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

Multiview Attenuation Estimation and Correction

Valentin Debarnot, Jonas Kahn and Pierre Weiss

Received: date / Accepted: date

Abstract Measuring attenuation coefficients is a fundamental problem that can be solved with diverse techniques such as X-ray or optical tomography and lidar. We propose a novel approach based on the observation of a sample from a few different angles. This principle can be used in existing devices such as lidar or various types of fluorescence microscopes. It is based on the resolution of a nonlinear inverse problem. We propose a specific computational approach to solve it and show the well-foundedness of the approach on simulated data. Some of the tools developed are of independent interest. In particular we propose an efficient method to correct attenuation defects, new robust solvers for the lidar equation as well as new efficient algorithms to compute the proximal operator of the logsumexp function in dimension 2.

Keywords Non-convex optimization, Bayesian estimation, lidar, fluorescence microscopy, multiview estimation, Poisson noise.

Acknowledgements This work was supported by the Fondation pour la Recherche Médicale (FRM grant number ECO20170637521 to V.D.) and by Plan CANCER, MIMMOSA project.
The authors wish to thank Juan Cuesta, Emilio Gualda, Jan Huisken, Philipp Keller, Théo Liu, Jürgen Mayer and Anne Sentenac for interesting discussions and feedbacks on the model. They thank the anonymous reviewers for pointing out reference [30], which is closely related to this paper.

Valentin Debarnot
ITAV, CNRS, France
E-mail: valentin.debarnot@gmail.com

Jonas Kahn
IMT and ITAV, CNRS and Université de Toulouse, France
E-mail: jokahn@phare.normalesup.org

Pierre Weiss
IMT and ITAV, CNRS and Université de Toulouse, France
E-mail: pierre.armand.weiss@gmail.com

1 Introduction

The ability to analyze the composition of gases in the atmosphere, the organization of a biological tissue, or the state of organs in the human body has invaluable scientific and societal repercussions. These seemingly unrelated issues can be solved thanks to a common principle: rays traveling through the sample are attenuated and this attenuation provides an indirect measurement of absorption coefficients. This is the basis of various devices such as X-ray and optical projection tomography [22, 31, 35] or lidar [36]. The aim of this paper is to provide an alternative approach based on the observation of the sample from a few different angles.

1.1 The basic principle

Let us provide a flavor of the proposed idea in an idealized 1D system. Assume that two measured signals u_1 and u_2 are formed according to the following model:

$$u_1(x) = \beta(x) \exp\left(-\int_0^x \alpha(t)\,dt\right) \text{ for } x \in [0,1] \tag{1}$$

and

$$u_2(x) = \beta(x) \exp\left(-\int_x^1 \alpha(t)\,dt\right) \text{ for } x \in [0,1]. \tag{2}$$

The function $\beta : [0,1] \to \mathbb{R}_+$ will be referred to as a density throughout the paper. It may represent different physical quantities such as backscatter coefficients in

lidar or fluorophore densities in microscopy. The function $\alpha : [0,1] \to \mathbb{R}_+$ will be referred to as the attenuation and may represent absorption or extinction coefficients. The signals u_1 and u_2 can be interpreted as measurements of the same scene under opposite directions. Equations (1) and (2) coincide with the Beer-Lambert law that is a simple model to describe attenuation of light in absorbing media. The question tackled in this paper is: *can we recover both α and β from the knowledge of u_1 and u_2?*

Under a positivity assumption $\beta(x) > 0$ for all $x \in [0,1]$, the answer is straightforwardly positive. Setting $v(x) = \log\left(\frac{u_2(x)}{u_1(x)}\right)$, equations (1) and (2) yield:

$$v(x) = \int_0^x \alpha(t)\,dt - \int_x^1 \alpha(t)\,dt. \tag{3}$$

Therefore

$$\alpha(x) = \frac{1}{2}\frac{\partial}{\partial_x} v(x) \tag{4}$$

and

$$\beta(x) = \frac{u_1(x)}{\exp\left(-\int_0^x \alpha(t)\,dt\right)}. \tag{5}$$

Unfortunately, formulas (4) and (5) only have a theoretical interest: they cannot be used in practice since computing the derivative of a log of a ratio is extremely sensitive to noise and thus very unstable from a numerical point of view. We will therefore design a numerical procedure based on a Bayesian estimator to retrieve the density β and attenuation coefficients α in a stable and efficient manner. It is particularly relevant when the data suffer from Poisson noise.

1.2 Contributions

This paper contains various contributions listed below.

- We show that it is possible to retrieve attenuation coefficients from multiview measurements in different systems such as lidar, confocal or SPIM microscopes. Figure 1 summarizes the proposed idea. The attenuation, which is usually considered as a nuisance in confocal microscopy is exploited to measure the absorption coefficients. The algorithm successfully retrieves estimates of the density and attenuation from two attenuated and noisy images. Let us also mention that some researchers already proposed to measure absorption and correct attenuation by combining optical projection tomography and SPIM imaging [21]. The principle outlined here shows that much simpler optical setups (a traditional confocal microscope) theoretically allows estimating the same quantities.
- We propose novel Bayesian estimators for the density α and the attenuation β based on a Poisson noise modeling.
- The proposed estimators are solutions of a nonconvex problem. We show that exact solutions of the problem can be obtained by using a trick making the problem convex.
- The resulting convex program is challenging from a numerical point of view and involves functions that are uncommon in imaging. This leads us to develop an efficient algorithm to compute the proximal operator of the logsumexp function in dimension 2.
- The proposed estimators also seem to be novel for the standard mono-view inverse problem in lidar and for correcting attenuation defects under a Poisson noise assumption with multiple views.
- We perform a numerical validation of the proposed ideas on synthetic data, showing the well-foundedness of the approach. The validation of the method on specific devices is left as an outlook for future works.

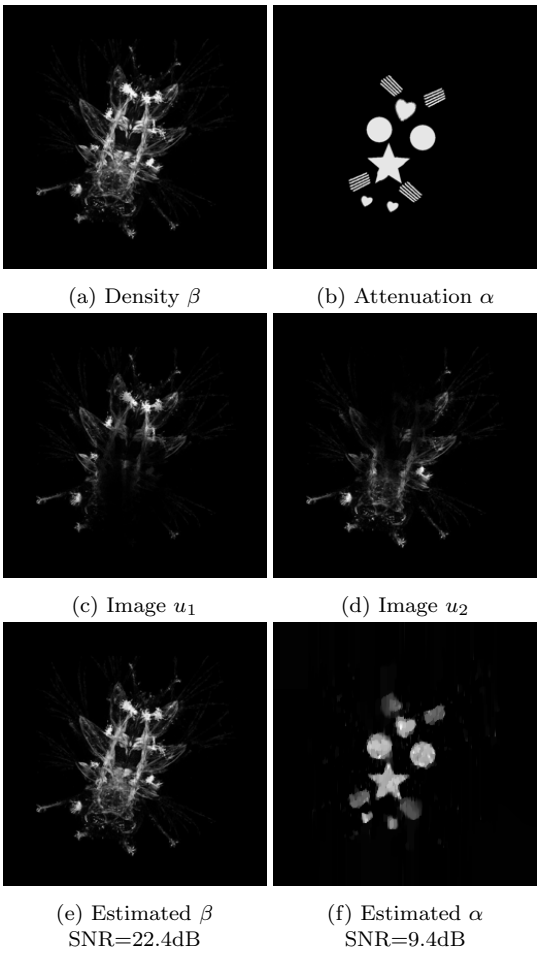
We found the general principle stated above independently, but became aware of a few papers proposing similar concepts after finishing the manuscript. In lidar, the idea was explored in the 1980's already [9, 15, 20]. In confocal microscopy, the recent paper [30] proposes a setting very similar to the one proposed here. From a practical point of view, the early papers [9, 15, 20] were based on simple Wiener filtering approaches. Our tests using these approaches on simulations led us to the conclusion that they were far too unstable and we will not report these results. On the other hand, the paper [30] presents a closely related framework: the authors use a maximum a posteriori estimator with a total variation regularizer, leading to a variational formulation of the problem. We will see however that its structure seems less amenable to an efficient numerical resolution. Our conclusion using the L-BFGS approach suggested in [30] is that unless a very good initial guess is provided, the method is unable to retrieve the attenuation, whereas our globally convergent approach is insensitive to the starting point.

2 Applications

In this section, we show various applications where the methodology proposed in this paper can be applied.

2.1 Lidar

In lidar, an object (atmosphere, gas,...) is illuminated with a laser beam. Particles within the object reflect



(a) Density β (b) Attenuation α

(c) Image u_1 (d) Image u_2

(e) Estimated β SNR=22.4dB (f) Estimated α SNR=9.4dB

Fig. 1: Illustration of the contribution. A sample (here an insect) has a fluorophore density β shown in Fig. 1a and an attenuation map α shown in Fig.1b. The two measured images u_1 and u_2 are displayed in Fig. 1c and 1d. As can be seen, they are attenuated differently (top to bottom and bottom to top) since the optical path is reversed. From these two images, our algorithm provides a reliable estimate of each map in Fig. 1e and 1f despite Poisson noise.

light. The time to return of the reflected light is then measured with a scanner. The received signal $u_1(x)$ is the backscattered mean power at altitude x for a specific wavelength. The density β corresponds to the backscattered coefficient, while α is called extinction coefficient. The equation relating u_1 to α and β is:

$$u_1(x) = \mathcal{P}\left(\frac{C}{x^2}\beta(x)\exp\left(-2\int_0^x \alpha(t)\,dt\right)\right), \qquad (6)$$

where C is independent of x. The notation $\mathcal{P}(z)$ stands for a Poisson distributed random variable of parameter z. The Poisson distribution is a rather good noise model in lidar, since measurements describe a number of detected photons. The term $\frac{C}{x^2}\beta(x)$ appears in the lidar equation (6) instead of simply β. The algorithm developed later will allow retrieving $\frac{C}{x^2}\beta(x)$ instead of β. This is not a problem since there is a direct known relationship between both.

Remark 1 In Raman lidar, the coefficient β corresponds to the molecular density of the atmosphere, while α is the sum of extinction coefficients at different wavelengths. The theory developed herein also applies to this setting.

When the backscatter coefficient β has a known analytical relationship with the extinction coefficient α, direct inversion is possible. A popular method is Klett's formula [18] for instance. Alternative formula exist [1] when the backscatter coefficient is known. The recent trend consists in using iterative methods coming from the field of inverse problems [12, 26, 32], leading to improved robustness. All these approaches crucially depend on a precise knowledge of the backscatter coefficient. This is a strong hypothesis that is often rough or unreasonable in practice.

To overcome this issue, a few authors proposed to use two opposite lidars and to retrieve the attenuation coefficients using equation (4) [9, 15, 20]. The stability to noise was ensured by linear filtering of the input and output data. Our simulations using them did not yield satisfactory results and we will not report them.

2.2 Fluorescence microscopy

The principle proposed herein can also be applied to some fluorescence microscopes. This idea was already proposed in confocal microscopy [30]. Here we show that it can be extended to other microscopes such as 4π or selective plane illumination microscopes (SPIM).

All fluorescence microscopes share a common principle: a source of illumination excites fluorophores within the sample, which in turn emit some light. This light is collected with a camera. Both the illumination and emission light can be absorbed along its optical path, which results in inhomogeneities in the image contrasts.

Depending on the imaging device, the way light absorption distorts images can be different. We illustrate this with two synthetic examples in Fig. 3. In a confocal microscope, the illumination and emission light both travel in the same direction, creating unidirectional absorption. Two images suffering from opposite contrast losses can be obtained by rotating the sample or by using a 4-pi microscope [8, 13], see Fig. 3a.

In the multi-view versions of the selective plane illumination microscopy SPIM (also called light sheet fluorescence microscopy) [6, 16, 19, 34], the illumination and

emission light travel in orthogonal directions, creating bi-directional contrast losses, see Fig. 3b.

The attenuation map α is wavelength dependent and to be precise, we should consider two attenuation maps α_i and α_e for the illumination and emission light respectively. In this work, we simply assume that the two are related through a linear relationship $\alpha = \alpha_e = \kappa\alpha_i$, where κ is a positive scalar. Under this assumption, the microscopes provide a set of images $(u_i)_{1\leq i\leq m}$, which can be modeled as follows:

$$u_i = \mathcal{P}\left(\beta \exp(-A_i\alpha)\right), \tag{7}$$

where A_i denotes a linear integral operator that depends on the geometry of the optical setup. Figure 2 provides a description of these operators in the case of a confocal microscope and of a multi-view SPIM microscope. For a point $x \in \mathbb{R}^d$, the expression of $(A_i\alpha)(x)$ is given by:

$$(A_i\alpha)(x) = \int_{S_1(x)} \alpha(y)\mathrm{d}y + \kappa \int_{S_2(x)} \alpha(y)\mathrm{d}y,$$

where $S_1(x)$ and $S_2(x)$ are the cones of light depicted in Fig. 2. In our numerical experiments, we use a simple version where the operators return line integrals (and not cone integrals). This amounts to assuming that the light rays are infinitely thin. However, the proposed approach may be extended to arbitrary geometries through the use of heavier linear algebra solvers.

3 MAP estimator and numerical evaluation

In this section, we describe our numerical procedure completely. We start by providing a discrete version of our image formation model. Then, we design a Bayesian estimator of the attenuation map α and of the density β. We finally design an effective optimization algorithm to compute our statistical estimator.

3.1 The discretized model

The discrete model we consider in this paper reads:

$$\begin{cases} u_1 &= \mathcal{P}\left(\beta \exp(-A_1\alpha)\right) \\ \vdots & \\ u_m &= \mathcal{P}\left(\beta \exp(-A_m\alpha)\right). \end{cases} \tag{8}$$

The signals u_i, β and α are assumed to be nonnegative and belong to $\mathbb{R}^n$, where $n = n_1 \dots n_d$ denotes the number of pixels and d is the space dimension. The value of a vector u_1 at location $i = (i_1, \dots, i_d)$ will be denoted either $u_1[i]$ or $u_1[i_1, \dots, i_d]$. The matrices A_i in $\mathbb{R}^{n\times n}$ are discretization of linear integral operators. In our numerical experiments, we use $m = 2$ views and the product A_1u_1 simply represents the cumulative sum of u_1 along one direction while the product A_2u_2 represents the cumulative sum of u_2 in the opposite direction. For instance, for a 1D signal, we set:

$$(A_1u)[i] = \sum_{j=1}^{i} u_1[j].$$

Therefore, the matrix A_1 has the following lower triangular shape:

$$A_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \ddots & \dots & 0 \\ 1 & 1 & 1 & 1 & \dots & 1 \end{pmatrix} \tag{9}$$

We are now ready to design a Bayesian estimator of α and β from model (8).

3.2 A Bayesian estimator

The Maximum A Posteriori (MAP) estimators $\hat{\alpha}$ and $\hat{\beta}$ of α and β are defined as the maximizers of the conditional probability density:

$$\max_{\alpha\in\mathbb{R}^n, \beta\in\mathbb{R}^n} \mathrm{p}(\alpha, \beta | (u_i)_{1\leq i\leq m}).$$

By using the Bayes rule and a negative log-likelihood, this is equivalent to finding the minimizers of:

$$\min_{\alpha\in\mathbb{R}^n, \beta\in\mathbb{R}^n} -\log(\mathrm{p}((u_i)_{1\leq i\leq m}|\alpha, \beta)) - \log(\mathrm{p}(\alpha, \beta)).$$

Let us evaluate $\mathrm{p}((u_i)_{1\leq i\leq m}|\alpha, \beta)$. To this end, set

$$\lambda_j = \beta \exp(-(A_j\alpha)).$$

Since the distribution of a Poisson distributed random variable with parameter λ has the following probability mass function:

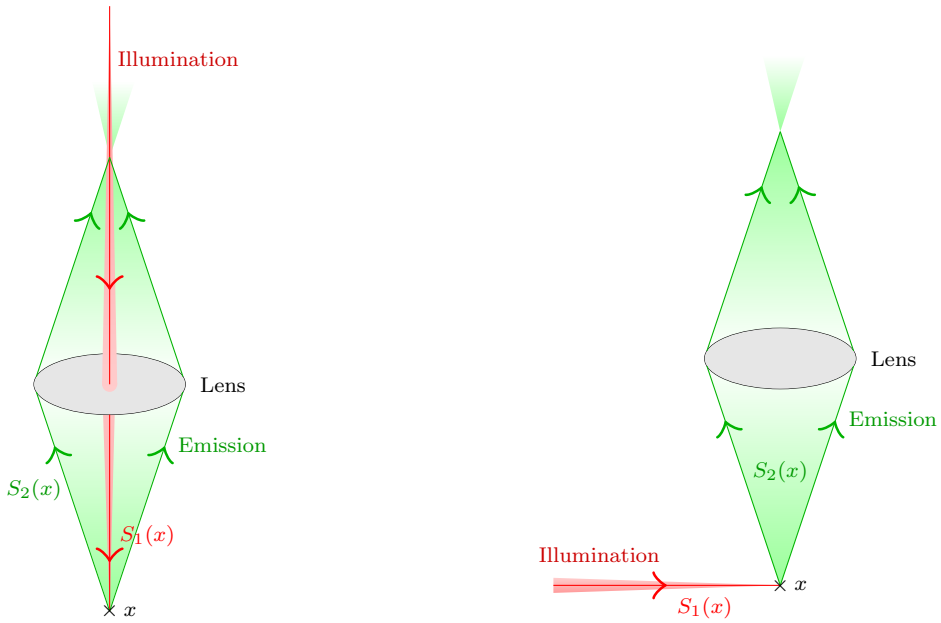
$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

we get:

$$\begin{aligned} &-\log\left(\mathrm{p}((u_i)_{1\leq i\leq m}|\alpha, \beta)\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} \lambda_j[i] - u_j[i]\log(\lambda_j[i]) + C, \end{aligned}$$

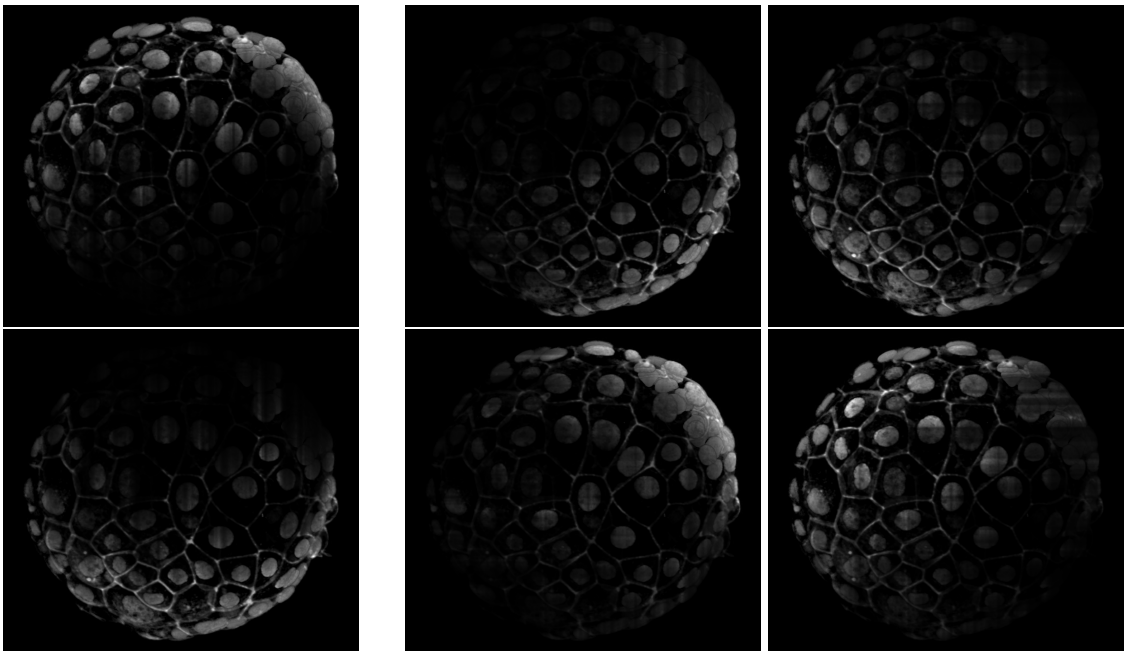
where C is a value that does not depend on α and β. Next, we assume that α and β are *independent random vectors* with probability distribution functions of type:

$$\mathrm{p}(\alpha) \propto \exp(-R_\alpha(\alpha)) \text{ and } \mathrm{p}(\beta) \propto \exp(-R_\beta(\beta)),$$



(a) Confocal (b) Multi-SPIM

Fig. 2: Path of the light from different microscopes.



(a) Confocal microscope (b) Multi-SPIM microscope

Fig. 3: Simulated contrast loss in a slice of mouse embryo with a confocal microscope (left) and a multi-view light sheet fluorescence microscope (left).

where $R_\alpha : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $R_\beta : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ are regularizers describing properties of the density and attenuation maps. Overall, the optimization problem characterizing the MAP estimates reads:

$$\min_{\alpha \in \mathbb{R}^n, \beta \in \mathbb{R}^n} F(\alpha, \beta) \tag{10}$$

where

$$F(\alpha,\beta) = R_\alpha(\alpha) + R_\beta(\beta) + \left\langle \sum_{j=1}^{m} \beta \exp\left(-A_j\alpha\right) + u_j \left(A_j\alpha - \log(\beta)\right), \mathbb{1} \right\rangle \tag{11}$$

and $\mathbb{1}$ stands for the vector in $\mathbb{R}^n$ with all components equal to 1.

Remark 2 For $m = 1$ view, the problem $\min_\alpha F(\alpha, \beta)$ allows recovering the attenuation knowing the density: this is the standard inverse problem met in lidar. To the best of our knowledge, the proposed formulation is novel for this problem.

Remark 3 The problem $\min_\beta F(\alpha, \beta)$ corresponds to correcting the attenuation on the density map. This is also a frequently met problem [3, 17, 27, 28] and to the best of our knowledge, the proposed approach - based on the MAP principle - is original, though it bear resemblances with [30] for instance.

3.3 Making the problem convex

Let us start by analyzing the convexity properties of the function F. To this end, let us introduce the function $G : \mathbb{R}^n \times \mathbb{R}^n_+ \to \mathbb{R}$ defined by:

$$G(\alpha,\beta) = \left\langle \sum_{j=1}^{m} \beta \exp\left(-A_j\alpha\right) + u_j \left(A_j\alpha - \log(\beta)\right), \mathbb{1} \right\rangle.$$

Notice that $F(\alpha, \beta) = G(\alpha, \beta) + R_\alpha(\alpha) + R_\beta(\beta)$. The following proposition provides the domain of convexity of G.

Proposition 1 *The function G*

- *is convex on each variable separately on $\mathbb{R}^n \times \mathbb{R}^n_+$.*
- *is non convex on $\mathbb{R}^n \times \mathbb{R}^n_+$.*
- *has a positive semidefinite Hessian on the (nonconvex) set:*

$$\left\{ (\alpha,\beta) \in \mathbb{R}^n \times \mathbb{R}^n_+, \; \beta \sum_{j=1}^{m} \exp(-A_j\alpha) \leq \sum_{j=1}^{m} u_j \right\}. \tag{12}$$

This proposition is proved in the appendix 8.1. It shows that if R_α and R_β are convex functions, then F is convex in each variable separately. Unfortunately, it is nonconvex on the product space unless the regularizers R_α and R_β compensate for the nonconvexity. The main observation in this paragraph is that it is possible to find a global minimizer of F when R_α is a standard convex regularizer and R_β is the indicator function of the positive orthant. This property is related to the third item in Proposition 1.

Proposition 2 *Set*

$$R_\beta(\beta) = \iota_{\mathbb{R}^n_+}(\beta) := \begin{cases} 0 & \text{if } \beta[i] \geq 0, \forall i \in \{1, \ldots, n\}, \\ +\infty & \text{otherwise.} \end{cases}$$

Then, if they exist, the solutions $(\hat\alpha, \hat\beta)$ of problem (10) are given by:

$$\hat\alpha \in \operatorname*{argmin}_{\alpha \in \mathbb{R}^n} R_\alpha(\alpha) + \left\langle \sum_{j=1}^{m} u_j \left[A_j\alpha + \log\left(\sum_{i=1}^{m} \exp(-(A_i\alpha)) \right) \right], \mathbb{1} \right\rangle \tag{13}$$

and

$$\hat\beta = \frac{\sum_{j=1}^{m} u_j}{\sum_{j=1}^{m} \exp\left(-A_j\hat\alpha\right)}. \tag{14}$$

In addition, Problem (13) is convex if the regularizer R_α is convex.

The expression (14) can be seen as a simple estimator of β knowing $(u_j)_{1 \leq j \leq m}$ and α. Notice that it coincides exactly with the boundary of the set (12). The existence and uniqueness of minimizers can also be shown by adding assumptions on R_α, such as strict convexity. We do not study this question further, since at this point we state results for arbitrary regularizers.

The convexity of problem (13) is critical: it shows that *global* minimizers of (14) can likely be computed if the regularizer R_α is chosen wisely. This observation motivates solving (13) to get an estimate of $\hat\alpha$. The only problem is that the estimated density $\hat\beta$ is regularized mildly using the sole non-negativity assumption. Hence, we propose an additional denoising step in the next paragraph.

3.4 Density estimation with a fixed attenuation

In order to remove the noise from the density, we use once again a MAP estimator with a more advanced regularizer R_β, assuming that the true attenuation α is actually equal to $\hat\alpha$.

Following Section 3.2, we get that the estimator $\hat{\hat{\beta}}$ is given by:

$$\hat{\hat{\beta}} \in \underset{\beta\in\mathbb{R}^n_+}{\operatorname{argmin}} F(\hat{\alpha},\beta) = \underset{\beta\in\mathbb{R}^n_+}{\operatorname{argmin}} R_\beta(\beta) + \left\langle \sum_{j=1}^m \beta \exp\left(-A_j\hat{\alpha}\right) + u_j\left(A_j\hat{\alpha} - \log\left(\beta\right)\right), \mathbb{1} \right\rangle, \tag{15}$$

where we assumed that β is a random vector with probability distribution function of type:

$$\mathrm{p}(\beta) \propto \exp\left(-R_\beta(\beta)\right).$$

Once again, the global solution of this problem can be computed by choosing a sufficiently simple convex regularizer R_β.

Proposition 3 *Problem* (15) *is convex for a convex regularizer* R_β.

Proof The proof derives directly from Proposition 1, first item.

4 Optimization methods

4.1 Recovering the attenuation

We now delve into the numerical resolution of (13). First, we need to choose a convex regularizer R_α. In this paper, we propose to simply use the total variation [29] together with a non-negativity constraint, which is well known to preserve sharp edges. Its expression is given by:

$$R_\alpha(\alpha) = \lambda_\alpha \sum_{i=1}^n \|(\nabla\alpha)[i]\|_2 + \iota_{\mathbb{R}^n_+}(\alpha),$$

where $\nabla : \mathbb{R}^n \to \mathbb{R}^{dn}$ is a discretization of the gradient, $\lambda_\alpha \geq 0$ is a regularization parameter and $\iota_{\mathbb{R}^n_+}$ is the indicator of the nonnegative orthant. We will use the standard discretization proposed in [4] in our numerical experiments.

Problem (13) is convex, but rather hard to minimize for various reasons listed below. First, the vectors α and β may be very high dimensional, preventing the use of an arbitrary black-box method. Second, the regularizer R_α is non differentiable. Third, the operators A_i have a spectral norm depending on the dimension n, preventing the use of gradient based methods since the Lipschitz constant of the gradient would be too high, see Proposition 4. Last, the proximal operator associated to the logsumexp function has no simple analytical formula.

Proposition 4 *Matrix* A_1 *in* (9) *satisfies* $\|A_1\|_{2\to 2} \gtrsim n$, *where* $\|\cdot\|_{2\to 2}$ *stands for the spectral norm.*

Proof

$$\|A_1\|^2_{2\to 2} \geq \left\| A_1 \begin{pmatrix} 1/\sqrt{n} \\ \vdots \\ 1/\sqrt{n} \end{pmatrix} \right\|_2^2 \geq \frac{1}{n} \left\| \begin{pmatrix} 1 \\ 2 \\ \vdots \\ n \end{pmatrix} \right\|_2^2 \gtrsim \frac{n^3}{n} = n^2.$$

A large number of splitting methods have been developed to solve problems of type (13), and we refer to the excellent review papers [5, 7] for an overview. Among them, the Simultaneous Direction Method of Multipliers (SDMM), a variant of the ADMM [11,23] is particularily adapted to the structure of our problem. This algorithm allows solving problems of type:

$$\min_{\alpha\in\mathbb{R}^n} g_1(L_1\alpha) + \ldots + g_m(L_m\alpha), \tag{16}$$

where functions $g_i : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ are convex closed and the operators $L_i : \mathbb{R}^n \to \mathbb{R}^{m_i}$ are linear and such that $Q = \sum_{i=1}^n L_i^T L_i$ is an invertible matrix. The SDMM then takes the algorithmic form described in Algorithm 1.

Algorithm 1 The SDMM algorithm to solve (16)

input: Nit, $\gamma > 0$, $(y_{i,0})_{1\leq i\leq m}$, $(z_{i,0})_{1\leq i\leq m}$
for $k = 1$ to Nit **do**
  $x_k = Q^{-1}\sum_{i=1}^m L_i^T(y_{i,k} - z_{i,k})$
  for $i = 1$ to m **do**
    $s_{i,k} = L_i x_k$.
    $y_{i,k+1} = \mathrm{prox}_{\gamma g_i}(s_{i,k} + z_{i,k})$
    $z_{i,k+1} = z_{i,k} + s_{i,k} - y_{i,k+1}$
  end for
end for

To cast problem (13) into form (16), we use the following choices. We set $L_2 = c_2\nabla$

$$\begin{aligned} L_1 : \mathbb{R}^n &\to \mathbb{R}^{2n} \\ \alpha &\mapsto c_1 \begin{pmatrix} A_1\alpha \\ A_2\alpha \end{pmatrix}, \end{aligned} \tag{17}$$

$$\begin{aligned} g_1 : \mathbb{R}^{2n} &\to \mathbb{R}\cup\{+\infty\} \\ \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} &\mapsto \sum_{i=1}^n \sum_{j=1}^2 u_j[i]\Big(z_j[i]/c_1 \\ &\quad + \log\Big(\sum_{j=1}^2 \exp\ (-z_j[i]/c_1)\Big)\Big), \end{aligned}$$

and

$$g_2 \begin{pmatrix} z_1 \\ \vdots \\ z_d \end{pmatrix} = \frac{\lambda}{c_2} \sum_{i=1}^n \sqrt{z_1^2[i] + \ldots + z_d[i]^2},$$

$L_3 = c_3 I_n$ and $g_3(z) = \iota_{\mathbb{R}^n_+}(z)$.

The numbers c_1, c_2, c_3 are positive constants allowing to accelerate the algorithm's convergence by balancing the relative importance of each term. This can also be seen as a simple diagonal preconditioner. In our numerical experiments, we set $c_1 = 1$ and tune c_2 and c_3 manually to accelerate convergence.

In order to apply Algorithm 1, we need to compute the proximal operators of each function g_i, defined by:

$$\text{prox}_{\gamma g_i}(z_0) = \underset{z \in \mathbb{R}^{m_i}}{\text{argmin}}\, \gamma g_i(z) + \frac{1}{2}\|z - z_0\|_2^2.$$

The proximal operators of g_2 and g_3 have closed form solutions found in nearly all recent total variation minimization solvers. We refer to [23] for instance. Unfortunately, the proximal operator of g_1 has no closed-form expression. In order to compute it, we propose using a non trivial Newton-based algorithm described in section 8.3. Finally, we need to evaluate matrix-vector products with Q^{-1}. This can be achieved using either a LU factorization or a conjugate gradient. The LU factorization is feasible if the operator acts independently on each column of the image, since the matrix then has a moderate size. It is not feasible in general for arbitrary integral operators due to the large image size. Hence, in our numerical experiments, we simply use a conjugate gradient (CG) algorithm. The precision of the resolution is fixed and the CG algorithm is initialized with the result at the previous iteration. In practice, we observe that 10 iterations are enough for the overall algorithm to converge.

To conclude this paragraph, we illustrate the results obtained by the described procedure in Fig. 4.

4.2 Recovering the density

In this paragraph, we focus on the resolution of problem (15). This amounts to simultaneously correcting the attenuation and denoising the resulting image. This is a rather simple inverse problem, but it seems original due to the noise statistics. A Poisson distributed variable multiplied by a positive constant different from 1 is not Poisson anymore. This makes the proposed algorithm similar, but different from existing approaches developed for Poisson noise in [10, 33] for instance.

A simple idea to regularize the problem is to use the total variation again, i.e. to set

$$R_\beta(\beta) = \lambda_\beta \sum_{i=1}^{n} \|(\nabla\beta)[i]\|_2.$$

Once again the resulting problem can be solved with the SDMM. Let us detail this procedure. Define

$$a[i] = \sum_{j=1}^{2} \exp(-(A_j\alpha)[i]) \text{ and } u[i] = \sum_{j=1}^{2} u_j[i].$$

The problem then reads:

$$\begin{aligned}&\min_{\beta \in \mathbb{R}^n_+} \sum_{i=1}^{n} a[i]\beta[i] - u[i]\log(\beta[i]) + \lambda_\beta \sum_{i=1}^{n} \|(\nabla\beta)[i]\|_2 \\ &= \min_{\beta \in \mathbb{R}^n} f_1(L_1\beta) + f_2(L_2\beta),\end{aligned}$$

with $L_1 = c_1 I_n$,

$$\begin{aligned}f_1 : \mathbb{R}^n &\to \mathbb{R} \cup \{+\infty\} \\ z &\mapsto \iota_{\mathbb{R}^n_+}(z) + \tfrac{1}{c_1}\textstyle\sum_{i=1}^{n} a[i]z[i] - u[i]\log(z[i]),\end{aligned}$$

$L_2 = c_2\nabla$ and

$$\begin{aligned}f_2 : \mathbb{R}^{2n} &\to \mathbb{R} \\ \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} &\mapsto \tfrac{\lambda_\beta}{c_2} \textstyle\sum_{i=1}^{n} \sqrt{z_1[i]^2 + z_2[i]^2}.\end{aligned}$$

The proximal operators of f_2 is standard and we do not detail it here. The proximal operator of f_1 is provided below:

Proposition 5 *We have:*

$$\text{prox}_{\gamma f_1}(z_0) = \frac{-(\gamma/c_1 a - z_0) + \sqrt{(\gamma/c_1 a - z_0)^2 + 4\gamma u}}{2}. \tag{18}$$

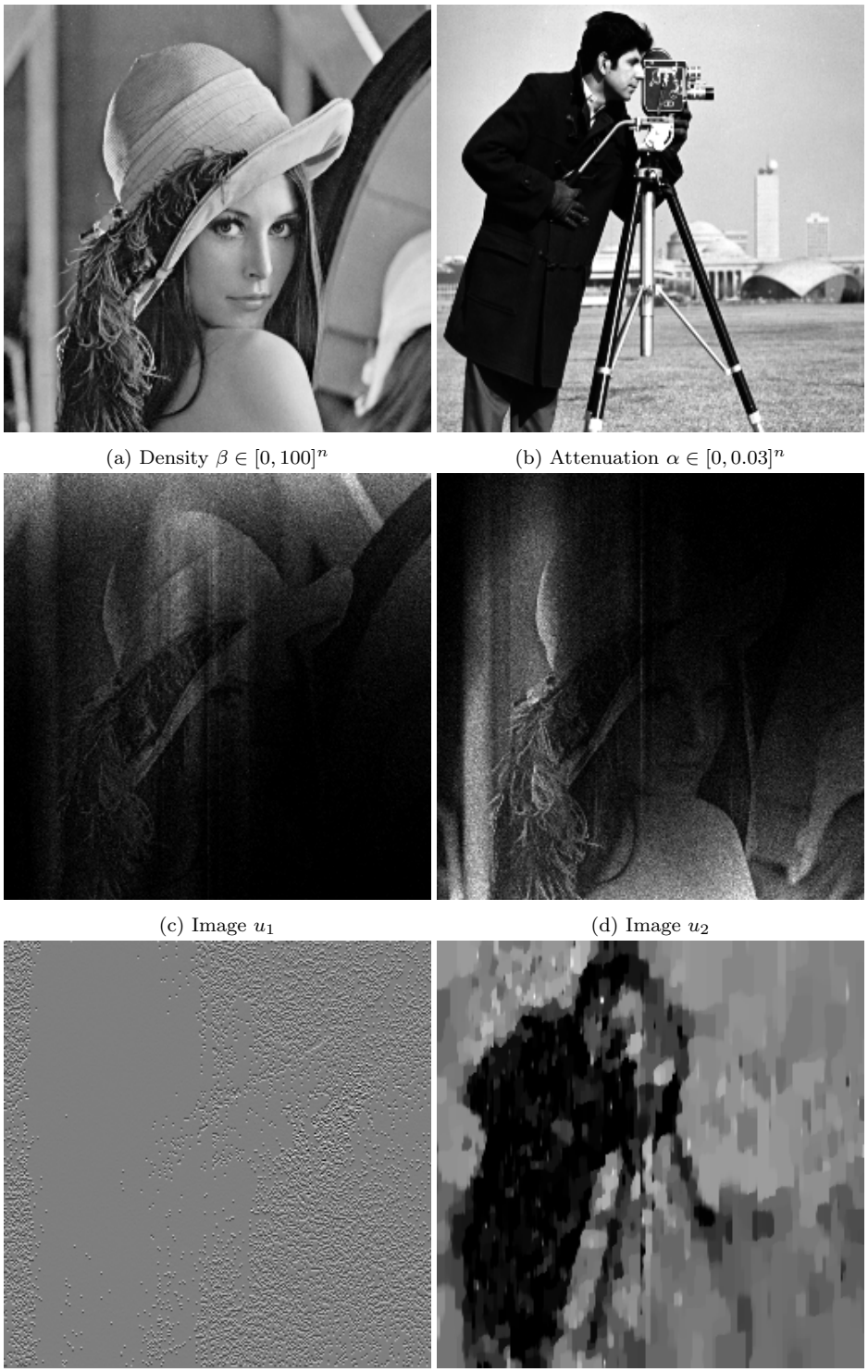
Proof It suffices to write the first order optimality conditions of $\min_{z \geq 0} 1/2\|z - z_0\|_2^2 + a/c_1 z - u\log(z/c_1)$. This shows that z is the root of a second order polynomial. Its only positive root is given in (18).

We show a typical result of total variation minimization in Fig. 5. Pamareter λ_β was chosen manually so as to maximize the SNR of the result.

5 Additional comments

5.1 Parameter selection

Data terms The two data term parameters are λ_α and λ_β. They specify the regularity of the attenuation and the density respectively. In all our experiments, we optimized them by trial and error. We observed experimentally, that similar results are obtained within a relatively large range, making a manual optimization quite easy. In addition, for a given measurement device, the same parameter is likely to be always the same, decreasing the interest of an automatized procedure such as SURE.



(a) Density $\beta \in [0, 100]^n$ (b) Attenuation $\alpha \in [0, 0.03]^n$

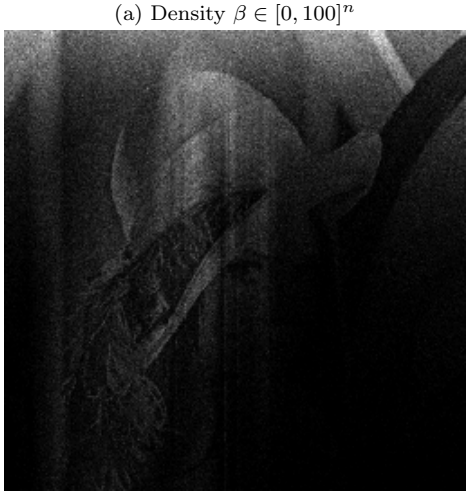
(c) Image u_1 (d) Image u_2

(e) (14): SNR=-53.1dB (f) TV: SNR=11.9dB

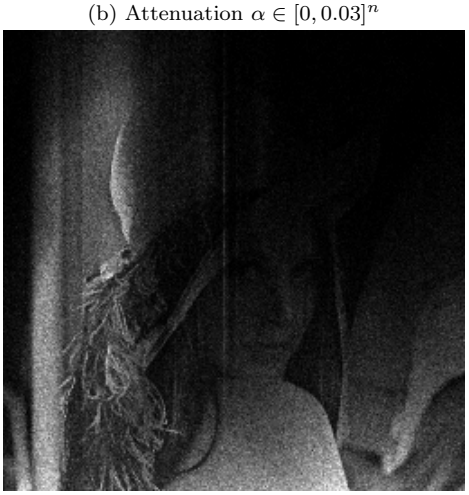
Fig. 4: Illustration of the limits of a direct estimate of the attenuation coefficients and output of a warm start initialization. (4a) and (4b) are the original density and attenuation. (4c) and (4d) are the observed signals. (4e) is the direct density estimate (4). As can be seen, the formula yields useless results since it is completely unstable to noise. (4f) is the density estimate using the total variation solver. It allows recovering the main details of the cameraman, despite a significant amount of noise.

(a) Density $\beta \in [0, 100]^n$

(b) Attenuation $\alpha \in [0, 0.03]^n$



(c) Image u_1



(d) Image u_2

(e) (14): SNR=12.9dB

(f) TV: SNR=21.9dB

Fig. 5: Recovering the density knowing the exact attenuation, with a non regularized estimator or a total variation solver. (5a) and (5b) are the original density and attenuation. (5c) and (5d) are the observed signals. (5e) is the direct density estimate (14). (5f) is the density estimate using a total variation solver.

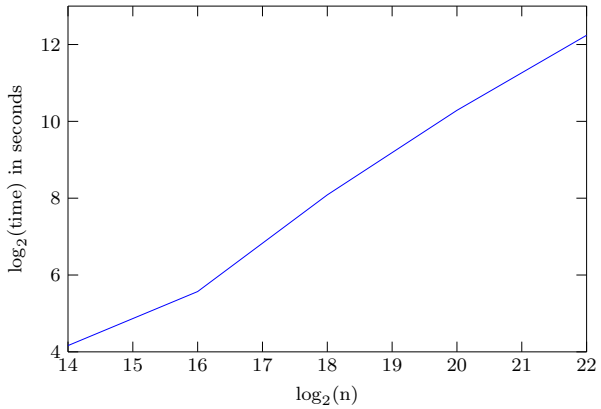


Fig. 6: Time needed to compute the warm start estimate and correct the attenuation with respect to the number n of pixels (in log-log scale). A linear regression indicates that the slope is roughly equal to 1, showing a linear dependency with respect to the number of pixels.

Algorithms parameters The optimization algorithms are based on the SDMM and their convergence rates depend a lot on the paramaters γ, c_1, c_2, c_3 and c_4. They may converge to a satisfactory solution rapidly (about 50 iterations) or slowly (more than 10000 iterations) depending on these choices. Unfortunately, we found no systematic method to choose them and also used a trial and error strategy in our numerical experiments. Our numerical experiments suggest that these parameters are suitable for a wide range of data (image size, maximum attenuation, image dynamics), so that the tedious tuning can be done once and for all for a given application.

5.2 Computing times

All the experiments of the paper were performed on a laptop with an Intel i7 processor with 4-cores. The codes were written mostly in Matlab (natively parallel), with some parts written in C with OpenMP support.

The complexity of the proposed algorithms scale roughly linearly with the number of pixels n, as shown in Fig. 6. We observed that the number of iterations of the SDMM to reach a given relative accuracy remains the same whatever the size n, while the cost per iteration scales linearly with it (at least for the cumulative sum integral operators considered herein).

As can be seen on Fig. 6 the algorithm takes around 48 seconds for a 256×256 image. Out of these, 45 seconds are spent to recover the attenuation, while the 3 remaining are dedicated to correct the density.

All codes can be easily parallelized on a GPU. A speed-up of 100 can be expected on such an architecture, making the proposed methods suitable for large 2D or 3D images.

5.3 Influence of attenuation and signal-to-noise ratio

Two parameters strongly influence the ability to recover the attenuation and density: the signals dynamics (or signal-to-noise ratio) and the attenuation dynamics.

As the signal-to-noise ratio decreases, it becomes impossible to recover fine details. For instance, the fine stripes are not recovered in Fig. 1e, but they are recovered for signals with a much higher amplitude. In Fig. 7, it can indeed be verified that a high dynamics of 10^5 allows recovering most of the stripes. This experiment shows that highly sensitive EMCCD cameras should be preferred over more standard devices for this specific application.

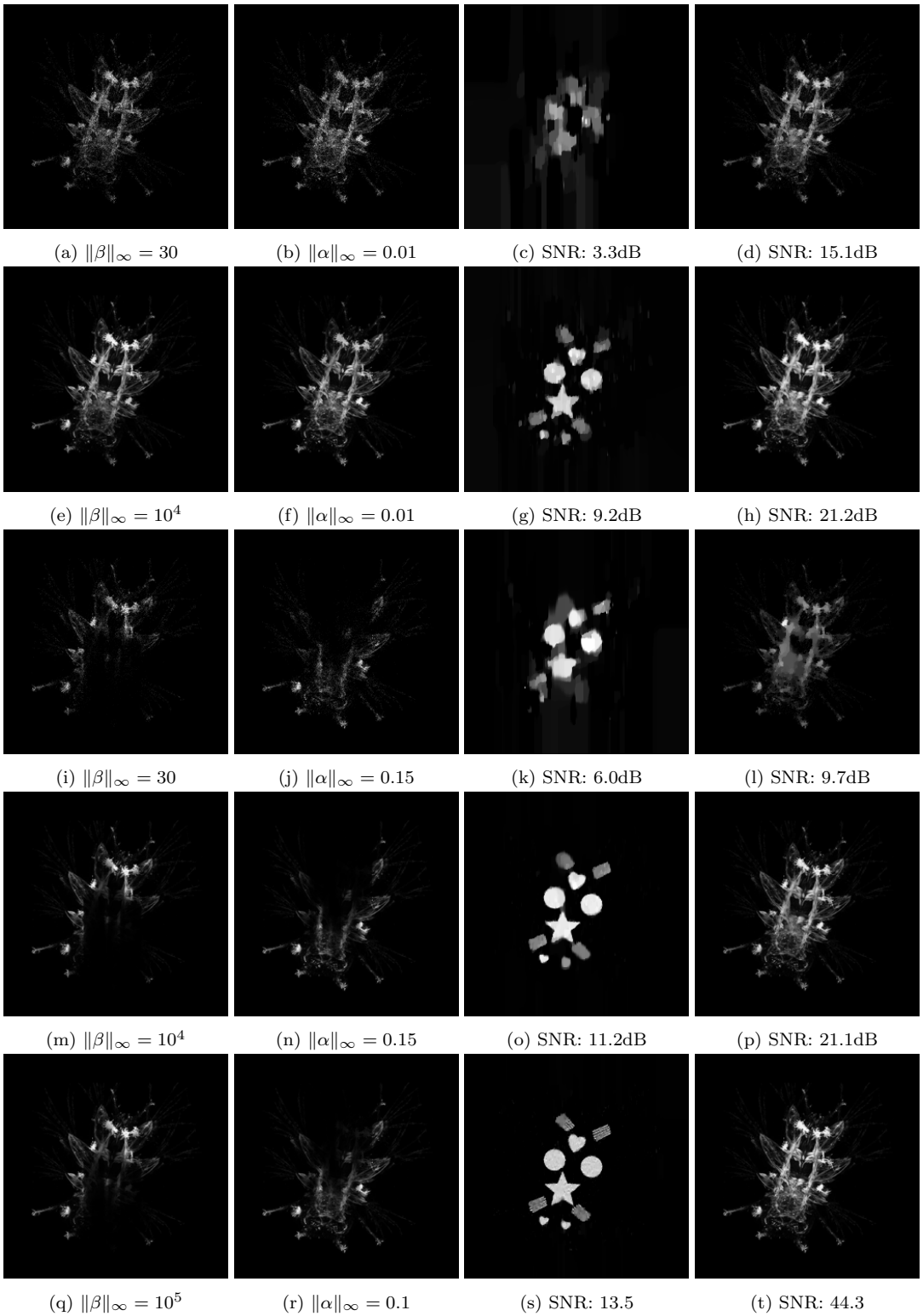
The attenuation amplitude also plays a key role: if it is too low, then no attenuation can be detected. On the contrary, if it is too high, then the signals u_1 and u_2 will vanish too rapidly, making it impossible to evaluate the attenuation. This is illustrated in Fig. 7. It is remarkable that the algorithm manages to recover the attenuation partially for very low signal-to-noise ratio. In Fig. 7c, we observe that the attenuation is partially recovered with no more than 30 expected photons per pixel!

5.4 Toolbox

A Matlab toolbox containing all the main algorithms described in the paper is provided on the website of the authors `https://www.math.univ-toulouse.fr/~weiss/` and on GitHub `https://github.com/pierre-weiss/MAEC`. The Lambert W function and the proximal operator of logsumexp have been implemented with C-mex files with OpenMP support for multicore acceleration. Demonstration scripts are available for testing.

6 Comparison with the approach [30]

In this section, we propose to compare our results with the approach suggested in [30]. In this paper, the authors concentrate on the case of confocal microscopy (i.e. two opposite views). They consider cone integrals to better model the light path. This is also possible with our approach, though we have not explored it yet. Following a maximum a posteriori principle, and using the notation of our paper, they propose to minimize the



(a) $\|\beta\|_\infty = 30$ (b) $\|\alpha\|_\infty = 0.01$ (c) SNR: 3.3dB (d) SNR: 15.1dB

(e) $\|\beta\|_\infty = 10^4$ (f) $\|\alpha\|_\infty = 0.01$ (g) SNR: 9.2dB (h) SNR: 21.2dB

(i) $\|\beta\|_\infty = 30$ (j) $\|\alpha\|_\infty = 0.15$ (k) SNR: 6.0dB (l) SNR: 9.7dB

(m) $\|\beta\|_\infty = 10^4$ (n) $\|\alpha\|_\infty = 0.15$ (o) SNR: 11.2dB (p) SNR: 21.1dB

(q) $\|\beta\|_\infty = 10^5$ (r) $\|\alpha\|_\infty = 0.1$ (s) SNR: 13.5 (t) SNR: 44.3

Fig. 7: Ability to recover the attenuation and density depending on the density and attenuation amplitude.

following functional:

$$\min_{\alpha\in\mathbb{R}^n_+,\beta\in\mathbb{R}^n} \sum_{j=1}^{2} \left\| \frac{u_j - \beta\exp(-A_j\alpha)}{\sqrt{\gamma^2 u_j + \sigma^2}} \right\|_2^2 + R_\alpha(\alpha) + \mu\|\alpha\|_1. \tag{19}$$

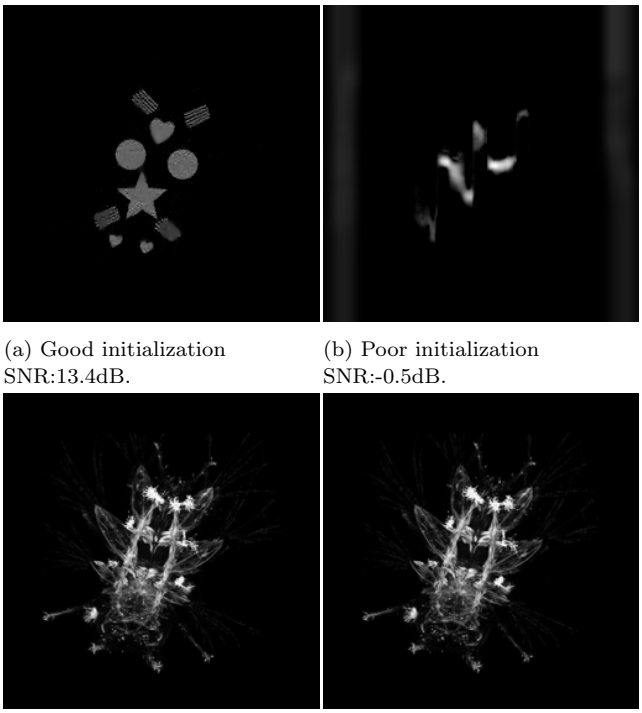
In this equation, $\lambda \geq 0$ and $\mu \geq 0$ are regularization parameters and $\gamma > 0$ and $\sigma > 0$ are parameters allowing to deal with a mixture of Poisson and Gaussian noise (in fact this energy is only an approximation of the MAP). In comparison with our work, the term $\mu\|\alpha\|_1$ is proposed to favor a black background.

The problem (19) is nonconvex and non differentiable. In order to minimize it, the authors propose to use smooth approximations of the functions $R_\alpha(\alpha)$ and $\|\alpha\|_1$, and to use a limited-BFGS-B approach [37], which allows to deal with bound constraints (the nonnegativity assumption). We reimplemented this algorithm with the line integrals used in the previous section and provide comparisons in Fig. 8. Since the problem is nonconvex the result depends on the initialization.

It takes around 30 seconds to run the limited-BFGS-B to solve the problem (19) on 256×256 images, which is on a par with the computing times required by our approach. The main differences with our approach are:

Parameters the energy (19) requires tuning 4 interdependent parameters: γ, σ, μ and the regularization parameter associated to the total variation. In comparison, our algorithm only requires tuning two *independent* parameters λ_α and λ_β, which is arguably much easier. That being said, we observed in our simulations that μ could be safely set to 0. A possible reason might be that the nonnegativity constraint already favors the background to be 0 [2]. In addition, we expect that the parameters might be tuned once and for all for a given device, limiting the importance of this drawback. In our experiments, we discretized the parameter space finely, and kept the parameters leading to the highest SNR. This is of course feasible only when dealing with simulations.

Initialization we observed that the most important difference between the two approaches comes from the initialization. Problem (19) is nonconvex and the iterative descent algorithms may hence converge to different points depending on the starting point (α_0, β_0). In practice, we observed that this was a serious limitation of the approach in [30]. Figure 8 illustrates this point. We reproduce the experiment of Fig. 7, fourth line, corresponding to the favorable case $\|\beta\|_\infty = 10^4$ and $\|\alpha\|_\infty = 0.15$. On the left, we initialize the algorithm with the true attenuation and density maps, i.e. $\alpha_0 = \alpha$ and $\beta_0 = \beta$. The algorithm converges to a satisfactory estimate of the attenuation map. On the right, we initialized the algorithm with $\alpha_0 = \text{mean}(\alpha)$ and $\beta_0 = \frac{1}{2}(u_1 + u_2)$, which looks quite natural, since in practice, nearly no information on the actual solution is available. As can be seen, the L-BFGS-B algorithm yields a very poor estimate of the true attenuation map. In comparison, the convex formulation proposed in this paper will converge to the same minimizer *whatever* the initialization. We believe that this is a distinctive trait and a real strength of the proposed approach.



(a) Good initialization SNR:13.4dB.

(b) Poor initialization SNR:-0.5dB.

(c) Good initialization SNR: 141.2dB.

(d) Poor initialization SNR:2.8dB.

Fig. 8: Example of results obtained with the approach in [30]. Parameters selected to give the highest SNR. Top: estimate of the attenuation map. Bottom: estimate of the density map. The good initialization corresponds to the exact (unknown) maps that we would like to recover. The poor initialization corresponds to a realistic initialization that could be achieved in practice.

7 Conclusion & outlook

We proposed a robust and efficient approach based on a Bayesian estimator to recover attenuation and correct density from multiview measurements. This principle was already known in the field of lidar and solved with simple filtering approaches, while the algorithms

proposed herein are based on a clear and versatile statistical framework. In confocal microscopy, Schmidt et al. [30] recently proposed a Bayesian formulation that shares many similarities with the proposed approach and applied it to real 3D data. The proposed approach differs in two regards: i) we consider a Poisson model for the noise, while [30] only uses an approximation of it and ii) we develop an algorithm that provably converges to the global minimizer of the cost function, while [30] is based on a nonlinear programming approach which leads to different results depending on the initialization. In practice, this distinctive feature seems to be of high importance, since good initial guesses are unavailable in the considered applications.

The approach seems promising for various devices such as lidar or some fluorescence microscopes. It is likely that its scope is much wider and we therefore provide a free Matlab toolbox on GitHub `https://github.com/pierre-weiss/MAEC`.

As a prospective, we plan to confront our algorithms with real data coming from lidar and microscopy. The total variation based algorithm to correct attenuation defects is somewhat disappointing since it is unable to recover fine textures. A promising step would be to use more advanced nonlocal denoisers such as convolutional neural networks. To conclude, let us mention a serious limitation of the proposed approach: it is not so common to find a couple optical system-sample, where attenuation dominates scattering. We do not know at the present time how many applications can reasonably be modeled by equation (7). This question is central to precisely understand the strengths and limits of the proposed approach.

8 Appendices

8.1 Proof Proposition 1

Proof The first item is obtained by direct inspection:

- for β fixed, $\alpha \mapsto \langle \exp(-A_j\alpha), \mathbb{1}\rangle$ is the composition of a linear operator with a convex function, hence it is convex. In addition $\alpha \mapsto \langle A_j\alpha, \mathbb{1}\rangle$ is a linear mapping.
- for α fixed, the first term in β is linear and $\beta \mapsto \langle -\log(\beta), \mathbb{1}\rangle$ is convex.

Let us now focus on the second and third points. The function G can be rewritten as a sum of functions:

$$G(\alpha,\beta) = \sum_{i=1}^{n} g_i\left((A_j\alpha[i])_{1\leq j\leq m}, \beta[i]\right),$$

where $g_i : \mathbb{R}^m \times \mathbb{R}_+ \to \mathbb{R}$ is defined as follows:

$$(x,y) \mapsto g_i(x,y) = \sum_{j=1}^{m} \exp(-x[j])y + u_j[i]\left(x[j] - \log(y)\right).$$

To prove the convexity of G, it suffices to study the convexity of each function g_i. From now on, we skip the indices i to lighten the notation.

Let us analyze the eigenvalues of the Hessian H_g:

$$H_g(x,y) = \begin{pmatrix} \operatorname{diag}(y\exp(-x)) & -\exp(-x) \\ -\exp(-x)^T & \sum_{j=1}^m u_j/y^2 \end{pmatrix}.$$

To study the positive semidefiniteness, let $(v,w) \in \mathbb{R}^{m+1}$, denote an arbitrary vector. We have:

$$\left\langle \begin{pmatrix} v \\ w \end{pmatrix}, H_g(x,y) \begin{pmatrix} v \\ w \end{pmatrix} \right\rangle = \sum_{j=1}^{m} v[j]\exp(-x[j])\left(yv[j] - 2w\right) + w^2 \frac{\sum_{j=1}^m u_j}{y^2}.$$

In the case $y > \frac{\sum_{j=1}^m u_j}{\sum_{j=1}^m \exp(-x[j])}$ and $w \neq 0$, we get that:

$$\begin{aligned}
&\left\langle \begin{pmatrix} v \\ w \end{pmatrix}, H_g(x,y) \begin{pmatrix} v \\ w \end{pmatrix} \right\rangle \\
&< \sum_{j=1}^{m} v[j]\exp(-x[j])\left(yv[j] - 2w\right) + w^2 \frac{\sum_{j=1}^m \exp(-x[j])}{y} \\
&= \sum_{j=1}^{m} \exp(-x[j])\left(yv[j]^2 - 2wv[j] + \frac{w^2}{y}\right) \\
&= \sum_{j=1}^{m} \exp(-x[j])\left(v[j]\sqrt{y} - \frac{w}{\sqrt{y}}\right)^2
\end{aligned}$$

where the last expression is equal to 0 for the particular choice $v[j]y = w$, for all $1 \leq j \leq m$. This implies $\left\langle \begin{pmatrix} v \\ w \end{pmatrix}, H_g \begin{pmatrix} v \\ w \end{pmatrix} \right\rangle < 0$, which shows that function g is not convex in $\mathbb{R}^n \times \mathbb{R}^n_+$ and proves the second item.

The same argument with $y \leq \frac{\sum_{j=1}^m u_j}{\sum_{j=1}^m \exp(-x[j])}$ proves the third item.

8.2 Proof Proposition 2

Proof With this specific choice, it is easy to check that the optimality conditions of problem (10) with respect to variable β yield (14). By replacing this expression in (11), we obtain the optimization problem shown in equation (13).

Checking convexity of this problem can be done by simple inspection. The term $\langle u_j(A_j\alpha), \mathbb{1}\rangle$ is linear, hence convex. The term $\log\left(\sum_{j=1}^m \exp(-(A_j\alpha)[i])\right)$ is the composition of the convex logsumexp function with a linear operator, hence it is convex.

8.3 Proximal operator of logsumexp in dimension 2

In this section, we propose a fast and accurate numerical algorithm based on Newton's method to solve the following problem:

$$\begin{aligned} w &= \operatorname{prox}_{\gamma g_1}(z) \\ &= \operatorname*{argmin}_{x\in\mathbb{R}^{2n}} \frac{1}{2}\|x-z\|_2^2 + \\ &\gamma \sum_{i=1}^{n}\sum_{j=1}^{2} u_j[i]\left[x_j[i] + \log\left(\sum_{j=1}^{2}\exp(-x_j[i])\right)\right], \end{aligned}$$

where $z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$ and $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ are vectors in $\mathbb{R}^{2n}$. This problem may seem innocuous at first sight, but turns out to be quite a numerical challenge. The first observation is that it can be decomposed as n *independent* problems of dimension 2 since:

$$\begin{aligned} w[i] = \operatorname*{argmin}_{(x_1,x_2)\in\mathbb{R}^2} &\frac{1}{2}(x_j - z_j[i])_2^2 \\ &+ \gamma\sum_{j=1}^{2} u_j[i]\left[x_j + \log\left(\sum_{j=1}^{2}\exp(-x_j)\right)\right]. \end{aligned} \tag{20}$$

To simplify the notation, we will skip the index i in what follows. The following proposition shows that our problem is equivalent to finding the proximal operator associated to the "logsumexp" function.

Proposition 6 *Define the logsumexp function* $\operatorname{lse}(x_1,x_2) = \log\left(\sum_{j=1}^2 \exp(x_j)\right)$. *The solution of problem* (20) *coincides with the opposite of the proximal operator of* lse*:*

$$w[i] = -\operatorname*{argmin}_{(x_1,x_2)\in\mathbb{R}^2} a\operatorname{lse}(x_1,x_2) + \frac{1}{2}((x_1-y_1)^2 + (x_2-y_2)^2) \tag{21}$$

$$= -\operatorname{prox}_{a\operatorname{lse}}(y_1,y_2), \tag{22}$$

where $a = \gamma(u_1+u_2)$ *and* $y_j = \gamma u_j - z_j$.

Proof The first order optimality conditions for problem (20) read

$$\begin{cases} \gamma u_1 - \frac{\gamma(u_1+u_2)\exp(-x_1)}{\exp(-x_1)+\exp(-x_2)} + x_1 - z_1 = 0 \\ \gamma u_2 - \frac{\gamma(u_1+u_2)\exp(-x_2)}{\exp(-x_1)+\exp(-x_2)} + x_2 - z_2 = 0. \end{cases} \tag{23}$$

By letting $a = \gamma(u_1+u_2)$ and $y_j = \gamma u_j - z_j$, this equation becomes

$$\begin{cases} -\frac{a\exp(-x_1)}{\exp(-x_1)+\exp(-x_2)} + x_1 + y_1 = 0 \\ -\frac{a\exp(-x_2)}{\exp(-x_1)+\exp(-x_2)} + x_2 + y_2 = 0. \end{cases} \tag{24}$$

It now suffices to make the change of variable $x_i' = -x_i$ to retrieve the optimality conditions of problem (22)

$$\begin{cases} \frac{a\exp(x_1')}{\exp(x_1')+\exp(x_2')} + x_1' - y_1 = 0 \\ \frac{a\exp(x_2')}{\exp(x_1')+\exp(x_2')} + x_2' - y_2 = 0. \end{cases} \tag{25}$$

Remark 4 To the best of our knowledge, this is the first attempt to find a fast algorithm to evaluate the prox of logsumexp. This function is important in many regards. In particular, it is a C^∞ approximation of the maximum value of a vector. In addition, its Fenchel conjugate coincides with the Shannon entropy restricted to the unit simplex. We refer to [14, §3.2] for some details. The algorithm that follows has potential applications outside the scope of this paper.

We now design a fast and accurate minimization algorithm for problem (22) or equivalently, a root finding algorithm for problem (25). This algorithm differs depending on whether $y_1 \geq y_2$ or $y_2 \geq y_1$. We focus on the case $y_1 \geq y_2$. The case $y_2 \geq y_1$ can be handled by symmetry.

Let $\lambda = \frac{\exp(x_1')}{\exp(x_1')+\exp(x_2')}$ and notice that

$$\frac{\exp(x_2')}{\exp(x_1')+\exp(x_2')} = 1-\lambda.$$

Therefore (25) becomes:

$$\begin{cases} x_1' = y_1 - a\lambda \\ x_2' = y_2 - a(1-\lambda). \end{cases} \tag{26}$$

Hence

$$\frac{1-\lambda}{\lambda} = \exp(x_2'-x_1') = \exp(y_2-y_1-a)\exp(2a\lambda). \tag{27}$$

Taking the logarithm on each side yields [1]:

$$\log(1-\lambda) - \log(\lambda) = y_2 - y_1 - a + 2a\lambda. \tag{28}$$

We are now facing the problem of finding the root λ^* of the following function:

$$f(\lambda) = y_2 - y_1 - a + 2a\lambda - \log(1-\lambda) + \log(\lambda). \tag{29}$$

There are two important advantages for this approach compared to the direct resolution of (25). First, we have to solve a 1D problem instead of a 2D problem. More

[1] Applying the logarithm is important for numerical purposes. When $y_2 - y_1 - a$ is very small, the exponential cannot be computed accurately in double precision.

importantly, we directly constrain x' to be of form $x' = y - a\delta$, where δ lives on the 2D simplex.

Let us collect a few properties of function f. First, we have:

$$f'(\lambda) = 2a + \frac{1}{1-\lambda} + \frac{1}{\lambda} > 0, \forall \lambda \in (0,1). \tag{30}$$

Therefore, f is increasing on $(0,1)$. To use convergence results of Newton's algorithm, we need to compute f'' as well:

$$f''(\lambda) = -\frac{1}{\lambda^2} + \frac{1}{(1-\lambda)^2}. \tag{31}$$

Proposition 7 *If $y_1 \geq y_2$, then $x'_1 \geq x'_2$ and*

$$\max\left(\frac{1}{2}, \frac{1}{1+\exp(y_2 - y_1 + a)}\right) \leq \lambda^* \leq \frac{1}{1+\exp(y_2 - y_1)}. \tag{32}$$

Proof The first statement can be proven by contradiction. Assume that $x'_2 > x'_1$, then equation (25) indicates that $y_2 > y_1$.

For the second statement, it suffices to evaluate f at the extremities of the interval since $f' > 0$. We get $f(1/2) = y_2 - y_1 \leq 0$ and $f\left(\frac{1}{1+\exp(y_2-y_1)}\right) = -a + \frac{2a}{1+\exp(y_2-y_1)} \geq 0$.

Proposition 8 *Set $\lambda_0 = \frac{1}{1+\exp(y_2-y_1)}$. Then, the following Newton's method*

$$\lambda_{k+1} = \lambda_k - \frac{f(\lambda_k)}{f'(\lambda_k)} \tag{33}$$

converges to the root λ^* of f, with a locally quadratic rate.

Proof First notice that $f''(\lambda) \geq 0$ on the interval $[1/2, 1)$. Hence f'' is also positive on $I = [\lambda^*, \lambda_0]$. This ensures that

$$\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda^*. \tag{34}$$

We prove this assertion by recurrence. Notice that $\lambda_0 \geq \lambda^*$ by Proposition 7. Now, assume that $\lambda_k \geq \lambda^*$, then

$$f(\lambda_k) = f(\lambda^*) + \int_{\lambda^*}^{\lambda_k} f'(t)\, dt \leq f'(\lambda_k)(\lambda_k - \lambda^*). \tag{35}$$

Hence, $\lambda_k - \lambda^* \geq \frac{f(\lambda_k)}{f'(\lambda_k)}$ and $\lambda_{k+1} \geq \lambda^*$. In addition $\frac{f(\lambda_k)}{f'(\lambda_k)} \geq 0$ on I, so that $\lambda_{k+1} \geq \lambda_k$.

The sequence $(\lambda_k)_{k\in\mathbb{N}}$ is monotonically decreasing and bounded below, therefore it converges to some value $\lambda' \geq \lambda^*$. Necessarily $\lambda' = \lambda^*$, since for $\lambda' > \lambda^*$, $\frac{f(\lambda')}{f'(\lambda')} > 0$.

To prove the locally quadratic convergence rate, we just invoke the celebrated Newton-Kantorovich's theorem [24,25], that ensures local quadratic convergence if f'' is bounded in a neighborhood of the minimizer.

Finally, let us mention that computing λ_0 on a computer is a tricky due to underflow problems: in double precision the command $1 + \exp(y_2 - y_1)$ will return 1 for $y_2 - y_1 < -37 \simeq \log(10^{-16})$. This may cause the algorithm to fail since f and its derivatives are undefined at $\lambda = 1$. In practice we therefore set $\lambda_0 = 1/(1+\exp(y_2 - y_1)) - 10^{-16}$. Similarly, by bound (32), we get $\lambda^* = 1$ up to machine precision whenever $y_2 - y_2 - a < \log(10^{-16})$. Algorithm 2 summarizes all the ideas described in this paragraph.

An attentive reader may have remarked that the convergence of Newton's algorithm depends only on the difference $y(1) - y(2)$ and a. A shift of $y(1)$ and $y(2)$ by the same value does not change Newton's iteration. In Fig. 9, we show that the algorithm behaves very well for a wide range of parameters. For $y(1) - y(2)$ and a varying in the interval $[2^{-10}, 2^{20}]$, the algorithm never requires more than 18 iterations to reach machine precision and needs 2.8 iterations in average.

Algorithm 2 An algorithm to compute $\text{prox}_{a\text{lse}}(y_1, y_2)$ with machine precision

Input: $(y_1, y_2) \in \mathbb{R}^2, a \in \mathbb{R}_+$.
Output: $(x_1, x_2) = \text{prox}_{a\text{lse}(y_1,y_2)}$.
Set $\epsilon = 10^{-16}$.
if $y_1 \geq y_2$ **then**
  if $y_2 - y_1 + a < \log(\epsilon)$ **then**
    Set $\lambda = 1$.
  else
    Set $\lambda = \frac{1}{1+\exp(y_2-y_1)} - \epsilon$.
    Define $d(\lambda) = \frac{y_2 - y_1 - a + 2a\lambda + \log(\lambda/(1-\lambda))}{2a + \frac{1}{\lambda(1-\lambda)}}$.
    while $|d(\lambda)| > \epsilon$ **do**
      Set $\lambda = \lambda - d(\lambda)$.
    end while
  end if
  Set $[x_1, x_2] = [y(1) - a\lambda, y(2) - a(1-\lambda)]$.
else if $y_1 < y_2$ **then**
  if $y_1 - y_2 + a < \log(\epsilon)$ **then**
    Set $\lambda = 1$.
  else
    Set $\lambda = \frac{1}{1+\exp(y_1-y_2)} - \epsilon$.
    Define $d = \frac{y_1 - y_2 - a + 2a\lambda + \log(\lambda/(1-\lambda))}{2a + \frac{1}{\lambda(1-\lambda)}}$.
    while $|d| > \epsilon$ **do**
      Set $\lambda = \lambda - d(\lambda)$.
    end while
  end if
  Set $[x_1, x_2] = [y(1) - a(1-\lambda), y(2) - a\lambda]$.
end if

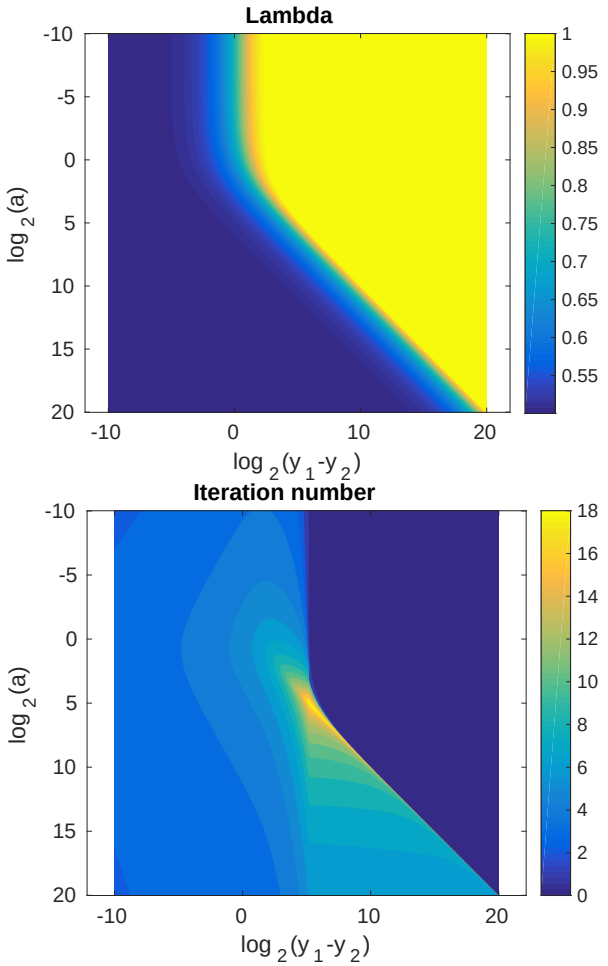


Fig. 9: Performance evaluation for Newton's algorithm. Left: λ^* depending on a and $y_1 - y_2$. Right: number of iterations of Newton's method to reach machine precision.

References

1. Albert Ansmann, Maren Riebesell, and Claus Weitkamp. Measurement of atmospheric aerosol extinction profiles with a Raman lidar. *Optics letters*, 15(13):746–748, 1990.
2. Claire Boyer, Antonin Chambolle, Yohann De Castro, Vincent Duval, Frédéric De Gournay, and Pierre Weiss. On representer theorems and convex regularization. *arXiv preprint arXiv:1806.09810*, 2018.
3. A Can, O Al-Kofahi, S Lasek, DH Szarowski, JN Turner, and B Roysam. Attenuation correction in confocal laser microscopes: a novel two-view approach. *Journal of microscopy*, 211(1):67–79, 2003.
4. Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20(1-2):89–97, 2004.
5. Antonin Chambolle and Thomas Pock. An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319, 2016.
6. Raghav K Chhetri, Fernando Amat, Yinan Wan, Burkhard Höckendorf, William C Lemon, and Philipp J Keller. Whole-animal functional and developmental imaging with isotropic spatial resolution. *Nature methods*, 2015.
7. Patrick L Combettes and Jean-Christophe Pesquet. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer, 2011.
8. Christoph Cremer and Thomas Cremer. Considerations on a laser-scanning-microscope with high resolution and depth of field. *Microscopica acta*, pages 31–44, 1974.
9. Juan Cuesta and Pierre H Flamant. Lidar beams in opposite directions for quality assessment of Cloud-Aerosol Lidar with Orthogonal Polarization spaceborne measurements. *Applied optics*, 49(12):2232–2243, 2010.
10. François-Xavier Dupé, Jalal M Fadili, and Jean-Luc Starck. A proximal iteration for deconvolving Poisson noisy images using sparse representations. *IEEE Transactions on Image Processing*, 18(2):310–321, 2009.
11. Michel Fortin and Roland Glowinski. *Augmented Lagrangian methods: applications to the numerical solution of boundary-value problems*, volume 15. Elsevier, 2000.
12. Sara Garbarino, Alberto Sorrentino, Anna Maria Massone, Alessia Sannino, Antonella Boselli, Xuan Wang, Nicola Spinelli, and Michele Piana. Expectation maximization and the retrieval of the atmospheric extinction coefficients by inversion of Raman lidar data. *Optics Express*, 24(19):21497–21511, 2016.
13. Stefan Hell and Ernst HK Stelzer. Properties of a 4Pi confocal fluorescence microscope. *JOSA A*, 9(12):2159–2166, 1992.
14. Jean-Baptiste Hiriart-Urruty. A note on the Legendre-Fenchel transform of convex composite functions. In *Nonsmooth Mechanics and Analysis*, pages 35–46. Springer, 2006.
15. Herbert G Hughes and Merle R Paulson. Double-ended lidar technique for aerosol studies. *Applied optics*, 27(11):2273–2278, 1988.
16. Jan Huisken and Didier YR Stainier. Even fluorescence excitation by multidirectional selective plane illumination microscopy (mSPIM). *Optics letters*, 32(17):2608–2610, 2007.
17. C Kervrann, D Legland, and L Pardini. Robust incremental compensation of the light attenuation with depth in 3D fluorescence microscopy. *Journal of Microscopy*, 214(3):297–314, 2004.
18. James D Klett. Stable analytical inversion solution for processing lidar returns. *Applied Optics*, 20(2):211–220, 1981.
19. Uros Krzic, Stefan Gunther, Timothy E Saunders, Sebastian J Streichan, and Lars Hufnagel. Multiview light-sheet microscope for rapid in toto imaging. *Nature methods*, 9(7):730–733, 2012.
20. Gerard J Kunz. Bipath method as a way to measure the spatial backscatter and extinction coefficients with lidar. *Applied optics*, 26(5):794–795, 1987.
21. Jürgen Mayer, Alexandre Robert-Moreno, Renzo Danuser, Jens V Stein, James Sharpe, and Jim Swoger. OPTiSPIM: integrating optical projection tomography in light sheet microscopy extends specimen characterization to nonfluorescent contrasts. *Optics letters*, 39(4):1053–1056, 2014.
22. Frank Natterer. *The mathematics of computerized tomography*, volume 32. Siam, 1986.
23. Michael K Ng, Pierre Weiss, and Xiaoming Yuan. Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods. *SIAM journal on Scientific Computing*, 32(5):2710–2736, 2010.
24. James M Ortega. The Newton-Kantorovich theorem. *The American Mathematical Monthly*, 75(6):658–660, 1968.
25. Boris T Polyak. Newton's method and its use in optimization. *European Journal of Operational Research*, 181(3):1086–1096, 2007.
26. Pornsarp Pornsawad, Christine Böckmann, Christoph Ritter, and Mathias Rafler. Ill-posed retrieval of aerosol extinction coefficient profiles from Raman lidar data by regularization. *Applied optics*, 47(10):1649–1661, 2008.

27. Jean Paul Rigaut and Jany Vassy. High-resolution three-dimensional images from confocal scanning laser microscopy. Quantitative study and mathematical correction of the effects from bleaching and fluorescence attenuation in depth. *Analytical and quantitative cytology and histology/the International Academy of Cytology [and] American Society of Cytology*, 13(4):223–232, 1991.
28. JBTM Roerdink and Miente Bakker. An FFT-based method for attenuation correction in fluorescence confocal microscopy. *Journal of microscopy*, 169(1):3–14, 1993.
29. Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.
30. Thorsten Schmidt, Jasmin Dürr, Margret Keuper, Thomas Blein, Klaus Palme, and Olaf Ronneberger. Variational attenuation correction in two-view confocal microscopy. *BMC bioinformatics*, 14(1):366, 2013.
31. James Sharpe, Ulf Ahlgren, Paul Perry, Bill Hill, Allyson Ross, Jacob Hecksher-Sørensen, Richard Baldock, and Duncan Davidson. Optical projection tomography as a tool for 3D microscopy and gene expression studies. *Science*, 296(5567):541–545, 2002.
32. Valery Shcherbakov. Regularized algorithm for Raman lidar data processing. *Applied optics*, 46(22):4879–4889, 2007.
33. Gabriele Steidl and Tanja Teuber. Removing multiplicative noise by Douglas-Rachford splitting methods. *Journal of Mathematical Imaging and Vision*, 36(2):168–184, 2010.
34. Raju Tomer, Khaled Khairy, Fernando Amat, and Philipp J Keller. Quantitative high-speed imaging of entire developing embryos with simultaneous multiview light-sheet microscopy. *Nature methods*, 9(7):755–763, 2012.
35. KA Vermeer, J Mo, JJA Weda, HG Lemij, and JF de Boer. Depth-resolved model-based reconstruction of attenuation coefficients in optical coherence tomography. *Biomedical optics express*, 5(1):322–337, 2014.
36. Claus Weitkamp. *Lidar: range-resolved optical remote sensing of the atmosphere*, volume 102. Springer Science & Business, 2006.
37. Ciyou Zhu, Richard H Byrd, Peihuang Lu, and Jorge Nocedal. Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS)*, 23(4):550–560, 1997.