widehat{\boldsymbol{p}},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{D}})\\
&\quad F(\widehat{a},\widehat{b},\widehat{\boldsymbol{p}},\boldsymbol{\mu},\widehat{\boldsymbol{D}}) \geq F(\widehat{a},\widehat{b},\widehat{\boldsymbol{p}},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{D}})\\
&\quad F(\widehat{a},\widehat{b},\widehat{\boldsymbol{p}},\widehat{\boldsymbol{\mu}},\boldsymbol{D}) \geq F(\widehat{a},\widehat{b},\widehat{\boldsymbol{p}},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{D}}).
\end{aligned} \tag{4.9}
$$

Moreover the sequence $(\boldsymbol{t}^{(i)})_{i\in\mathbb{N}}$ has a finite length.

Proof In (4.1), it appears that, if $\boldsymbol{p}\notin[0,+\infty)^N\cap\mathscr{C}$ or $\boldsymbol{D}\notin\mathscr{S}^+(Q)$, then $F(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D}) = +\infty$, whereas, if $\boldsymbol{p}\in[0,+\infty)^N\cap\mathscr{C}$ and $\boldsymbol{D}\in\mathscr{S}^+(Q)$,

$$
\begin{aligned}
&G(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D})\\
&\quad\geq\lambda\sum_{n=1}^{N}\frac{p_n}{2}\Big((\boldsymbol{x}_n-\boldsymbol{\mu})^\top(\boldsymbol{D}+\varepsilon\boldsymbol{I}_Q)(\boldsymbol{x}_n-\boldsymbol{\mu})+\varphi(\boldsymbol{D})\Big)\\
&\quad\geq\lambda\frac{\Delta^{-1}}{2}\left(\varepsilon\inf_{n\in\{1,\dots,N\}}\|\boldsymbol{x}_n-\boldsymbol{\mu}\|^2-Q\log\varepsilon\right),
\end{aligned} \tag{4.10}
$$

$$
f_1(a)\geq 0, f_2(b)\geq 0, f_4(\boldsymbol{D}) = 0, \tag{4.11}
$$

$$
f_3(\boldsymbol{p})\geq\lambda\sum_{n=1}^{N}\mathrm{ent}(p_n)\geq -Ne^{-1}\lambda. \tag{4.12}
$$

Hence

$$
\begin{aligned}
&F(a,b,\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{D})\\
&\quad\geq\lambda\frac{\Delta^{-1}}{2}\left(\varepsilon\inf_{n\in\{1,\dots,N\}}\|\boldsymbol{x}_n-\boldsymbol{\mu}\|^2-Q\log\varepsilon\right)-Ne^{-1}\lambda.
\end{aligned} \tag{4.13}
$$

This shows that F is bounded from below. Moreover, since FIGARO alternates proximal steps, $\big(F(\boldsymbol{t}^{(i)})\big)_{i\in\mathbb{N}}$ is a decaying convergent sequence. It then follows from (4.13) that $(\boldsymbol{\mu}^{(i)})_{i\in\mathbb{N}}$ is bounded (otherwise the function value sequence would be divergent). Since $(a^{(i)})_{i\in\mathbb{N}}$, $(b^{(i)})_{i\in\mathbb{N}}$ and $(\boldsymbol{D}^{(i)})_{i\in\mathbb{N}}$ are bounded sequences, $(\boldsymbol{t}^{(i)})_{i\in\mathbb{N}}$ is bounded. Moreover, according to Lemma 3, G is C^2 on $\mathbb{R}\times\mathbb{R}\times\mathbb{R}^N\times\mathbb{R}^Q\times\mathscr{S}(Q)$, which implies that G is C^1 with locally Lipschitz gradient on $\mathbb{R}\times\mathbb{R}\times\mathbb{R}^N\times\mathbb{R}^Q\times\mathscr{S}(Q)$. Consequently, all the conditions in [4, Theorem 6.2] are met to guarantee that $(\boldsymbol{t}^{(i)})_{i\in\mathbb{N}}$ is a finite length sequence converging to a critical point of F. We then deduce (4.9) from the fact that F is convex with respect to each of its argument.

Remark 2 Note that the assumption on the boundedness of $(\boldsymbol{D}^{(i)})_{i\in\mathbb{N}}$ becomes unnecessary if an upper bound on $\boldsymbol{D}$ is introduced in the formulation of the optimization problem. This however was not observed to influence the practical behaviour of the algorithm.

5 Experiments on Synthetic Data

In order to validate the good performance of the **FIGARO** Algorithm 1, we generate 3D synthetic data $\boldsymbol{y} = (y(\boldsymbol{x}_n))_{1\leq n\leq N}$ where $(\boldsymbol{x}_n)_{1\leq n\leq N}$ are coordinates in $\mathbb{R}^3$ regularly spaced on a grid with size $N = 15\times 15\times 50$ and voxel dimension $0.05\times 0.05\times 0.1\mu m^3$. For every $n\in\{1,\ldots,N\}$, $y_n = \overline{a}+\overline{b}\widetilde{p}(\boldsymbol{x}_n)+v_n$. In order to illustrate the robustness of our formulation, we define $\widetilde{p}$ from the multivariate generalized Gaussian probability density function:

$$(\forall n\in\{1,\ldots,N\})\ \widetilde{p}(\boldsymbol{x}_n) = \sqrt{|\overline{\boldsymbol{C}}|}\frac{\rho\Gamma(\frac{3}{2})}{\pi^{\frac{3}{2}}2^{\frac{3}{2\rho}}\Gamma(\frac{3}{2\rho})\varsigma^{\frac{3}{2}}}\exp\left(-\frac{1}{2\varsigma^{\rho}}((\boldsymbol{x}_n-\overline{\boldsymbol{\mu}})^{\top}\overline{\boldsymbol{C}}(\boldsymbol{x}_n-\overline{\boldsymbol{\mu}}))^{\rho}\right),$$

with scale and shape parameters $(\varsigma,\rho)\in(0,+\infty)^2$. Various values will be tested for ρ, and for each of them, the scale parameter ς is adjusted such that most of the probability mass lies in the observation grid. When $\rho = 1$, we recover the standard multivariate Gaussian distribution. We set $\overline{a} = \overline{b} = 1$, $\overline{\boldsymbol{\mu}} = [0.3, 0.4, 2]^{\top}$, and

$$\overline{\boldsymbol{C}} = \left(\mathbf{R}\,\mathrm{Diag}([0.1, 0.05, 0.5]^{\top})\mathbf{R}^{\top}\right)^{-1}$$

with $\mathbf{R}\in\mathbb{R}^{3\times 3}$ the rotation matrix associated with angles equal to $\left(0.2, 10^{-4}, 0.1\right)$ radians. Note that these values for the distribution parameters $(\overline{b},\overline{C},\overline{\boldsymbol{\mu}})$ have been chosen in our tests in order to correspond to typical values encountered in our target application to PSF estimation in microscopy. Finally, $\mathbf{v} = (v_n)_{1\leq n\leq N}$ is the realization of a zero-mean Gaussian noise, with standard deviation σ chosen so as to obtain a given input signal-to-noise ratio (SNR).

The regularization parameter $\lambda > 0$ in **FIGARO** is set automatically thanks to a golden bisection search, so as to satisfy the χ^2 criterion $\|\boldsymbol{y}-\widehat{a}-\widehat{b}\widehat{\boldsymbol{p}}\| = \sigma\sqrt{N}$ [22]. We set $a_{\min} = b_{\min} = 0$, $a_{\max} = b_{\max} = 10^5$, $\varepsilon = 10^{-8}$. The initialization of the algorithm is of particular matter, as the cost function is nonconvex. Here, we observed that a good initialization strategy is to take $a^{(0)} = \min_{n\in\{1,\ldots,N\}} y_n, b^{(0)} = 1, \boldsymbol{p}^{(0)} = \boldsymbol{y}$, $\boldsymbol{\mu}^{(0)}$ as the position of the maximum intensity in $\boldsymbol{y}$, and $\boldsymbol{C}^{(0)}$ a diagonal matrix with entries equal to the voxel size in each direction. The algorithm iterations are stopped as soon as the relative residual between two consecutive iterates on the fitting model $(a+bg(\boldsymbol{x}_n,\boldsymbol{\mu},\boldsymbol{D}+\varepsilon\mathbf{I}))_{1\leq n\leq N}$ is below 10^{-5}.

We provide in Fig. 1 the performance of our approach, in terms of the Percent Root Mean Square Difference (PRD) between the estimated $(\widehat{a}+\widehat{b}\widehat{p}(\boldsymbol{x}_n))_{1\leq n\leq N}$ and the true vector $(\overline{a}+\overline{b}\widetilde{p}(\boldsymbol{x}_n))_{1\leq n\leq N}$, averaged on 50 noise realizations. The range of values for the standard deviations (std) is indicated in the figure caption. We also provide the averaged PRD and associated std range, obtained when solving the problem with the nonlinear least squares approach based on Levenberg-Macquardt (**LM**) algorithm. We use the `lsqcurvefit` function available in Matlab software with the same initialization as **FIGARO**[2]. It is important to emphasize that, even in the case when $\rho\neq 1$, we still assume a Gaussian model in both fitting approaches in order to assess their robustness to an imperfect model.

The plots show that **FIGARO** outperforms **LM**, in all scenarios in terms of averaged PRD. **FIGARO** is, in addition, very stable to a model mismatch (i.e., $\rho\neq 1$), while **LM** performance highly decreases as soon as the data are not generated by using the Gaussian model. This clearly highlights the advantage of our formulation, relying on the extra variable $\boldsymbol{p}$ whose shape is controlled by the KL divergence penalty term. Finally, it is noticeable that **FIGARO** is much more stable to noise fluctuations, as confirmed by the low values of std on the PRD. In contrast, the PRD values for **LM** are highly dispersed, which questions its reliability for the systematic treatment of real datasets.

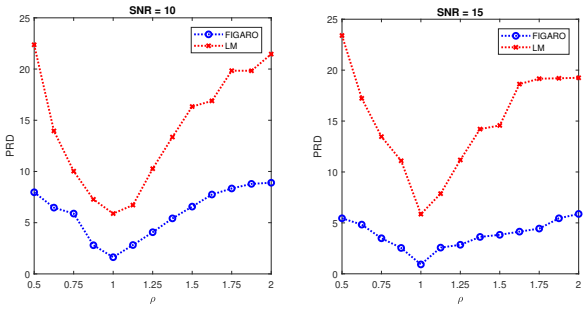


Fig. 1 Quality of 3D fitting results in terms of PRD, using **FIGARO** and **LM** strategies, for different shape parameters ρ and SNR values (in dB). Averaged values over 50 noise realizations. For **FIGARO**, the std varies between 0.05 and 1.72, while for LM, it lies between 7.91 and 20.1.

6 Application of FIGARO to Two-photon Microscopy

The objective of this part is to illustrate experimentally the good performance of our fitting strategy in the context of computational imaging. Multiphoton microscopy (MPM) is a popular method for biomedical imaging at the micron scale, able to generate 3D images *in vivo* and in depth, starting from a superposition of 2D image stacks. However, the instrumental PSF in MPM has a particularly negative impact on the resulting images especially when a sub-micrometer resolution is searched (about less than 0.5 μm) or when the sample emits a low level multiphoton signal. These situations represent most of the cases encountered in MPM where

[2] Our implementation relies on the extension to the 3D case of the 2D Gaussian fitting software publicly available at: `https://fr.mathworks.com/matlabcentral/fileexchange/41938-fit-2d-gaussian-with-optimization-toolbox`.

the PSF is responsible for the resolution and contrast deteriorations, with an increase of the image blur and noise. We propose to apply our multivariate Gaussian fitting strategy **FIGARO** to experimental MPM 3D images of fluorescent microbeads, with the aim to better analyze the instrumental PSF of this modality and to get high quality restoration results. This section is organized as follows. First, the experimental and algorithmic setup is described in Subsection 6.1. Numerical results obtained with **FIGARO** are presented in Subsection 6.2, and Subsection 6.3 shows a comparison with the state-of-the-art **MetroloJ** plugin based on 1D Gaussian fitting on marginalized data, which is highly employed in many platforms as a routine tool for analysis of microscopes resolution power. Finally, Subsection 6.4 illustrates restoration results obtained by using our estimated PSF model.

6.1 Presentation of the Experimental Setup

The experimental dataset has been recorded from a commercial multiphoton microscope (Olympus, BX61WI) employed in a routine protocol for two-photon fluorescence imaging. A standard femtosecond titan sapphire laser source, (Chameleon Ultra II, Coherent Inc., 800 nm, 150 fs, 10 nm, 82 MHz, 4 W) is coupled to the working station ended by a 25× water immersion microscope objective (Olympus, XLPLN 25× WMP, 1.05 numerical aperture). In order to characterize experimentally the optical performance of the microscope and especially its response function, images of fluorescent spherical latex microbeads, having a known diameter smaller than the resolution spot, are generated. The retained microbeads have been provided by Molecular Probes, and have a diameter of 0.2 μm. Such a small diameter of the beads allows us to consider each observed one as the (space-variant) instrument PSF at the bead center coordinates. Microbeads are diluted into liquid gelatin and, after a short period at frig, the gelatin is solidified. The imaged sample thus constitutes the microbeads homogeneously distributed and immobilized into a bulk and solid volume. Their fluorescence emission at 515 nm is detected with a photomultiplier tube coupled with an optical filter between 495 and 540 nm. A dichroic mirror at 690 nm splits the excitation beam from the laser source and the back-fluorescence from the volume of microbeads which is the exclusive one directed to the detection module.

2D image slices are generated, with a dimension of 1600×1600 squared pixels. 230 slices with a pixel size of 0.053 μm are realized in deepness and spaced 0.1 μm apart; the superposition of the 230 slices consequently results in a 3D image having the following dimensions in XYZ: $85 \times 85 \times 23$ μm^3. From this 3D image, forty volumes of interest (VOIs) are selected, each of them corresponding to the noisy and blurry observation of a single bead. For each selected VOI, the **FIGARO** algorithm is ran, using the same settings than those provided in Section 5, giving rise to a set of estimated parameters $(\widehat{a},\widehat{b},\widehat{\boldsymbol{p}},\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{C}})$ (with $\widehat{\boldsymbol{C}} = \widehat{\boldsymbol{D}} + \varepsilon \boldsymbol{I}_Q$) directly related to the position, size and orientation of the PSF.

6.2 3D Estimation Results

Figure 2 shows an illustration of the 3D fitting results for four VOIs. Dots represent the raw data acquired experimentally while red spheres with their axis represent the reconstructed 3D image of each microbead inside its VOI, resulting from our multivariate Gaussian fitting strategy. Here, the contour plots delimit the full-width at the half maximum (FWHM) region, i.e., where $\boldsymbol{x}_n$ is such that $\widehat{a}+\widehat{b}g(\boldsymbol{x}_n,\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{C}}) = 0.5 \times \max(\widehat{a}+\widehat{b}g(\boldsymbol{x}_n,\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{C}}))_{1\leq n\leq N}$.

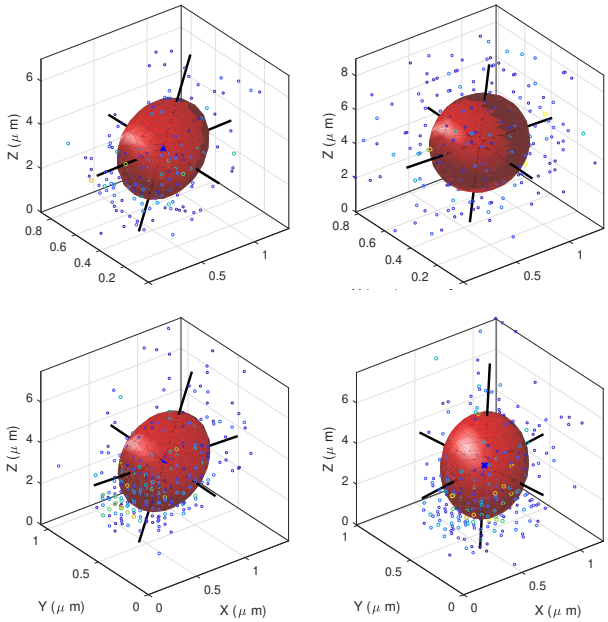


Fig. 2 Example of 3D fitting results using FIGARO on two-photon microscopy data.

In biomedical MPM, the carrying medium has often scattering and absorbing properties not well-known or well-characterized. The more the imaged medium is scattering or absorbing the light (laser excitation or fluorescence emission), the more the image will be deteriorated. This phenomenon is often increasing with the imaging depth. **FIGARO** fitting results allow us to quantify this PFS variation along the depth of the sample. To this aim, we compute the FWHM along the 3 main axes of the Gaussian shapes for each VOIs, defined as $(2\sqrt{2\log 2}s_i)_{1\leq i\leq 3}$ where $(s_i)_{1\leq i\leq 3}$ are the eigenvalues of $\widehat{\boldsymbol{C}}^{-1}$.

An analysis of these results for the whole set of VOIs shows that, for this dataset and this range of depths, the planar width, related to the FWHM associated to the second and

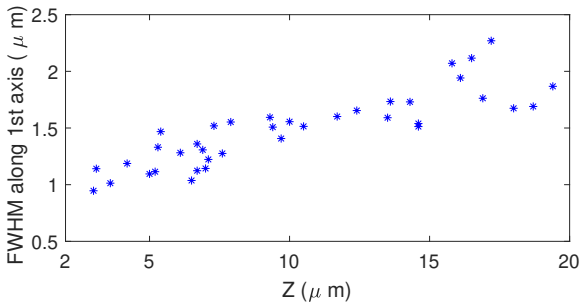


Fig. 3 Evolution of the estimated FWHM along the axial axis of the fitted 3D Gaussian shapes, with respect to the bead center depth.

third eigenvalues of $\widehat{\boldsymbol{C}}^{-1}$, does not vary much with respect to the beads location. Here, the averaged FWHM of the estimated Gaussian shapes is of (0.21, 0.27) μm, which appears to be consistent with the theoretical limit of optical planar resolution of 0.2 μm for this emission wavelength and numerical aperture. The axial PSF width values, related to the maximum eigenvalue of $\widehat{\boldsymbol{C}}^{-1}$, are displayed in Figure 3 as a function of the depth of bead centers. The origin of the abscissa axis is related to the surface of the sample, it is not represented here as the beads employed for these measurements are only present in depths between 3 μm and 20 μm under the surface of the sample. One can observe that the axial PSF width is slightly increasing when the depth of the bead center increases, as it is expected from optical theory [26]. The averaged axial resolution is of 1.49 μm which fits well the theoretical resolution limit of 1.5 μm displayed in the literature [36]. Consequently, **FIGARO** appears as a solution very well adapted for estimating the 3D variability of the PSF of a system.

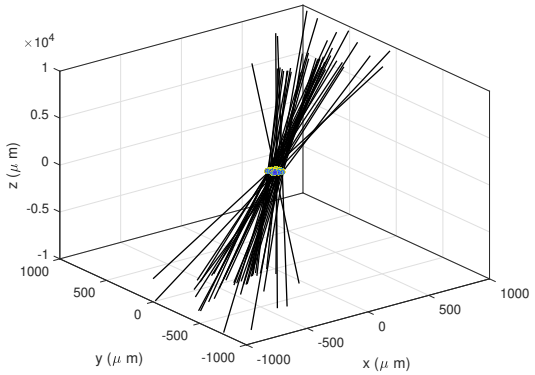


Fig. 4 3D representation of the estimated PSF main axis.

Additionally to the relevant and reliable measurement of PSF widths, our computational strategy gives also access to the orientation of each PSF inside its corresponding VOI. Of particular interest is the computation of the Euler angles (Φ_1, Φ_2) characterizing the slope of the main direction of the PSF, i.e., the eigenvector of $\boldsymbol{C}$ associated with its largest eigenvalue. We represent a 3D representation of the PSF main axes regarding its center position in Figure 4. Due to the presence of optical aberrations, the PSF orientations measured with **FIGARO** change according to the beads location. In particular, the tilt angle quantifying the angle between the Z axis and the main PSF direction (i.e., Φ_2) varies for this dataset between 0.6° and 7.7°.

6.3 Comparison with A Standard Procedure

Let us now present the comparison of our results with those obtained from the **MetroloJ** plugin [3] of Fiji. **MetroloJ** presents several interests in microscopy: it is a free plugin of an open source software, allowing to have a precise idea of the PSF of the microscope, and is now a routine tool for tracking microscope performances. Unfortunately, like other available Fiji plugins for PSF analysis in fluorescence microscopy (eg, **QuickPALM** [27] and **rapidSTORM** [39]), it only performs 1D shape fitting, and thus only allows to treat marginalized versions of the datasets. Thus, one may expect that such dimension reduction comes at the price of a loss in modeling accuracy and thus restoration quality.

For the sake of our comparisons, we have selected four samples from the VOI set. The experimental results are gathered in Table 1. For each VOI, we provide the estimation of the center coordinates of the fitted Gaussian shapes, the FWHM, and the orientation (Euler angles) resulting from **FIGARO** and **MetroloJ** approaches. Since the latter is based on 1D Gaussian fitting on the 3 marginals, only center position and FWHM along the axis XYZ of the image are available as outputs. In contrast, for **FIGARO**, the FWHM is estimated along the actual bead axis, accounting for its inclination angles.

As already observed in the previous section, the PSF orientations measured with **FIGARO** change according to the beads location. The 1D-based analysis of **MetroloJ** does not have access to such a precise estimation of the tilt angle, yet of main importance for an efficient computational processing of the microscope images. Concerning the estimated center positions, they are quite similar for both methods, mainly because of the small size of the VOIs. But results from Table 1 highlight substantial differences in FWHM estimations of the PSF between the two ways of calculations. With **MetroloJ** method, several estimations of FWHM are not significant since the computed sizes are highly below the true bead dimension. The high variability of the estimated FWHM by **MetroloJ** probably results from (i) the ignorance of 3D inclination of the PSF shape, (ii) a high sensitivity to

[3] Available at: `http://imagejdocu.tudor.lu/doku.php?id=plugin:analysis:metroloj:start`

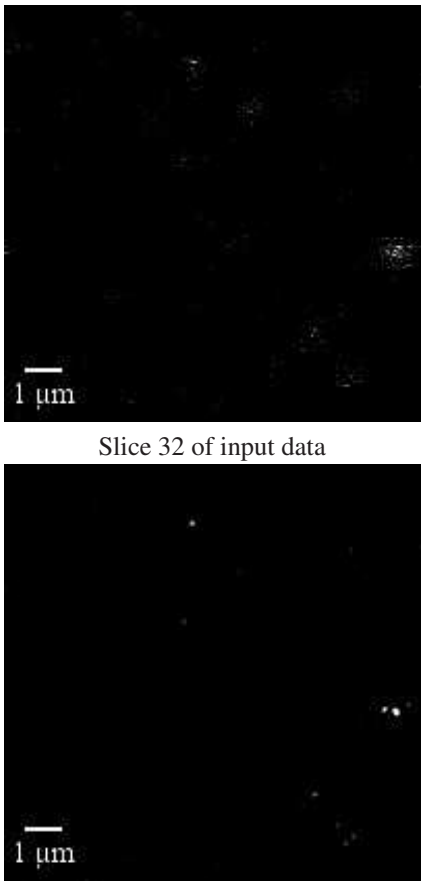
Table 1 Example of fitting results on 4 VOIs for our approach, and the MetroloJ plugin from Fiji.

Volume of Interest		n°1	n°2	n°3	n°4
MetroloJ	Center (μm)	(62.77, 18.59, 5.46)	(41.62, 65.69, 5.50)	(66.01, 0.35, 13.82)	(10.24, 66.96, 10.46)
	FWHM (μm)	(0.32, 0.03, 0.05)	(0.29, 0.03, 0.001)	(0.028, 0.19, 0.1)	(0.05, 0.04, 0.57)
FIGARO	Center (μm)	(62.78, 19.19, 7.57)	(41.71, 66.27, 6.10)	(66.22, 1.03, 14.61)	(10.29, 67.59, 11.72)
	FWHM (μm)	(0.192, 0.247, 1.275)	(0.201, 0.307, 1.282)	(0.198, 0.252, 1.539)	(0.205, 0.259, 1.601)
	Angles (°)	(73.1, 2.38)	(67.3, 5.63)	(87.2, 1.54)	(105.6, 2.24)

noise and model mismatch, both reasons making impossible a correct estimation of the PSF width. This emphasizes the importance of robustly and directly dealing with 3D models, for which FIGARO is able to give reliable and relevant results.

6.4 Increasing the Resolving Power

We finalize this experimental section by presenting restoration results of a section of the same acquired dataset with size $200 \times 200 \times 50$ voxels, corresponding to a field of view of $10 \times 10 \times 5$ μm^3. A constant 3D Gaussian PSF shape is considered in this region, whose width and orientation are deduced from our previously described fitting results by interpolation. The deblurring step is performed using the **OPTIMISM** toolbox from Fiji [4] [16]. Figure 5 illustrates one 2D slice extracted from the input dataset (top) and the corresponding restored image (bottom). In Figure 5(top), the presence of approximately seven microbeads is supposed in this 2D image. For the biggest and brightest one, its diameter is about 1 μm on the raw image, exceeding highly the expected 0.2 μm. No conclusion can be drawn from such poor observation quality. When applying **OPTIMISM** with **FIGARO** fitted PSF, this halo of light appears in fact as a bunch of microbeads as it is visible on Figure 5(bottom). These microbeads were too small and too close to each other to be individually identified with the multiphoton microscope alone and the help of a suitable 3D PSF model, as the one resulting from **FIGARO**, is thus mandatory for increasing numerically the resolving power of the MPM device.



Slice 32 of input data

Restored volume

Fig. 5 Deblurring results.

7 Conclusion

In this paper, a new algorithm has been proposed for multivariate Gaussian fitting of observed data corrupted by additive Gaussian noise. Our approach relies on the proposal of an original hybrid cost function combining a Kullback-Leibler divergence regularizer, a least-squares data fidelity term and range constraints on the parameters. An efficient proximal alternating iterative resolution scheme, grounded on solid mathematical foundations, has been proposed for the resolution of the underlying nonconvex minimization problem. The interest of this strategy named **FIGARO** has been illustrated by means of experiments in fitting synthetic data when a model mismatch is present. We have also presented experimental results in the context of computational fluorescence microscopy. The objective was to characterize the instrumental space-varying 3D PSF of a two-photon fluorescence microscope from raw observations of microbeads. Our numerical tests have shown the efficiency of our method

[4] Available at: `http://sites.imagej.net/Dbenielli/`

for PSF model determination. Future work will address the cases of more general multivariate models and noise statistics.

Acknowledgements

This work was supported by the CNRS under grant MI-AAP Interne2018-SupRéMA.

References

1. Anthony, S.M., Granick, S.: Image analysis with rapid and accurate two-dimensional Gaussian fitting. Langmuir **25**(14), 8152–8160 (2009). DOI 10.1021/la900393v
2. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality. Math. Oper. Res. **35**(2), 438–457 (2010). DOI 10.1287/moor.1100.0449
3. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality. Mathematics of Operations Research **35**(2), 438–457 (2010). DOI 10.1287/moor.1100.0449
4. Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. Math. Prog. **137**(1), 91–129 (2013). DOI 10.1007/s10107-011-0484-9
5. Basseville, M., Cardoso, J.F.: On entropies, divergences, and mean values. In: Proceedings of 1995 IEEE International Symposium on Information Theory, pp. 330– (1995). DOI 10.1109/ISIT.1995.550317
6. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces, 2nd edn. Springer International Publishing (2017)
7. Bertsekas, D.P.: Nonlinear Programming, 2nd edn. Athena Scientific, Belmont, MA (1999)
8. Bolte, J., Combettes, P.L., Pesquet, J.C.: Alternating proximal algorithm for blind image recovery. In: Proc. IEEE Int. Conf. Image Process. (ICIP 2010), pp. 1673–1676. Hong-Kong, China (2010)
9. Bolte, J., Daniilidis, A., Lewis, A., Shiota, M.: Clarke subgradients of stratifiable functions. SIAM Journal on Optimization **18**(2), 556–572 (2007). DOI 10.1137/060670080
10. Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Mathematical Programming **146**(1), 459–494 (2014). DOI 10.1007/s10107-013-0701-9
11. Burger, M., Sawatzky, A., Steidl, G.: First Order Algorithms in Variational Image Processing, pp. 345–407. Springer International Publishing, Cham (2016)
12. Caruana, R., Searle, R., Heller, T., Shupack, S.: Fast algorithm for the resolution of spectra. Anal. Chem. **58**(6), 1162–1167 (1986)
13. Chan, R.H., Chan, T.F., Shen, L., Shen, Z.: Wavelet deblurring algorithms for spatially varying blur from high-resolution image reconstruction. Linear Algebra and its Applications **366**, 139 – 155 (2003). Special issue on Structured Matrices: Analysis, Algorithms and Applications
14. Chaux, C., Pesquet, J., Pustelnik, N.: Nested iterative algorithms for convex constrained image recovery problems. SIAM Journal on Imaging Sciences **2**(2), 730–762 (2009). DOI 10.1137/080727749. URL `https://doi.org/10.1137/080727749`
15. Chen, Y.C., Furenlid, L.R., Wilson, D.W., Barrett, H.H.: Calibration of scintillation cameras and pinhole SPECT imaging systems, pp. 195–202. 12. Springer (2005)
16. Chouzenoux, E., Lamassé, L., Chaux, C., Jaouen, A., Vanzetta, I., Debarbieux, F.: Approche variationnelle pour la déconvolution rapide de données 3d en microscopie biphotonique. In: Actes du 25e colloque GRETSI (2015)
17. Chouzenoux, E., Pesquet, J.C., Repetti, A.: A block coordinate variable metric forward–backward algorithm. Journal of Global Optimization **66**(3), 457–485 (2016)
18. Combettes, P.L., Pesquet, J.C.: Proximal splitting methods in signal processing. In: H.H. Bauschke, R.S. Burachik, P.L. Combettes, V. Elser, D.R. Luke, H. Wolkowicz (eds.) Fixed-Point Algorithms for Inverse Problems in Science and Engineering, pp. 185–212. Springer New York (2011)
19. Corless, R.M., Gonnet, G.H., Hare, D.E.G., Jeffrey, D.J., Knuth, D.E.: On the Lambert W function. Advances in Computational mathematics **5**(1), 329–359 (1996)
20. van den Dries, L., Miller, C.: Geometric categories and o-minimal structures. Duke Math. J. **84**(2), 497–540 (1996)
21. Friesen, W.I., Michaelian, K.H.: Deconvolution and curve-fitting in the analysis of complex spectra: The CH stretching region in infrared spectra of coal. Appl. Spectrosc. **45**(1), 50–56 (1991)
22. Galatsanos, N.P., Katsaggelos, A.K.: Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation. IEEE Trans. Image Process. **1**(3), 322–336 (1992)
23. Guo, H.: A simple algorithm for fitting a Gaussian function [DSP tips and tricks]. IEEE Signal Proc. Mag. **28**(5), 134–137 (2011). DOI 10.1109/MSP.2011.941846
24. Hagen, N., Dereniak, E.L.: Gaussian profile estimation in two dimensions. Appl. Opt. **47**(36), 6842–6851 (2008). DOI 10.1364/AO.47.006842
25. Hagen, N., Kupinski, M., Dereniak, E.L.: Gaussian profile estimation in one dimension. Appl. Opt. **46**(22), 5374–5383 (2007). DOI 10.1364/AO.46.005374
26. Helmchen, F., Denk, W.: Deep tissue two-photon microscopy. Nat. Methods **2**(12) (2005)
27. Henriques, R., Lelek, M., Fornasiero, E.F., Valtorta, F., Zimmer, C., Mhlanga, M.M.: QuickPALM: 3D real-time photoactivation nanoscopy image processing in ImageJ. Nat. Methods **7**(5), 339–340 (2010)
28. Kazovsky, L.G.: Beam position estimation by means of detector arrays. Opt. Quantum Electron. **13**, 201–208 (1981)
29. Kincaid, D., Cheney, E.: Numerical Analysis: Mathematics of Scientific Computing, 3th edn. Pure and applied undergraduate texts. American Mathematical Society (2002)
30. Kirshner, H., Ahuet, F., Sage, D., Unser, M.: 3-D PSF fitting for fluorescence microscopy: implementation and localization application. J. Microsc. **249**(1), 13–25 (2013). DOI 10.1111/j.1365-2818.2012.03675.x
31. Landman, D.A., Roussel-Dupré, R., Tanigawa, G.: On the statistical uncertainties associated with line profile fitting. Astrophys. J. **261**, 732–735 (1982)
32. Marim, M., Zhang, B., Olivo-Marin, J.C., Zimmer, C.: Improving single particle localization with an empirically calibrated Gaussian kernel. In: 5th IEEE Int. Symp. Biomed. Imag.: From Nano to Macro (ISBI 2008), pp. 1003–1006. Paris, France (2008). DOI 10.1109/ISBI.2008.4541168
33. Rockafellar, R., Wets, R.J.B.: Variational Analysis. Springer Verlag (1998). DOI 10.1007/978-3-642-02431-3
34. Roonizi, E.K.: A new algorithm for fitting a Gaussian function riding on the polynomial background. IEEE Signal Process. Lett. **20**(11), 1062–1065 (2013). DOI 10.1109/LSP.2013.2280577
35. Sarder, P., Nehorai, A.: Estimating locations of quantum-dot-encoded microparticles from ultra-high density 3-d microarrays. IEEE Trans. NanoBioscience **7**(4), 284–297 (2008)

36. Tal, E., Oron, D., Silverberg, Y.: Improved depth resolution in video-rate line-scanning multiphoton microscopy using temporal focusing. Opt. Lett. (30), 1686–1688 (2005)
37. Thompson, R.E., Larson, D.R., Webb, W.W.: Precise nanometer localization analysis for individual fluorescent probes. Biophys. J. **82**, 2775–2783 (2002)
38. Tseng, P.: Convergence of a block coordinate descent method for nondifferentiable minimization. Journal of Optimization Theory and Applications **109**(3), 475–494 (2001)
39. Wolter, S., Löschberger, A., Holm, T., Aufmkolk, S., Dabauvalle, M.C., van de Linde, S., Sauer, M.: rapidSTORM: accurate, fast open-source software for localization microscopy. Nat. Methods **9**, 1040–1041 (2012)
40. Xu, Y., Yin, W.: A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. SIAM Journal on Imaging Sciences **6**(3), 1758–1789 (2013). DOI 10.1137/120887795
41. Zhang, B., Zerubia, J., Olivo-Marin, J.C.: Gaussian approximations of fluorescence microscope point-spread function models. Appl. Opt. **46**(10), 1819–1829 (2007). DOI 10.1364/AO.46.001819
42. Zhu, X., Zhang, D.: Efficient parallel Levenberg-Marquardt model fitting towards real-time automated parametric imaging microscopy. PLOS ONE **8**(10), 1–9 (2013). DOI 10.1371/journal.pone.0076665