

Bayesian estimation of turbulent motion

Patrick Héas¹, Cédric Herzet¹, Etienne Mémin¹, Dominique Heitz² and Pablo D. Mininni³

¹Inria, Rennes Bretagne-Atlantique, F-35042 Rennes, France,

²Irstea, UR TERE, F-35044 Rennes, France,

³Universidad de Buenos Aires, Departamento de Física, Argentina

Patrick.Heas@inria.fr

Abstract—Based on physical laws describing the multi-scale structure of turbulent flows, this article proposes a regularizer for fluid motion estimation from an image sequence. Regularization is achieved by imposing some scale invariance property between histograms of motion increments computed at different scales. By reformulating this problem from a Bayesian perspective, an algorithm is proposed to jointly estimate motion, regularization hyper-parameters, and to select the most likely physical prior among a set of models. Hyper-parameter and model inference is conducted by posterior maximization, obtained by marginalizing out non-Gaussian motion variables. The Bayesian estimator is assessed on several image sequences depicting synthetic and real turbulent fluid flows. Results obtained with the proposed approach exceed the state of the art results in fluid flow estimation.

I. INTRODUCTION

The inverse modeling of turbulent motion in images is an important issue in various application areas like meteorology, oceanography, or turbulence studies and experimental fluid mechanics. Motion characterization from experimental images is particularly important since a complete physical theory is still missing for turbulence phenomenology. Optic flow techniques estimating dense velocity fields are interesting in this context since they can describe precisely such complex and continuous flows. For these dense representations, regularization models are commonly used to remove the motion ambiguities and achieve inversion. However, the actual regularizers proposed in the literature are all insufficient since they impose in a small spatial neighborhood a prior smoothness which improperly describes the regularity of turbulent flows [2] [8] [9] [17] [20] [37]. Moreover, going further and assuming a set of proper models is given, choosing the most appropriate one and fixing its hyper-parameters is a tricky problem, which has only been studied for Gaussian models in the optic flow literature.

In this paper, we propose a novel motion estimation procedure dedicated to turbulent flows dealing with such limitations. Our work exploits simultaneously two ideas: *i*) the definition of physically-consistent priors for fluid flows relying on scale invariance properties characterizing turbulent flows; *ii*) the use of Bayesian modeling for robust optic flow estimation.

Our approach for the definition of physically-consistent priors has been driven by the important advances made in the field of statistical modeling of turbulence since the precursor work of *Kolmogorov* in 1941 [21]. *Kolmogorov* hypothesized, using exact properties followed by the fundamental Navier-Stokes equations, that one can expect some universal scaling

properties on the empirical variance of motion increments at different scales. Although limitations to his universality assumption are now known, his theory agrees up to a good degree with observations from experiments and numerical simulations. Therefore, this point of view of turbulence has been the starting point of many recent studies [10]. In this paper, we exploit these results to design a family of physically-sound priors for our motion estimation problem.

We place the problem of model selection within a Bayesian framework. Bayesian analysis has been intensively studied in the past for hyper-parameter estimation and for model selection [19] [28] [30] [33]. However, in optic flow estimation, state-of-the-art inference techniques remained very limited. Indeed, neither proper model selection, nor hyper-parameter estimation [23] with model deviations from Gaussianity, have been considered, except in the very recent work on Bayesian inference for optic flow [12]. Such non-Gaussian models are nevertheless very common in computer vision, for instance to cope with observation outliers due to noise or varying lighting conditions. In this paper, we propose a Bayesian methodology for the selection of proper models in the context of fluid flow estimation from image sequences. The proposed inference is a generalization of preliminary work [13], [14] to non-Gaussian model of unknown hyper-parameters.

The paper is organized as follows. In section III, we present the problem of fluid motion estimation from an image sequence and highlight the paper's contributions. In section IV, we introduce the scale invariance properties derived from the physics of fluids. Then, in section V, we define and solve a constrained motion estimation problem where the motion estimate is forced to possess these scale invariance properties. In section VI, by reformulating the constrained estimation problem from a Bayesian perspective, we show how we can achieve model selection and hyper-parameter estimation. Finally, the experimental evaluation of the Bayesian method is presented in section VIII.

II. NOTATIONS

Except if otherwise specified, the notational conventions adopted in this paper are as follows. Italic lowercase indicates a scalar quantity, as in a ; boldface lowercase indicates a vector quantity, as in \mathbf{a} ; the k th element of vector \mathbf{a} is denoted $\mathbf{a}(k)$; capital boldface letters indicate matrices, as in \mathbf{A} ; the element corresponding to the i th row and j th column of \mathbf{A} is denoted as $\mathbf{A}(i, j)$; the trace (resp. determinant) of \mathbf{A} writes $\text{tr}\mathbf{A}$ (resp. $\det\mathbf{A}$); $\mathcal{R}(\mathbf{A})$ is the range of the columns

of \mathbf{A} ; we will use the notation $\Lambda_{\mathbf{a}}$ to define a diagonal matrix whose elements are those of vector \mathbf{a} ; calligraphic letters, e.g., \mathcal{A} , represent the set of values that a variable or a vector can take on; capital normal letters, as A , denote random variables. The distribution of a random variable A evaluated at a is denoted $p_A(a)$; when clear from the context, we will usually use the short-hand notation: $p_A(a) \triangleq p(a)$. The gradient and the Laplacian operators are respectively defined by $\nabla_{\mathbf{s}}A(\mathbf{s}', t') \triangleq [\frac{\partial A}{\partial s(1)}(\mathbf{s}', t') \dots \frac{\partial A}{\partial s(n)}(\mathbf{s}', t')]^T$ and by $\Delta_{\mathbf{s}}A(\mathbf{s}', t') \triangleq \frac{\partial^2 A}{\partial s(1)^2}(\mathbf{s}', t') + \dots + \frac{\partial^2 A}{\partial s(n)^2}(\mathbf{s}', t')$. The symbol \propto will denote equality up to a multiplicative factor.

III. PROBLEM AND CONTRIBUTIONS

A. Problem formulation

Estimating the motion of a physical system from the observation of a sequence of images is a well-known problem in the computer vision community. At the heart of state-of-the-art motion estimation algorithms is the definition of an *observation* and a *prior* model on the sought velocity field. The observation model relates the motion of the physical system to the spatial and temporal variations of the image intensity. The prior model defines the spatio-temporal constraints that the motion has to satisfy. In a nutshell, state-of-the-art algorithms intend to find the motion field as the “best” compromise between these two models. We provide a formal description of this approach hereafter.

Let $I : (\mathbf{s}, t) \in \mathcal{S} \times \mathcal{T} \rightarrow I(\mathbf{s}, t) \in \mathbb{R}$ be an image intensity function where $\mathcal{S} \subseteq \mathbb{R}^2$ (resp. $\mathcal{T} \subseteq \mathbb{R}$) is the image spatial (resp. temporal) domain. Let the motion of the physical system be defined by a function $v : (\mathbf{s}, t) \in \mathcal{S} \times \mathcal{T} \rightarrow v(\mathbf{s}, t) \in \mathbb{R}^2$ which associates a two-dimensional motion vector to every spatio-temporal position. Many motion estimation algorithms can be regarded as procedures intending to solve a constrained optimization problem of the form:

$$\begin{aligned} \hat{v} &= \arg \min_v f(v, I) \\ &\text{subject to } h(v) = 0. \end{aligned} \quad (1)$$

For example, one standard way to relate $v(\mathbf{s})$ to $I(\mathbf{s}, t)$ is by means of the well-known “*Optical Flow Constraint*” (OFC),

$$g(v, I, \mathbf{s}, t) = \frac{\partial I}{\partial t}(\mathbf{s}, t) + \nabla_{\mathbf{s}}^T I(\mathbf{s}, t) v(\mathbf{s}, t) = 0, \quad \forall \mathbf{s} \in \mathcal{S} \times \mathcal{T} \quad (2)$$

which is valid under intensity conservation along motion trajectories. Quadratic penalization of OFC discrepancies yields [17]:

$$f(v, I) = \int_{\mathcal{T}} \int_{\mathcal{S}} g(v, I, \mathbf{s}, t)^2 ds dt. \quad (3)$$

For other configurations, many other models have been proposed in the literature to relate the image intensity function to the sought motion fields [8], [26]. $h(v)$ defines a set of constraints that have to be satisfied by the sought velocity field. In the standard computer vision literature, dealing with the motion analysis of natural scenes, these constraints are usually chosen to enforce some spatial coherence of the solution, see e.g., [17], [27]. In the problem considered in this paper, we

have the advantage of a perfect knowledge of the equations governing the fluid motion, i.e., the Navier-Stokes equations. Hence, a perfect characterization of the velocity field for incompressible flows should, in principle, solve a problem of the form of (1) with the following constraints:

$$\begin{cases} \nabla_{\mathbf{s}}^T v = 0 \\ \frac{\partial(\rho v)}{\partial t} + (v^T \nabla_{\mathbf{s}})v = -\frac{1}{\rho} \nabla_{\mathbf{s}} p + \nu \Delta_{\mathbf{s}} v + \mathcal{F} \end{cases} \quad (4)$$

where v denotes here an unknown 3-dimensional velocity field, p and ρ are scalar variables representing the 3-dimensional pressure and density fields, \mathcal{F} and ν represent an external force acting on the fluid flow and its cinematic viscosity. Unfortunately, this choice of constraints leads to an optimization problem which is computationally very demanding and often under-determined since p , ρ , \mathcal{F} and the motion component of v perpendicular to the image plane are usually unknown [16]. Some recent works propose to alleviate this problem by using simplified constraints of the Navier-Stokes equations for global [32] or sequential [15] [35] regularization.

Alternatively, in this paper we propose a novel *physically-sound* prior model on the velocity field and derive an efficient algorithm to solve the corresponding optimization problem. Let us remark that the proposed prior can in principle be combined with the former simplified dynamical constraints. In practice, the estimation of the motion field is performed over a *finite* set $\mathcal{S}_r \times \mathcal{T}_r$ of $\mathcal{S} \times \mathcal{T}$. Hereafter, except if otherwise specified, we will consider the scenario where the subset \mathcal{T}_r only contains one elements and, accordingly, we will drop the velocity field temporal index. Moreover, we will assume that \mathcal{S}_r collects the spatial samples on a two-dimensional cartesian grid. We will then use the notation \mathbf{v} to denote the m -dimensional vector made up of the concatenation of $v(\mathbf{s})$'s $\forall \mathbf{s} \in \mathcal{S}_r$, row by row of the cartesian grid. Hence, in the sequel we will consider an optimization problem with N constraints of the form:

$$\begin{aligned} \hat{\mathbf{v}} &= \arg \min_{\mathbf{v}} f(\mathbf{v}, I) \\ &\text{subject to } h_i(\mathbf{v}) = 0, \quad i \in [1, \dots, N]. \end{aligned} \quad (5)$$

The data term, obtained by discretizing (2) on the grid \mathcal{S}_r , has the following quadratic form:

$$f(\mathbf{v}, I) = \|\mathbf{A}\mathbf{v} + \mathbf{b}\|_2^2 \quad (6)$$

where \mathbf{A} and \mathbf{b} represent respectively a rectangle matrix and a vector depending on the image intensity function I .

B. Overview on the proposed solution

Instead of using the original Navier-Stokes equations describing a fluid flow trajectory, we will rather rely on a simple set of constraints $\{h_i(\mathbf{v}), i \in [1, \dots, N]\}$ which only involve the instantaneous flow \mathbf{v} . These constraints rely on theoretical and experimental results obtained as the continuation of Kolmogorov's works at the beginning of the 20th century. They describe some scale invariance properties of turbulent flows. We will see in section IV, that these constraints can be defined in the following quadratic form:

$$h_i(\mathbf{v}) = \mathbf{v}^T \mathbf{A}_i \mathbf{v} - c_i, \quad i \in [1, \dots, N], \quad (7)$$

where \mathbf{A}_i denotes a semi-positive definite matrix and where $c_i \in \mathbb{R}_+$ denotes a scalar parameter. In the sequel, we will consider problem (5) where $f(\mathbf{v}, I)$ and $h(\mathbf{v})$ are defined in (6) and (7). This leads to a particular instance of quadratic programming under quadratic constraints (QPQC) optimization problem. In section V, we will then consider a *dual* optimization algorithm to solve this problem. Since constraints (7) are parametric, we address in section VI-C the problem of estimating parameters c_i , $i \in [1, \dots, N]$. We will deal with this issue by exploiting tools from Bayesian model selection theory.

IV. DEFINITION OF PHYSICALLY CONSISTENT CONSTRAINTS

A. Physical model for structure functions

In turbulence, a quantity of interest is the longitudinal velocity increment function $\delta v_{\parallel} : (\ell, \mathbf{s}, \theta) \in \mathbb{R}_+ \times \mathcal{S} \times [0, 2\pi] \rightarrow \delta v_{\parallel}(\ell, \mathbf{s}, \theta) \in \mathbb{R}$, which is given by:

$$\delta v_{\parallel}(\ell, \mathbf{s}, \theta) \triangleq [v(\mathbf{s} + \ell \mathbf{n}_{\theta}) - v(\mathbf{s})]^T \mathbf{n}_{\theta}, \quad (8)$$

where the scalar ℓ represents a spatial scale and where \mathbf{n}_{θ} denotes the unitary vector in the direction θ . The increment function (8) is thus defined as a velocity difference vector between two points distant of ℓ in the direction of vector \mathbf{n}_{θ} , which is then projected onto vector \mathbf{n}_{θ} . Let us give an example to illustrate the definition. We denote the two components of the discrete velocity field by $\mathbf{v} = (\mathbf{v}_1^T, \mathbf{v}_2^T)^T$, where each component is a m -dimensional vector $\mathbf{v}_i = (v_i(1), \dots, v_i(m))^T$. Then, if we consider that an angle $\theta = 0$ corresponds to a vector $\mathbf{n}_{\theta=0}$ collinear to the velocity field component v_1 , we have:

$$\delta v_{\parallel}(1, \mathbf{s}, 0) = v_1(\mathbf{s}(1) + 1, \mathbf{s}(2)) - v_1(\mathbf{s}) \quad (9)$$

$$\delta v_{\parallel}(1, \mathbf{s}, \pi/2) = v_2(\mathbf{s}(1), \mathbf{s}(2) + 1) - v_2(\mathbf{s}). \quad (10)$$

The spatial average of the velocity increment function raised to some power p , so-called the p^{th} longitudinal structure function, is given by:

$$\overline{\delta v_{\parallel}(\ell)^p} = \frac{1}{2\pi|\mathcal{S}|} \int_{\mathcal{S}} \int_{[0, 2\pi]} \delta v_{\parallel}(\ell, \mathbf{s}, \theta)^p d\theta ds, \quad (11)$$

where $|\mathcal{S}|$ represents the surface of spatial domain \mathcal{S} . Since a finite grid \mathcal{S}_r with $|\mathcal{S}_r|$ nodes is considered, we need to approach the spatial integral by a sum over the finite subset \mathcal{S}_r , while the integral with respect to θ can be approximated by a sum over the finite angular subset $\mathcal{A}_r \subset [0, 2\pi]$ of $|\mathcal{A}_r|$ elements:

$$\overline{\delta v_{\parallel}(\ell)^p} \approx \frac{1}{|\mathcal{A}_r||\mathcal{S}_r|} \sum_{\mathbf{s} \in \mathcal{S}_r} \sum_{\theta \in \mathcal{A}_r} |\delta v_{\parallel}(\ell, \mathbf{s}, \theta)|^p. \quad (12)$$

For the particular case $p=2$, the longitudinal structure function is proportional to the two-point cross-correlation function of the velocity, and therefore related (through the convolution theorem) with the power spectrum of the velocity field. Both the structure functions and the power spectrum are often used in experimental and numerical studies to quantify scaling properties of turbulent flows. Moreover, in fluid mechanics,

several exact relations for the third-order longitudinal structure function ($p = 3$) have been demonstrated for turbulent flows [31]. In particular, for three-dimensional isotropic and homogeneous flows, Kolmogorov [21] demonstrated from the Navier-Stokes equations that the third-order moment is linear with respect to scale, *i.e.*, $\overline{\delta v_{\parallel}(\ell)^3} \propto \ell$. This linear relation theoretically arises in the so-called *inertial* scale range of the flow, that is to say from the smallest scale of turbulence, also known as Kolmogorov scale, to the scale of the energy-containing vortices of the flow. Analogously, for bidimensional turbulence, Kraichnan [22] showed that there exists at small scales a phenomenology implying that $\overline{\delta v_{\parallel}(\ell)^3} \propto \ell^3$. More generally, for any turbulent flow, the p^{th} -order structure function can be described in some scale range $\mathcal{I} \subset \mathbb{R}_+$, so-called inertial range, by a power-law:

$$\overline{\delta v_{\parallel}(\ell)^p} = \gamma_p \ell^{\zeta_p}, \quad \ell \in \mathcal{I}, \quad (13)$$

where γ_p and ζ_p denote the prefactor and the exponent of the power-law.

B. Constraints definition

We now propose constraints for the motion estimation problem (5), derived from (13). We choose to consider the second-order structure function $\overline{\delta v_{\parallel}(\ell)^2}$, *i.e.*, our set of constraints will be derived from the relation:

$$\overline{\delta v_{\parallel}(\ell)^2} - \gamma_2 \ell^{\zeta_2} = 0, \quad \ell \in \mathcal{I}, \quad (14)$$

for ℓ sufficiently small. In practice, only a finite number of longitudinal velocity increments can be computed from \mathbf{v} since the latter vector collects the values of the velocity field $v(\mathbf{s})$ on a *discrete* spatial grid \mathcal{S}_r . It can then easily be observed that (14) can be rewritten in the following matrix form:

$$\mathbf{v}^T \mathbf{D}_{\ell}^T \mathbf{D}_{\ell} \mathbf{v} - \gamma_2 \ell^{\zeta_2} = 0, \quad \ell \in \mathcal{I}, \quad (15)$$

for some ℓ compatible with the considered spatial grid. For example, in the case of $\ell = 1$, a 2×2 spatial grid \mathcal{S}_r and periodic boundary conditions, using the subset $\mathcal{A}_r = \{0, \frac{\pi}{2}\}$, *i.e.*, motion increments defined in (9)-(10), we obtain:

$$\overline{\delta v_{\parallel}(1)^2} - \gamma_2 1^{\zeta_2} \approx \frac{1}{|\mathcal{A}_r||\mathcal{S}_r|} \sum_{\mathcal{S}_r} \sum_{\theta \in \mathcal{A}_r} |\delta v_{\parallel}(1, \mathbf{s}, \theta)|^2 \quad (16)$$

$$= \mathbf{v}^T \mathbf{D}_1^T \mathbf{D}_1 \mathbf{v} - \gamma_2 \quad (17)$$

with

$$\mathbf{D}_1 = \frac{1}{\sqrt{8}} \begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

For experimental evaluation, in section VIII, we will choose the following discretization subsets: $\mathcal{A}_r = \{-\pi, -\frac{\pi}{2}, 0, \frac{\pi}{2}\}$ for $\ell \in \mathbb{N}_+ \cap \mathcal{I}$ and $\mathcal{A}_r = \{-\frac{3\pi}{4}, -\frac{\pi}{4}, \frac{\pi}{4}, \frac{3\pi}{4}\}$ for $\ell \in \sqrt{2}\mathbb{N}_+ \cap \mathcal{I}$. Obviously, choosing larger subsets \mathcal{S}_r and \mathcal{A}_r would result in more accurate approximations of integral (11).

Before concluding this section, let us point out that parameters γ_2 , ζ_2 are usually unknown. In practice, these parameters should therefore be estimated. We will however postpone this problem to section VI-C. In the next section, we will assume that γ_2 , ζ_2 are perfectly known and will address the problem of motion estimation under constraints (15).

V. FLUID FLOW ESTIMATION UNDER PHYSICAL CONSTRAINTS

A. A dual approach for constrained motion estimation

In this section, we address the constrained motion estimation problem:

$$\begin{aligned} \mathbf{v}^* &= \arg \min_{\mathbf{v}} f(\mathbf{v}, I) \\ &\text{subject to } \mathbf{v}^T \mathbf{D}_\ell^T \mathbf{D}_\ell \mathbf{v} - \gamma_2 \ell^{\zeta_2} = 0, \quad \forall \ell \in \mathcal{I}, \end{aligned} \quad (18)$$

where \mathcal{I} denotes the set of constrained scales. In this section, we will particularize our development to the case of a quadratic objective function, *i.e.*,

$$f(\mathbf{v}, I) = \|\mathbf{A}\mathbf{v} + \mathbf{b}\|_2^2. \quad (19)$$

Problem (18) thus belong to the general family of QPQC optimization problems. The extension to the more general case where $f(\mathbf{v}, I)$ has the structure of an ‘‘M-estimator’’ will be discussed in section V-B.

We consider a *dual* optimization algorithm, where the solution of (18) is searched as the minimum of a Lagrangian function:

$$L(\mathbf{v}, \boldsymbol{\lambda}) = \|\mathbf{A}\mathbf{v} + \mathbf{b}\|_2^2 + \sum_{\ell} \lambda_{\ell} (\mathbf{v}^T \mathbf{D}_\ell^T \mathbf{D}_\ell \mathbf{v} - \gamma_2 \ell^{\zeta_2}).$$

The equivalence between the points minimizing the Lagrangian, *i.e.*,

$$\hat{\mathbf{v}}_{\boldsymbol{\lambda}} = \arg \min_{\mathbf{v}} L(\mathbf{v}, \boldsymbol{\lambda}), \quad (20)$$

and the solution of (18) is not straightforward and requires additional conditions [5]. In particular, $\hat{\mathbf{v}}_{\boldsymbol{\lambda}^*}$ is a solution of (18) if

$$\hat{\mathbf{v}}_{\boldsymbol{\lambda}^*}^T \mathbf{D}_\ell^T \mathbf{D}_\ell \hat{\mathbf{v}}_{\boldsymbol{\lambda}^*} - \gamma_2 \ell^{\zeta_2} = 0, \quad \forall \ell \in \mathcal{I}, \quad (21)$$

$$\boldsymbol{\lambda}^* = \arg \max_{\boldsymbol{\lambda} \in \mathcal{D}} q(\boldsymbol{\lambda}), \quad (22)$$

where $q(\boldsymbol{\lambda}) \triangleq \inf_{\mathbf{v}} L(\mathbf{v}, \boldsymbol{\lambda})$ is the *dual function* of (18) and $\mathcal{D} \triangleq \{\boldsymbol{\lambda} | q(\boldsymbol{\lambda}) > -\infty\}$ is its domain.

The optimization procedure considered in this paper is then as follows:

- 1) Compute $\boldsymbol{\lambda}^*$ as the solution of (22),
- 2) Evaluate $\hat{\mathbf{v}}_{\boldsymbol{\lambda}^*}$ from (20),
- 3) Check that $\hat{\mathbf{v}}_{\boldsymbol{\lambda}^*}$ satisfies conditions (21).

We discuss hereafter each step of the procedure:

1) *Resolution of the dual problem* (22): The dual function $q(\boldsymbol{\lambda})$ is concave and its domain is convex [5]. Moreover, it can readily be shown that $q(\boldsymbol{\lambda})$ is differentiable in the interior of its domain. We can thus access the global maximum of $q(\boldsymbol{\lambda})$ with a gradient ascent method¹:

$$\boldsymbol{\lambda}^{(n+1)} = \boldsymbol{\lambda}^{(n)} + \alpha^{(n)} \nabla_{\boldsymbol{\lambda}} q(\boldsymbol{\lambda}^{(n)}), \quad (23)$$

where $\alpha^{(n)}$ is a ‘‘properly-chosen’’ step factor and the components of the gradient write

$$\frac{\partial q(\boldsymbol{\lambda})}{\partial \lambda_{\ell}} = -\hat{\mathbf{v}}_{\boldsymbol{\lambda}}^T \mathbf{D}_\ell^T \mathbf{D}_\ell \hat{\mathbf{v}}_{\boldsymbol{\lambda}} + \gamma_2 \ell^{\zeta_2}. \quad (24)$$

¹When $\boldsymbol{\lambda}^{(n)}$ is a boundary point of \mathcal{D} , the *subgradient* of $q(\boldsymbol{\lambda})$ should in principle be considered. However, we avoid such details hereafter for the sake of keeping the discussion as simple as possible.

Note that a global maximizer of $q(\boldsymbol{\lambda})$ may not exist, *i.e.*, there may be no $\boldsymbol{\lambda}^*$ such that

$$q(\boldsymbol{\lambda}^*) = \sup_{\boldsymbol{\lambda}} q(\boldsymbol{\lambda}). \quad (25)$$

This scenario may for example occur when \mathcal{D} is an open or unbounded set². We have therefore to distinguish between several cases.

If a global maximizer $\boldsymbol{\lambda}^*$ exists and is an interior point of \mathcal{D} , the (unique) fixed point of the gradient-ascent procedure corresponds to the global maximizer $\boldsymbol{\lambda}^*$ and satisfies

$$\nabla_{\boldsymbol{\lambda}} q(\boldsymbol{\lambda}^*) = \mathbf{0}. \quad (26)$$

If $\boldsymbol{\lambda}^*$ exists but is not an interior point of \mathcal{D} , a gradient procedure can still access the maximum but the latter does not necessarily satisfy (26).

Finally, if a global maximizer does not exist, the dual optimization procedure (20) fails in solving problem (18) since (22) can not be satisfied. We can however find an approximation of the solution. To do so, we proceed exactly in the same way as if a global maximizer exists but return instead the value of $\boldsymbol{\lambda}^{(n)}$ achieved when $\nabla_{\boldsymbol{\lambda}} q(\boldsymbol{\lambda}^{(n)})$ drops below a given threshold. With a slight abuse of notation, we also denote the latter value as $\boldsymbol{\lambda}^*$ in the sequel. We will discuss the quality of the resulting approximation below.

Note that resorting to an approximation of the global maximizer does not lead to any particular problem for the rest of the optimization procedure. In particular, as will be shown shortly, a solution to (20) exists for any $\boldsymbol{\lambda}^{(n)}$ provided that $\boldsymbol{\lambda}^{(0)} \in \mathcal{D}$.

2) *Resolution of the dual problem* (20): First, let us emphasize that any feasible $\boldsymbol{\lambda}$ (*i.e.*, $\boldsymbol{\lambda} \in \mathcal{D}$) has to satisfy

$$\begin{cases} \mathbf{A}^T \mathbf{A} + \sum_{\ell} \lambda_{\ell} \mathbf{D}_\ell^T \mathbf{D}_\ell \succeq 0, \\ \mathbf{A} \mathbf{b} \in \mathcal{R}(\mathbf{A}^T \mathbf{A} + \sum_{\ell} \lambda_{\ell} \mathbf{D}_\ell^T \mathbf{D}_\ell), \end{cases} \quad (27)$$

where \succeq denotes the fact the matrix is semi-positive definite. Indeed, if (27) does not hold for some $\boldsymbol{\lambda}$, it can be seen that

$$q(\boldsymbol{\lambda}) \triangleq \inf_{\mathbf{v}} L(\mathbf{v}, \boldsymbol{\lambda}) = -\infty, \quad (28)$$

so that $\boldsymbol{\lambda} \notin \mathcal{D}$.

From this observation, it follows that $\forall \boldsymbol{\lambda} \in \mathcal{D}$, problem (20) has an analytic solution, namely:

$$\hat{\mathbf{v}}_{\boldsymbol{\lambda}} = - \left(\mathbf{A}^T \mathbf{A} + \sum_{\ell} \lambda_{\ell} \mathbf{D}_\ell^T \mathbf{D}_\ell \right)^{\dagger} \mathbf{A} \mathbf{b}, \quad (29)$$

where \dagger denotes the Moore-Penrose pseudo inverse of the matrix.

Coming back to the gradient procedure (23), we can note that, if $\boldsymbol{\lambda}^{(0)} \in \mathcal{D}$, then $\boldsymbol{\lambda}^{(n)} \in \mathcal{D} \forall n$ since the gradient algorithm (23) can only increase the value of $q(\boldsymbol{\lambda})$ throughout the iterations. The solution of (20) is therefore well-defined for any value of $\boldsymbol{\lambda}$ generated by the gradient algorithm (23).

²It is however not possible to predict the existence of a solution for general QPQC optimization problems.

3) *Constraint satisfaction* (21): Let us first consider the case where a global maximizer of $q(\boldsymbol{\lambda})$ exists and is an interior point of \mathcal{D} . Then, the unique fixed point of the gradient algorithm must satisfy (26). Hence, taking the expression of the gradient elements (24) into account, we have that (21) is necessarily satisfied.

In the general case (*i.e.*, if a global maximizer of $q(\boldsymbol{\lambda})$ does not exist or is not an interior point), $\mathbf{v}_{\boldsymbol{\lambda}^*}$ is not necessarily ensured to satisfy constraints (21). If needed, it is possible to compute a feasible solution and assess its quality as follows. We first project $\mathbf{v}_{\boldsymbol{\lambda}^*}$ onto the set of feasible points and obtain a feasible solution $\mathbf{v}_{\boldsymbol{\lambda}^*}^\perp$. We then obtain the following bounds on the optimal solution $f(\mathbf{v}^*)$,

$$f(\mathbf{v}_{\boldsymbol{\lambda}^*}^\perp) \stackrel{(a)}{\geq} f(\mathbf{v}^*) \stackrel{(b)}{\geq} \sup_{\boldsymbol{\lambda}} q(\boldsymbol{\lambda}) \stackrel{(c)}{\geq} q(\boldsymbol{\lambda}^*), \quad (30)$$

by using the following observations:

- (a) follows from the fact that \mathbf{v}^* is a global minimizer of f over all feasible points,
- (b) is a consequence of the Weak Duality Theorem [5],
- (c) follows from the definition of the supremum.

The left-hand (resp. right-hand) side of (30) provides an upper (resp. lower) bound on the value of $f(\mathbf{v}^*)$. In other words, knowing $\mathbf{v}_{\boldsymbol{\lambda}^*}^\perp$ and $\boldsymbol{\lambda}^*$, we can derive an objective measure of the quality of the proposed solution. Note, in particular, that if $\boldsymbol{\lambda}^*$ is an interior point of \mathcal{D} , all the inequalities in (30) become equalities.

B. Extension to “M-estimators”

We now extend the procedure described in section V-A to the case where the objective function f is an “M-estimator”:

$$f(\mathbf{v}, \mathbf{I}) = \rho(\mathbf{A}\mathbf{v} + \mathbf{b}). \quad (31)$$

This type of function is broadly used in the computer vision literature and enjoys several desirable properties. In particular, it can be shown [11] that any M-estimator ρ can be expressed as the solution of a convex optimization problem, *i.e.*,

$$\rho(\mathbf{y}) = \min_{\mathbf{z}} \mathbf{y}^T \Lambda_{\mathbf{z}} \mathbf{y} + \psi(\mathbf{z}), \quad (32)$$

where $\psi(\mathbf{z})$ is the Fenchel-Legendre dual function of ρ .

Thus, problem (18) can be rewritten as

$$\begin{aligned} (\mathbf{v}^*, \mathbf{z}^*) &= \arg \min_{\mathbf{v}, \mathbf{z}} \{ (\mathbf{A}\mathbf{v} + \mathbf{b})^T \Lambda_{\mathbf{z}} (\mathbf{A}\mathbf{v} + \mathbf{b}) + \psi(\mathbf{z}) \} \\ &\text{subject to } \mathbf{v}^T \mathbf{D}_\ell^T \mathbf{D}_\ell \mathbf{v} - \gamma_2 \ell^{\zeta_2} = 0, \quad \forall \ell \in \mathcal{I}, \end{aligned} \quad (33)$$

and a block-coordinate minimization procedure can be used to access to the solution, *i.e.*,

$$\mathbf{z}^{(n+1)} = \arg \min_{\mathbf{z}} \{ (\mathbf{A}\mathbf{v}^{(n)} + \mathbf{b})^T \Lambda_{\mathbf{z}} (\mathbf{A}\mathbf{v}^{(n)} + \mathbf{b}) + \psi(\mathbf{z}) \} \quad (34)$$

$$\begin{aligned} \mathbf{v}^{(n+1)} &= \arg \min_{\mathbf{v}} \{ (\mathbf{A}\mathbf{v} + \mathbf{b})^T \Lambda_{\mathbf{z}^{(n)}} (\mathbf{A}\mathbf{v} + \mathbf{b}) \} \\ &\text{subject to } \mathbf{v}^T \mathbf{D}_\ell^T \mathbf{D}_\ell \mathbf{v} - \gamma_2 \ell^{\zeta_2} = 0, \quad \forall \ell \in \mathcal{I}, \end{aligned} \quad (35)$$

Problem (34) has the following analytic solution [11]

$$\mathbf{z}^{(n+1)} = \frac{1}{2} \Lambda_{\mathbf{A}\mathbf{v}^{(n)} + \mathbf{b}}^{-1} \nabla \rho(\mathbf{A}\mathbf{v}^{(n)} + \mathbf{b}), \quad (36)$$

which is typically easy to evaluate. Problem (35) is a quadratic optimization problem with quadratic equality constraints.

<p>From an initial point $\mathbf{z}^{(i=0)}$, iterate until convergence:</p> <ul style="list-style-type: none"> • Reach the fixed point $\boldsymbol{\lambda}^*$ solution of (22) by gradient ascent; <i>i.e.</i>, from an initial point $\boldsymbol{\lambda}^{(n=0)} \in \mathcal{D}$, iterate until convergence: <ul style="list-style-type: none"> – gradient computation (24) with $\mathbf{v}_{\boldsymbol{\lambda}^{(n)}}^{(i)}$ given by (29) at $\mathbf{z}^{(i)}$ – gradient ascent step (23): $\boldsymbol{\lambda}^{(n+1)} = \boldsymbol{\lambda}^{(n)} + \alpha^{(n)} \nabla_{\boldsymbol{\lambda}} q(\boldsymbol{\lambda}^{(n)})$ • Compute $\mathbf{v}_{\boldsymbol{\lambda}^*}^{(i)}$ using (29) • Compute $\mathbf{z}^{(i+1)}$ using $\mathbf{v}_{\boldsymbol{\lambda}^*}^{(i)}$ in (36)

TABLE I

FLOW CHART OF THE OPTIMIZATION ALGORITHM

Hence, a solution can be searched by applying the procedure described in section V-A.

A flow chart of the overall optimization procedure is summarized in Table I

VI. A BAYESIAN FRAMEWORK FOR SELECTING PHYSICAL CONSTRAINTS

In this section, we address the problem of estimating the parameters γ_2, ζ_2 of model (14). Estimation will be addressed by exploiting Bayesian model selection theory.

A. Bayesian estimation theory

Let us first recall the basics of Bayesian estimation theory [33]. The latter provides nice answers to the general problematic of inference of the value of some vector $\boldsymbol{\theta}_0$ from some observations \mathbf{b} , which may be noisy and incomplete. To that aim, this theory makes the fundamental assumption that the observations, but also the unknown parameter $\boldsymbol{\theta}_0$, have a probabilistic nature, *i.e.*, are described by inter-dependent random variables. The Bayesian approach then builds an estimator exploiting the knowledge of the conditional law $p(\boldsymbol{\theta}_0|\mathbf{b})$, so-called *a posteriori* probability distribution. The estimator noted $\hat{\boldsymbol{\theta}}_0$ is designed according to the optimization of some criteria. Common Bayesian estimators are for example the conditional-mean estimator: $\hat{\boldsymbol{\theta}}_0 = \int \boldsymbol{\theta}_0 p(\boldsymbol{\theta}_0|\mathbf{b}) d\boldsymbol{\theta}_0$; or the well-known Maximum A Posteriori (MAP) estimator: $\hat{\boldsymbol{\theta}}_0 = \arg \max_{\boldsymbol{\theta}_0} p(\boldsymbol{\theta}_0|\mathbf{b})$. However, in practice, the observations \mathbf{b} rarely only depend on $\boldsymbol{\theta}_0$ but often also on a set of other variables $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots\}$. Let us restrict ourselves to the set of extra variables $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$. By making analogous assumptions on their probabilistic nature, it is possible to design the joint posterior probability distribution $p(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{b})$. Thus, the construction of the posterior $p(\boldsymbol{\theta}_0|\mathbf{b})$ now requires to integrate the joint probability over $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$:

$$p(\boldsymbol{\theta}_0|\mathbf{b}) = \int \int p(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{b}) d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2. \quad (37)$$

The random variables $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ often live in high dimensional spaces. Consequently, the computation of the latter integral may be quite cumbersome. Trackable alternatives consist in relying on:

- 1) the joint MAP solution of the problem:

$$\arg \max_{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2} p(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{b}), \quad (38)$$

- 2) the evidence framework [28] which solves the following

maximization problem:

$$\arg \max_{\theta_0} p(\theta_0 | \theta_1^{\text{map}}, \theta_2^{\text{map}}, \mathbf{b}), \quad (39)$$

where

$$\begin{cases} \theta_1^{\text{map}} = \arg \max_{\theta_1} p(\theta_1 | \theta_2^{\text{map}}, \mathbf{b}) \\ \theta_2^{\text{map}} = \arg \max_{\theta_2} p(\theta_2 | \mathbf{b}). \end{cases} \quad (40)$$

It is well known that the joint MAP criterion (38) yields an estimate of θ_0 overfitting the observations [33]. The evidence framework usually provides a better approximation of the marginalized MAP estimate of θ_0 , *i.e.*, the maximizer of (37). In particular, for Gaussian variables, solving (39) provides the exact marginalized MAP estimate [33]. However, the two maximization problems in (40) may remain potentially difficult due to high-dimensional non-Gaussian integrals. Nevertheless, as we shall see, there exist trackable algorithms yielding relevant approximated solutions for the evidence framework. Thus, in the following, we will stick to this estimation framework.

B. Probabilistic model and inference problem

Now, coming back to our specific problem, we consider a joint probability distribution on the following random variables: image observations \mathbf{b} , the motion field variable \mathbf{v} and parameters \mathbf{z} , β , γ_2 and ζ_2 . The purpose of the introduction of a new parameter β will be clarified latter. We propose to use the evidence framework (39) in order to solve a problem of the form:

$$\arg \max_{\mathbf{v}, \mathbf{z}} p(\mathbf{v}, \mathbf{z} | \beta^{\text{map}}, \gamma_2^{\text{map}}, \zeta_2^{\text{map}}, \mathbf{b}), \quad (41)$$

where β^{map} , γ_2^{map} and ζ_2^{map} are the marginalized MAP estimates given by:

$$\beta^{\text{map}} = \arg \max_{\beta} \iint p(\mathbf{v}, \mathbf{z}, \beta | \gamma_2^{\text{map}}, \zeta_2^{\text{map}}, \mathbf{b}) d\mathbf{v} d\mathbf{z}, \quad (42)$$

$$(\gamma_2^{\text{map}}, \zeta_2^{\text{map}}) = \arg \max_{\gamma_2, \zeta_2} \iiint p(\mathbf{v}, \mathbf{z}, \beta, \gamma_2, \zeta_2 | \mathbf{b}) d\mathbf{v} d\mathbf{z} d\beta. \quad (43)$$

Clearly, system (41)-(43) corresponds to the evidence framework defined in (39)-(40) with $\theta_0 = \{\mathbf{v}, \mathbf{z}\}$, $\theta_1 = \{\beta\}$ and $\theta_2 = \{\gamma_2, \zeta_2\}$. The practical implementation (41)-(43) requires the definition of an explicit joint probability distribution. To this aim, we propose the following factorization:

$$p(\mathbf{b}, \mathbf{v}, \mathbf{z}, \beta, \gamma_2, \zeta_2) = p(\mathbf{b} | \mathbf{v}, \mathbf{z}, \beta) p(\mathbf{v} | \gamma_2, \zeta_2) p(\mathbf{z} | \beta) p(\gamma_2, \zeta_2, \beta) \quad (44)$$

and we define the following probabilistic models:

$$p(\mathbf{b} | \mathbf{z}, \mathbf{v}, \beta) \propto \exp \left\{ -\frac{(\mathbf{b} + \mathbf{A}\mathbf{v})^T \Lambda_{\mathbf{z}} (\mathbf{b} + \mathbf{A}\mathbf{v})}{2\beta^{-1}} \right\}, \quad (45)$$

$$p(\mathbf{z} | \beta) \propto \exp \left\{ -\frac{\psi(\mathbf{z})}{2\beta^{-1}} \right\} \sqrt{\det(\beta \Lambda_{\mathbf{z}})^{-1}}, \quad (46)$$

$$p(\mathbf{v} | \gamma_2, \zeta_2) \propto \exp \left\{ -\frac{\mathbf{v}^T \Gamma(\gamma_2, \zeta_2)^{-1} \mathbf{v}}{2} \right\}, \quad (47)$$

$$p(\beta, \gamma_2, \zeta_2) = p(\beta) p(\gamma_2) p(\zeta_2). \quad (48)$$

In order to obtain factorization (44), we have exploited obvious conditional independences: $p(\mathbf{b} | \mathbf{v}, \mathbf{z}, \beta) = p(\mathbf{b} | \mathbf{v}, \mathbf{z}, \beta, \gamma_2, \zeta_2)$, $p(\mathbf{v} | \gamma_2, \zeta_2) = p(\mathbf{v} | \mathbf{z}, \beta, \gamma_2, \zeta_2)$ and $p(\mathbf{z} | \beta) = p(\mathbf{z} | \beta, \gamma_2, \zeta_2)$. We consider non-informative probability distributions for $p(\beta)$, $p(\gamma_2)$ and $p(\zeta_2)$ in (48). Models (45) and (47) are Gaussian, with β denoting a positive parameter and the covariance matrix:

$$\Gamma(\gamma_2, \zeta_2) \triangleq (\beta \sum_{\ell} \lambda_{\ell}^*(\gamma_2, \zeta_2) \mathbf{D}_{\ell}^T \mathbf{D}_{\ell})^{-1}, \quad (49)$$

where $\lambda_{\ell}^*(\gamma_2, \zeta_2)$ denotes the solution of (22) when the value of the parameters appearing in the constraints in (18) are equal to γ_2, ζ_2 . To alleviate notations, we will omit in the following the dependence of the covariance to γ_2, ζ_2 . Now, we need to clarify the choice of models (44)-(48). To this aim, let us emphasize the equivalence between the dual formulation (20) of optimization problem (18) and the MAP problem (41). We consider the general case, where the cost function f is an M-estimator. The Lagrangian function corresponding to (33) then writes:

$$\begin{aligned} L(\mathbf{v}, \mathbf{z}, \boldsymbol{\lambda}^*) &= (\mathbf{A}\mathbf{v} + \mathbf{b})^T \Lambda_{\mathbf{z}} (\mathbf{A}\mathbf{v} + \mathbf{b}) + \psi(\mathbf{z}) \\ &\quad + \sum_{\ell} \lambda_{\ell}^* (\mathbf{v}^T \mathbf{D}_{\ell}^T \mathbf{D}_{\ell} \mathbf{v} - \gamma_2 \ell^{\zeta_2}). \end{aligned} \quad (50)$$

Then, using the probabilistic models (45)-(47), it is easy to see (by direct substitution) that

$$(\hat{\mathbf{v}}_{\boldsymbol{\lambda}^*}, \hat{\mathbf{z}}_{\boldsymbol{\lambda}^*}) \triangleq \arg \min_{\mathbf{v}, \mathbf{z}} L(\mathbf{v}, \mathbf{z}, \boldsymbol{\lambda}^*), \quad (51)$$

$$= \arg \max_{\mathbf{v}, \mathbf{z}} \log p(\mathbf{b}, \mathbf{v}, \mathbf{z} | \beta, \gamma_2, \zeta_2). \quad (52)$$

Hence, relying on model (45)-(47), we obtain that the solution of the Lagrangian optimization problem (51) and of problem (52), *i.e.*, of the MAP estimation problem (41), are identical. The MAP estimation problem is conditioned by the parameters β, γ_2, ζ_2 . Note however that the variable β does not affect the equivalence between (51) and (52) since it represents a multiplicative constant. Therefore, the problem of estimating $(\hat{\mathbf{v}}_{\boldsymbol{\lambda}^*}, \hat{\mathbf{z}}_{\boldsymbol{\lambda}^*})$ does not depend on β . However, we will see in the following that this variable is important for model selection.

C. Estimation method

Considering the MAP interpretation of the Lagrangian optimization problem, we now propose to infer γ_2 and ζ_2 by solving the marginalized MAP problem (43). Nevertheless, we avoid the complex calculation of the non-Gaussian integral with respect to \mathbf{z} by substituting (43) with the following problem:

$$(\gamma_2^{\text{map}}, \zeta_2^{\text{map}}) = \arg \max_{\gamma_2, \zeta_2} \log p(\gamma_2, \zeta_2 | \hat{\mathbf{z}}_{\boldsymbol{\lambda}^*}, \mathbf{b}) \quad (53)$$

conditioned from (36) by:

$$\hat{\mathbf{z}}_{\boldsymbol{\lambda}^*} \triangleq \frac{1}{2} \Lambda_{\mathbf{A}\hat{\mathbf{v}}_{\boldsymbol{\lambda}^*} + \mathbf{b}}^{-1} \nabla \rho(\mathbf{A}\hat{\mathbf{v}}_{\boldsymbol{\lambda}^*} + \mathbf{b}). \quad (54)$$

The probability to be maximized in (53) is defined through two successive marginalizations of the joint distribution:

$$p(\gamma_2, \zeta_2 | \hat{\mathbf{z}}_{\boldsymbol{\lambda}^*}, \mathbf{b}) \triangleq \int p(\beta, \gamma_2, \zeta_2 | \hat{\mathbf{z}}_{\boldsymbol{\lambda}^*}, \mathbf{b}) d\beta, \quad (55)$$

$$p(\beta, \gamma_2, \zeta_2 | \hat{\mathbf{z}}_{\boldsymbol{\lambda}^*}, \mathbf{b}) \triangleq \int p(\beta, \gamma_2, \zeta_2, \mathbf{v} | \hat{\mathbf{z}}_{\boldsymbol{\lambda}^*}, \mathbf{b}) d\mathbf{v}. \quad (56)$$

One can provide an exact analytical expression of the integral (56) over \mathbf{v} since the integrand is a m -dimensional Gaussian. The Laplace's method yields an approximation to the mono-dimensional integral (55) over variable β . The Laplace's method approximates the integral of a function by fitting a Gaussian at its maximum and computing the volume under the Gaussian. For a uni-dimensional variable x and a function $f(x)$, the Laplace's approximation reads:

$$\int f(x)dx \simeq f(\hat{x})\sqrt{2\pi} \left(-\frac{\partial^2}{\partial x^2} \log f(x) \right)_{|\hat{x}}^{-1/2}, \quad (57)$$

where $\hat{x} = \arg \max_x f(x)$. Hence, in order to evaluate (55) we consider the marginal approximation:

$$\begin{aligned} p(\gamma_2, \zeta_2 | \hat{\mathbf{z}}_{\lambda^*}, \mathbf{b}) &= \int p(\mathbf{b} | \hat{\mathbf{z}}_{\lambda^*}, \beta, \gamma_2, \zeta_2) p(\beta, \gamma_2, \zeta_2) d\beta, \\ &\simeq p(\mathbf{b} | \hat{\mathbf{z}}_{\lambda^*}, \hat{\beta}_{\lambda^*}, \gamma_2, \zeta_2) \sqrt{2\pi} \sigma_{\beta}(\hat{\beta}_{\lambda^*}) p(\hat{\beta}_{\lambda^*}, \gamma_2, \zeta_2), \end{aligned} \quad (58)$$

where

$$\hat{\beta}_{\lambda^*} = \arg \max_{\beta} \log p(\mathbf{b} | \hat{\mathbf{z}}_{\lambda^*}, \beta, \gamma_2, \zeta_2). \quad (59)$$

$$\sigma_{\beta}(\hat{\beta}_{\lambda^*}) = \left(-\frac{\partial^2}{\partial \beta^2} \log p(\mathbf{b} | \hat{\mathbf{z}}_{\lambda^*}, \beta, \gamma_2, \zeta_2) \right)_{|\hat{\beta}_{\lambda^*}}^{-1/2} \quad (60)$$

Note that, contrarily to problem (52), estimate $\hat{\beta}_{\lambda^*}$ is no longer a multiplicative constant and appears in the evaluation of the marginal probabilities.

In summary, in order to solve problem (53), we search the element of a finite set $\Upsilon \subset \mathbb{R}^2$ of parameters (γ_2, ζ_2) maximizing model probability $p(\gamma_2, \zeta_2 | \hat{\mathbf{z}}_{\lambda^*}, \mathbf{b})$, using the following procedure:

$$(\gamma_2^{\text{map}}, \zeta_2^{\text{map}}) = \arg \max_{(\gamma_2, \zeta_2) \in \Upsilon} \log p(\mathbf{b} | \hat{\mathbf{z}}_{\lambda^*}, \gamma_2, \zeta_2) p(\gamma_2, \zeta_2) \quad (61)$$

where, in order to evaluate $p(\mathbf{b} | \hat{\mathbf{z}}_{\lambda^*}, \gamma_2, \zeta_2)$ for some $(\gamma_2, \zeta_2) \in \Upsilon$, we solve:

$$(\hat{\mathbf{v}}_{\lambda^*}, \hat{\mathbf{z}}_{\lambda^*}) = \arg \max_{\mathbf{v}, \mathbf{z}} \log p(\mathbf{b}, \mathbf{v}, \mathbf{z} | \beta, \gamma_2, \zeta_2), \quad (62)$$

$$\hat{\beta}_{\lambda^*} = \arg \max_{\beta} \log p(\mathbf{b} | \hat{\mathbf{z}}_{\lambda^*}, \beta, \gamma_2, \zeta_2). \quad (63)$$

We detail hereafter the strategy we considered to implement each step of the proposed algorithm:

1) *Robust constraint motion estimation* (62): Given some power-law parameters (γ_2, ζ_2) , the problem is equivalent to the constraint optic flow estimation problem (18) discussed and solved in section V. It can be solved by the proposed dual optimization algorithm, which uses iterative conditional maximizations to estimate $(\hat{\mathbf{v}}_{\lambda^*}, \hat{\mathbf{z}}_{\lambda^*})$, and a gradient algorithm to converge towards λ^* related to (γ_2, ζ_2) .

2) *Parameter estimation* (63): Given the power-law parameters (γ_2, ζ_2) , the estimate $\hat{\beta}_{\lambda^*}$ is obtained by canceling out the gradient of the marginalized logarithm likelihood $\log p(\mathbf{b} | \hat{\mathbf{z}}_{\lambda^*}, \beta, \gamma_2, \zeta_2)$. Let $\langle \cdot \rangle$ to denote the expectation with respect to $p(\mathbf{v} | \mathbf{b}, \hat{\mathbf{z}}_{\lambda^*}, \beta, \gamma_2, \zeta_2)$. By noticing that:

$$\begin{aligned} \nabla_{\beta} \log p(\mathbf{b} | \hat{\mathbf{z}}_{\lambda^*}, \beta, \gamma_2, \zeta_2) &= \nabla_{\beta} \langle \log p(\mathbf{b}, \mathbf{v} | \hat{\mathbf{z}}_{\lambda^*}, \beta, \gamma_2, \zeta_2) \rangle \\ &\int p(\mathbf{v} | \mathbf{b}, \hat{\mathbf{z}}_{\lambda^*}, \beta, \gamma_2, \zeta_2) \nabla_{\beta} \log p(\mathbf{b}, \mathbf{v} | \hat{\mathbf{z}}_{\lambda^*}, \beta, \gamma_2, \zeta_2) d\mathbf{v} \end{aligned} \quad (64)$$

and using the definition of the joint distribution in (45)-(47), the gradient of the marginalized log likelihood can be efficiently computed as:

$$\begin{aligned} \nabla_{\beta} \log p(\mathbf{b} | \hat{\mathbf{z}}_{\lambda^*}, \beta, \gamma_2, \zeta_2) &= \\ &\frac{1}{2} \langle \beta^{-1} |\mathcal{S}_r| - (\mathbf{b} - \mathbf{A}\mathbf{v})^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} (\mathbf{b} - \mathbf{A}\mathbf{v}) \rangle. \end{aligned} \quad (65)$$

Then, after some manipulation of (65) described in Appendix I, one can show that the gradient vanishes for:

$$\hat{\beta}_{\lambda^*} = \frac{|\mathcal{S}_r| - \text{tr}((\mathbf{A}^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} \mathbf{A} + \sum_{\ell} \lambda_{\ell}^* \mathbf{D}_{\ell}^T \mathbf{D}_{\ell})^{-1} \mathbf{A}^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} \mathbf{A})}{(\mathbf{b} - \mathbf{A}\hat{\mathbf{v}}_{\lambda^*})^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} (\mathbf{b} - \mathbf{A}\hat{\mathbf{v}}_{\lambda^*})}. \quad (66)$$

$\hat{\beta}_{\lambda^*}$ is the unique global maximum since it can be shown that the function $\log p(\mathbf{b} | \hat{\mathbf{z}}_{\lambda^*}, \beta, \gamma_2, \zeta_2)$ is concave. The computation of (66) only requires tractable linear operations and the MAP estimates $(\hat{\mathbf{v}}_{\lambda^*}, \hat{\mathbf{z}}_{\lambda^*})$, which are the solutions of (62). Details are given in section VII.

3) *Prior model inference* (61): The decision on the prior model parameter (γ_2, ζ_2) is made by maximizing $\log p(\mathbf{b} | \hat{\mathbf{z}}_{\lambda^*}, \gamma_2, \zeta_2)$. Since we assume that (γ_2, ζ_2) takes its values in a finite set Υ , solving (61) only requires the evaluation of $\log p(\mathbf{b} | \hat{\mathbf{z}}_{\lambda^*}, \gamma_2, \zeta_2)$ for these values. Since $p(\gamma_2)$ and $p(\zeta_2)$ are assumed to be non-informative priors, the logarithm of the Laplace approximation of this probability (58) reads:

$$\log p(\mathbf{b} | \hat{\mathbf{z}}_{\lambda^*}, \gamma_2, \zeta_2) \propto \log \sigma_{\beta}(\hat{\beta}_{\lambda^*}) + \log p(\mathbf{b} | \hat{\mathbf{z}}_{\lambda^*}, \hat{\beta}_{\lambda^*}, \gamma_2, \zeta_2), \quad (67)$$

where, after some calculations detailed in appendix II, the logarithm of the marginalized likelihood can be rewritten as:

$$\begin{aligned} \log p(\mathbf{b} | \hat{\mathbf{z}}_{\lambda^*}, \hat{\beta}_{\lambda^*}, \gamma_2, \zeta_2) &= \frac{1}{2} \log \det \left(\sum_{\ell} \lambda_{\ell}^* \mathbf{D}_{\ell}^T \mathbf{D}_{\ell} \right) \\ &\frac{1}{2} \log \det \hat{\beta}_{\lambda^*} \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} - \frac{1}{2} (\mathbf{b} - \mathbf{A}\hat{\mathbf{v}}_{\lambda^*})^T \hat{\beta}_{\lambda^*} \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} (\mathbf{b} - \mathbf{A}\hat{\mathbf{v}}_{\lambda^*}) \\ &- \frac{1}{2} \log \det \left(\mathbf{A}^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} \mathbf{A} + \sum_{\ell} \lambda_{\ell}^* \mathbf{D}_{\ell}^T \mathbf{D}_{\ell} \right), \end{aligned}$$

and the logarithm of its Hessian evaluated at $\hat{\beta}_{\lambda^*}$ can be expressed as:

$$\log \sigma_{\beta}(\hat{\beta}_{\lambda^*}) = \frac{1}{2} \log \left(\frac{2\hat{\beta}_{\lambda^*}}{(\mathbf{b} - \mathbf{A}\hat{\mathbf{v}}_{\lambda^*})^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} (\mathbf{b} - \mathbf{A}\hat{\mathbf{v}}_{\lambda^*})} \right). \quad (68)$$

VII. ALGORITHM COMPLEXITY

The flow chart of the algorithm for the selection of the power-law prior and the estimation of the instantaneous velocity field from an image couple is presented in Table II.

Let κ denote the conditioning number of the posterior Hessian matrices, which have been preconditioned by an incomplete LU decomposition. The complexity of the overall algorithm presented in Table II is of $\mathcal{O}(\kappa m)$, where we recall that m denotes the dimension of the motion vector.

More precisely, the complexity of two computation steps dominates the algorithm. First, the complexity of calculating

- For all power-laws $(\gamma_2, \zeta_2) \in \Upsilon$:
 - Compute MAP estimate $(\hat{\mathbf{v}}_{\lambda^*}, \hat{\mathbf{z}}_{\lambda^*})$ using algorithm of Table I
 - Access $\hat{\beta}_{\lambda^*}$ with (66)
 - Evaluate the probability $p(\mathbf{b}|\hat{\mathbf{z}}_{\lambda^*}, \gamma_2, \zeta_2)$ of the power-law (γ_2, ζ_2) with (67)
- Find among the elements of Υ the MAP estimate $(\gamma_2^{\text{map}}, \zeta_2^{\text{map}})$ satisfying (61)

TABLE II

FLOW CHART OF THE OVERALL ESTIMATION ALGORITHM

(29), *i.e.* solving a linear system using for instance a conjugate gradient squared (CGS) algorithm, is of $\mathcal{O}(\kappa m)$. The additional use of the incomplete LU preconditioner enables: 1) to lower the conditioning number κ and therefore to reduce the complexity of (29), 2) to access with a negligible complexity to the determinants appearing in (67). Indeed the determinant is given by the product of the diagonal elements u_{ii} of the upper triangular matrix U:

$$\det(\text{LU}) = \prod_i u_{ii}, \quad (69)$$

where the incomplete LU decomposition is performed on the posterior and prior Hessian matrix. Second, a straightforward calculation of the trace appearing in (66) can be very demanding. Instead, we rely on a trace randomization technique [36] which is efficiently achieved using the CGS algorithm. Details are given in Appendix I. The method complexity for the calculation of the trace is also $\mathcal{O}(\kappa m)$

VIII. EXPERIMENTAL EVALUATION

A. Simulated 2D turbulence

The Bayesian method is first assessed on a synthetic image sequence database of numerically generated 2D homogeneous, isotropic and incompressible turbulent flow fields [7]. This database has proven its relevance in previous studies in the field of fluid flow estimation [13], [15], [32], [37]. It relies on a Direct Numerical Simulation (DNS) of Navier-Stokes equations at Reynolds number $Re = 3000$. The output of this simulation is then used to compute the Lagrangian trajectories of particle tracers transported by the flow or, in the case of scalar images, to solve an advection-diffusion equation. The latter equation reads:

$$\frac{\partial I}{\partial t}(\mathbf{s}, t) + \nabla_{\mathbf{s}}^T I(\mathbf{s}, t) \mathbf{v}(\mathbf{s}) = \kappa \Delta_{\mathbf{s}} I(\mathbf{s}, t), \quad \forall \mathbf{s} \in \mathcal{S}_r, \forall t \in \mathcal{T}_r, \quad (70)$$

where $\kappa \in \mathbb{R}_+$ denotes a diffusion coefficient. Motion and scalar transport equations are solved in Fourier space using dealiased Fourier expansions in two directions with periodic boundary conditions. In order to sustain turbulence, the flow was stirred by an external random force in Fourier space acting at low wavenumbers. The time integration is done using a third-order Runge-Kutta scheme. The simulation resulted in two synthetic sequences of 100 images of 256×256 pixels depicting scalar or particles, together with the true underlying motion fields. A precise description of the simulation can be found in [7]. Fig. 1 presents three samples of the synthetic scalar or particle image sequences and a color representation of the underlying velocity fields. In these visualizations color

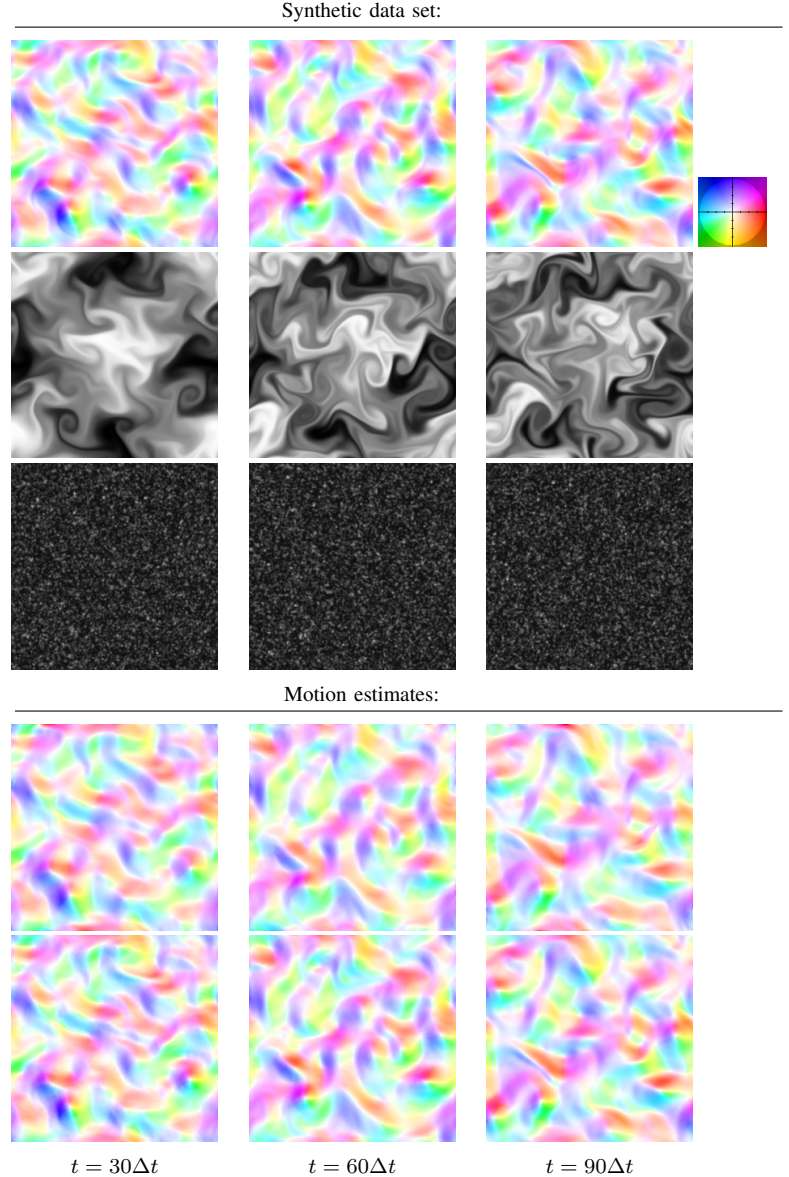


Fig. 1. 2D turbulence: synthetic data set and motion estimates. From top to bottom at time $t = 30\Delta t$ (left), $t = 60\Delta t$ (middle) and $t = 90\Delta t$ (right): simulated velocity fields visualized using the color system of [1], scalar and particle images, motion estimates in the case of scalar and particle images .

and intensity code vector orientations and magnitudes [1]. In the case of scalar images (resp. particle images), the observation model was the advection-diffusion equation (70) (resp. the optic flow constraint equation (2)) penalized by the robust M-estimators approaching the L_1 norm (see *e.g.*, [18]). A multi-resolution approach was used to cope with large displacements [3]. We chose to perform model inference relying on motion field increment estimated at the last level of this pyramidal approach.

The inference of the prior power-law model $(\gamma_2^{\text{map}}, \zeta_2^{\text{map}})$ is performed for the two experiences, in the scale range of 1 to 4 pixels and at the arbitrary chosen time of $t = 20\Delta t$. Fig. 6 shows, for both experiments, the shape (up to an irrelevant constant) of the a posteriori probability distribution of the power-law model with respect to its parameters (γ_2, ζ_2) , that is $p(\gamma_2, \zeta_2 | \mathbf{b}, \hat{\mathbf{z}}_{\lambda^*})$. The global maximum of the probability is

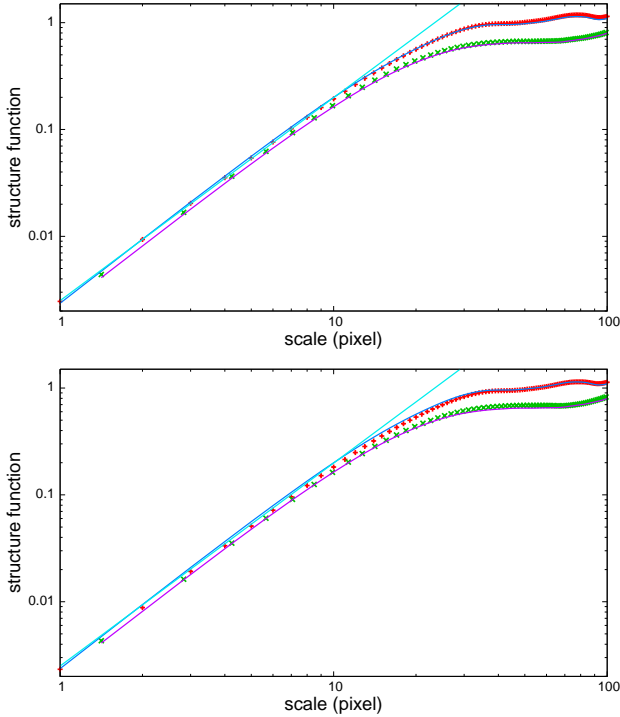


Fig. 2. 2D turbulence: second order structure function estimation $\overline{\delta v_{\parallel}(\ell)^2}$ at $t=20\Delta t$. Above: inferred power-law model (straight line), true (continuous curves) and estimated (stars and crosses) structure function in horizontal-vertical and diagonal direction for particle images. Below: identical legend for scalar imagery.

located for both the cases of scalar or particle imagery, at the coordinates $(\gamma_2^{\text{map}}, \zeta_2^{\text{map}}) = (2.5e-3, 1.9)$. It is important to note that, as shown in Fig. 6, these coordinates also correspond to an “optimal” choice of parameters in the sense of the minimum of MBA error. Moreover, we can observe in Fig. 2 that the inferred power-law describes perfectly the true structure function computed with the ground truth motion field. Finally, the inferred scaling exponent is consistent with the theoretical value of 2 predicted for 2D turbulence in the inertial or dissipative range [22].

Assuming constant second order statistics along time, the selected power-law model $(\gamma_2^{\text{map}}, \zeta_2^{\text{map}})$ is used to process the entire image sequence. The evolutions of the resulting mean barron angular (MBA) error and root mean square (RMS) error are plotted in Fig. 7 and compared to recent state of the art approaches³. Note that MBA and RMS errors were computed with divergence-free motion components obtained by Helmholtz decomposition [20]. Indeed, this decomposition was used here since a subset of state-of-the-art methods code implicitly an additional divergence-free constraint [20] [37], leading possibly to a bias in the regularization scheme evaluation.

The proposed estimation method reaches the level of the best state-of-the-art algorithms in the case of the particle image sequence, while for scalar imagery it outperforms them all. In

³Let us mention that, for consistency of the evaluation, we have compared the proposed method to algorithms implying neither temporal regularization schemes (see *e.g.*, [15] [35] [32]) nor specific configurations of fluid flows (see *e.g.*, [34] for the particular case of stokes flows). Note that, if necessary, the proposed Bayesian method could in principle be extended without any major difficulties to include an additional temporal regularization term.

particular, although second order regularizations [37] [20] or higher-order regularization schemes [9] yield some improvements in comparison to standard first order regularization [17], there is a significant gap in accuracy with the proposed method in the case of scalar imagery. This gap is due to the weak spatial gradients of scalar images (*i.e.*, poorly conditioned data terms), which make crucial the information brought by an appropriate prior. For particle imagery, the data term is better conditioned and penalization of high-order derivatives is sufficient to obtain relevant turbulent flow approximations. Let us also mention that classical correlation-based techniques or the adaptive correlation method proposed by [2] perform reasonably well in the case of particle images, while such local schemes usually fail in the case of scalar imagery.

B. Plane components of simulated 3D turbulence

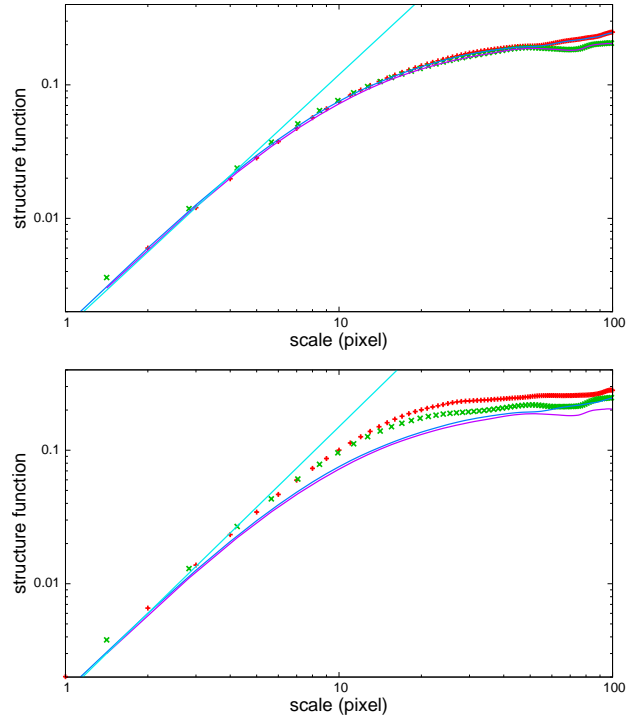


Fig. 3. 3D turbulence: second order structure function estimation $\overline{\delta v_{\parallel}(\ell)^2}$. Above: inferred power-law model (straight line), true (continuous curves) and estimated (stars and crosses) structure function in horizontal-vertical and diagonal direction for particle images. Below: identical legend for scalar imagery.

The method is then evaluated on two synthetic image couples obtained by deformation of instances of the scalar and particle image sequences presented in the last section. The velocity field used for deformation is the plane components of 3D motion fields computed by numerical simulation of 3D homogeneous, isotropic and incompressible turbulence [29] at Reynolds number $Re = 667$. On the contrary to the incompressible 2D case, the apparent motion field is here characterized by divergence-free and irrotational components of comparable magnitude. Fig. 8 presents the scalar or particle synthetic image couples and a color representation of the underlying velocity field. For both the scalar and particle cases, the data model used in these experiments is the optic flow

constraint equation (2) penalized by the robust M-estimators approaching the L_1 norm (see *e.g.*, [18]). A multi-resolution approach is also used to cope with large displacements [3].

The inference of the prior power-law model $(\gamma_2^{\text{map}}, \zeta_2^{\text{map}})$ is performed for the two experiences, in the scale range of 1 to 4 pixels. Fig. 9 depicts, for both experiments, the shape of the a posteriori probability distribution of the power-law model parameters (γ_2, ζ_2) . The global maximum of the probability is located for the case of particle imagery at $(\gamma_2^{\text{map}}, \zeta_2^{\text{map}}) = (1.5\text{e-}3, 2.0)$ and for the case of scalar imagery at $(\gamma_2^{\text{map}}, \zeta_2^{\text{map}}) = (1.5\text{e-}3, 1.9)$. It is important to note that, as shown in this figure, these coordinates also correspond to a nearly “optimal” choice of parameters in the sense of the minimum of MBA error for both experiments. Moreover, we can observe in Fig. 3 that the inferred power-law describes almost perfectly the true structure function computed with the ground truth motion field.

A comparison with recent state of the art approaches in terms of MBA and RMS errors is given in the table of Fig. 10. The proposed estimation method reaches the level of the best state of the art algorithms in the scalar case or outperforms them in the case of particle images. Let us mention that, conversely to the case of 2D incompressible turbulence, the adaptive correlation method of [2] performs relatively well in the case of particle images.

C. Experimental 3D turbulence

The Bayesian approach is also assessed on a real image sequence recorded in a wind tunnel⁴ by particle image velocimetry. This image sequence shows three-dimensional homogeneous and isotropic turbulent motion at Reynolds number $Re = 2300$ in the wake of a biplane grid (see [24]). The size of images is 512×512 pixels corresponding to $34 \times 34 \text{ mm}^2$. A spatial subset of an image couple is depicted in Fig. 4. A tiny time delay between frames was chosen in order to limit the loss of particles due to the “out of plane” component.

The data model used in these experiments was the OFC equation (2) penalized by the robust M-estimators approaching the L_1 norm (see *e.g.*, [18]). The inference of the power-law parameters $(\gamma_2^{\text{map}}, \zeta_2^{\text{map}})$ was performed on an image couple of the sequence in the scale range of 1 to 4 pixels. The global maximum of the model probability is located near the coordinates $(\gamma_2^{\text{map}}, \zeta_2^{\text{map}}) = (1\text{e-}3, 2.0)$. The inferred scaling exponent seems to be consistent with physics: the scales of the power-law mainly belongs to the dissipative range which is known to be ruled theoretically by the exponent $\gamma_2 = 2.0$. Indeed, the dissipative scales was estimated below 3.3 pixels with hot wire anemometry measurements.

The selected prior power-law yielded highly structured motion fields as shown in the zoom in Fig. 4. The strength of multi-scale regularization is particularly striking in this case: tiny turbulent structures are revealed. Some vortex structures are smaller than 10 pixels diameter. It can be noted that for this flow with energy level sustained across the small scales, classical optical flow methods and correlation approaches estimated only inaccurate vector fields. At larger scales, differences are

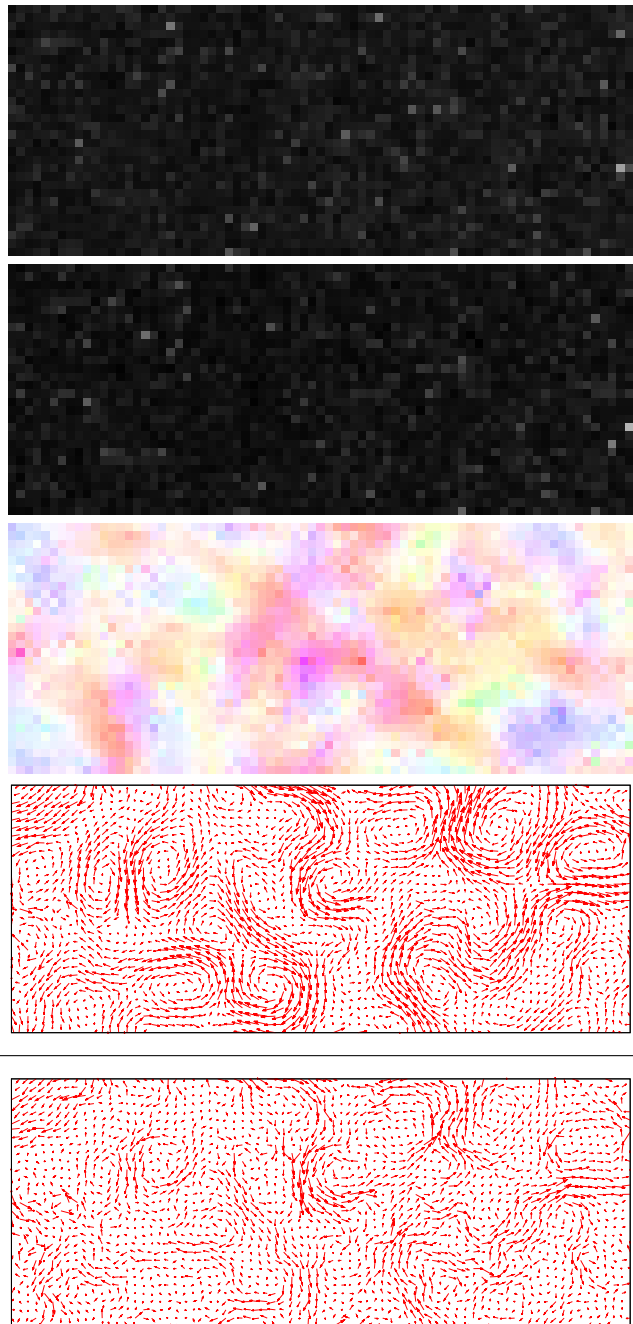


Fig. 4. 3D experimental turbulence. From top to bottom, spatial subset of the two input images, estimated velocity in color representation and divergence-free component obtained by a Helmholtz decomposition [20] for the proposed method, divergence-free component given by a first-order regularizer [6] (below). The inferred power-laws parameters are $(\gamma_2^{\text{map}}, \zeta_2^{\text{map}}) = (1\text{e-}3, 2.0)$.

less prominent and the proposed regularization seems coherent with a first-order scheme [6].

D. Atmospheric turbulence

We finally test the proposed modeling on a satellite image sequence. The analysis was done using METEOSAT Second Generation meteorological images acquired above the north Atlantic Ocean at a rate of an image every 15 min. The spatial resolution is $3 \times 3 \text{ km}^2$ at the center of the Earth image disk. Following the methodology in [15], a set of sparse pressure difference images of 256×256 pixels associated to a stack of layers at various altitude was derived. As detailed in the

⁴Irstea research institute, center of Rennes, France

latter work, each image of the stack corresponds to a given altitude and can be related to a vertically averaged horizontal wind field *via* an observation model relying on layer mass conservation $\forall \mathbf{s} \in \mathcal{S}_r, \forall t \in \mathcal{T}_r$:

$$\frac{\partial I}{\partial t}(\mathbf{s}, t) + \nabla_{\mathbf{s}}^T I(\mathbf{s}, t) v(\mathbf{s}) + I(\mathbf{s}, t) \nabla_{\mathbf{s}} \cdot v(\mathbf{s}) = 0. \quad (71)$$

A spatial subset of the very sparse image observations used in these experiments is displayed in Fig. 5. These images represent incomplete pressure difference maps of an atmospheric layer at intermediate altitude. We used in this case the data model proposed in [15], which describes discrepancies from the mass conservation equation (71) only in observed areas. Discrepancies are given by the Leclerc’s robust M-estimators [8].

The inference of the power-law parameters $(\gamma_2^{\text{map}}, \zeta_2^{\text{map}})$ was performed in the scale range of 1 to 4 pixels. A multi-resolution approach was used again to cope with large displacements. A clear maximum is located at $(\gamma_2^{\text{map}}, \zeta_2^{\text{map}}) = (2.5e-5, 2.0)$. According to physical consideration, the inferred scaling law exponent suggests two-dimensional turbulence, *i.e.*, scaling exponent $\gamma_2 = 2.0$. Although some controversy remain on the spectrum of atmospheric turbulence at intermediate scales [25], this result seems reasonable since the atmosphere can be modeled as a quasi two-dimensional flow in a wide range of scales.

As illustrated in a zoom displayed in Fig. 5, the proposed modeling produced motion fields visually consistent with the sparse observations and realistic in non-observed regions. The velocity field at this resolution appears to be very smooth and structured with vortices of only a few pixels size. Improvements brought by the method is in this case particular striking when compared to the second order regularization scheme proposed in [15].

IX. CONCLUSION

We have presented a method for fluid flow estimation exploiting physical prior knowledge on the scale invariant properties of turbulent flows. Multi-scale regularization is achieved by imposing some power-law behavior to the second order structure function of motion. Bayesian modeling treating motion fields, prior power-law models together with their hyper-parameters as interdependent random variables allows us to simultaneously provide solutions for three problems: robust motion estimation, hyper-parameter estimation, and power-law model selection. Optimal model and hyper-parameter estimates are derived in analytical forms in a marginalized a posteriori sense. Consequently, motion estimation does not involve neither tuning regularization hyper-parameter nor choosing an arbitrary power-law model. Experiments on synthetic and real image sequences prove that the proposed Bayesian inference scheme succeeds to select adapted model and hyper-parameters and enhances state of the art in fluid flow estimation. Extending the Bayesian method to the modeling of non-homogeneous motion fields constitutes a promising research direction. Nevertheless, this extension is likely to face complexity bottlenecks, necessitating probably the introduction of variational Bayesian inference schemes [4].

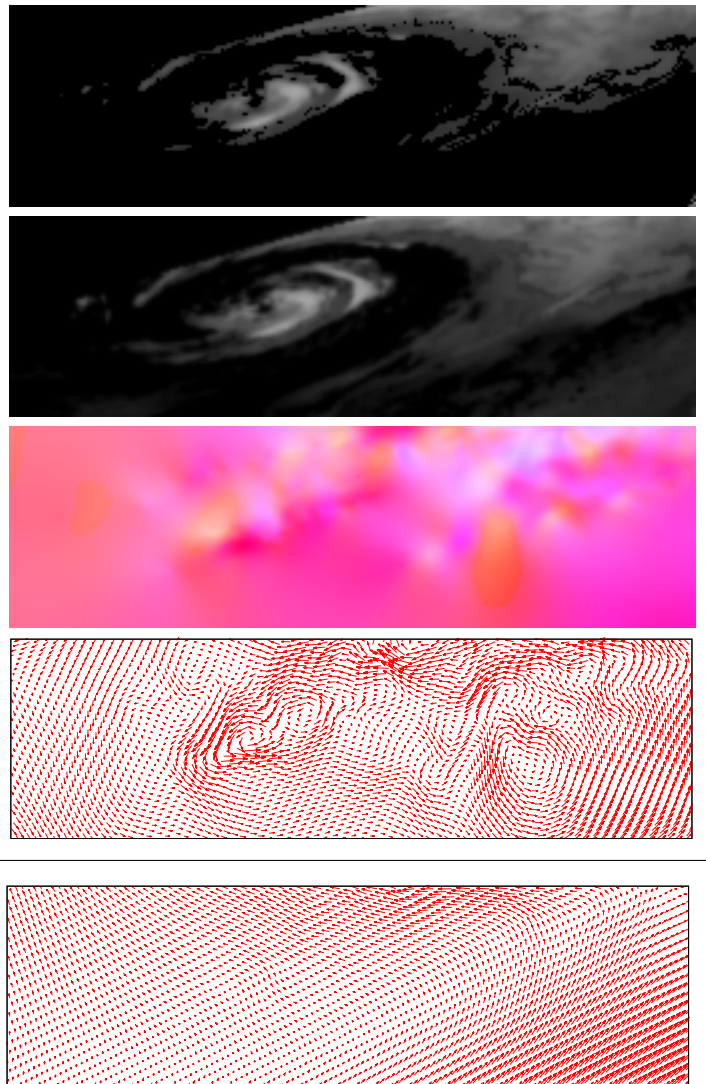


Fig. 5. Horizontal wind at intermediate altitude. From top to bottom, spatial subset of the two input incomplete images, estimated wind field in color representation and divergent-free component obtained by Helmholtz decomposition [20] for the proposed method, divergence-free component given by a second-order regularizer [15] (below). The inferred power-laws parameters are $(\gamma_2^{\text{map}}, \zeta_2^{\text{map}}) = (2.5e-5, 2.0)$.

REFERENCES

- [1] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski. A database and evaluation methodology for optical flow. In *Int. Conf. on Comp. Vis., ICCV 2007*, 2007.
- [2] F. Becker, B. Wieneke, S. Petra, A. Schroeder, and C. Schnoerr. Variational adaptive correlation method for flow estimation. *Image Processing, IEEE Transactions on*, PP(99):1, 2011.
- [3] J. Bergen, P. Burt, R. Hingorani, and S. Peleg. A 3-frame algorithm for estimating two-component image motion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(9):886–895, Sept. 1992.
- [4] J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, M. W. (eds, M. J. Beal, and Z. Ghahramani. The variational bayesian em algorithm for incomplete data: with application to scoring graphical model structures. *Publisher: Oxford University Press, USA*, 7(7):453, 2003.
- [5] D. Bertsekas, A. Nedic, and A. Ozdaglar, editors. *Convex Analysis and Optimization*. Athena Scientific, 2003.
- [6] M. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104, 1996.
- [7] J. Carlier and B. Wieneke. Report 1 on production and diffusion of fluid mechanics images and data. *Fluid project deliverable 1.2*. <http://www.fluid.irisa.fr>, 2005.

- [8] T. Corpetti, E. Mémin, and P. Pérez. Dense estimation of fluid flows. *Pattern Anal Mach Intel*, 24(3):365–380, 2002.
- [9] P. Dérian, P. Héas, C. Herzet, and E. Mémin. Wavelet-based fluid motion estimation. In *Scale Space Methods and Variational Methods (SSVM) in Computer Vision*, Israel, June 2011.
- [10] U. Frisch. *Turbulence : the legacy of A.N. Kolmogorov*. Cambridge university press, 1995.
- [11] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(3):367–383, 1992.
- [12] P. Heas, C. Herzet, and E. Memin. Bayesian inference of models and hyper-parameters for robust optic-flow estimation. *Image Processing, IEEE Transactions on*, PP(99):1, 2011.
- [13] P. Heas, E. Memin, D. Heitz, and P. Mininni. Bayesian selection of scaling laws for motion modeling in images. In *International Conference on Computer Vision (ICCV'09)*, Kyoto, Japan, October 2009.
- [14] P. Heas, E. Memin, D. Heitz, and P. Mininni. Power laws and inverse motion modeling: application to turbulence measurements from satellite images. *TellusA*, 64, DOI: 10.3402/tellusA.v64i0.10962, 2012.
- [15] P. Heas, E. Memin, N. Papadakis, and A. Szantai. Layered estimation of atmospheric mesoscale dynamics from satellite imagery. *IEEE trans. on Geo. and Rem. Sensing*, 45(12):4087–4104, 2007.
- [16] D. Heitz, E. Memin, and C. Schnoerr. Variational fluid flow measurements from image sequences: synopsis and perspectives. *Exp. Fluids*, 48(3):369–393, 2010.
- [17] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.
- [18] J. Idier. Convex half-quadratic criteria and interacting auxiliary variables for image restoration. *Image Processing, IEEE Transactions on*, 10(7):1001–1009, jul 2001.
- [19] E. T. Jaynes. Bayesian methods: General background, 1986.
- [20] S. Kadri Harouna, P. Dérian, P. Héas, and E. Memin. Divergence-free Wavelets and High Order Regularization. hal-00646104, January 2012.
- [21] A. Kolmogorov. The local structure of turbulence in incompressible viscous fluid for very large reynolds number. *Dolk. Akad. Nauk SSSR*, 30:301–5, 1941.
- [22] R. Kraichnan. Inertial ranges in two-dimensional turbulence. *Phys. Fluids*, 10:1417–1423, 1967.
- [23] K. Krajsek and R. Mester. Bayesian model selection for optical flow estimation. In *DAGM-Symposium*, pages 142–151, 2007.
- [24] P. Lavoie, L. Djenidi, and R. A. Antonia. Effects of initial conditions in decaying turbulence generated by passive grids. *Journal of Fluid Mechanics*, 585(1):395–420, 2007.
- [25] E. Lindborg and J. Cho. Horizontal velocity structure functions in the upper troposphere and lower stratosphere 2. theoretical considerations. *J. Geophysical Research.*, 106:233–241, 2001.
- [26] T. Liu and L. Shen. Fluid flow and optical flow. *Journal of Fluid Mechanics*, 614:253, Oct. 2008.
- [27] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereovision. In *Int. Joint Conf. on Artificial Intel. (IJCAI)*, pages 674–679, 1981.
- [28] D. J. C. MacKay. Bayesian interpolation. *Neural Computation*, 4(3):415–447, 1992.
- [29] P. D. Mininni, D. O. Gomez, and S. M. Mahajan. Direct simulations of helical hall-mhd turbulence and dynamo action. *The Astrophysical Journal*, 619:1019, 2005.
- [30] R. Molina, A. K. Katsaggelos, and J. Mateos. Bayesian and regularization methods for hyperparameter estimation in image restoration. *IEEE Trans. Image Processing*, 8:231–246, 1999.
- [31] A. Monin and A. Yaglom. *Statistical Fluid Mechanics: Mechanics of Turbulence*. J Dover Pubns, 1971.
- [32] N. Papadakis and E. Memin. Variational assimilation of fluid motion from image sequences. *SIAM Journal on Imaging Science*, 1(4):343–363, 2008.
- [33] C. P. Robert. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation (Springer Texts in Statistics)*. Springer Verlag, New York, 2nd edition, June 2007.
- [34] P. Ruhnau and C. Schnörr. Optical stokes flow estimation: An imaging-based control approach. *Exp. in Fluids*, 42:61–78, 2007.
- [35] P. Ruhnau, A. Stahl, and C. Schnörr. Variational estimation of experimental fluid flows with physics-based spatio-temporal regularization. *Measurement Science and Technology*, 18:755–763, 2007.
- [36] J. Skilling. The eigenvalues of mega-dimensional matrices. In *Maximum Entropy and Bayesian Methods*, pages 455–466, 1989.
- [37] J. Yuan, C. Schnoerr, and E. Memin. Discrete orthogonal decomposition and variational fluid flow estimation. *Journ. of Math. Imaging & Vision*, 28:67–80, 2007.



Patrick Héas received his Ph.D. in 2005 in informatics and telecommunication from the National Aerospace Engineering School (SUPAERO), Toulouse, France. Since 2008, he has been a permanent research scientist at INRIA, Rennes, France. His research focuses on image sequence analysis applied to inverse motion modeling in geophysics. In particular, his current interests include Bayesian inference, probabilistic motion modeling, turbulent statistical models and scale-space representations.



Cédric Herzet received the Ph.D. in Applied Science from the Université catholique de Louvain (UCL), Belgium, in 2006. From May to August 2006, he was a post-doctoral researcher with the Ecole Normale Supérieure de Cachan, France and, from September 2006 to September 2007, with the university of California, Berkeley. Since then, he has been a permanent research fellow at INRIA, Rennes, France. His current research interests include sparse representations, Bayesian inference and image processing applied to geophysics and fluid mechanics.



Etienne Mémin is Research Director at INRIA. He leads the Fluminance group jointly affiliated with INRIA and IRSTEA. Fluminance is dedicated to the analysis of fluid flows from image sequences. The research axes of Etienne Mémin focus on the coupling between image sequences analysis techniques and physical models. This concerns the study of data assimilation strategies, data driven dynamical models or fluid flows motion measurement and tracking approaches.



Dominique Heitz Dominique Heitz holds a Ph.D. in fluid mechanics from the University of Poitiers, France. After receiving his Ph.D., he joined in 1999 the National Research Institute of Science and Technology for Environment and Agriculture (Irstea) as a permanent research scientist, working on turbulent shear flows analysis and modeling. His current research is mostly focused on methods for 2D and 3D spatio-temporal turbulent flows reconstruction from image sequences.



Pablo D. Mininni received his Ph.D. in 2003 in physics from the University of Buenos Aires (UBA) in Argentina. From 2004 to 2007, he lived in the United States in Boulder, Colorado, where he was a postdoc at National Center for Atmospheric Research (NCAR). Now, he is professor and head of the Physics Department at the University of Buenos Aires, and scientist at National Center for Atmospheric Research. Dr. Mininni works on the numerical and theoretical study of turbulent flows, with applications in geophysics, astrophysics.

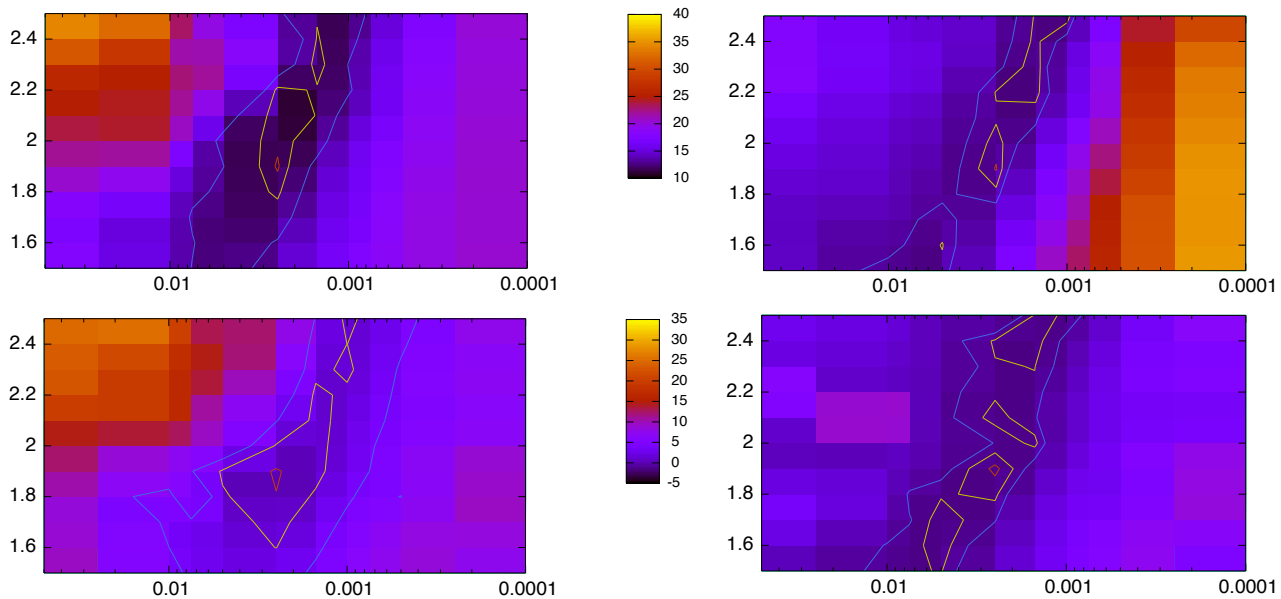


Fig. 6. 2D turbulence: comparison at $t=20\Delta t$ between minus of the logarithm of unnormalized power-law model probability (left) and MBA error (right) with respect to power-law exponent ζ_2 (y-axis) and prefactor γ_2 (x-axis) for particle (above) or scalar (below) imagery. Iso-contours of decreasing values are plotted around the minimum $(\gamma_2^{\text{map}}, \zeta_2^{\text{map}}) = (2.5e-3, 1.9)$.

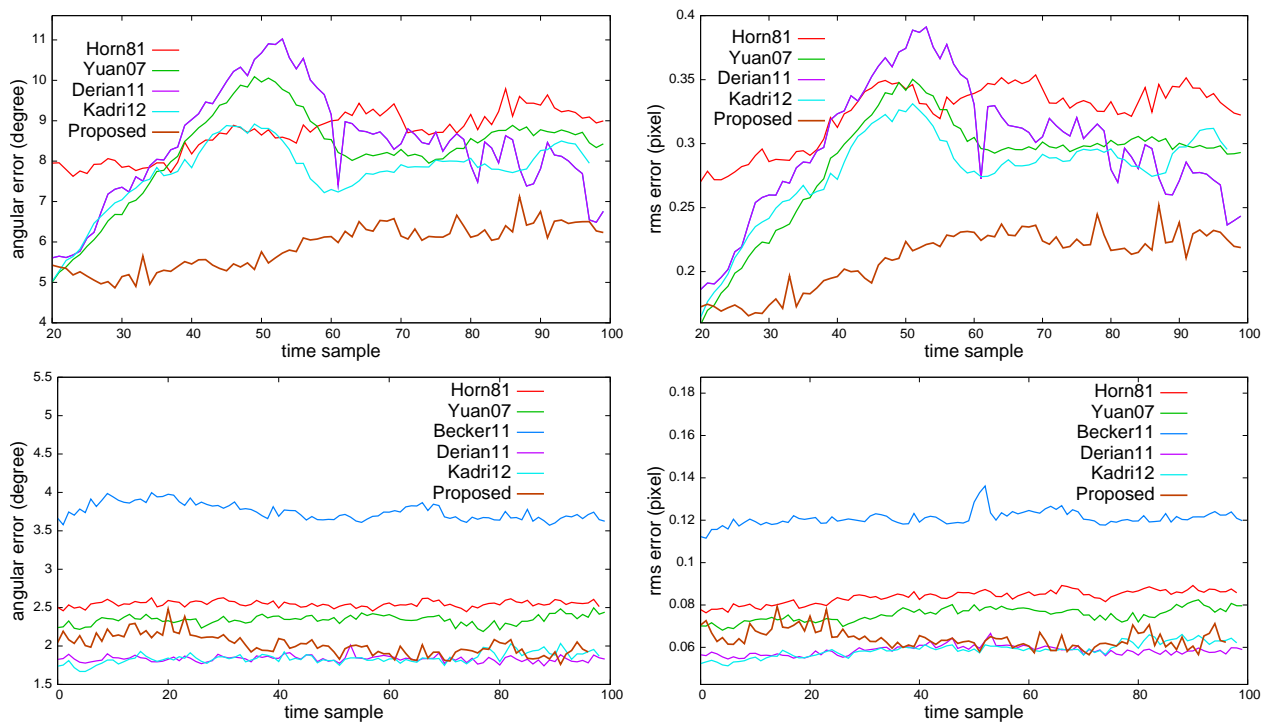


Fig. 7. 2D turbulence: evolution with respect to time index, in the case of scalar (above) or particle (below) imagery, of MBA (left) and RMS (right) errors of the divergence-free motion component relative to motion estimate with adaptive correlation [2], with first order regularization [17], with second order regularizations [37] [20], with higher-order regularization [9], or with the proposed method. Divergence-free motion components are obtained by Helmholtz decomposition [20]

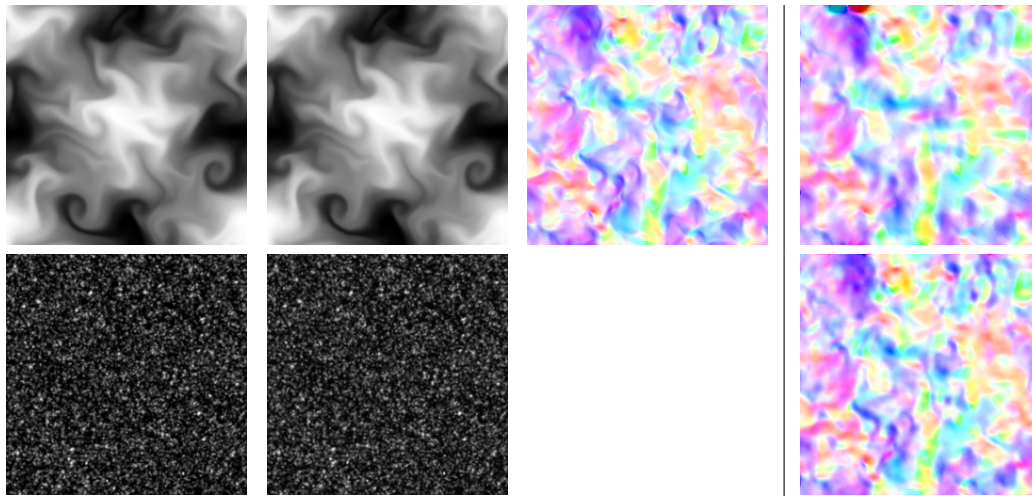


Fig. 8. 3D turbulence: synthetic data set and motion estimates. From left to right, scalar (above) and particle (below) image couple, simulated velocity field (above) and motion estimates for scalar (above) and particle (below) images visualized using the color system of [1].

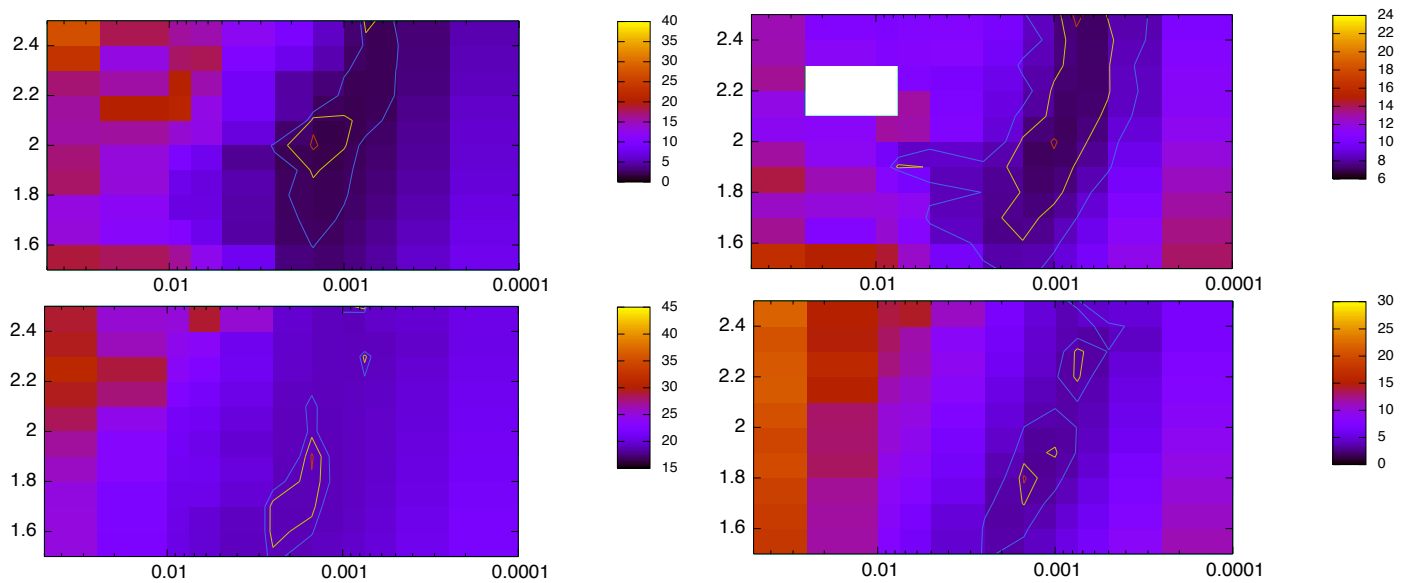


Fig. 9. 3D turbulence: comparison between minus of the logarithm of unnormalized power-law model probability (left) and MBA error (right) with respect to power-law exponent ζ_2 (y-axis) and prefactor γ_2 (x-axis) for scalar (above) or particle (below) imagery. Iso-contours of decreasing values are plotted around the MBA error minimum and the probability minimum, that is to say $(\gamma_2^{\text{map}}, \zeta_2^{\text{map}}) = (1.5e-3, 2.0)$ for particle imagery or $(\gamma_2^{\text{map}}, \zeta_2^{\text{map}}) = (1.5e-3, 1.9)$ for scalar imagery.

	Becker11 [2]	Horn81 [17]	Derian11 [9]	Kadri12 [20]	proposed
	particles				
RMS	0.08474	0.09332	0.09747	0.1117	0.07147
MBA	3.46075	3.90757	4.04505	4.58077	3.06059
	scalar				
RMS		0.19104	0.17693	0.2212	0.18909
MBA		8.07831	7.48561	8.8059	7.58903

Fig. 10. 3D turbulence: comparison in the case of scalar or particle imagery, of MBA and RMS errors relative to motion estimate with adaptive correlation [2], with first order regularization [17] [9], with second order regularizations [20], or with the proposed method.

APPENDIX I

GRADIENT OF THE MARGINALIZED LIKELIHOOD

We want to compute the gradient given in (65) which writes $\nabla_{\beta} \log p(\mathbf{b}|\hat{\mathbf{z}}_{\lambda^*}, \beta, \gamma_2, \zeta_2)$:

$$= \frac{1}{2} \langle \beta^{-1} |\mathcal{S}_r| - (\mathbf{b} - \mathbf{A}\mathbf{v})^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} (\mathbf{b} - \mathbf{A}\mathbf{v}) \rangle, \quad (72)$$

where we recall that $\langle \cdot \rangle$ is the expectation with respect to $p(\mathbf{v}|\mathbf{b}, \hat{\mathbf{z}}_{\lambda^*}, \beta, \gamma_2, \zeta_2)$. Note that in this expression, \mathbf{v} only appears in linear and quadratic forms. As a consequence, the latter derivative is only a function of the mean and covariance of $p(\mathbf{v}|\mathbf{b}, \hat{\mathbf{z}}_{\lambda^*}, \beta, \gamma_2, \zeta_2)$. Now, it is easy to see that this distribution is a Gaussian distribution with mean and covariance defined respectively as:

$$\hat{\mathbf{v}}_{\lambda^*} \triangleq \langle \mathbf{v} \rangle = \mathbf{\Gamma} (\beta \mathbf{A}^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} \mathbf{b}), \quad (73)$$

$$\mathbf{\Gamma} \triangleq \langle (\mathbf{v} - \hat{\mathbf{v}}_{\lambda^*})(\mathbf{v} - \hat{\mathbf{v}}_{\lambda^*})^T \rangle = \left(\beta \mathbf{A}^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} \mathbf{A} + \beta \sum_{\ell} \lambda_{\ell}^* \mathbf{D}_{\ell}^T \mathbf{D}_{\ell} \right)^{-1}. \quad (74)$$

Let us write the second term of (72) as follows:

$$\begin{aligned} \langle (\mathbf{b} - \mathbf{A}\mathbf{v})^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} (\mathbf{b} - \mathbf{A}\mathbf{v}) \rangle &= \\ \langle \mathbf{v}^T \mathbf{A}^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} \mathbf{A} \mathbf{v} \rangle - 2\mathbf{b}^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} \mathbf{A} \hat{\mathbf{v}}_{\lambda^*} + \mathbf{b}^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} \mathbf{b}. \end{aligned} \quad (75)$$

The first term on the right hand side of (75) can be rewritten as:

$$\langle \mathbf{v}^T \mathbf{A}^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} \mathbf{A} \mathbf{v} \rangle = \text{tr}(\Omega^{1/2} \mathbf{U}^T \langle \mathbf{v} \mathbf{v}^T \rangle \mathbf{U} \Omega^{1/2}), \quad (76)$$

where \mathbf{U} and Ω are defined by the eigen-value decomposition of $\mathbf{A}^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} \mathbf{A}$, *i.e.*,

$$\mathbf{A}^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} \mathbf{A} = \mathbf{U} \Omega \mathbf{U}^T. \quad (77)$$

Now,

$$\langle \mathbf{v} \mathbf{v}^T \rangle = \mathbf{\Gamma} + \hat{\mathbf{v}}_{\lambda^*} \hat{\mathbf{v}}_{\lambda^*}^T. \quad (78)$$

Plugging this expression in (76), we obtain

$$\begin{aligned} &\text{tr}(\Omega^{1/2} \mathbf{U}^T \langle \mathbf{v} \mathbf{v}^T \rangle \mathbf{U} \Omega^{1/2}) \\ &= \text{tr}(\Omega^{1/2} \mathbf{U}^T \mathbf{\Gamma} \mathbf{U} \Omega^{1/2}) + \text{tr}(\Omega^{1/2} \mathbf{U}^T \hat{\mathbf{v}}_{\lambda^*} \hat{\mathbf{v}}_{\lambda^*}^T \mathbf{U} \Omega^{1/2}) \end{aligned} \quad (79)$$

$$= \text{tr}(\mathbf{\Gamma} \mathbf{U} \Omega^{1/2} \Omega^{1/2} \mathbf{U}^T) + \hat{\mathbf{v}}_{\lambda^*}^T \mathbf{U} \Omega \mathbf{U}^T \hat{\mathbf{v}}_{\lambda^*} \quad (80)$$

$$= \text{tr}(\mathbf{\Gamma} \mathbf{A}^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} \mathbf{A}) + \hat{\mathbf{v}}_{\lambda^*}^T \mathbf{A}^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} \mathbf{A} \hat{\mathbf{v}}_{\lambda^*} \quad (81)$$

where we use the fact that $\text{tr}(\mathbf{A}\mathbf{B}) = \text{tr}(\mathbf{B}\mathbf{A})$ for square invertible matrices. Combining (65), (75) and (81), we finally get:

$$\begin{aligned} 2\nabla_{\beta} \log p(\mathbf{b}|\hat{\mathbf{z}}_{\lambda^*}, \beta, \gamma_2, \zeta_2) &= \beta^{-1} |\mathcal{S}_r| \\ &- \text{tr}(\mathbf{\Gamma} \mathbf{A}^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} \mathbf{A}) - (\mathbf{b} - \mathbf{A}\hat{\mathbf{v}}_{\lambda^*})^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} (\mathbf{b} - \mathbf{A}\hat{\mathbf{v}}_{\lambda^*}). \end{aligned} \quad (82)$$

The trace operator in (82) applies on a large and sparse matrix. However, it can easily be evaluated by randomization. Let \mathbf{I}^n represent a n -dimensional identity matrix. More precisely, for $J \gg 1$ random samples of a n -dimensional normalized and centered Gaussian distribution $\mathbf{r}_j \sim \mathcal{N}(0, \mathbf{I}^n)$, by the weak law of large numbers one gets the approximation that:

$$\begin{aligned} \text{tr}\{\mathbf{\Gamma} \mathbf{A}^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} \mathbf{A}\} &= \text{tr}\langle \mathbf{\Gamma} \mathbf{r}_j \mathbf{r}_j^T \mathbf{A}^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} \mathbf{A} \rangle \\ &= \langle (\mathbf{\Gamma} \mathbf{r}_j)^T \mathbf{A}^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} \mathbf{A} \mathbf{r}_j \rangle \\ &\simeq \frac{1}{J} \sum_{j=1}^J (\mathbf{\Gamma} \mathbf{r}_j)^T \mathbf{A}^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} \mathbf{A} \mathbf{r}_j, \end{aligned} \quad (83)$$

where $\langle \cdot \rangle$ denotes here the expectation with respect to the n -dimensional Gaussian and where the vector $\mathbf{\Gamma} \mathbf{r}_j$ is the solution provided by the CGS algorithm of the problem $\mathbf{\Gamma}^{-1} \mathcal{X} = \mathbf{r}_j$, where \mathcal{X} is the unknown.

APPENDIX II

EXPRESSION OF THE MARGINALIZED LIKELIHOOD

Introducing the definitions (45)-(47) in the integral (56), we obtain:

$$p(\mathbf{b}|\hat{\mathbf{z}}_{\lambda^*}, \beta, \gamma_2, \zeta_2) = \int p(\mathbf{b}|\mathbf{v}, \hat{\mathbf{z}}_{\lambda^*}, \beta, \gamma_2, \zeta_2) p(\mathbf{v}|\hat{\mathbf{z}}_{\lambda^*}, \beta, \gamma_2, \zeta_2) d\mathbf{v} \quad (84)$$

$$= Z_{\mathbf{b}}^{-1} Z_{\mathbf{v}}^{-1} \int \exp \left\{ -\frac{(\mathbf{b} + \mathbf{A}\mathbf{v})^T \beta \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} (\mathbf{b} + \mathbf{A}\mathbf{v})}{2} \right\} \exp \left\{ -\frac{\beta \mathbf{v}^T (\sum_{\ell} \lambda_{\ell}^* \mathbf{D}_{\ell}^T \mathbf{D}_{\ell}) \mathbf{v} - \gamma_2 \ell^{\zeta_2}}{2} \right\} d\mathbf{v}. \quad (85)$$

Let us rewrite the integrant as

$$\begin{aligned} &\exp \left\{ -\frac{1}{2} (\mathbf{v} - \hat{\mathbf{v}}_{\lambda^*})^T \mathbf{\Gamma}^{-1} (\mathbf{v} - \hat{\mathbf{v}}_{\lambda^*}) \right\} \\ &\exp \left\{ \frac{1}{2} \hat{\mathbf{v}}_{\lambda^*}^T \mathbf{\Gamma}^{-1} \hat{\mathbf{v}}_{\lambda^*} \right\} \exp \left\{ -\frac{1}{2} (\beta \mathbf{b}^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} \mathbf{b} - \gamma_2 \ell^{\zeta_2}) \right\}, \end{aligned} \quad (86)$$

where the posterior covariance $\mathbf{\Gamma}$ has been defined in (74). Now, the last two factors in (86) do not depend on \mathbf{v} , and the integration of the first term alone gives:

$$\int \exp \left\{ -\frac{1}{2} (\mathbf{v} - \hat{\mathbf{v}}_{\lambda^*})^T \mathbf{\Gamma}^{-1} (\mathbf{v} - \hat{\mathbf{v}}_{\lambda^*}) \right\} d\mathbf{v} = \sqrt{\det \mathbf{\Gamma}}. \quad (87)$$

Therefore, we have:

$$\begin{aligned} \log p(\mathbf{b}|\hat{\mathbf{z}}_{\lambda^*}, \beta, \boldsymbol{\theta}) &= \frac{1}{2} \log \det \beta \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} + \frac{1}{2} \log \det \left(\beta \sum_{\ell} \lambda_{\ell}^* \mathbf{D}_{\ell}^T \mathbf{D}_{\ell} \right) \\ &- \frac{1}{2} \log \det \mathbf{\Gamma}^{-1} + \frac{1}{2} \hat{\mathbf{v}}_{\lambda^*}^T \mathbf{\Gamma}^{-1} \hat{\mathbf{v}}_{\lambda^*} \\ &- \frac{1}{2} (\beta \mathbf{b}^T \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} \mathbf{b} - \gamma_2 \ell^{\zeta_2}). \end{aligned} \quad (88)$$

Since we have (21), using the definition (73)-(74), this can also be rewritten as:

$$\begin{aligned} \log p(\mathbf{b}|\hat{\mathbf{z}}_{\lambda^*}, \beta, \gamma_2, \zeta_2) &= \\ &\frac{1}{2} \log \det \left(\beta \sum_{\ell} \lambda_{\ell}^* \mathbf{D}_{\ell}^T \mathbf{D}_{\ell} \right) - \frac{1}{2} \log \det \mathbf{\Gamma}^{-1} \\ &+ \frac{1}{2} \log \det \beta \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} - \frac{1}{2} (\mathbf{b} - \mathbf{A}\hat{\mathbf{v}}_{\lambda^*})^T \beta \Lambda_{\hat{\mathbf{z}}_{\lambda^*}} (\mathbf{b} - \mathbf{A}\hat{\mathbf{v}}_{\lambda^*}). \end{aligned} \quad (89)$$

The prior used here is a degenerated distribution and its Hessian is not a full rank matrix *i.e.*, its determinant is equal to zero. We need to impose some Dirichlet boundary conditions before evaluating the logarithm of its determinant appearing in the first term on the right hand side of (89). Note that the precise value on the boundaries do not need to be specified. Indeed, considering some Dirichlet boundary conditions results in slightly modifying the form of the Hessian matrix but its content remains independent of the boundary function.