



HAL
open science

Algorithmic unfolding for image reconstruction and localization problems in fluorescence microscopy

Silvia Bonettini, Luca Calatroni, Danilo Pezzi, Marco Prato

► **To cite this version:**

Silvia Bonettini, Luca Calatroni, Danilo Pezzi, Marco Prato. Algorithmic unfolding for image reconstruction and localization problems in fluorescence microscopy. 2024. hal-04527400

HAL Id: hal-04527400

<https://hal.science/hal-04527400>

Preprint submitted on 30 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Algorithmic unfolding for image reconstruction and localization problems in fluorescence microscopy

Silvia Bonettini¹, Luca Calatroni², Danilo Pezzi^{1*}, Marco Prato¹

^{1*}Dipartimento di Scienze Fisiche, Informatiche e Matematiche,
Università degli Studi di Modena e Reggio Emilia, Via Campi 213/b,
Modena, 41125, Italy.

²CNRS, CNRS, Laboratoire I3S, UMR 7271, UniCa, 2000, Route des
Lucioles, Sophia-Antipolis, 06903, France.

*Corresponding author(s). E-mail(s): danilo.pezzi@unimore.it;
Contributing authors: silvia.bonettini@unimore.it;
calatroni@i3s.unice.fr; marco.prato@unimore.it;

Abstract

We propose an unfolded accelerated projected-gradient descent procedure to estimate model and algorithmic parameters for image super-resolution and molecule localization problems in image microscopy. The variational lower-level constraint enforces sparsity of the solution and encodes different noise statistics (Gaussian, Poisson), while the upper-level cost assesses optimality w.r.t. the task considered. In more detail, a standard ℓ_2 cost is considered for image reconstruction (e.g., deconvolution/super-resolution, semi-blind deconvolution) problems, while a smoothed ℓ_1 is employed to assess localization precision in some exemplary fluorescence microscopy problems exploiting single-molecule activation. Several numerical experiments are reported to validate the proposed approach on synthetic and realistic ISBI data.

Keywords: Algorithmic unfolding, imaging inverse problems, localization, reconstruction, fluorescence microscopy imaging.

1 Introduction

Imaging ill-posed inverse problems are ubiquitous in many applied fields where the quantities observed vary from cells to galaxies [1]. In the field of fluorescence microscopy, for instance, common sources of degradation are light diffraction and noise interference due to photon-counting processes and the electronics of the acquisition device. Due to these degradation effects, observed data typically reveal little amount of information, being limited in spatial resolution and presenting several artefacts caused by possible optical aberrations and/or background (or auto-)fluorescence effects. Mathematically, the image formation model mapping the unknown informative image $u^{\text{true}} \in \mathbb{R}^n$ onto its corresponding noisy and low-resolution version $f \in \mathbb{R}^m$, $m < n$ can be naturally formulated as:

$$f = N(Au^{\text{true}} + b), \quad (1)$$

where $A = SH \in \mathbb{R}^{m \times n}$ is the product of a convolution matrix $H \in \mathbb{R}^{n \times n}$ describing the convolutional action of the Point Spread Function (PSF) of the instrument on u and $S \in \mathbb{R}^{m \times n}$ is a down-sampling operator modelling the loss of resolution, $b \in \mathbb{R}^m$ is the (possibly space-variant) background image containing out-of-focus fluorescent molecules, while $N : \mathbb{R}^m \rightarrow \mathbb{R}^m$ models an interference process introducing noise. In the case of additive white Gaussian noise, for instance, $N(z) = z + n$ with $n \sim \mathcal{N}(0, \sigma^2 \text{Id})$ while in the case of signal-dependent Poisson noise $N(z) = \text{Pois}(z)$, that is the realisation of a multi-variate Poisson process with mean and variance $z \in \mathbb{R}^m$. Those noise distributions are the ones classically considered in a microscopy setting [2, 3], although more relevant mixed [4–7] and/or multiplicative processes can also be considered.

Variational regularization: modelling and optimization.

To counteract the instabilities arising when attempting to solve (1) by direct inversion, the field of Bayesian approaches and variational regularisation has proved effective in a plethora of applications, see, e.g., [8, 9] for surveys. In a Bayesian setting, the reconstruction procedure is formulated as an optimisation process in the form

$$\underset{u}{\operatorname{argmin}} (D(u; f, A, \theta_1) + R(u; \theta_2) =: E_f(u; \theta)), \quad (2)$$

where $D(u; f, A, \theta_1)$ is a data fidelity term whose expression depends on the statistics of the noise, while $R(u; \theta_2)$ is a regularization term which models *a-priori* knowledge on the solutions (such as sparsity or smoothness). The hyperparameters $\theta \in \mathbb{R}^\Theta$ represents all the (hyper)parameters the energy depends on. The first and most natural of them is the regularization parameter, which we indicate with $\rho \in \mathbb{R}$, and balances the tradeoff between the two terms of the energy.

From an optimization perspective, solving the composite problem (2) may be quite challenging. Convex data and regularization terms (such as, e.g., $D(u; f, A, \theta_1) = \frac{1}{2}\|Au - f\|^2$ and $R(u; \theta_2) = \theta_2\|u\|_1$) have been extensively employed over the last years, motivated by the success of compressed sensing in this field [10, 11]. Standard

first-order algorithms solving such composite problem (possibly incorporating further convex constraints) rely on the use of proximal-based algorithms. Among them, iterative proximal-gradient [12], primal-dual [13] and splitting-based approaches [14] have been successfully employed (see also [15] for a review), possibly coupled with acceleration techniques improving the convergence speed [16, 17]. Whenever proximal points can be expressed in closed-form (as it happens in the case of ℓ_1 regularization, possibly combined with orthogonal transformations), proximal-gradient or forward-backward algorithms are often regarded as the simplest approaches for solving structured, convex composite problems in the form (2). For standard ℓ_2 - ℓ_1 problems the use of such algorithm dates back to [18] under the name of Iterative Soft-Thresholding Algorithm (ISTA) which was used as an effective regularization procedures for linear ill-posed inverse problems with sparsity constraints. Given an initialization point $u^0 \in \mathbb{R}^n$ and a step-size parameter $\alpha \leq 1/\|A\|^2$ guaranteeing convergence, the ISTA iteration defined for $k \geq 0$ reads:

$$u^{(k+1)} = \mathcal{T}_{\alpha\theta_2} \left(u^{(k)} - \alpha A^T (A u^{(k)} - f) \right) = \mathcal{T}_{\alpha\theta_2} \left((\text{Id} - \alpha A^T A) u^{(k)} + \alpha A^T f \right), \quad (3)$$

where $\mathcal{T}_{\alpha\theta_2} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the soft-thresholding operator defined component-wise by $\mathcal{T}_{\alpha\theta_2}(z) = \max(|z| - \alpha\theta_2, 0) \text{sign}(z)$.

Parameter estimation.

A crucial challenge in the design of variational models in the form (2) is the choice of the fidelity/regularization parameters $\theta = (\theta_1, \theta_2)$ which, potentially, may lie in a very large dimensional space \mathbb{R}^Θ , $|\Theta| \gg 1$. They make the regularization functional $R(\cdot; \theta_2)$ expressive enough to encode the desired solution properties and balance its action against the data fit which, depending on $\theta_1 \in \mathbb{R}^{\Theta_1}$ may encode possible local noise dependence. Classical approaches addressing hyper-parameter estimation problems are based on prior knowledge of the noise level [19, 20], (heuristic) study of regularization paths/Pareto fronts and cross-validation, see, e.g. [21]. Over the last decade, however, the increasing availability and access to data, favoured the development of machine- and deep-learning based approaches for parameter estimation. The general idea of such methods consist in optimizing model parameters (even when the regularizer is parametrized as a neural-network) by minimizing some task-dependent quality metric assessing proximity (in some sense) to reference image data, see [22–24] for a review. Having in mind a close connection with the optimization framework of optimization-based iterative schemes such as (3), among the plethora of data-driven approaches for imaging, we focus here on the two particular classes of bilevel optimization and algorithmic unrolling techniques, see [25] and [24] for review papers, respectively.

Bilevel optimization and algorithmic unrolling.

As it will be described in more detail in Section 2, bilevel learning approaches compute optimal hyper-parameters θ in (2) by comparing the solution $u^*(\theta)$ of the problem with ground truth data w.r.t. some assessment metric. They thus naturally read as a nested optimization problem where the variational model serves to constrain the space of possible solutions. Bilevel approaches have been widely used in the context

of variational image reconstruction [25–30] and machine learning [31–34] and offer an interesting and sound perspective for learning hyperparameters in regularized inverse problem formulations. *Algorithmic unrolling* is a somehow similar approach where an explicit writing of the first $K \in \mathbb{N}$ iterations of an iterative algorithm as a function of the initial point u^0 and the model parameters is interpreted as the output $u^{(k)}(\theta)$ of a K -layer neural network (with suitable activation functions) to be matched with ground-truth data. Taking the ISTA iteration (3) as a toy example, in [35] an unrolled version of ISTA (therein called Learned ISTA, LISTA) was considered. Using the notation above, the main idea there consists in learning parameters $\{\alpha, \theta_2, W_e, W_t\}$ such that for $k = 0, \dots, K$

$$u^{(k+1)} = \mathcal{T}_{\alpha\theta_2} \left(W_e u^{(k)} + W_t f \right), \quad (4)$$

thus resembling (3) but being potentially more expressive being *a-priori* not connected to the minimization of any underlying variational problem.

Algorithmic unrolling is nowadays a popular strategy to solve inverse problems, see, e.g., [24, 36–38] for some relevant references and [39, 40] for connections with bilevel optimization and Deep Equilibrium Models. Depending on the particular problem and algorithm at hand, choosing to learn *all* problem parameters at once in an end-to-end fashion may be not ideal especially in applications involving known physical models (typically, encoded in the choice of the operator A or in the modelling of the data-fit D) which one would like to keep fixed.

A case study: reconstruction and localization in fluorescence microscopy.

In fluorescence microscopy imaging, the physical limitations imposed by light diffraction makes the accurate reconstruction and analysis of small biological samples very challenging. For standard microscopes structures closer to ~ 250 nm in the x - y plane cannot be distinguished. Super-resolution fluorescence microscopy techniques [41–43] aim to overcome such barrier, allowing resolution (i.e., minimal distance between two distinct objects) up to 20 nm [44, 45]. In this context, photons emitted by photo-activable fluorescent molecules are captured by detectors after passing through special lenses and optical devices, which limit spatial resolution due to light diffraction. Such process can be modelled by (1) as a convolution of the emitters with the microscope PSF, thus producing blurred and low-resolution measurements featuring also noise distortions due to electronic interference and a (possibly space-variant) background term accounting for out-of-focus molecules.

Among the many super-resolution techniques proposed in the microscopy community to address this challenge, some have attracted the attention of the applied mathematics community working on signal and image processing. The former has been popularized under the name of Single Molecule Localization Microscopy (SMLM) approaches, see, e.g., [46] for a review. Here, the idea consists in acquiring a temporal sequence of images $\{f_t\}_{t=1}^T$ where at each frame t only a small percentage of special photo-activable fluorescent molecules is active, thus making the detection of nonzero elements in the corresponding image u_t (that is, the localization process) easier. A super-resolved image u_{SR} is thus obtained by averaging $u_{\text{SR}} = \frac{1}{T} \sum_{t=1}^T u_t$. A

mathematical modelling of this problem relies, essentially, on the use of sparsity-based regularization techniques for accurate localization, either in a convex [47] or nonconvex [48, 49] regime. Parameter selection plays here a crucial role and their brute-force optimization could be a tedious task. Other approaches to SMLM make use of end-to-end deep learning procedures for accurate localization upon suitable training [50].

Some other techniques require less specific fluorescent molecules being based on the estimation of a super-resolved image by exploiting only a temporal sequence of images whose intensity fluctuations are analyzed typically in terms of their second-order statistics [51–55]. Those techniques are better suited to standard biological set ups where the repeated ON-OFF processes typical of SMLM may be harmful for the sample under observation. In these settings, a modelling similar to (1) but reformulated in a covariance domain is considered (see Section 3.2 for more details).

Contribution.

In this work, we unroll an accelerated projected gradient descent (APGD) scheme to estimate optimal parameters for a (smoothed) ℓ_2 - ℓ_1 regularization problem endowed with a non-negativity constraint. In the framework of fluorescence microscopy imaging, this simple regularization model can be effectively used for both reconstruction and molecule localization purposes. Given the intrinsic difference between these two tasks, we consider in the following task-adapted evaluation metrics estimating optimal algorithmic parameters depending on the specific objective considered. We report several numerical experiments confirming the validity of the approach on exemplary deconvolution/super-resolution problems classically encountered in fluorescence microscopy, also in a semi-blind scenario in the context of fluctuation-based microscopy for estimating the shape of the blurring function encoded by the operator A .

2 Models and algorithms

We review the mathematical formulation of the bilevel optimization schemes considered and consider their unrolled versions, making precise the models considered to deal with the reconstruction/localization problems encountered in the microscopy applications considered.

2.1 From bilevel optimization to algorithmic unrolling

Bilevel optimization approaches [25–34] rely on a supervised data-driven approach to estimate model and algorithmic hyperparameters. Given a training dataset $\mathcal{D} = \{(f_t, g_t) : t = 1, \dots, T\}$, where f_t is a corrupted (noisy, blurred, low-resolution) version of the ground truth image g_t , the problem of estimating optimal parameters $\hat{\theta}$ can be formulated

$$\begin{cases} \hat{\theta} \in \operatorname{argmin}_{\theta \in \mathbb{R}^p} \sum_{t=1}^T \mathcal{L}(u_t^*(\theta), g_t) \\ \text{s.t. } u_t^*(\theta) \in \operatorname{argmin}_u E_{f_t}(u; \theta) \quad t = 1, \dots, T. \end{cases} \quad (5)$$

Essentially, it consists in two nested optimization problems, where the inner (or lower) one feeds its result to the outer (or upper) one which is used to assess optimality of the estimation w.r.t. task-dependent evaluation loss. The ultimate goal of the scheme is to find an optimal set of parameters $\hat{\theta}$ which optimizes the performance of the underlying variational model.

Solving (5) requires the development of (gradient-based) optimization methods searching for the optimal parameters configuration minimizing the outer loss function. For that, the computation of the minimizers $u_t^*(\theta)$ for $t = 1, \dots, T$ of lower-level functional is required, which is often unfeasible. Most of the time, it is thus necessary to use iterative schemes solving the variational lower-level constraint under a fixed computational budget (that is fixed amount of iterations), thus leading to approximate minimizers and, ultimately, inexact computations of the gradient of the outer loss. This idea stands at the very basis of algorithmic unfolding [24, 36]. Mathematically this can be modelled by replacing (5) with the unrolling over $K \geq 1$ iterations of an optimization algorithm \mathcal{A} , such that:

$$u_t^*(\theta) \approx \mathcal{A}^K(f_t; \theta) \quad (6)$$

so that optimization is performed for a fixed number of iterations K . By expanding the expression of the computed quantity from $k = K$ till $k = 0$ it is thus possible to have a compact expression of $u_t^K(\theta)$ which can be used for differentiating over the desired hyperparameters θ , see, e.g., [28]. Note that, depending on the particular choice of \mathcal{A} , in addition to model parameters (such as regularization parameters and/or quantities related to the forward model), algorithmic parameters (such as algorithmic step-sizes) can also be learned. Once the training phase has been completed and an optimal configuration of parameters $\hat{\theta}$ has been found, one can simply apply the trained algorithm $\mathcal{A}^K(\cdot; \hat{\theta})$ on new unseen data.

2.2 A model for sparse reconstruction/localization

We describe in this section the variational model employed to solve localization and reconstruction problems often encountered in the framework of microscopy as in the case, for instance, of Single Molecule Localization methods [47–49]. In order to consider an unrolled bilevel strategy (5)–(6), some details on the the algorithm \mathcal{A} considered and on the training procedure employed (with a particular focus on the choice of the loss function \mathcal{L} are also given.

Localization and reconstruction by sparse optimization.

We consider the following ℓ_1 sparse reconstruction approach with a non-negativity constraint:

$$E_f(u; \theta) = D(Au; f, bg) + \rho \|u\|_1 + \iota_{\mathbb{R}_+^n}(u), \quad \iota_{\mathbb{R}_+^n}(u) := \begin{cases} 0 & u \in \mathbb{R}_+^n \\ +\infty & \text{otherwise} \end{cases}, \quad (7)$$

where $\mathbb{R}_+^n = \{u \in \mathbb{R}^n : u_i \geq 0, i = 1, \dots, n\}$ and $bg \in \mathbb{R}$ represents a constant positive background term which can be used both for modelling, see, e.g., [56]. Note that the

non-negativity constraint makes the ℓ_1 -norm differentiable. In this work we make use of the ℓ_1 -norm as a convex approximation of the ℓ_0 pseudo-norm in a compressed sensing fashion [10, 11].

The fidelity term D enforces proximity between the physical modelling and the observed data depending on the particular modelling of the noise statistics. In particular we consider the two choices:

- (i) (Additive white Gaussian noise) $D(Au; f) = \frac{1}{2} \|Au - f\|_2^2$
- (ii) (Poisson noise) $D(Au; f, bg) = KL(Au + bg; f) = \sum_{i=1}^m (Au)_i + bg - f_i - f_i \log \left(\frac{(Au)_i + bg}{f_i} \right)$.

Fidelity (i) is the quadratic penalty classically employed to model the presence of additive white Gaussian noise [2], while fidelity (ii) is the Kullback-Leibler (KL) used to model signal-dependent Poisson noise [56, 57]. Note that both functionals are convex and Lipschitz differentiable on the non-negative orthant. However, the KL divergence might present some numerical instability issues due to the presence of the logarithm for very small values of original (f_i) and/or reconstructed $(Au)_i + bg$ components of the data. We consider in the following an algorithmic unfolding procedure estimating optimal hyperparameters for (7) coupled for the two data terms above to assess and compare the performance of both models on localization and reconstruction problems.

Solving the lower-level problem via accelerated projected gradient descent.

By solving (5) via algorithmic unrolling (6), the computation of the loss gradient can simply be done via backpropagation, provided that the single iteration of \mathcal{A} is differentiable. Given the composite structure of problem (7), we consider as $\mathcal{A}^K(f; \theta)$ K iterations of an accelerated projected gradient descent endowed with a smoothed differentiable projection initialized by taking f as starting point:

Algorithm 1: Accelerated Projected Gradient Descent

INPUT: $f \in \mathbb{R}^m$.
 $u^{(0)} = u^{(-1)} = f$
 FOR $k = 0, \dots, K - 1$
 $\left[\begin{array}{l} \bar{v}^{(k)} = u^{(k)} + \beta_k(u^{(k)} - u^{(k-1)}) \\ v^{(k)} = \Pi(\bar{v}^{(k)}) \\ w^{(k)} = v^{(k)} - \alpha_k \nabla_u E_f(v^{(k)}; \theta) \\ u^{(k+1)} = \Pi(w^{(k)}) \end{array} \right.$
 $u^* = u^{(K)}$
 OUTPUT: u^* .

The map $\Pi : \mathbb{R}^n \rightarrow \mathbb{R}_+^n$ has been used in [30]: it is the smoothed version of the standard (non-differentiable) Euclidean projection. In particular, for an arbitrarily

small $\varepsilon > 0$, Π acts component-wise on $u \in \mathbb{R}^n$ as:

$$(\Pi(u))_i = \begin{cases} \left(2 - \frac{u_i}{\varepsilon}\right) \frac{u_i^2}{\varepsilon}, & u_i \in (0, \varepsilon) \\ \max(u_i, 0) & \text{otherwise} \end{cases}.$$

By definition Π acts as the Euclidean projection outside the interval $(0, \varepsilon)$, while it coincides with a third degree polynomial interpolating the points $(0, 0)$ and $(\varepsilon, \varepsilon)$.

Algorithm 1 is an instance of Accelerated Projected Gradient descent [16, 17] with a smoothed projection. The sequence $\{\beta_k\}_{k=0}^{K-1}$ of inertial parameters can be chosen as

$$\beta_k = \begin{cases} 0 & \text{if } k = 0 \\ \frac{k-1}{k+2} & \text{if } k \geq 1 \end{cases},$$

which was shown in [58] to guarantee weak convergence of the iterates. The sequence of step-size parameters $\{\alpha_k\}_{k=0}^{K-1}$ can be set either to a constant equal to the inverse of the Lipschitz constant of $\nabla_u E_f$, or adaptively chosen using, e.g., monotone Armijo-type rules [17]. In the following we let this sequence to be learned.

Since the general k -th iteration of \mathcal{A} is differentiable, it is possible to compute the gradient of a differentiable loss function by simple backpropagation procedure. Namely, for one training sample (f, g) and one parameter θ_j we have:

$$\frac{\partial \mathcal{L}(u^{(K)}(\theta), g)}{\partial \theta_j} = \left(\frac{\partial u^{(K)}(\theta)}{\partial \theta_j} \right)^T \frac{\partial \mathcal{L}(u^{(K)}(\theta), g)}{\partial u}. \quad (8)$$

The chain rule can then be applied through $\frac{\partial u^{(K)}(\theta)}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \mathcal{A}(u^{(K-1)}; \theta)$. With a small abuse of notation and denoting by $\Pi'(v)$ the diagonal Jacobian matrix of Π evaluated at v , we have

$$\frac{\partial u^{(K)}(\theta)}{\partial \theta_j} = \Pi'(w^{(K-1)}) \frac{\partial}{\partial \theta_j} \left(v^{(K-1)} - \alpha_{K-1} \nabla_u E_f(v^{(K-1)}; \theta) \right) \quad (9)$$

$$= \Pi'(w^{(K-1)}) \left(I_n - \alpha_{K-1} \nabla_{uu}^2 E_f(v^{(K-1)}; \theta) \right) \frac{\partial v^{(K-1)}}{\partial \theta_j} \quad (10)$$

$$- \alpha_{K-1} \Pi'(w^{(K-1)}) \frac{\partial}{\partial \theta_j} \nabla_u E_f(v^{(K-1)}; \theta), \quad (11)$$

where I_n is the identity matrix of size $n \times n$ and

$$\frac{\partial v^{(K-1)}}{\partial \theta_j} = \Pi'(\bar{v}^{(K-1)}) \left((1 + \beta_{K-1}) \frac{\partial u^{(K-1)}(\theta)}{\partial \theta_j} - \beta_{K-1} \frac{\partial u^{(K-2)}(\theta)}{\partial \theta_j} \right), \quad (12)$$

so that the computations can be repeated for $\frac{\partial u^{(K-1)}(\theta)}{\partial \theta_j}$ and $\frac{\partial u^{(K-2)}(\theta)}{\partial \theta_j}$ until $u^{(0)} = f$.

In the computations above, the quantities $\nabla_{uu}^2 E_f(v^{(k)}; \theta)$ and $\nabla_{\theta u}^2 E_f(v^{(k)}; \theta)$ are present, which requires the conditions:

- E is twice continuously differentiable, with $\nabla_{uu}^2 E$ invertible;
- $\nabla_u E$ is differentiable with respect to θ .

Both properties are satisfied by the energy function (7). Note that such conditions are quite restrictive for more effective sparsity-promoting regularization terms approximating the ℓ_0 pseudo-norm, such the continuous exact relaxations CEL0 studied in [59, 60]. Indeed, just like the ℓ_1 -norm, such penalty is differentiable with respect to u when the non-negativity constraint is taken into account, but its gradient is not differentiable with respect to the regularization parameter that governs it (see Appendix A).

The training loss functions.

In order to adapt the hyperparameter optimization to the particular task at hand (reconstruction/localization), we consider in the following experiments two different assessment metrics:

- (Reconstruction loss) a standard ℓ_2 -norm (corresponding to SNR optimization, as done, e.g., in [26, 27]):

$$\mathcal{L}_2(u^*(\theta); g_t) = \frac{1}{2} \|u^*(\theta) - g_t\|_2^2, \quad (13)$$

- (Localization loss) a smoothed ℓ_1 -norm with image binarization assessing localization precision:

$$\mathcal{L}_1(u^*(\theta); g_t) = \sum_{i=1}^n \psi_\gamma \left((B_{\delta,c,\epsilon}(u^*(\theta)) - \tilde{g}_t)_i^2 \right). \quad (14)$$

Note that in (13) g_t is a ground-truth image from the training dataset while in (14) \tilde{g}_t is a binarized version of g_t where each nonzero element is set as 1 or 255, depending on the scaling. As observed in [61], in the field of microscopy (and in particular of single-molecule localization approaches) is important to define tailored assessment metrics adapted to the task at hand which may aim at either estimating the precise intensity of the light they emit (and for such case (13) is a natural choice) or detect whether model (7) has computed a reconstruction in the correct position, for which a loss in the form (14) seems more natural. To allow the computation of derivatives, we employ in definition (14) a function ψ_γ which is the Huber function defined, for $\gamma > 0$, by:

$$\psi_\gamma(s) := \begin{cases} \frac{s}{\gamma} & s \leq \gamma^2 \\ 2\sqrt{s} - \gamma & s > \gamma^2. \end{cases} \quad (15)$$

The function $B_{\delta,c,\epsilon} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ acts pixel-wise by binarizing the tentative reconstruction $u_t^*(\theta)$ for the current estimation of θ in order to compare it with \tilde{g}_t . For a

given image $s \in \mathbb{R}^n$, by setting $\bar{p} = \max_i s_i$, it is defined as

$$B_{\delta,c,\epsilon}(s) := \begin{cases} 0 & s \leq \delta, \\ \left(2 - \frac{s-\delta}{\epsilon}\right) \frac{(s-\delta)^2}{\epsilon} \frac{\bar{p}}{2(c-\delta)} & \delta < s < \delta + \epsilon, \\ \frac{\bar{p}}{2(c-\delta)}(s - \delta) & \delta + \epsilon \leq s \leq 2c - \delta - \epsilon, \\ \bar{p} - \left(2 - \frac{2(c-\delta)-s+\delta}{\epsilon}\right) \frac{(2(c-\delta)-s+\delta)^2}{\epsilon} \frac{\bar{p}}{2(c-\delta)} & 2c - \delta - \epsilon < s < 2c - \delta, \\ \bar{p} & s \geq 2c - \delta. \end{cases} \quad (16)$$

We observe that such definition depends on three parameters: $\delta > 0$ is the thresholding parameter used to binarize into zeros and the value \bar{p} , so that the point $s = c$ is the one getting transformed into $\frac{\bar{p}}{2}$. The parameter ϵ controls the smoothing intervals, which are symmetric w.r.t. c , in the same way as in the projection Π previously described. Figure 1 shows various shapes of $B_{\delta,c,\epsilon}$ for three different values of δ . Ideally, we would like to consider situations where $c \approx \delta$ to have a sharp transition between the two different regimes. Note that while this choice may look at a first sight a complicated way of assessing localization precision (as it introduces further parameters to estimate), in practice ϵ does not need any fine-tuning provided it is sufficiently small. To obtain a steep slope one could further choose $c = \delta + c_0$, with $c_0 > \epsilon$, still small. Under this choice, the binarization actually becomes differentiable with respect to δ , which allows for a strategy to learn δ as well. There holds indeed:

$$\frac{\partial \mathcal{L}_1}{\partial \delta} = \sum_{i=1}^n \psi'_\gamma((B_{\delta,c,\epsilon}(u^*(\theta)) - \tilde{g}_t)_i) \frac{\partial B_{\delta,c,\epsilon}(u^*(\theta))_i}{\partial \delta}. \quad (17)$$

where

$$\frac{\partial B_{\delta,c,\epsilon}(s)}{\partial \delta} = \begin{cases} 0 & \delta \geq s \\ \frac{\bar{p}}{2c_0} \left(3 \frac{(s-\delta)^2}{\epsilon^2} - 4 \frac{(s-\delta)}{\epsilon}\right) & s > \delta > s - \epsilon \\ -\frac{\bar{p}}{2c_0} & s - \epsilon \geq \delta \geq s + \epsilon - 2c_0 \\ -\frac{\bar{p}}{2c_0} \left(4 \frac{2c_0 - s + \delta}{\epsilon} - 3 \frac{(2c_0 - s + \delta)^2}{\epsilon^2}\right) & s + \epsilon - 2c_0 > \delta > s - c_0 \\ 0 & \delta \leq s - 2c_0 \end{cases}, \quad (18)$$

which can be used for defining gradient-type updates.

3 Numerical experiments

In this Section we illustrate the results obtained by unrolling Algorithm 1 applied to model (7) for exemplar reconstruction/localization problems using different modelling for the noise statistics and loss functions (13)–(14) for training the model.

Unless specified otherwise, we evaluate the quality of the images by averaging, over all the T frame frames, the Peak Signal to Noise Ratio (PSNR) and the Jaccard index

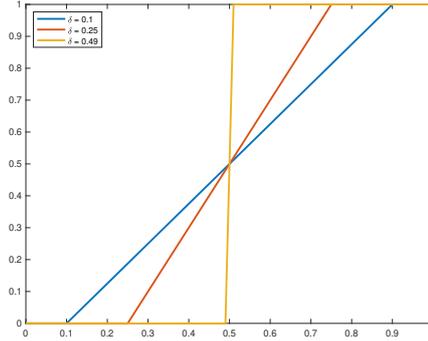


Fig. 1: Binarization in $[0,1]$ with $c = 0.5$, $\epsilon = 10^{-4}$ for three different values of δ : 0.1 (blue), 0.25 (red), 0.49 (yellow).

defined by:

$$J_{\tilde{\delta}} = \frac{\# \text{ TP}}{(\# \text{ TP}) + (\# \text{ FN}) + (\# \text{ FP})} \in [0, 1] \quad (19)$$

where TP, FN and FP denote True Positives, False Negatives and False Positives, respectively. The parameter $\tilde{\delta} > 0$ is a tolerance parameter acting as follows: a reconstructed pixel is counted as TP if it lies within a ball of radius $\tilde{\delta}$ from a true molecule. Figure 2 contains a visual representation of how the size of the tolerance value $\tilde{\delta}$ influences the value of the Jaccard index. Given the a ground truth pixel $a \neq 0$, let $\mathcal{B}_{\tilde{\delta}}(a)$ the ball of radius $\tilde{\delta}$ around a . A point reconstructed within $\mathcal{B}_{\tilde{\delta}}(a)$ is then counted as TP, outside it is counted as a FP. A TN pixel corresponds to a GT point does not corresponding to any reconstruction within $\mathcal{B}_{\tilde{\delta}}(a)$. In Figure 2a ($\tilde{\delta} = 0$) only a is counted as a TP, whereas for $\tilde{\delta} = 2, 4$ (Figures 2b–2c) all points lying within the orange region are counted as TP. Note, however, that if two elements are computed within $\mathcal{B}_{\tilde{\delta}}(a)$, only the closest to the GT point will be counted as a TP, with the other being a FP, see Figure 2d.

In all experiments, the optimization of the upper level problem *w.r.t.* θ is carried out via the Scaled Gradient Projection method (SGP) [62], so that we can impose reasonable bounds on the parameters. Some of such bounds are natural (such as the non-negativity of step-size parameters), but for some others some comments are needed. In particular, as far as the choice of the regularization parameters is concerned, in the case of a least squares term, we followed [63] and chose an upper bound on ρ given by

$$\rho_{\max} = \min_t \|A^T f_t\|_{\infty}. \quad (20)$$

It is easy to check that for all $\rho > \rho_{\max}$ the trivial zero solution is obtained. For the Poisson data term, such upper bound was chosen empirically. The upper bound on δ was set at half the maximum pixel value to avoid null reconstructions. The lower bounds were all set to 10^{-10} .

Due to the nonconvex nature of the bilevel optimization problem (5), the initial set of parameters $\theta^{(0)}$ may have a major impact on the final results. For all set of

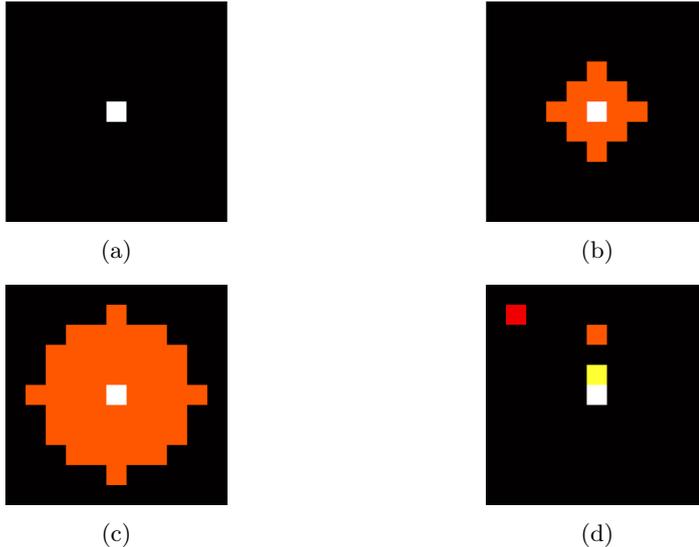


Fig. 2: (a) Radius of correct detection for $\tilde{\delta} = 0$. (b) Radius of correct detection for $\tilde{\delta} = 2$. (c) Radius of correct detection for $\tilde{\delta} = 4$. (d) Example in the case of $\tilde{\delta} = 0$: the white pixel is the true molecule, the yellow pixel is counted as a TP, the red pixel is FP as it is outside of the range, while the orange pixel is still counted as a FP even if it is in range because the yellow one is closer to the true molecule.

experiments the initial configurations are reported in the tables with the results. Note that the steplengths $\{\alpha_k\}_{K=0}^{K-1}$ are all initialized to one same value α .

3.1 Single molecule localization microscopy.

3.1.1 Simulated data

We now consider a Single Molecule Localization Microscopy framework where taking as a reference model (7), the unrolling strategy is used to learn optimal parameters $\theta = (\rho, \alpha_0, \dots, \alpha_{K-1})$ when considering the loss in (13) and $\theta = (\rho, \delta, \alpha_0, \dots, \alpha_{K-1})$ for the loss in (14).

We first test our strategy on simulated data in order to check whether the Localization loss (14) promotes better localization properties than (13). We consider a data set of 10 ground truth images g_t containing each between 75 and 150 randomly activated pixels with values in the range $U = [100, 255]$. The size of these (vectorized) images is $n \times n = mL \times mL$, where $L = 4$ is the super-resolution factor and $m = 64$ is the size of the downsampled image. The underlying linear model is described by an operator $A = SH$ being the composition of a convolution operator $H \in \mathbb{R}^{n^2 \times n^2}$ describing the convolution of a Gaussian PSF with standard deviation 2.5, and $S \in \mathbb{R}^{m^2 \times n^2}$, a down-sampling operator summing up pixel intensities for each 4×4 patch of the fine grid. Denoting by $\mathbf{1}_L$ and $\mathbf{0}_L$ the $L \times 1$ arrays with constant entries equal to 1 and 0,

respectively, and considering the matrix

$$S_L = \begin{pmatrix} \mathbf{1}_L & \mathbf{0}_L & \cdots & \mathbf{0}_L \\ \mathbf{0}_L & \mathbf{1}_L & \cdots & \mathbf{0}_L \\ \vdots & & & \\ \mathbf{0}_L & \mathbf{0}_L & \cdots & \mathbf{1}_L \end{pmatrix} \in \mathbb{R}^{m \times n},$$

then the action of the operator S on a vectorized image x can be written as

$$Sx = \text{vec}(S_L \bar{x} S_L^T),$$

where $\text{vec}(\bar{x}) = x$. The resulting blurred and downsampled images are then corrupted with additive white Gaussian noise with standard deviation $\sigma = 0.15$. As for the unrolling, $K = 190$ iterates were used and the value $c = \delta + 0.01$ is used in (16).

Table 1 shows the values of the Jaccard index (19) obtained on the 25 test images for different tolerance values $\tilde{\delta} \in \{0, 2, 4\}$. At a first glance we notice that training the model with (14) corresponds indeed to an improvement of the Jaccard values w.r.t. training with (13). We remark that, in order to obtain meaningful results with (13), a tailored dataset had to be considered. Our first tests showed that using, during the training, images with the same density of molecules as those in the test set, the quadratic loss would struggle to find a value of ρ big enough so that the images would not be sufficiently sparse. To avoid that, we created a special dataset of very sparse images containing only 4 non-zero pixels with random values in U and used it to train the model with the \mathcal{L}_2 loss. This was done to assess whether reducing the density of the activated pixels would lead to an improvement in terms of Jaccard index. Some exemplary results are reported in Figure 3.

Table 1: Results for Experiment 1. R = Number of examples used in the training; RS = Number of special examples used in the training. The test set is made by 25 images. For each quality measure, the best score across the models is in bold. The last column reports the initial parameter configuration

Loss function	R / RS	J_0	J_2	J_4	Avg. J	PSNR	Initial ρ, α, δ
\mathcal{L}_2	0 / 8	0.3878	0.4179	0.4179	0.4079	35.76	0.1, 0.1, n.a.
\mathcal{L}_1	10 / 0	0.7452	0.8480	0.8583	0.8172	33.09	0.1, 0.1, 50

3.1.2 ISBI super-resolution datasets.

We now test the unrolled localization/reconstruction procedures on the ISBI 2013 SMLM challenge data¹. The dataset features 361 ground truth images. Each one has approximately 200 pixels activated, with values in $\{1, 2, 3\}$. For our experiments we

¹<https://srm.epfl.ch/srm/index.php>

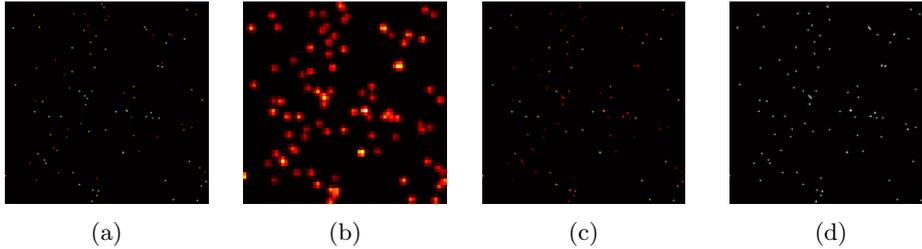


Fig. 3: (a) Ground Truth g_t . (b) Acquired f_t . (c) Optimal reconstruction $u(\hat{\theta})$ obtained by optimizing \mathcal{L}_2 (13). (d) Optimal reconstruction $B_{\delta, c, \epsilon}(u^{(K)}(\hat{\theta}))$ obtained by optimizing \mathcal{L}_1 (14).

rescaled them so that the maximum value was 255. The PSF was obtained using information provided by the challenge organizers: the observed data have size $m \times m = 64 \times 64$ pixels of side $100nm$. The PSF considered is a Gaussian convolution kernel with Full Width at Half Maximum (FWHM) of $258.21nm$. We consider $L = 4$ as a super-resolution factor.

We created a training dataset $\{g_t, f_t\}$ simulating different levels and types of noise:

- Gaussian noise with standard deviation 0.15.
- Poisson noise with constant low background emission $b = 0.1$.
- Poisson noise with constant high background emission $b = 12.75$.

Given the signal-dependence of Poisson noise on the signal, the higher is the background b and the higher is the noise. The two scenarios considered thus model low- and high-photon counts and serve us to assess the performance of the approach over different noise statistics. Reconstruction and localization performances were thus tested in this experiment for both upper-level loss functions (13)–(14) and for the Gaussian/Poisson data fidelity functions.

In the following experiments, we set $c = \delta + 0.01$ and used only 20 images of the stack for training. The number of inner iterations was set to $K = 300$. All the results are reported in Tables 2, 3 and 4. Both the widefield-type images of the ground truth as well as their the acquired data/reconstructed ones for each dataset are reported in Figure 4a/4, respectively. As a comparison, we report in Figure 4 (third column) the reconstructions computed by using DeepSTORM [50], an end-to-end deep-learning based procedure properly trained using the available Google COLAB notebook² under a physical parameter setting compatible with the datasets observed. Note that for the DeepSTORM reconstructions the PSNR was computed on the final super-resolved image and not frame-wise as the software outputs only the coordinates of the localizations (used to compute the Jaccard index values) and one final intensity image. We note in this experiment that no apparent optimal choice between the Gaussian/Poisson data term can be drawn as similar PSNR and Jaccard scores are obtained on the corresponding reconstructions. We argue that this can be due to the biases introduced

²https://colab.research.google.com/github/HenriquesLab/ZeroCostDL4Mic/blob/master/Colab_notebooks/Deep-STORM_2D_ZeroCostDL4Mic.ipynb

by the regularization (which make less important the precise modelling of the noise statistics) and by the overall non-convexity of the learning problem. As far as the DeepSTORM reconstructions are concerned, we observe a general tendency to overestimate the reconstruction support with severe artefacts appearing in the background in the case of high background.

Table 2: Results in the case of additive white Gaussian noise dataset. For each quality measure, the best score across the models is in bold. The last column reports the initial parameter configuration.

Model	J_0	J_2	J_4	Avg. J	PSNR	Initial ρ, α, δ
\mathcal{L}_2	0.1147	0.2175	0.2194	0.1839	34.27	4, 0.1, -
\mathcal{L}_1	0.1309	0.4889	0.5921	0.4040	22.45	4, 0.1, 5
DeepSTORM	0.0453	0.4563	0.5688	0.3568	20.10	-

Table 3: Results in the case of low Poisson noise dataset. For each column, the best score across the models is in bold. The last column reports the initial parameter configuration.

Model	Fidelity	J_0	J_2	J_4	Avg. J	PSNR	Initial ρ, α, δ
\mathcal{L}_2	KL	0.0921	0.2633	0.3159	0.2238	33.94	50, $5 \cdot 10^{-4}$, -
\mathcal{L}_2	$\ \cdot\ _2$	0.1057	0.1205	0.1211	0.1158	34.74	4, 0.075, -
\mathcal{L}_1	KL	0.0907	0.4250	0.5924	0.3694	22.66	50, $5 \cdot 10^{-4}$, 0.05
\mathcal{L}_1	$\ \cdot\ _2$	0.1287	0.4494	0.5358	0.3713	22.52	4, 0.1, 5
DeepSTORM	-	0.0381	0.4568	0.5890	0.3613	19.52	-

Table 4: Results in the case of high Poisson noise dataset. For each column, the best score across the models is in bold. The last column reports the initial parameter configuration.

Loss function	Fidelity	J_0	J_2	J_4	Avg. J	PSNR	Initial ρ, α, δ
\mathcal{L}_2	KL	0.0589	0.0909	0.0987	0.0828	34.35	1, 1, n.a.
\mathcal{L}_2	$\ \cdot\ _2$	0.0544	0.0778	0.0829	0.0717	34.36	20, 0.05, -
\mathcal{L}_1	KL	0.0982	0.3684	0.4771	0.3146	25.31	0.5, 0.5, 1
\mathcal{L}_1	$\ \cdot\ _2$	0.0950	0.3902	0.5012	0.3288	25.74	20, 0.01, 3
DeepSTORM	-	0.0106	0.1949	0.3185	0.1747	17.56	-

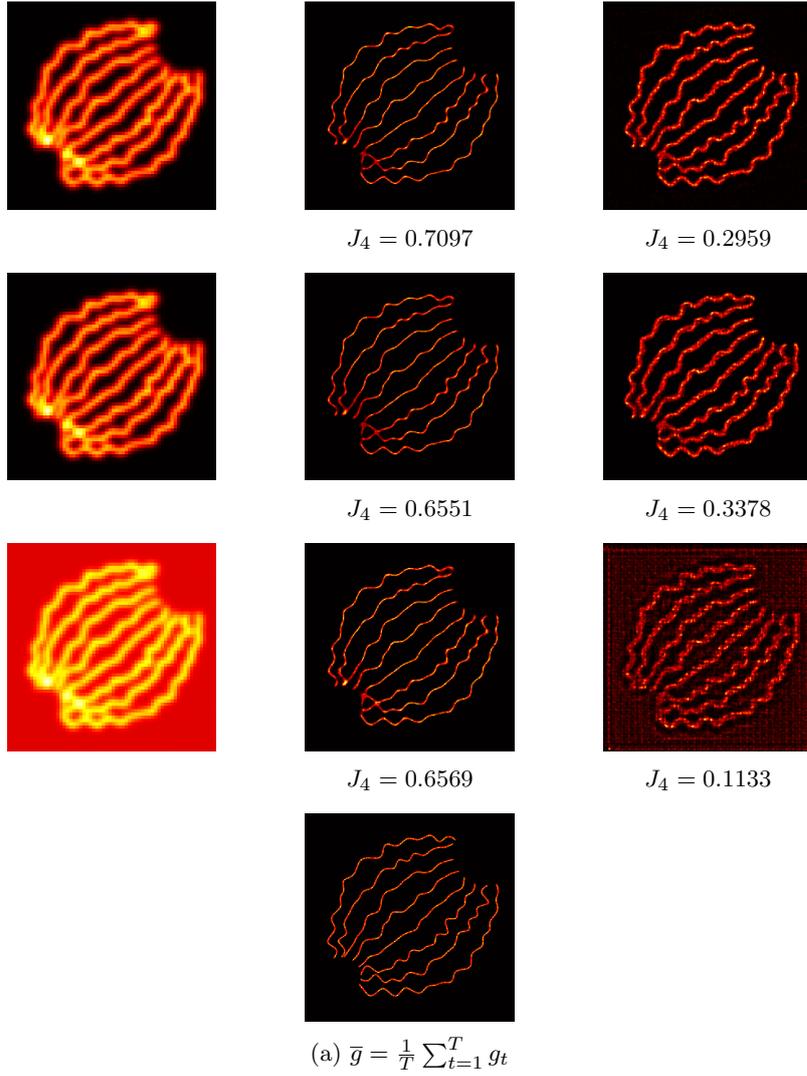


Fig. 4: Columns from left to right: low-resolution (widefield) image, reconstruction computed by our approach, DeepSTORM [50] reconstruction. Datasets, from top to bottom: Gaussian, Poisson noise with low background and Poisson noise with high background. (a) represents the ground truth stack.

We now consider a different ISBI dataset ³, where ground truth images are obtained by discretizing 2D positions of 8731 fluorescent molecules among which approximately 20 molecules are randomly activated at each frame for 436 ground truth training images. To simulate a long emission (ON) time τ_{em} , we considered in this experiment

³<https://srm.epfl.ch/DatasetPage?name=MT1.N1.LD>

a value $\tau_{\text{em}} > \tau_{\text{acq}}$, corresponding to the temporal resolution of the acquisitions. When active, a fluorescent molecule is assigned a value in $u = \{0, 1, 2, 3, 4\}$ with probability density $p = [0.1, 0.25, 0.3, 0.25, 0.1]$ so that for $i \in u$, $p(i) = p_i$. The same Gaussian PSF as in the previous experiment was considered with a medium Poisson noise modelling with a background intensity $b = 10$. The binarization was fully learned again with $c = \delta + 0.01$. The number of unrolled iterations of \mathcal{A} was set to $K = 300$.

Table 5 shows the Jaccard index and PSNR values for this test, while in Figure 6 we report the reconstructed images. In this test the use of the KL modelling outperforms a Gaussian fidelity. In Figure 5 we report the decay of the loss function (14) and the value of δ through the outer iterations. In the other graph, we report the decay of the lower-level energy functional during testing. We observe that the choice $K = 300$ iterations is good enough to guarantee the increasing of the Jaccard index.

Table 5: Results for medium Poisson noise dataset. For each column, the best score across the models is in bold. The last column reports the initial parameter configuration.

Loss function	Fidelity	J_0	J_2	J_4	Avg. J	PSNR	Initial ρ, α, δ
\mathcal{L}_2	KL	0.0619	0.1285	0.1363	0.1089	39.88	1, 0.5, n.a.
\mathcal{L}_2	$\ \cdot\ _2$	0.0545	0.0812	0.0837	0.0731	39.97	10, 0.01, n.a.
\mathcal{L}_1	KL	0.0776	0.3687	0.4995	0.3153	27.39	0.1, 0.5, 0.1
\mathcal{L}_1	$\ \cdot\ _2$	0.0661	0.2291	0.3259	0.2070	28.00	1, 0.05, 0.1

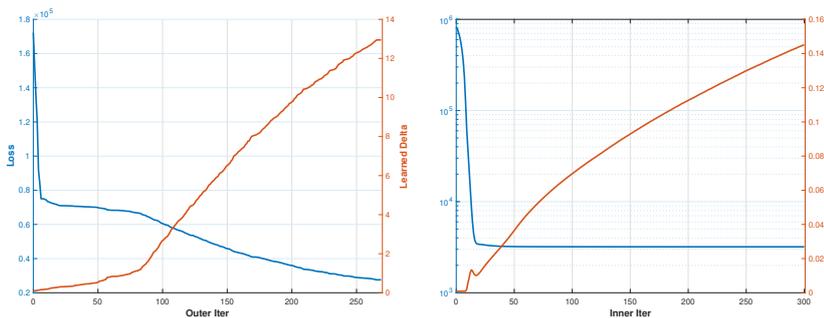


Fig. 5: Left: decrease of the loss functions \mathcal{L}_1 through the outer iterations (blue) and value of the binarization parameter δ (orange). Right: decrease of the inner energy functional (blue) and increase of Jaccard index (orange), through the inner iterations, computed as the mean across the 416 test samples, without applying the binarization.

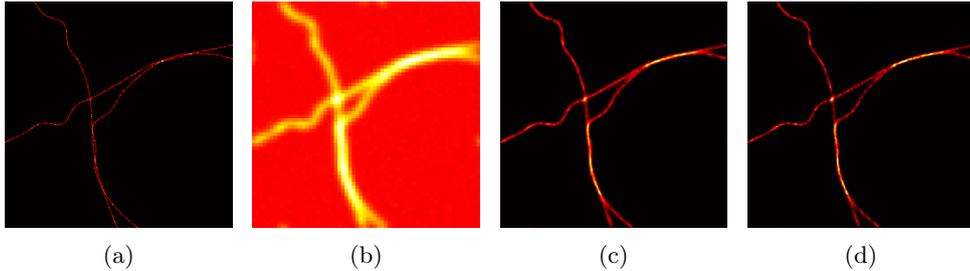


Fig. 6: (a) Ground truth image. (b) Blurred, noisy and low-resolution measurements (average over time). (c) Reconstruction computed using the \mathcal{L}_2 loss for training. (d) Reconstruction computed using the \mathcal{L}_1 loss for training.

3.2 Semi-blind image deconvolution in fluctuation-based microscopy.

We now consider a slightly different image formation model and test the proposed unrolling strategy on an image deconvolution problem based on the study of second-order image fluctuation statistics [52, 54, 55, 64] with semi-blind estimation of the underlying PSF. Starting from [51], this approach has proved indeed effective when dealing with biological samples featuring a high density of standard fluorescent molecules which, differently from the ones employed in SMLM applications, are less harmful for the samples observed and thus more suited for *in vivo* analyses. By exploiting the mutual independence of the intensity fluctuations at each pixel and computing second- (or higher-) order statistical information, such approaches reformulate the underlying inverse problem in a covariance domain, which, by the modelling assumptions, enforces a shrinkage of the optical PSF and where sparse regularization can be enforced.

The training dataset $\mathcal{D} = \{(V_G^s, V_F^s)\}_{s=1}^S$ is generated as follows:

- Step 1: for each simulated spatial pattern $s = 1, \dots, S$ a collection of $T = 1000$ fluctuating images is obtained for the molecular structure of interest, $\{g_s^{(t)}\}_{t=1}^T$. For simulating fluorescence fluctuations over time the SOFI simulator tool [51] was employed.
- Step 2: for each $s = 1, \dots, S$ and each frame $t = 1, \dots, T$, $g_s^{(t)}$ is blurred with a Gaussian PSF of standard deviation $\varsigma = 3$ and additive white Gaussian noise $n \sim \mathcal{N}(0, \sigma^2 \text{Id})$ with $\sigma = 3$ thus obtaining a measured image:

$$f_s^{(t)} = H(\varsigma)g_s^{(t)} + n. \quad (21)$$

- Step 3: for each $s = 1, \dots, S$ the empirical auto-covariances (that is, the variances) V_G^s and V_F^s of the stack are computed for both the clean and noisy data, respectively,

by the formulas:

$$M_G^s = \frac{1}{T} \sum_{t=1}^T g_s^{(t)}, \quad V_G^s = \frac{1}{T-1} \sum_{t=1}^T \left(g_s^{(t)} - M_G^s \right)^2;$$

$$M_F^s = \frac{1}{T} \sum_{t=1}^T f_s^{(t)}, \quad V_F^s = \frac{1}{T-1} \sum_{t=1}^T \left(f_s^{(t)} - M_F^s \right)^2.$$

For each spatial pattern $s = 1, \dots, S$ the pair (V_G^s, V_F^s) thus has as a ground-truth image V_G^s the variance of the fluctuating sample and its corrupted version V_F^s where by (21)

$$V_F^s = H(\varsigma) \odot H(\varsigma) V_G^s + V_N, \quad (22)$$

where \odot denotes the point-wise Hadamard product and, by definition, the vector $V_N = \sigma^2 \mathbf{e}$ with $\mathbf{e} = (1, \dots, 1)$ denotes the variance of the Gaussian noise considered. In Figure 7 we report a diagram of such modelling.

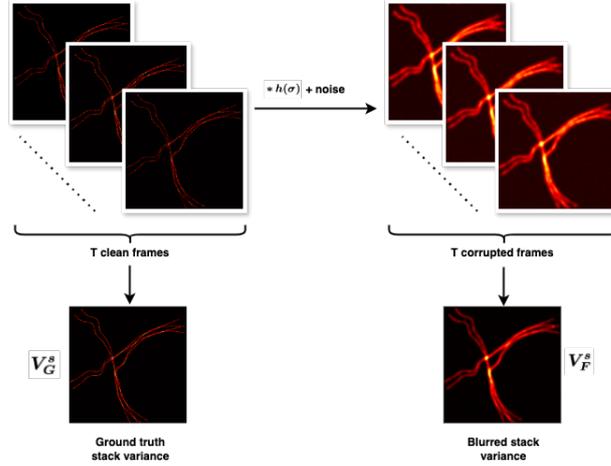


Fig. 7: Visualization of the acquisition process for one data sample in the Experiment 4 scenario.

Note that V_F^s is the second-order SOFI image [51] associated to $\left\{ f_s^{(t)} \right\}_{t=1}^T$. By (22) we notice that the ‘squaring’ of the PSF shrinks its spread by a factor $\sqrt{2}$, thus resulting in a better resolution. Defining for simplicity $H^2(\varsigma) := H(\varsigma) \odot H(\varsigma)$ the convolution matrix corresponding the kernel $h^2(\varsigma)$ we assume in the following that the Gaussian kernel is defined in terms of an unknown parameter σ which we incorporate within the learning procedure, thus considering as lower-level reconstruction functional the model:

$$(V_u^s)^*(\varsigma, \rho) \in \operatorname{argmin}_{V_u \in \mathbb{R}^{n^2}} E_{V_F} (V_u; \theta) := \|H^2(\varsigma)V_u + \sigma^2 \mathbf{e} - V_F^s\|_2^2 + \rho \|V_u\|_1 + \iota_{\{V_u \geq 0\}}, \quad (23)$$

where θ here includes both the parameters (ς, ρ) and the algorithmic step-sizes $\{\alpha_0, \dots, \alpha_{K-1}\}$ over the K unrolled iterations. Note that to backpropagate over ς we have to compute:

$$\begin{aligned} \frac{\partial}{\partial \varsigma} \nabla_{V_u} E_{V_F}(V_u; \theta) &= \frac{\partial}{\partial \varsigma} ((H^2(\varsigma))^T (H^2(\varsigma)V_u + \sigma^2 \mathbf{e} - V_F^s)) \\ &= \frac{\partial}{\partial \varsigma} (H^2(\varsigma))^T (H^2(\varsigma)V_u + \sigma^2 \mathbf{e} - V_F^s) + (H^2(\varsigma))^T \left(\frac{\partial}{\partial \varsigma} H^2(\varsigma)V_u \right), \end{aligned}$$

Further computations show that

$$\frac{\partial}{\partial \varsigma} H^2(\varsigma)V_u = \left(\frac{\partial}{\partial \varsigma} H^2(\varsigma) \right) V_u = \frac{\partial}{\partial \varsigma} h^2(\varsigma) * V_u,$$

with

$$\frac{\partial}{\partial \varsigma} h^2(\varsigma)[x, y] = \frac{e^{-\frac{(x-\frac{n+1}{2})^2 + (y-\frac{n+1}{2})^2}{\varsigma^2}}}{2\pi^2 \varsigma^5} \left(\frac{1}{\varsigma^2} \left(\left(x - \frac{n+1}{2} \right)^2 + \left(y - \frac{n+1}{2} \right)^2 \right) - 2 \right).$$

For constructing the dataset \mathcal{D} we considered $S = 30$ different spatial patterns. Each pattern is obtained by super-position/rotation and translation of filament structures similar to the ones in Figure 6a. The dataset was split into a training set with 20 elements and a test set with 10 images. The same experimental setup was used for the lower-level solver, with a number of $K = 300$ unrolled iterations. The \mathcal{L}_1 penalty (14) was used for training. The binarization function was fully learned with $c = \delta + 0.01$. The initial values for the parameters of this experiment were: $\rho = 10^{-5}$, $\delta = 25$, $\alpha = 1000$ and $\varsigma = 5$.

The average Jaccard index computed over the whole test set is reported in Table 6. Values of the Jaccard index computed before the binarization is applied are also reported to show the beneficial effect of introducing such function within the training phase. The learned value for ς was $\varsigma^* = 2.87$. In Figure 8 we illustrate one reconstruction computed from the test dataset corresponding to the optimal parameters learned. Note that the deconvolution performance appears quite accurate. Moreover, despite the slight underestimation of the PSF, no ringing effect is observed, probably due to the sparsity induced by the use of the ℓ_1 regularization with the non-negativity constraint.

4 Conclusions

We proposed an unfolded accelerated projected gradient descent scheme to solve both reconstruction and localization inverse problems often arising in the field of fluorescence microscopy.

Table 6: Localization assessment for the semi-blind deconvolution problem, with/without binarization.

Binarization	J_0	J_2	J_4	Avg. J
No	0.7096	0.7164	0.7208	0.7156
Yes	0.7922	0.8362	0.8582	0.8289

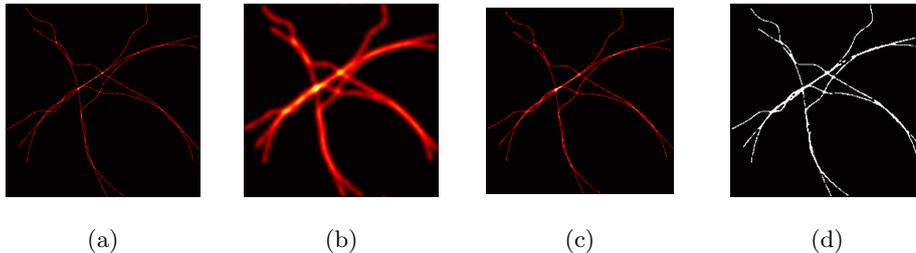


Fig. 8: (a) Ground truth image V_G^s . (b) Noisy variance image V_F^s . (c) Computed reconstruction. (d) Computed reconstruction post-binarization.

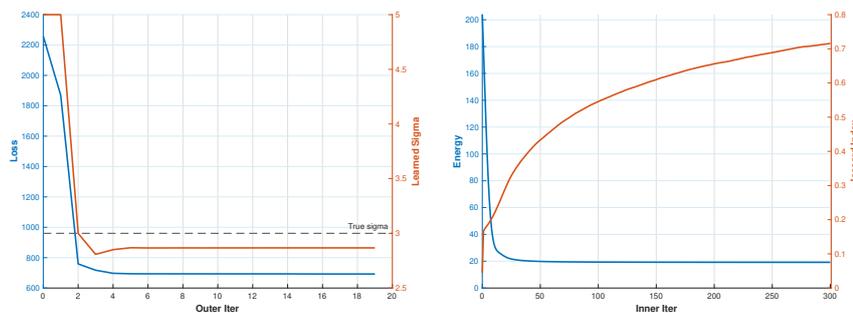


Fig. 9: Left: decrease of the \mathcal{L}_1 loss function through the outer iterations (blue) and evolution of the learned ς (orange). Right image: decrease of the inner energy functional (23) (blue) and increase of Jaccard index (orange) along inner iterations, computed as the mean across the 10 test samples.

By considering ℓ_1 -regularized criteria featuring suitable (Gaussian, Poisson) data terms, the proposed strategy learn optimal regularization and algorithmic parameters by means of two different loss functions assessing reconstruction and localization performance by suitable comparisons with given ground truth data.

The framework proposed adapts to both SMLM and fluctuation-based approaches as showed by several simulated and realistic numerical results, even in the presence of partially known physical models which can be learned within the training procedure.

Further work should address and compare the performance of unrolled schemes when used either in a physics-driven or in an end-to-end fashion as well as potentially combined with more effective, even neural network-based, sparsity promoted regularizations, such in the case of Plug and Play regularization procedures.

Acknowledgments

The authors acknowledge Vasiliki Stergiopoulou for kindly providing the datasets and the specifics for the tests run in Section 3.2. Silvia Bonettini, Danilo Pezzi and Marco Prato are all members of the INdAM research group GNCS, which is kindly acknowledged.

Declarations

Funding.

SB, MP, DP are partially supported by the PRIN projects 2022ANC8HL and 20225STXSB, under the National Recovery and Resilience Plan (NRRP) funded by the European Union – NextGenerationEU. LC acknowledges the support received by the ANR JCJC TASKABILE project (ANR-22-CE48-0010) and by the ANR PRC MICROBLIND (ANR-21-CE48-0008).

Competing interests.

The authors have no competing interests to declare that are relevant to the content of this article.

Ethics approval.

The authors declare that research ethics approval was not required for this study.

Consent.

The authors declare that informed consents were not required for this study.

Data availability statement.

The images of Section 3.1.2 can be downloaded from the websites <https://srm.epfl.ch/srm/index.php> and <https://srm.epfl.ch/DatasetPage?name=MT1.N1.LD>. All the synthetic datasets used in the making of this paper can be found at <https://github.com/DaniloPezzi/AGD-unfolding-for-microscopy-imaging>.

Conflict of interests.

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Authors' contributions.

All authors contributed to the study conception and design of the paper. The unrolled scheme has been created by S. Bonettini, D. Pezzi and M. Prato. The application to the SMLM problem has been handled by L. Calatroni. Code implementation, data collection and analysis were performed by L. Calatroni and D. Pezzi. The first draft of the manuscript was written by D. Pezzi, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Appendix A Non-differentiability of CEL0 w.r.t. the regularization parameter

We mentioned in Section 2.2 that the differentiability w.r.t. the regularization parameter denoted by ρ in (7) is required to apply the proposed unrolling strategy. Here we show such property does not hold for the CEL0 regularizer continuous and non-convex regularization functional R_{CEL0} studied in [59, 60] which, given its performance in localization microscopy problems (see, e.g., [48, 49]) would be a natural choice for localization purposes.

Such functional takes the form

$$R_{\text{CEL0}}(u) = \sum_{p=1}^n \phi(u_p; \rho, \|a_p\|_2), \quad (\text{A1})$$

where $\|a_p\|_2$ denotes the norm of the p -th column of the linear modeling operator A and ρ is the regularization parameter whose dependence here is highly non-linear. Enforcing non-negativity one has

$$\phi(s; \rho, a) = \begin{cases} \rho - \frac{a^2}{2} \left(s - \frac{\sqrt{2\rho}}{a} \right)^2, & 0 \leq s < \frac{\sqrt{2\rho}}{a} \\ \rho, & s \geq \frac{\sqrt{2\rho}}{a} \end{cases} \quad (\text{A2})$$

showing that the function $\phi(s; \rho, a)$ is differentiable w.r.t. s with derivative

$$\phi'(s; \rho, a) = \begin{cases} -a^2 \left(s - \frac{\sqrt{2\rho}}{a} \right), & 0 \leq s < \frac{\sqrt{2\rho}}{a} \quad \left(\text{equiv. } \rho > \frac{a^2 s^2}{2} \right) \\ 0, & s \geq \frac{\sqrt{2\rho}}{a} \quad \left(\text{equiv. } \rho \leq \frac{a^2 s^2}{2} \right) \end{cases}. \quad (\text{A3})$$

Since

$$\lim_{\epsilon \downarrow 0^+} \frac{\phi'(s; \frac{a^2 s^2}{2} + \epsilon, a) - \phi'(s; \frac{a^2 s^2}{2}, a)}{\epsilon} = \lim_{\epsilon \downarrow 0^+} -\frac{a^2}{\epsilon} \left(s - \frac{\sqrt{a^2 s^2 + 2\epsilon}}{a} \right) = \frac{1}{s} \neq 0, \quad (\text{A4})$$

the function $\phi'(s; \rho, a)$ is not differentiable w.r.t. ρ .

References

- [1] Bertero, M., Boccacci, P., Desiderà, G., Vicidomini, G.: Image deblurring with Poisson data: From cells to galaxies. *Inverse Probl.* **25**(12), 123006 (2009) <https://doi.org/10.1088/0266-5611/25/12/123006>
- [2] Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phen.* **60**(1), 259–268 (1992) [https://doi.org/10.1016/0167-2789\(92\)90242-F](https://doi.org/10.1016/0167-2789(92)90242-F)
- [3] Bertero, M., Boccacci, P., Ruggiero, V.: *Inverse Imaging with Poisson Data*. IOP Publishing, Bristol (2018). <https://doi.org/10.1088/2053-2563/aae109>
- [4] Luisier, F., Blu, T., Unser, M.: Image denoising in mixed Poisson-Gaussian noise. *IEEE Trans. Image Process.* **20**(3), 696–708 (2011) <https://doi.org/10.1109/TIP.2010.2073477>
- [5] Jezierska, A., Chaux, C., Pesquet, J., Talbot, H., Engler, G.: An EM approach for time-variant Poisson-Gaussian model parameter estimation. *IEEE Trans. on Sign. Process.* **62**(1), 17–30 (2014) <https://doi.org/10.1109/TSP.2013.2283839>
- [6] Calatroni, L., De Los Reyes, J.C., Schönlieb, C.-B.: Infimal convolution of data discrepancies for mixed noise removal. *SIAM J. Imaging Sci.* **10**(3), 1196–1233 (2017) <https://doi.org/10.1137/16M1101684>
- [7] Toader, B., Boulanger, J., Korolev, Y., Lenz, M.O., Manton, J., Schönlieb, C.-B., Mureşan, L.: Image reconstruction in light-sheet microscopy: Spatially varying deconvolution and mixed noise. *J. Math. Imaging Vis.* **64**, 968–992 (2022) <https://doi.org/10.1007/s10851-022-01100-3>
- [8] Stuart, A.M.: Inverse problems: A Bayesian perspective. *Acta Numer.* **19**, 451–559 (2010) <https://doi.org/10.1017/S0962492910000061>
- [9] Chan, T., Shen, J.: *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*. SIAM, Philadelphia (2005). <https://doi.org/10.1137/1.9780898717877>
- [10] Candes, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory* **52**(2), 489–509 (2006) <https://doi.org/10.1109/TIT.2005.862083>
- [11] Donoho, D.L.: Compressed sensing. *IEEE Transactions on Information Theory* **52**(4), 1289–1306 (2006) <https://doi.org/10.1109/TIT.2006.871582>
- [12] Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.* **4**(4), 1168–1200 (2005) <https://doi.org/10.1137/050626090>

- [13] Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**(1), 120–145 (2011) <https://doi.org/10.1007/s10851-010-0251-1>
- [14] Boyd, S.P., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**(1), 1–122 (2011) <https://doi.org/10.1561/2200000016>
- [15] Chambolle, A., Pock, T.: An introduction to continuous optimization for imaging. *Acta Numer.* **25**, 161–319 (2016) <https://doi.org/10.1017/S096249291600009X>
- [16] Nesterov, Y.: *Introductory Lectures on Convex Optimization: a Basic Course*. Springer, New York (2004). <https://doi.org/10.1007/978-1-4419-8853-9>
- [17] Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**, 183–202 (2009) <https://doi.org/10.1137/080716542>
- [18] Daubechies, I., Defrise, M., De Mol, C.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.* **57**(11), 1413–1457 (2004) <https://doi.org/10.1002/cpa.20042>
- [19] Morozov, V.A.: On the solution of functional equations by the method of regularization. *Dokl. Akad. Nauk SSSR* **167**(3), 510–512 (1966)
- [20] Bertero, M., Boccacci, P., Talenti, G., Zanella, R., Zanni, L.: A discrepancy principle for Poisson data. *Inverse Probl.* **26**(10), 105004 (2010) <https://doi.org/10.1088/0266-5611/26/10/105004>
- [21] Hansen, P.C.: Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Rev.* **34**(4), 561–580 (1992) <https://doi.org/10.1137/1034115>
- [22] Lucas, A., Iliadis, M., Molina, R., Katsaggelos, A.K.: Using deep neural networks for inverse problems in imaging: Beyond analytical methods. *IEEE Signal Process. Mag.* **35**(1), 20–36 (2018) <https://doi.org/10.1109/MSP.2017.2760358>
- [23] Arridge, S., Maass, P., Öktem, O., Schönlieb, C.-B.: Solving inverse problems using data-driven models. *Acta Numer.* **28**, 1–174 (2019) <https://doi.org/10.1017/S0962492919000059>
- [24] Monga, V., Li, Y., Eldar, Y.C.: Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Process. Mag.* **38**(2), 18–44 (2021) <https://doi.org/10.1109/MSP.2020.3016905>
- [25] Crockett, C., Fessler, J.A.: Bilevel methods for image reconstruction. *Found. Trends Signal Process.* **15**(2-3), 121–289 (2022) <https://doi.org/10.1561/2000000111>

- [26] Reyes, J.C., Schönlieb, C.-B.: Image denoising: Learning the noise model via nonsmooth PDE-constrained optimization. *Inverse Probl. Imaging* **7**(4), 1183–1214 (2013) <https://doi.org/10.3934/ipi.2013.7.1183>
- [27] Kunisch, K., Pock, T.: A bilevel optimization approach for parameter learning in variational models. *SIAM J. Imaging Sci.* **6**(2), 938–983 (2013) <https://doi.org/10.1137/120882706>
- [28] Ochs, P., Ranftl, R., Brox, T., Pock, T.: Techniques for gradient-based bilevel optimization with non-smooth lower level problems. *J. Math. Imaging Vis.* **56**(2), 175–194 (2016) <https://doi.org/10.1007/s10851-016-0663-7>
- [29] Calatroni, L., Cao, C., Reyes, J.C., Schönlieb, C.-B., Valkonen, T.: Bilevel approaches for learning of variational imaging models, vol. 18, pp. 252–290. De Gruyter, Berlin (2017). <https://doi.org/10.1515/9783110430394-008>
- [30] Bonettini, S., Franchini, G., Pezzi, D., Prato, M.: Explainable bilevel optimization: An application to the Helsinki deblur challenge. *Inverse Probl. Imaging* **17**(5), 925–950 (2023) <https://doi.org/10.3934/ipi.2022055>
- [31] Pedregosa, F.: Hyperparameter optimization with approximate gradient. In: *Proceedings of The 33rd International Conference on Machine Learning*, vol. 48, pp. 737–746. PMLR, New York (2016)
- [32] Ghadimi, S., Wang, M.: Approximation methods for bilevel programming. *arXiv:1802.02246* (2018). <https://doi.org/10.48550/arxiv.1802.02246>
- [33] Grazi, R., Pontil, M., Salzo, S.: Convergence properties of stochastic hypergradients. In: *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, vol. 130, pp. 3826–3834. PMLR, New York (2021)
- [34] Dagréou, M., Ablin, P., Vaiter, S., Moreau, T.: A framework for bilevel optimization that enables stochastic and global variance reduction algorithms. In: *Advances in Neural Information Processing Systems*, vol. 35, pp. 26698–26710. Curran Associates, Inc., New York (2022)
- [35] Gregor, K., LeCun, Y.: Learning fast approximations of sparse coding. In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*, pp. 399–406. Omnipress, Madison (2010)
- [36] Hershey, J.R., Roux, J.L., Wenginger, F.: Deep unfolding: Model-based inspiration of novel deep architectures. *arXiv:1409.2574* (2014). <https://doi.org/10.48550/arxiv.1409.2574>
- [37] Adler, J., Öktem, O.: Learned primal-dual reconstruction. *IEEE Trans. Med. Imaging* **37**(6), 1322–1332 (2018) <https://doi.org/10.1109/TMI.2018.2799231>

- [38] Bertocchi, C., Chouzenoux, E., Corbineau, M.-C., Pesquet, J.-C., Prato, M.: Deep unfolding of a proximal interior point method for image restoration. *Inverse Probl.* **36**(3), 034005 (2020) <https://doi.org/10.1088/1361-6420/ab460a>
- [39] Brauer, C., Breustedt, N., Wolff, T., Lorenz, D.A.: Learning variational models with unrolling and bilevel optimization. arxiv:2209.12651 (2022). <https://doi.org/10.48550/arxiv.2209.12651>
- [40] Riccio, D., Ehrhardt, M.J., Benning, M.: Regularization of inverse problems: deep equilibrium models versus bilevel learning. *Numer. Algebr. Control Optim.* (in press) <https://doi.org/10.3934/naco.2023026>
- [41] Dickson, R.M., Cubitt, A.B., Tsien, R.Y., Moerner, W.E.: On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature* **388**(6640), 355–358 (1997) <https://doi.org/10.1038/41048>
- [42] Hell, S.W., Wichmann, J.: Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.* **19**(11), 780–782 (1994) <https://doi.org/10.1364/OL.19.000780>
- [43] Betzig, E., Patterson, G.H., Sougrat, R., Lindwasser, O.W., Olenych, S., Bonifacio, J.S., Davidson, M.W., Lippincott-Schwartz, J., Hess, H.F.: Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**(5793), 1642–1645 (2006) <https://doi.org/10.1126/science.1127344>
- [44] Huang, B., Wang, W., Bates, M., Zhuang, X.: Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319**(5864), 810–813 (2008) <https://doi.org/10.1126/science.1153529>
- [45] Shtengel, G., Galbraith, J.A., Galbraith, C.G., Lippincott-Schwartz, J., Gillette, J.M., Manley, S., Sougrat, R., Waterman, C.M., Kanchanawong, P., Davidson, M.W., Fetter, R.D., Hess, H.F.: Interferometric fluorescent super-resolution microscopy resolves 3D cellular ultrastructure. *Proc. Natl. Acad. Sci. U. S. A.* **106**(9), 3125–3130 (2009) <https://doi.org/10.1073/pnas.0813131106>
- [46] Lelek, M., Gyparaki, M.T., Beliu, G., Schueder, F., Griffié, J., Manley, S., Jungmann, R., Sauer, M., Lakadamyali, M., Zimmer, C.: Single-molecule localization microscopy. *Nat. Rev. Method. Prim.* **1**(1), 39 (2021) <https://doi.org/10.1038/s43586-021-00038-x>
- [47] Aritake, T., Hino, H., Namiki, S., Asanuma, D., Hirose, K., Murata, N.: Fast and robust multiplane single-molecule localization microscopy using a deep neural network. *Neurocomputing* **451**, 279–289 (2021) <https://doi.org/10.1016/j.neucom.2021.04.050>
- [48] Gazagnes, S., Soubies, E., Blanc-Féraud, L.: High-density molecule localization for super-resolution microscopy using CEL0-based sparse approximation. In: 2017

- IEEE 14th International Symposium on Biomedical Imaging, pp. 28–31 (2017). <https://doi.org/10.1109/ISBI.2017.7950460>
- [49] Lazzaretti, M., Calatroni, L., Estatico, C.: Weighted-CELO sparse regularisation for molecule localisation in super-resolution microscopy with Poisson data. In: 2021 IEEE 18th International Symposium on Biomedical Imaging, pp. 1751–1754 (2021). <https://doi.org/10.1109/ISBI48211.2021.9434014>
- [50] Nehme, E., Weiss, L.E., Michaeli, T., Shechtman, Y.: Deep-storm: super-resolution single-molecule microscopy by deep learning. *Optica* **5**(4), 458–464 (2018) <https://doi.org/10.1364/OPTICA.5.000458>
- [51] Dertinger, T., Colyer, R., Iyer, G., Weiss, S., Enderlein, J.: Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI). *Proc. Natl. Acad. Sci. U. S. A.* **106**(52), 22287–22292 (2009) <https://doi.org/10.1073/pnas.0907866106>
- [52] Solomon, O., Eldar, Y.C., Mutzafi, M., Segev, M.: SPARCOM: Sparsity based super-resolution correlation microscopy. *SIAM J. Imaging Sci.* **12**, 392–419 (2019) <https://doi.org/10.1137/18M1174921>
- [53] Gustafsson, N., Culley, S., Ashdown, G., Owen, D.M., Pereira, P.M., Henriques, R.: Fast live-cell conventional fluorophore nanoscopy with ImageJ through super-resolution radial fluctuations. *Nat. Commun.* **7**(1), 12471 (2016) <https://doi.org/10.1038/ncomms12471>
- [54] Stergiopoulou, V., Morais Goulart, J.H., Schaub, S., Calatroni, L., Blanc-Féraud, L.: COLORME: Covariance-based ℓ_0 super-resolution microscopy with intensity estimation. In: 2021 IEEE 18th International Symposium on Biomedical Imaging, pp. 349–352 (2021). <https://doi.org/10.1109/ISBI48211.2021.9433976>
- [55] Stergiopoulou, V., Calatroni, L., Morais Goulart, H., Schaub, S., Blanc-Féraud, L.: COLORME: Super-resolution microscopy based on sparse blinking/fluctuating fluorophore localization and intensity estimation. *Biolog. Imaging* **2**, 1 (2022) <https://doi.org/10.1017/S2633903X22000010>
- [56] Harmany, Z.T., Marcia, R.F., Willett, R.M.: This is spiral-tap: Sparse poisson intensity reconstruction algorithms—theory and practice. *IEEE Transactions on Image Processing* **21**(3), 1084–1096 (2012) <https://doi.org/10.1109/TIP.2011.2168410>
- [57] Bertero, M., Lantéri, H., Zanni, L.: Iterative image reconstruction: a point of view. In: *Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy (IMRT)*, pp. 37–63. Birkhauser-Verlag, Pisa, Italy (2008)
- [58] Chambolle, A., Dossal, C.: On the convergence of the iterates of the “Fast Iterative Shrinkage/Thresholding Algorithm”. *J. Optim. Theory Appl.* **166**(3), 968–982 (2015) <https://doi.org/10.1007/s10957-015-0746-4>

- [59] Soubies, E., Blanc-Féraud, L., Aubert, G.: A continuous exact ℓ_0 penalty (CEL0) for least squares regularized problem. *SIAM J. Imaging Sci.* **8**(3), 1607–1639 (2015) <https://doi.org/10.1137/151003714>
- [60] Soubies, E., Blanc-Féraud, L., Aubert, G.: A unified view of exact continuous penalties for ℓ_2 - ℓ_0 minimization. *SIAM J. Optim.* **27**(3), 2034–2060 (2017) <https://doi.org/10.1137/16M1059333>
- [61] Sage, D., Pham, T.-A., Babcock, H., Lukes, T., Pengo, T., Chao, J., Velmurugan, R., Herbert, A., Agrawal, A., Colabrese, S., Wheeler, A., Archetti, A., Rieger, B., Ober, R., Hagen, G.M., Sibarita, J.-B., Ries, J., Henriques, R., Unser, M., Holden, S.: Super-resolution fight club: assessment of 2D and 3D single-molecule localization microscopy software. *Nat. Methods* **16**(5), 387–395 (2019) <https://doi.org/10.1038/s41592-019-0364-4>
- [62] Bonettini, S., Zanella, R., Zanni, L.: A scaled gradient projection method for constrained image deblurring. *Inverse Probl.* **25**(1), 015002 (2008) <https://doi.org/10.1088/0266-5611/25/1/015002>
- [63] Koulouri, A., Heins, P., Burger, M.: Adaptive superresolution in deconvolution of sparse peaks. *IEEE Trans. Signal Process.* **69**, 165–178 (2021) <https://doi.org/10.1109/TSP.2020.3037373>
- [64] Stergiopoulou, V., Mukherjee, S., Calatroni, L., Blanc-Féraud, L.: Fluctuation-based deconvolution in fluorescence microscopy using plug-and-play denoisers. In: *SSVM 2023: Scale Space and Variational Methods in Computer Vision*, vol. 14009, pp. 498–510. Springer, New York (2023). https://doi.org/10.1007/978-3-031-31975-4_38