

Bayesian design of control space for optimal assimilation of observations. Part II: Asymptotic solutions

M. Bocquet* and L. Wu

Université Paris-Est, CERA, Joint laboratory École des Ponts ParisTech and EDF R&D, Champs-sur-Marne, France and INRIA, Paris Rocquencourt Research Centre, France

*Correspondence to: M. Bocquet, CERA, École des Ponts ParisTech, 6–8 avenue Blaise Pascal, Cité Descartes Champs-sur-Marne 77455 Marne la Vallée Cedex, France. E-mail: bocquet@cerea.enpc.fr

A consistent formalism for a Bayesian design of control space for an optimal assimilation of observations was proposed in Part I of this two-part article. This optimal discretization of control space leads to an efficient data assimilation scheme implementation. However, the construction of the grid itself, prior to its use for data assimilation, requires an optimization that may be challenging for high-dimensional systems. This paper derives analytical solutions for these optimal grids in the limit where the discretization of control space has a large number of grid cells. Analytical solutions for the density of grid cells are obtained for the so-called tilings, qtrees and frees, that represent different types of adaptive grids, with more or fewer degrees of freedom. These analytical solutions are explicit and the algorithms that allow densities to be converted into discrete adaptive grids are costless.

The approach is tested with a simplified physics in the Jacobian matrix in a tracer dispersion context in which radionuclides are monitored by the global observation network operated by the Comprehensive Nuclear Test Ban Treaty Organisation of the United Nations. The asymptotic solutions are then compared to the optimal grids obtained from the methodology perfected in Part I. In this example, and using qtree representations, the discrepancy between the approximate solution and the exact solution almost vanishes when the number of grid cells represents as few as 1% of the total number of grid cells in the finest grid. This opens the way to the application of this multiscale data assimilation framework to computationally challenging problems. Copyright © 2011 Royal Meteorological Society

Key Words: multiscale data assimilation; control space; tracer dispersion

Received 30 September 2010; Revised 1 April 2011; Accepted 6 April 2011; Published online in Wiley Online Library 17 May 2011

Citation: Bocquet M, Wu L. 2011. Bayesian design of control space for optimal assimilation of observations. Part II: Asymptotic solutions. *Q. J. R. Meteorol. Soc.* 137: 1357–1368. DOI:10.1002/qj.841

1. Introduction

The concept of an optimal representation of control space was introduced by Bocquet (2009), building on the standpoint developed in the seminal work of Rodgers (2000) in the context of remote sounding. The idea is to define a discretization of large parameter spaces that best accounts for the observations. Those spaces are typically met in geophysical and environmental problems where fields of

forcing parameters are uncertain, such as in atmospheric chemistry, where the emissions are poorly known. The theory was applied to the inverse modelling of source of atmospheric tracers. In many data assimilation experiments, such as the inversions of air quality pollutant sources or greenhouse gases fluxes, one is interested in the reduction of uncertainty achieved by the assimilation of observations. It was shown that optimal adaptive grids of control space can yield a reduction of uncertainty equivalent to a highly

resolved regular grid, but with a far fewer number of grid cells.

1.1. Selected results from Part I

In Part I of this work (Bocquet *et al.*, 2011), the optimal representation theory was perfected. The multiscale aspect of the theory was made Bayesian, allowing for a consistent use of background information on control space parameters.

One considers the typical inverse modelling problem

$$\boldsymbol{\mu} = \mathbf{H}\boldsymbol{\sigma} + \boldsymbol{\epsilon}. \quad (1)$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$ is the vector of observations, \mathbf{H} is the Jacobian matrix of the problem (linear or linearized), $\boldsymbol{\sigma} \in \mathbb{R}^N$ is the vector of parameters, which is defined in control space, and $\boldsymbol{\epsilon}$ is the vector of observational error. The typical data assimilation problem related to this equation assumes some prior statistical information on the errors that follow a Gaussian distribution $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$, and on the parameters that follow the Gaussian background-error statistics $\boldsymbol{\sigma} \sim \mathcal{N}(\boldsymbol{\sigma}_b, \mathbf{B})$. For a general representation (or discretization) of control space ω , the observation equation would read:

$$\boldsymbol{\mu} = \mathbf{H}_\omega \boldsymbol{\sigma}_\omega + \boldsymbol{\epsilon}_\omega. \quad (2)$$

Such representation ω is an adaptive discretization made of cells of various form and sizes, each one representing a scalar variable, that compose a partition of the domain Ω of control space. Coarsening $\boldsymbol{\Gamma}_\omega$ and prolongation $\boldsymbol{\Gamma}_\omega^*$ operators are used to scale up or down these grid cells. The prolongation operator is derived using all available information from the background.

Assume these grid cells are aggregations of smaller grid cells defined on a regular finest grid with N_{fg} grid cells. Then the prolongation of the representation ω (with $N \leq N_{\text{fg}}$ grid cells) to the finest grid, followed by a coarsening back to ω , should correspond to the identity operator: $\boldsymbol{\Gamma}_\omega \boldsymbol{\Gamma}_\omega^* = \mathbf{I}_N$. However the reverse, coarsening from the finest grid to ω , and prolongating back to the finest grid, implies a loss of information, so that the resulting (affine) operator is not the identity but

$$\boldsymbol{\Gamma}_\omega^* \boldsymbol{\Gamma}_\omega = (\mathbf{I}_N - \boldsymbol{\Pi}_\omega) \boldsymbol{\sigma}_b + \boldsymbol{\Pi}_\omega, \quad (3)$$

where

$$\boldsymbol{\Pi}_\omega = \mathbf{B} \boldsymbol{\Gamma}_\omega^T (\boldsymbol{\Gamma}_\omega \mathbf{B} \boldsymbol{\Gamma}_\omega^T)^{-1} \boldsymbol{\Gamma}_\omega. \quad (4)$$

Aggregation errors that account for representativeness errors are taken into account in this framework. We called them scale-covariant errors because they follow

$$\boldsymbol{\epsilon}_\omega = \boldsymbol{\epsilon} + \mathbf{H} (\mathbf{I}_N - \boldsymbol{\Pi}_\omega) (\boldsymbol{\sigma} - \boldsymbol{\sigma}_b). \quad (5)$$

New objective functions for the design of the optimal representations were introduced: the degrees of freedom for the signal (DFS) $\mathcal{J}_\omega = N - \text{Tr}(\mathbf{B}_\omega^{-1} \mathbf{P}_\omega^d)$, which is a well-used criterion in data assimilation though not used for that purpose. A data-dependent criterion was also introduced.

In conjunction with scale-covariant errors, the DFS criterion takes the simple scale-dependent form

$$\mathcal{J}_\omega = \text{Tr} \left\{ \boldsymbol{\Pi}_\omega \mathbf{B} \mathbf{H}^T (\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T)^{-1} \mathbf{H} \right\}. \quad (6)$$

We shall use this objective function for the design of representations in the rest of this article.

The adaptive grids are optimized on a dictionary of representations. A large dictionary was composed of general tilings. Each tiling is a set of rectangles (or tiles) that partition control space. Definition and implementation can be found in Part I. Figure 4 (this paper) also offers an illustration. By construction the subset of *qtrees* provides adaptive grids that are less efficient than optimal tilings (Figure 2 in the current paper provides an illustration). However, it was shown in Part I that the discrepancy is small. Besides the *qtrees* are expected to be computationally more efficient than the general tilings, as explained below.

1.2. Computational costs

The numerical optimization of the grid of control space entails the minimization of a functional. This functional depends on Lagrange parameters. Among them, N_{fg} parameters enforce the one-point one-tile constraints and a single one enforces the number of tiles of the representation. The optimizations are carried out with the L-BFGS-B quasi-Newton minimizer (Byrd *et al.*, 1995), on $N_{\text{fg}} + 1$ variables. It is difficult to estimate *a priori* the number of iterations of the minimization since it is problem-dependent, and since it depends on choices made by the operator such as the stopping criterion. As a general rule though, the minimization of a quadratic functional to machine precision has cubic dependence in the number of variables ($(N_{\text{fg}} + 1)^3$ here). With the iterative BFGS minimization, each iteration computational cost scales like $(N_{\text{fg}} + 1)^2$ multiplications, whereas it scales like $L(N_{\text{fg}} + 1)$ multiplications for the limited memory L-BFGS, where L is the memory length. (Typically $L \simeq 10\text{--}30$ for high-dimensional applications.)

However, this does not account for the evaluation of the cost function and of its gradient which is a vector of $N_{\text{fg}} + 1$ components. Such evaluations are required by the quasi-Newton algorithm and they are needed at each iteration. For high-dimensional geophysical systems, most of the computational time would be spent there. The cost function has the form of a sum over all tiles of the multiscale structure. Hence, the computational cost of the functional is linear in the total number of tiles. The total number of tiles scales at most like $4N_{\text{fg}}$ in the general 2D tiling case, and at most like $4/3N_{\text{fg}}$ in the 2D *qtree* case. This explains why the *qtrees* are faster to optimize on than the general tilings, even though the number of grid cells in the finest grid N_{fg} is the same. In the examples of Part I, the optimization over the dictionary of *qtrees* were at least twice as fast as the optimization on the dictionary of general tilings. However, it does not have to match perfectly the 1/3 scaling since the sum of the cost function is parallelised in both cases with communication overheads. Also note that the regularisation of the functional used in Part I requires functions such as logarithm and exponential which are more costly than matrix-vector multiplications.

As a clearly distinct problem, one needs to compute Jacobian \mathbf{H} . For high-dimensional problems, most of the computational power can be dedicated to \mathbf{H} , running geophysical numerical models. But once \mathbf{H} is computed and stored, optimizations can be performed without the need to re-compute it, except if models are nonlinear.

Finally the storage requirements of the multiscale Jacobian scales like the total number of tiles of the multiscale structure,

which was an argument in favour of the qtrees over the general tilings (three times more costly for a 2D domain) put forward in Bocquet (2009).

1.3. Objectives

For an application where the observation locations and schedule are known *a priori*, the optimization on the dictionary of grids can be performed *a priori* once and for all subsequent data assimilation analyses. However, even for moderately high-dimensional Jacobians, the optimization can be computationally challenging.

The objective of this Part II article is to introduce sub-optimal analytical solutions to the problem of the construction of optimal adaptive grids. A continuum or asymptotic limit of the problem will be first defined. An optimization will be analytically performed in the continuum limit framework. As a result, a density of tiles will be obtained. A discretization algorithm will then be needed to build discrete representations of control space, using those continuum densities.

Note that this is a constructive approach. The overall interest of the theory must be judged on the quantitative performance of the representation that it yields. This performance is objectively measured by an objective function such as the DFS.

1.4. Outline

The results of this article will be illustrated using a problem of interest for the Comprehensive Nuclear Test Ban Treaty Organisation (CTBTO) of the United Nations, but using simplified physics in the Jacobian. Details of the setup are given in section 2. The asymptotic analytical solutions are derived in section 3. They are introduced with increasing complexity. The continuum limit will be first derived in the one-dimensional (1D) case, because the limiting density of tiles is expected to be asymptotically exact. The multidimensional case is then treated but depends on the type of dictionary employed: ftrees, qtrees, or tilings. Focussing on the general tilings and qtrees, the construction of a discrete representation of control space using these analytical densities is then discussed, and simple algorithms are proposed. In some cases, the analytical densities may be improper (they cannot be normalized to one). This corresponds to a problem uncovered by Bocquet (2005), and it is dealt with in this context. In section 4, several of these results are illustrated on the CTBTO test case. We summarize the results and conclude in section 5.

2. The CTBTO atmospheric monitoring network

The formalism recalled above and the following theoretical development will be illustrated on a tracer dispersion problem at global scale.

The Comprehensive Test Ban Treaty signed by 182 states bans nuclear explosions (United Nations, 1996). The verification of the treaty is implemented by the United Nations CTBTO, based in Vienna, Austria. It operates an International Monitoring System (IMS) and collects seismic, infrasound, and hydro-acoustic data as well as radionuclide (particulate matter and noble gases) activity concentrations. These observations could be used for inverse modelling

purposes to track the origin of a nuclear test, and characterize it (signature, intensity).

As for the CarboEurope-IP test case of Part I, a much simpler prototype of the radionuclide IMS network is used here, with only 79 annual-mean measurements, one for each of the 79 radionuclide monitoring stations. The Jacobian matrix is computed similarly, assuming a power law for each of the adjoint solutions

$$c_i^*(r) = \frac{1}{r^\alpha}, \tag{7}$$

where r is the great-circle distance separating the observation point from the point where this sensitivity is being computed. The exponent $\alpha \simeq 2.4$ is chosen heuristically following Roustan and Bocquet (2006), and comes from the rate of decrease of an average footprint from the observation point. It must be understood as an average midlatitude value. Other choices of α will be made in section 4, to account for different diffusion conditions. The Jacobian entries are given by $[\mathbf{H}]_{ik} = [c_i^*]_k$, where c_i^* is a discretized adjoint solution.

The multiscale structure is built on a finest regular Mercator grid on the globe. It has dimensions $N_x^f = 512$ and $N_y^f = 256$. The number of cells of size $\Delta_x = \Delta_y = 0.703125^\circ$ in this grid is $N_{fg} = 131\,072$. A multiscale structure of eight levels for each direction is defined. It means that the finest grid cells can be coarse-grained 8 times following a dyadic hierarchy. Hence the coarser regular grid, made of the coarser tiles, has dimensions $N_x^c = 4$ and $N_y^c = 2$.

3. Approximate solution from a continuum limit

We would like to investigate the optimization of the DFS criterion Eq. (6) in the asymptotic limit where the grid spacing tends to zero (continuum limit). The criterion can also be written

$$\mathcal{J}_\omega = \text{Tr} \left\{ \tilde{\mathbf{\Pi}}_\omega \tilde{\mathbf{H}}^T (\mathbf{I}_p + \tilde{\mathbf{H}} \tilde{\mathbf{H}}^T)^{-1} \tilde{\mathbf{H}} \right\}, \tag{8}$$

where $\tilde{\mathbf{H}} = \mathbf{R}^{-1/2} \mathbf{H} \mathbf{B}^{1/2}$ is the Jacobian for which the statistics are normalized and uncorrelated in both state and observation space (following e.g. Rodgers, 2000), and where $\tilde{\mathbf{\Pi}}_\omega = \mathbf{B}^{-1/2} \mathbf{\Pi}_\omega \mathbf{B}^{1/2}$ is a symmetric projector.

The singular value decomposition of the normalized Jacobian matrix is

$$\tilde{\mathbf{H}} = \mathbf{R}^{-1/2} \mathbf{H} \mathbf{B}^{1/2} = \mathbf{V} \mathbf{D} \mathbf{U}^T, \tag{9}$$

where $\mathbf{V} \in \mathbb{R}^{p \times p}$ satisfies $\mathbf{V} \mathbf{V}^T = \mathbf{V}^T \mathbf{V} = \mathbf{I}_p$, $\mathbf{U} \in \mathbb{R}^{N_{fg} \times p}$ satisfies $\mathbf{U}^T \mathbf{U} = \mathbf{I}_{N_{fg}}$, and $\mathbf{D} \in \mathbb{R}^{p \times p}$ is the diagonal matrix of positive singular values $\{\lambda_i\}_{i=1, \dots, p}$ (assuming $p \leq N_{fg}$). As a result, the objective function can be reduced to

$$\begin{aligned} \mathcal{J}_\omega &= \text{Tr} \left\{ \mathbf{D}^2 (\mathbf{I}_p + \mathbf{D}^2)^{-1} \mathbf{U}^T \mathbf{\Pi}_\omega \mathbf{U} \right\} \\ &= \sum_{i=1}^p \frac{\lambda_i^2}{1 + \lambda_i^2} \mathbf{u}_i^T \mathbf{\Pi}_\omega \mathbf{u}_i, \end{aligned} \tag{10}$$

where the \mathbf{u}_i are the p singular vectors in $\mathbb{R}^{N_{fg}}$. In the following we will denote

$$f(\lambda_i) = \frac{\lambda_i^2}{1 + \lambda_i^2} \tag{11}$$

as the spectral signature of the criterion. Another example is given by the objective function $\text{Tr}(\mathbf{B}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})$ mentioned in Part I, and which we have called the Fisher criterion; it has signature $f(\lambda_i) = \lambda_i^2$.

Assuming the restriction operator has been redefined according to Part I ($\Gamma_\omega\mathbf{B}^{-1/2}$), then the projector operator $\tilde{\Pi}_\omega$ can be written:

$$\tilde{\Pi}_\omega = \sum_1 \sum_{k=1}^{m_1} \alpha_{1,k}^\omega \frac{\mathbf{v}_{1,k} \mathbf{v}_{1,k}^T}{\mathbf{v}_{1,k}^T \mathbf{v}_{1,k}}, \quad (12)$$

where the $\mathbf{v}_{1,k}$ are vectors of $\mathbb{R}^{N_{\text{fig}}}$ that represent tiles of any scale (defined by \mathbf{I}) and position (defined by k), in the multiscale structure. They have been defined more precisely in Part I.

3.1. Asymptotic limit in the one-dimensional case

As a first cautious step, we consider the 1D case. The interval $[x_b, x_f]$ is partitioned with

$$x_b = x_0 < x_1 < x_2 < \dots < x_{N-1} < x_N = x_f.$$

The number N of intervals $[x_k, x_{k+1}]$ is fixed, whereas the nodes' positions x_k for $k = 1, \dots, N - 1$ are to be optimized. Here, there is no need to assume any discrete structure, because of the natural order in $[x_b, x_f]$. But it can be done here for the sake of simplicity: the x_k are supposed to take values in a fine regular grid

$$\left\{ x_b + (x_f - x_b) \frac{j}{N_{\text{fig}}} \right\}_{j=0, \dots, N_{\text{fig}}}.$$

Using Eqs (10) and (12), one obtains

$$\mathcal{J}_\omega = \sum_{i=1}^p f(\lambda_i) \mathbf{u}_i^T \tilde{\Pi}_\omega \mathbf{u}_i = \sum_{i=1}^p f(\lambda_i) \sum_{k=1}^N \frac{(\mathbf{v}_k^T \mathbf{u}_i)^2}{\mathbf{v}_k^T \mathbf{v}_k}, \quad (13)$$

where the vector $\mathbf{v}_k \in \mathbb{R}^{N_{\text{fig}}}$ are defined similarly to the $\mathbf{v}_{1,k}$. They have components equal to 1 for entries between x_{k-1} and x_k , and have null components elsewhere.

Assuming at this stage that $\tilde{\mathbf{H}}$ (and its singular vector decomposition) has a continuum limit, the singular vector $\mathbf{u}_i \in \mathbb{R}^{N_{\text{fig}}}$ has the functional limit $u_i(x)$. It is assumed that it is normalized:

$$\int_{x_b}^{x_f} dx u_i^2(x) = 1, \quad (14)$$

provided it can be (which will be questioned later). Then, one shows

$$\mathcal{J}_\omega = \sum_{i=1}^p f(\lambda_i) \sum_{k=0}^{N-1} \frac{1}{x_{k+1} - x_k} \left(\int_{x_k}^{x_{k+1}} dx u_i(x) \right)^2. \quad (15)$$

If u_i is assumed sufficiently regular, one obtains the following asymptotic expansion when N is large, or, more precisely, when $0 \leq N_{\text{fig}} - N \ll N_{\text{fig}}$:

$$\begin{aligned} \left(\int_{x_k}^{x_{k+1}} dx u_i(x) \right)^2 &\simeq (x_{k+1} - x_k) \int_{x_k}^{x_{k+1}} dx u_i^2(x) \\ &- \frac{1}{12} (x_{k+1} - x_k)^4 (\partial_x u_i)^2 \Big|_{\frac{x_k+x_{k+1}}{2}}. \end{aligned} \quad (16)$$

Then, at the next leading order, one gets

$$\begin{aligned} \mathcal{J}_\omega &= \sum_{i=1}^p f(\lambda_i) \sum_{k=0}^{N-1} \frac{1}{x_{k+1} - x_k} \left(\int_{x_k}^{x_{k+1}} dx u_i(x) \right)^2 \\ &\simeq \sum_{i=1}^p f(\lambda_i) \int_{x_b}^{x_f} dx u_i^2(x) \\ &- \frac{1}{12} \sum_{i=1}^p f(\lambda_i) \sum_{k=0}^{N-1} (x_{k+1} - x_k)^3 (\partial_x u_i)^2 \Big|_{\frac{x_k+x_{k+1}}{2}}. \end{aligned} \quad (17)$$

The first term of this expansion is merely the maximum reachable value of the cost function

$$\sum_{i=1}^p f(\lambda_i). \quad (18)$$

It does not depend on the grid. The second term depends on the positions of the grid's nodes x_k . To define a continuum limit of this term, we first define an ancillary variable ξ taking values in the interval $[0, 1]$. Then, for each singular value $i = 1, \dots, p$

$$\begin{aligned} &\sum_{k=0}^{N-1} (x_{k+1} - x_k)^3 (\partial_x u_i)^2 \Big|_{\frac{x_k+x_{k+1}}{2}} \\ &\simeq \sum_{k=0}^{N-1} \left(x \left[\frac{k+1}{N} \right] - x \left[\frac{k}{N} \right] \right)^3 (\partial_x u_i)^2 \left(x \left[\frac{k}{N} \right] \right) \\ &\simeq \frac{1}{N^3} \sum_{k=0}^{N-1} \left\{ N \left(x \left[\frac{k+1}{N} \right] - x \left[\frac{k}{N} \right] \right) \right\}^3 (\partial_x u_i)^2 \left(x \left[\frac{k}{N} \right] \right) \\ &\underset{N \rightarrow \infty}{\sim} \frac{1}{N^3} \int_0^1 d\xi \left(\frac{dx}{d\xi} \right)^3 (\partial_x u_i)^2 \{x(\xi)\} \\ &\underset{N \rightarrow \infty}{\sim} \frac{1}{N^3} \int_{x_b}^{x_f} dx \left(\frac{dx}{d\xi} \right)^2 (\partial_x u_i)^2(x). \end{aligned} \quad (19)$$

The coordinate x has been considered a function of the parameter ξ . Note that $dx/d\xi$ measures the resolution of the network in the interval $[x_b, x_f]$. Its inverse measures the density of the partition points x_k in the asymptotic limit $N \rightarrow \infty$:

$$\rho(x) = \frac{d\xi}{dx}. \quad (20)$$

Keeping only the leading-order term that depends on this density, and which scales *a priori* like N^{-3} , the cost function \mathcal{J}_ω has the continuum limit:

$$\mathcal{J}_\rho = \int_{x_b}^{x_f} dx \rho^{-2}(x) \left(\sum_{i=1}^p f(\lambda_i) (\partial_x u_i)^2(x) \right). \quad (21)$$

Because of the change of sign, it ought to be minimized if \mathcal{J}_ω is to be maximized. One is looking for a normalized density such that

$$\int_{x_b}^{x_f} dx \rho(x) = 1.$$

The functional minimization of \mathcal{J}_ρ under that constraint yields the density

$$\rho^*(x) \propto \left(\sum_{i=1}^p f(\lambda_i) (\partial_x u_i)^2(x) \right)^{1/3}. \quad (22)$$

We expect that this result is the exact asymptotic density, since no approximation was used in this limit. The multidimensional case is less simple, since it depends on the type of discrete dictionary the continuum limit is taken from.

3.2. Asymptotic limit in the multidimensional case

The 2D case $d = 2$ will mainly be investigated, but a selection of generalizing results will be given in higher dimension.

3.2.1. Ftrees

Let us first assume that the tiling belongs to the subset defined as the direct product of two binary trees, one for each direction. It belongs to the dictionary of the *ftrees*. It has been shown in Part I that they offer poor performances compared with other representation classes. Nevertheless, in this analytical study, they correspond to the missing link between a presumably asymptotically exact solution in the 1D case and approximate solutions in the multidimensional cases, as the direct product of two 1D structures.

Then, in the asymptotic limit, one can define a density $\rho_x(x)$ and a density $\rho_y(y)$ that measure the density of the mesh in each of the two directions. It is possible to reason as in the 1D case so that one can re-use Eq. (21). An asymptotic cost function, constrained by the normalization conditions of the densities is

$$\begin{aligned} \mathcal{J}_\rho = & \int_{x_b}^{x_f} \int_{y_b}^{y_f} dx dy \rho_x^{-2} \left(\sum_{i=1}^p f(\lambda_i) (\partial_x u_i)^2 \right) \\ & + \int_{x_b}^{x_f} \int_{y_b}^{y_f} dx dy \rho_y^{-2} \left(\sum_{i=1}^p f(\lambda_i) (\partial_y u_i)^2 \right) \\ & + \gamma_x \left(\int_{x_b}^{x_f} dx \rho_x - 1 \right) + \gamma_y \left(\int_{y_b}^{y_f} dy \rho_y - 1 \right), \end{aligned} \quad (23)$$

where the Lagrange multipliers γ_x and γ_y impose the normalization of the densities. This leads to the optimal densities defined on the domain $[x_b, x_f] \times [y_b, y_f]$:

$$\left. \begin{aligned} \rho_x^* & \propto \left[\int_{y_b}^{y_f} dy \left(\sum_{i=1}^p f(\lambda_i) (\partial_x u_i)^2 \right) \right]^{1/3}, \\ \rho_y^* & \propto \left[\int_{x_b}^{x_f} dx \left(\sum_{i=1}^p f(\lambda_i) (\partial_y u_i)^2 \right) \right]^{1/3}. \end{aligned} \right\} \quad (24)$$

In dimension $d > 2$, the calculation is very similar and leads, for any $i = 1, \dots, d$, to

$$\rho_{x_i}^* \propto \left[\int \prod_{j=1, j \neq i}^d dx_j \left(\sum_{l=1}^p f(\lambda_l) (\partial_{x_i} u_l)^2 \right) \right]^{1/3}. \quad (25)$$

Applied to the power law influence function $c^*(r) = r^{-\alpha}$ and in dimension $d = 2$, the total density is

$$\rho^* \propto \frac{\sqrt{|x||y|}}{r^{\frac{2}{3}(\alpha+2)}}, \quad (26)$$

reasonably far from the singularity $r = 0$. Even though the influence function is rotationally invariant, the solution for the density of tiles is not. This is not surprising, since all *ftrees* are very anisotropic, with quite a limited number of degrees of freedom.

3.2.2. Qtrees

Let us assume that we wish to extract the asymptotic limit from a quaternary tree structure in dimension $d = 2$. We consider one square of this *qtree* of size $[a, b] \times [a, b]$ in the local coordinates' system. The contribution of this square and of the singular vector u_i to the cost function \mathcal{J}_ω has the expansion

$$\begin{aligned} & (b-a)^2 \int_{[a,b]^2} dx dy u_i^2(x, y) \\ & - \frac{1}{12} (b-a)^6 \left\{ (\partial_x u_i)^2 + (\partial_y u_i)^2 \right\}_{\left| \left(\frac{a+b}{2}, \frac{a+b}{2} \right) \right.}, \end{aligned} \quad (27)$$

resulting from a similar Taylor expansion to the 1D case (Eq. (16)). Let us denote $\{S_k\}_{k=1, \dots, N}$, the N squares that partition the grid. The area of square S_k is $|S_k|$. Using Eq. (27), let us sum their contribution:

$$\begin{aligned} \mathcal{J}_\omega \simeq & \sum_{i=1}^p f(\lambda_i) \int_{\Omega} dx dy u_i^2(x, y) \\ & - \frac{1}{12} \sum_{i=1}^p f(\lambda_i) \sum_{k=1}^N |S_k|^2 \|\nabla u_i\|^2. \end{aligned} \quad (28)$$

Defining the 2D density ρ by the inverse of the area of the local square, $\rho \propto |S_k|^{-1}$, one obtains the asymptotic cost function

$$\begin{aligned} \mathcal{J}_\rho = & \int_{\Omega} dx dy \rho^{-1} \left(\sum_{i=1}^p f(\lambda_i) \|\nabla u_i\|^2 \right) \\ & + \gamma \left(\int_{\Omega} dx dy \rho - 1 \right), \end{aligned} \quad (29)$$

where γ enforces the normalization condition of the density. As a result, the optimal density is

$$\rho^* \propto \left(\sum_{i=1}^p f(\lambda_i) \|\nabla u_i\|^2 \right)^{1/2}. \quad (30)$$

In dimension d , a quaternary tree would be generalized to 2^d -tree, that is a tree with potentially 2^d daughter nodes for each mother node. The straightforward generalization of the 2D result to $d > 2$ yields

$$\begin{aligned} \mathcal{J}_\rho = & \int_{\Omega} d^d \mathbf{x} \rho^{-\frac{2}{d}} \left(\sum_{i=1}^p f(\lambda_i) \|\nabla u_i\|^2 \right) \\ & + \gamma \left(\int_{\Omega} d^d \mathbf{x} \rho - 1 \right). \end{aligned} \quad (31)$$

As a result, the d -dimensional optimal density is

$$\rho^* \propto \left(\sum_{i=1}^p f(\lambda_i) \|\nabla u_i\|^2 \right)^{\frac{d}{2+d}}. \quad (32)$$

Applied to the power law influence function $c^*(r) = r^{-\alpha}$ in dimension $d = 2$, where $r = \|\mathbf{x}\|$, the *radial* density equals $\rho^* \propto r^{-\alpha}$. Contrary to the *ftrees*, it is rotationally invariant. It could also be checked directly on Eq. (32), since the density is defined through the norm of a gradient.

3.2.3. General tilings

When the tiles are elements of the Kronecker product of two binary trees, then, in the asymptotic limit, ρ_x and ρ_y are functions of both x and y . In this case, the total density $\rho_x \rho_y$ ought to be normalized to one (there is a finite number of tiles), but the directional densities are not. We conjecture that the limiting optimal density over the domain Ω is inferred from the minimum of the following functional of the densities:

$$\begin{aligned} \mathcal{J}_\rho = & \int_\Omega dx dy \rho_x^{-2} \left(\sum_{i=1}^p f(\lambda_i) (\partial_x u_i)^2 \right) \\ & + \int_\Omega dx dy \rho_y^{-2} \left(\sum_{i=1}^p f(\lambda_i) (\partial_y u_i)^2 \right) \\ & + \gamma \left(\int_\Omega dx dy \rho_x \rho_y - 1 \right). \end{aligned} \tag{33}$$

This leads to the optimal density

$$\rho_x^* \rho_y^* \propto \left[\left(\sum_{i=1}^p f(\lambda_i) (\partial_x u_i)^2 \right) \left(\sum_{i=1}^p f(\lambda_i) (\partial_y u_i)^2 \right) \right]^{1/4}. \tag{34}$$

Note that, along the derivation, one also gets

$$\left. \begin{aligned} \rho_x^* & \propto \left[\sum_{i=1}^p f(\lambda_i) (\partial_x u_i)^2 \right]^{3/8} \left[\sum_{i=1}^p f(\lambda_i) (\partial_y u_i)^2 \right]^{-1/8}, \\ \rho_y^* & \propto \left[\sum_{i=1}^p f(\lambda_i) (\partial_x u_i)^2 \right]^{-1/8} \left[\sum_{i=1}^p f(\lambda_i) (\partial_y u_i)^2 \right]^{3/8}. \end{aligned} \right\} \tag{35}$$

More generally, in dimension d , assuming that the dictionary of representations is the Kronecker product of binary trees, one for each direction, the optimal density has the form

$$\rho^* \propto \left[\prod_{q=1}^d \left(\sum_{i=1}^p f(\lambda_i) (\partial_{x_q} u_i)^2 \right) \right]^{\frac{1}{2+d}}. \tag{36}$$

Applied to the power law influence function $c^*(r) = r^{-\alpha}$ and in dimension $d = 2$, the total density is

$$\rho^* \propto \frac{\sqrt{|x||y|}}{r^{\alpha+2}}. \tag{37}$$

Contrary to the qtrees, this density is not rotationally invariant.

Our understanding is that the general tilings allow singular objects in the continuum limit, which are directional. These are stretched tiles, which resembles lines or segments in the continuum limit. But, by construction, these can only be created in the Ox and Oy directions. Because of these directional objects, optimal solutions in the continuum limit may not be rotationally invariant, even though the underlying continuum problem is. On the contrary, for a qtree, each split or merging of tiles in one direction is accompanied by the same operation in the second direction of the qtree, so as to form coarser squares. Therefore the singular tiles that can be found in the continuum limit of a tiling cannot be created in the continuum limit of a qtree.

3.3. Construction of a discrete adaptive grid

Assume that an approximation ρ^* of the optimal discretization density is known for each point of the domain, such as the asymptotic densities derived earlier. One would like to build a qtree or a general tiling, with a fixed number of tiles, based on the continuum density ρ^* , or in the case of a general tiling, based on the two continuum densities ρ_x^* and ρ_y^* . This would be an alternative approach to the non-approximate but demanding variational method of Part I.

For the sake of simplicity and because of the examples below, it is assumed that the domain has two space dimensions, and no time dimension. The ftrees class will not be considered here because of its insufficient overall performance, which was demonstrated in Part I. For the qtrees and the general tilings, we shall adopt the following heuristic algorithms.

3.3.1. Qtree

Firstly, assume that the hierarchical structure follows a quaternary tree. Because there is a natural tree structure (the qtree itself), the algorithm can be conceived recursively, from the coarsest tiles to the finest ones. For each tile of size $[a, b] \times [a, b]$, one can define an index which corresponds to the integrated density ρ^* over that tile:

$$\int_{[a,b] \times [a,b]} dx dy \rho^*. \tag{38}$$

We start with the coarsest single tile which for the sake of simplicity is assumed to cover the full 2D domain. Its index is greater than one. Otherwise, the final optimal tiling identifies with this single coarse tile. Then the cell is split into four sub-tiles, if there exists a finer level (otherwise one ascribes the coarse tile to the final tiling). For each one of the four resulting tiles, the index can be computed. Take one of the tiles. If its index is lower or equal to 1, then keep this tile as a piece of the final tiling. Then take the second, third and fourth tiles and do the same. If the index is greater than 1, then one should proceed in the subdivision for each one of these three tiles.

3.3.2. General tiling

Secondly, assume that the hierarchical structure follows the Kronecker product of two binary trees. For each tile in the following, typically a box $[a, b] \times [c, d] \in \Omega$, one can define an index which corresponds to the integrated density in the Ox direction ρ_x^* or in the Oy direction ρ_y^* over that tile. Assume again for simplicity that the full 2D domain corresponds to a single coarsest tile. First, if the index product

$$\int_{[a,b] \times [c,d]} dx dy \rho_x^* \rho_y^*, \tag{39}$$

which is the total tile density, is lower than 1, the final optimal tiling identifies with this single coarse tile. Otherwise, two integrated indices can be computed for each one of the two directions. Then, the cell is split into two sub-tiles in the direction which corresponds to the greater index, either

$$\frac{1}{d-c} \int_{[a,b]} dx \rho_x^* \quad \text{or} \quad \frac{1}{b-a} \int_{[c,d]} dy \rho_y^*, \tag{40}$$

provided such a split is allowed in the hierarchical structure. For both resulting tiles, the index

$$\int_{[a,b] \times [c,d]} dx dy \rho_x^* \rho_y^* \tag{41}$$

can be computed. Take one of the tiles. If its index is lower or equal to 1, then keep this tile as a piece of the final near-optimal tiling. Otherwise, one should proceed in the subdivision. Compute the integrated indices related to ρ_x^* and ρ_y^* . Then, split the tile in the direction with the biggest index. Next, take the second tile and do the same, and so on.

The direct implementation of these two algorithms yields adaptive grids with the targeted density in most of the domain, but the total number of tiles can be far from the targeted number N . This phenomenon is explained and remedied in the following subsections.

3.4. Divergent and convergent cases

The algorithm described above entails that the normalization of the optimal densities are known, whereas the asymptotic solutions are unnormalized densities. To obtain an approximate normalization, one can estimate the normalization factor by summing the contributions of the unnormalized discrete density on each grid cell of the mesh where the singular vectors are defined.

In the previous sections, we have assumed that the integrals

$$\int_{\Omega} u_i^2(x_1, \dots, x_d) \prod_{q=1}^d dx_q \tag{42}$$

are proper (for $i = 1, \dots, d$), and densities like Eq. (32) and Eq. (36) are integrable. It turns out that there are relevant cases where this is not true. In fact, the examples that we have taken with an average adjoint solution of the form Eq. (7) and with $\alpha \simeq 2.4$, are divergent. This type of divergence has been explained by Bocquet (2005, 2009) and Saide *et al.* (2011), and is likely to affect data assimilation in air quality and atmospheric dispersion applications, where pointwise measurements are involved (which is so in most cases).

Using the findings of Bocquet (2005), a criterion of convergence for these integrals can be established. The singular vectors of the Jacobian matrix \mathbf{H} are combinations of the adjoint solutions. Close to the divergence, every singular vector identifies with one of the adjoint solutions attached to one of the observations. In this limit, it was shown that the diagonal elements of the innovation statistics behave like an integral over the full domain punctured by a small ball Ω_i of radius r_i around the observation point:

$$\int_{\Omega_i} d^d \mathbf{x} u_i^2(\mathbf{x}) \sim \int_{r_i}^{\infty} dr r^{d-1-2\alpha}. \tag{43}$$

Applied to the toy model Eq. (7), this leads to the convergence criterion $\alpha < d/2$. These integrals also appear as the first term of the expansions of the cost functions computed earlier.

The divergence can also be characterized on the grid density ρ^* directly from Eq. (32): the behaviour of ρ^* close to the observation site i is dominated by

$$\rho^*(r) \sim r^{\frac{d}{2+\alpha}(d-2\alpha)}, \tag{44}$$

which leads to the same criterion.

In practice, it means that two regimes can be expected when computing an optimal grid. Firstly, if the data assimilation problem is non-divergent ($\alpha < d/2$ in the toy model case), then one should not expect any divergence in the way the domain is gridded, and the above asymptotic results hold. Secondly, if the data assimilation problem is divergent ($\alpha > d/2$ in the toy model case), then the optimization will tend to refine the grid as much as possible in the vicinities of the observation sites. Then, the above asymptotic results do not hold as such.

The criterion of convergence is violated in the examples of Parts I and II. This is clear in the CarboEurope-IP and CTBTO cases, and assessed from the experiments of Bocquet (2005, 2009) in the case of ETEX-I, which uses a realistic dispersion model.

However, because there is an underlying finest regular grid, in these divergent cases the density will in fact saturate to the density of the finest grid.

3.5. Saturation zone

For the sake of simplicity, assume that the dictionary of representations stems from a 2^d -tree. Besides, the regime may be divergent. In this case, the asymptotic solution Eq. (32) does not hold. Instead the optimal density is the optimum of the functional

$$\begin{aligned} \mathcal{J}_{\rho} = & \int_{\Omega} d^d \mathbf{x} \rho^{-\frac{2}{d}} \left(\sum_{i=1}^p f(\lambda_i) \|\nabla u_i\|^2 \right) \\ & + \gamma \left(\int_{\Omega} d^d \mathbf{x} \rho - 1 \right) + \int_{\Omega} d^d \mathbf{x} \xi(\rho - \nu), \end{aligned} \tag{45}$$

where γ is the Lagrange multiplier enforcing the normalization of the density, and $\xi(\mathbf{x}) \geq 0$ is a function of multipliers that enforces a higher bound ν of the targeted density ρ . Let us define φ as a level of the function

$$\psi(\mathbf{x}) = \left(\sum_{i=1}^p f(\lambda_i) \|\nabla u_i(\mathbf{x})\|^2 \right)^{\frac{d}{2+d}}. \tag{46}$$

Then we define the sub-domain of Ω related to φ

$$\Omega_{\varphi} = \{\mathbf{x} \in \Omega, \psi(\mathbf{x}) \leq \varphi\}, \quad \Omega \setminus \Omega_{\varphi} = \{\mathbf{x} \in \Omega, \psi(\mathbf{x}) > \varphi\}. \tag{47}$$

In subdomain Ω_{φ} , the solution of the density is proportional to $\psi(\mathbf{x})$, whereas in the saturation zone $\Omega \setminus \Omega_{\varphi}$, it is equal to ν since $\xi(\mathbf{x})$ is non-zero on the saturation zone, and null elsewhere. One needs to determine the right level φ . The normalization of the density and the continuity of the density determine the right φ . Define the function

$$\Xi(\varphi) = \frac{1 - \nu |\Omega \setminus \Omega_{\varphi}|}{\int_{\Omega_{\varphi}} d^d \mathbf{x} \psi(\mathbf{x})}, \tag{48}$$

where $|\Omega \setminus \Omega_{\varphi}|$ is the area of $\Omega \setminus \Omega_{\varphi}$. Then the solution φ^* satisfies the equation

$$\Xi(\varphi)\varphi = \nu, \tag{49}$$

which enforces the continuity of the density. Since it is straightforward to compute Ξ numerically, the solution of this nonlinear 1D problem is easily obtained numerically. Once φ^* is determined, the solution reads

$$\left. \begin{aligned} \forall \mathbf{x} \in \Omega \setminus \Omega_{\varphi^*} : \rho(\mathbf{x}) &= \nu, \\ \forall \mathbf{x} \in \Omega_{\varphi^*} : \rho(\mathbf{x}) &= \Xi(\varphi^*) \psi(\mathbf{x}). \end{aligned} \right\} \quad (50)$$

3.6. Practical algorithms

Although this asymptotic solution is simple enough to implement, we found that a more practical approach can be used to construct a discretized representation that respects this asymptotic result. To obtain a discrete representation, even in the divergent case, one can apply directly the algorithms detailed in section 3.3. In the non-divergent case, the effective number of tiles will be equal or close to the target number of tiles. However, in the divergent case, this effective number could differ significantly from the target number. Yet, obviously, it is an increasing function of the target number. Therefore, one can use a dichotomy and vary the target number until the effective number matches the original target number. Because the construction of a representation from the asymptotic limit is costless, it is cheap to run such a dichotomic search. This turns out to be as efficient as simple so that it is used in the following.

4. Illustration on the CTBTO case

Let us illustrate the relevance and usefulness of this asymptotic approximations, but also explore their range of validity, on the CTBTO example.

4.1. Qtree

Optimal qtree grids are built from a wide range of tile numbers $N = 2^l$, with $l = 0, 2, \dots, 14$, using the DFS criterion Eq. (6). The DFS of each optimal qtree are plotted against the number of tiles in Figure 1(a) ($\alpha = 2.4$). The **B** and **R** diagonal entries are chosen such that the total DFS in the finest grid is 78.9936 rather than the ideal noise-free value 79 (because there are 79 observations, one for each station).

DFS values also given for regular grids to stress the optimality of these qtrees.

The asymptotic solutions are obtained within a few seconds, whereas the set of the 15 optimal representations is obtained within 12 hours on a standard 8-core computer. However the latter duration is very case-dependent, and depends considerably on the stopping criterion of the optimization and the required precision of the solution.

Comparing the DFS of corresponding adaptive grids, there is a clear discrepancy between the asymptotic qtree and the optimal one, especially for a low value of the total tile number N . This is clearly consistent with the fact that these approximations are asymptotic. It is however remarkable that for $N = 2048$, which represents only 1.6% of the total number of grid cells in the finest grid, the DFS of the asymptotic and the optimal qtree are very close: 78.58 and 78.79 respectively, which represent 99.5% and 99.7% respectively of the maximum DFS (which can only be achieved in the finest grid). Figure 2 shows that the

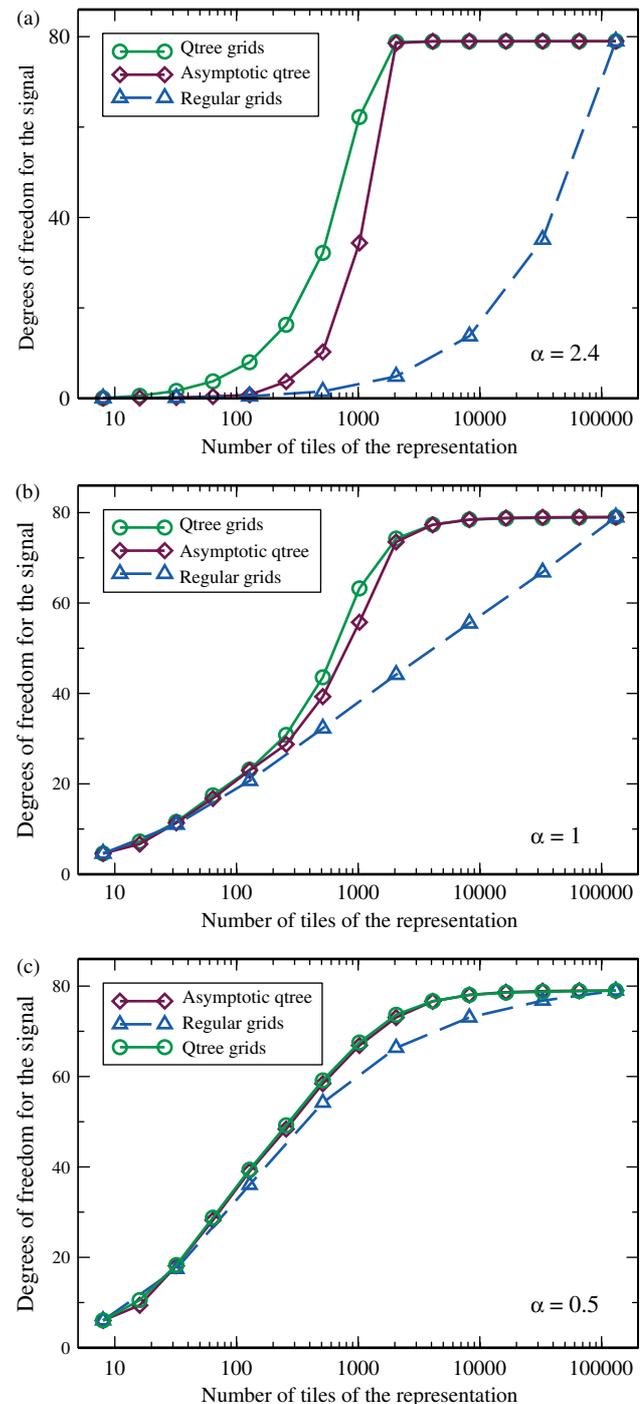


Figure 1. Degrees of freedom for the signal for optimal qtrees, obtained from the asymptotic limit approximation, and regular grids against the number of grid cells in the representation (CTBTO example). (a) corresponds to a Jacobian generated with the power law exponent $\alpha = 2.4$, (b) with $\alpha = 1$, and (c) with $\alpha = 0.5$.

corresponding qtrees are quite similar, not only in terms of DFS, but the structures are also quite close. We notice that 84% of the tiles in the two qtrees are shared.

Even though it cannot be seen on Figure 1, some of the asymptotic solutions slightly outperform the optimal grids obtained from the method of Part I for large N . This is due to the fact that the discrete optimization algorithm is performed at a large but nevertheless finite value of a regularisation parameter β , and is therefore slightly suboptimal, as explained in Part I.

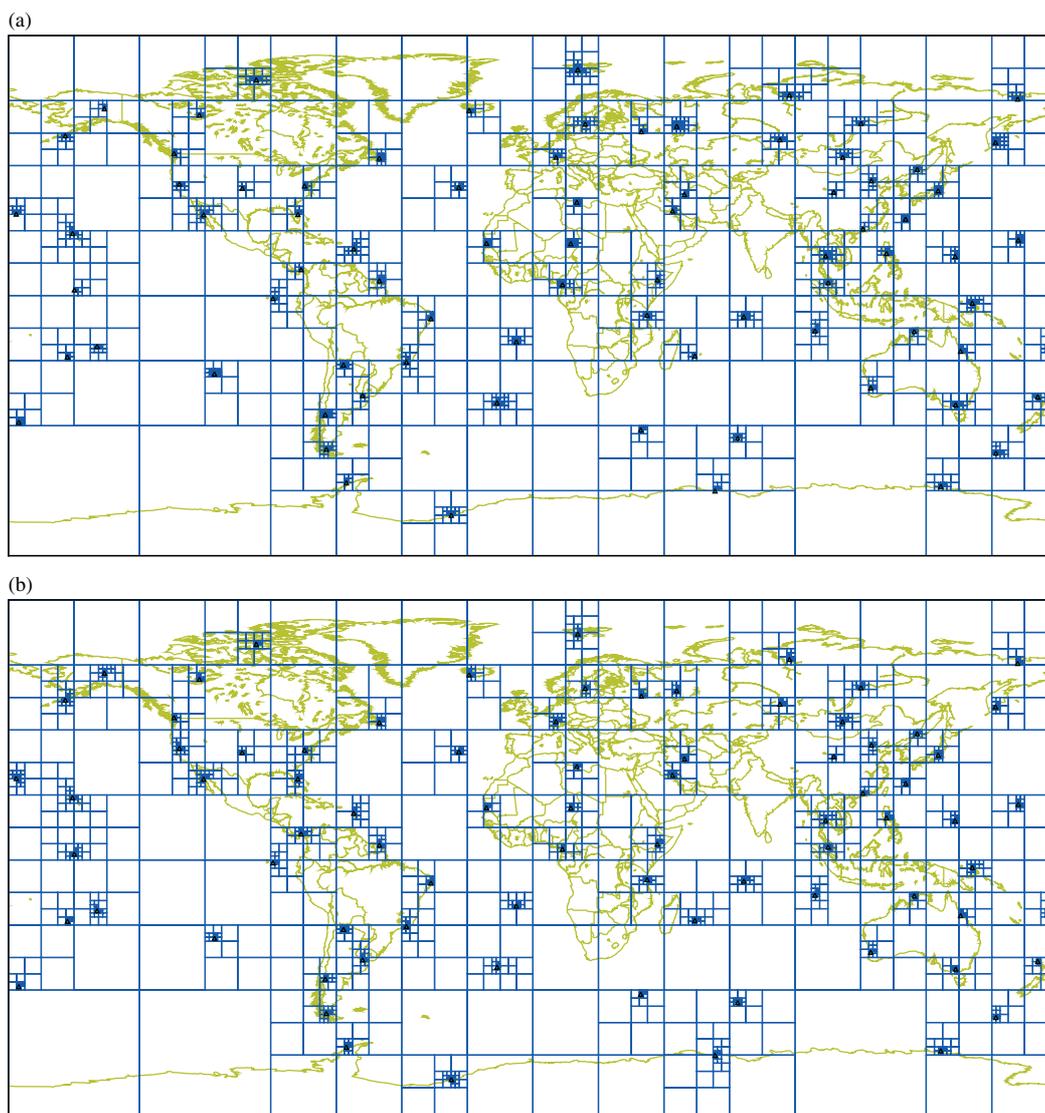


Figure 2. Visual comparison of optimal or near-optimal qtrees with $N = 2048$ obtained (a) from the plain optimization and (b) from the asymptotic result.

Similar results can be obtained but for different values of the power law exponent α of the influence function defined by Eq. (7). Figure 1(b, c) illustrate cases $\alpha = 0.5$, which is a typical convergent case, and $\alpha = 1$, which corresponds to the borderline case between convergence and divergence for $d = 2$. In the convergent case, the asymptotic grid and the optimal grid are almost identical for every N , whereas the regular grid performs well. Indeed, the influence function fields are much smoother than in the $\alpha = 2.4$ case, and the asymptotic theory is expected to perform well as a consequence.

In the borderline case, a discrepancy between the asymptotic grid and the optimal grid appears. The regular grid curve is almost linear with the logarithm of the number of tiles. In this case too, the asymptotic and optimal qtrees offer very similar performances in the asymptotic regime $N > 2048$.

4.2. General tiling

The same comparison is carried out for optimal general tilings and their asymptotic approximations. The results are reported in Figure 3, for $\alpha = 2.4, 1.0$ and 0.5 .

For $N = 2048$, the optimal tiling captures 78.96 DFS while the asymptotics tiling captures 73.33 DFS, which is worse than the qtree asymptotic result (78.58). Nevertheless 49% of the tiles in the asymptotic and optimal tiling are shared. Still, in this regime and this value of N , the *simpler* asymptotic qtree outperforms the asymptotic tiling. Therefore, for such stringent regime ($\alpha = 2.4$), optimizing on a qtree dictionary might be simpler, faster, and actually more efficient. Yet, a real non-simplistic Jacobian representing realistic dispersion physics might not be that stringent. Also, realistic averaged dispersion may exhibit strong anisotropy which may give an advantage to general tilings over qtrees.

For $N = 4096$, the asymptotic tiling (78.985 DFS) does outperform the asymptotic qtree (78.975 DFS). The optimal tiling captures 78.988 DFS. The optimal and the asymptotic tiling of control space share 60% of their tiles. The two corresponding tilings at $N = 4096$ are shown in Figure 4.

4.3. When are regular grids good enough?

In Figures 1 and 3, the DFS curves of the regular grids with respect to $\ln(N)$ are depending on the effective diffusion regime, characterized by α . There is a transition of behaviour

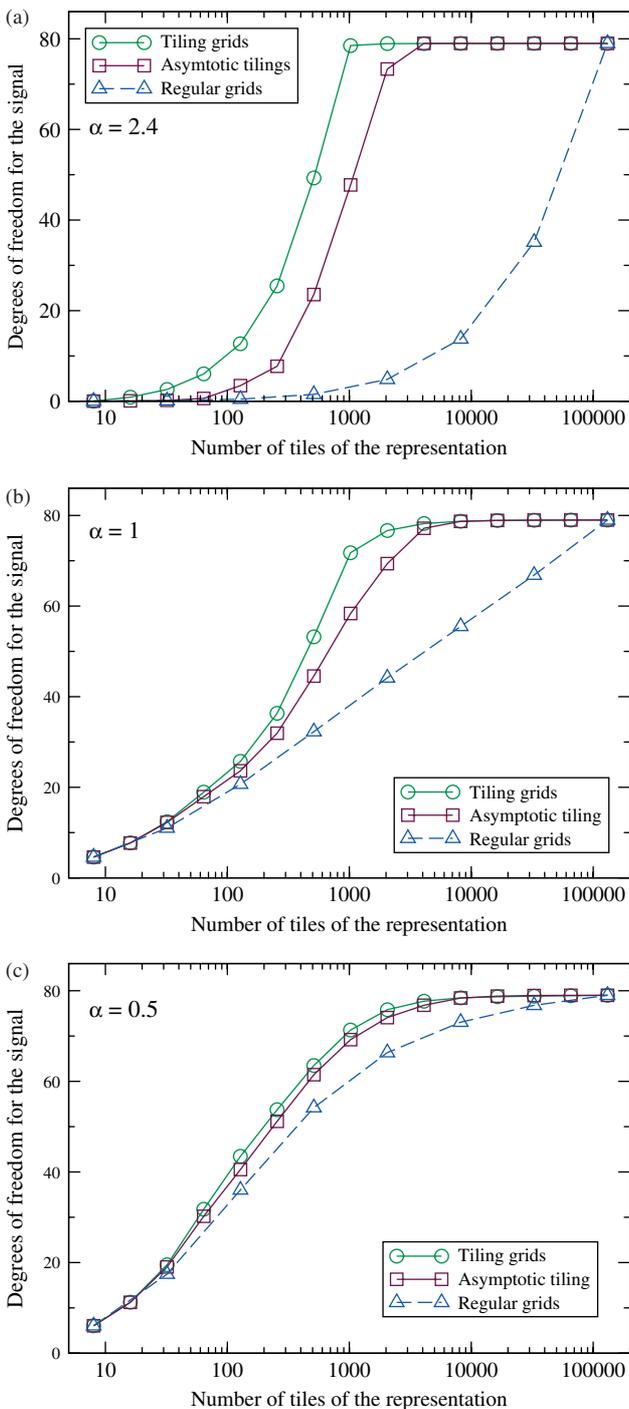


Figure 3. As Figure 1, but for optimal general tilings.

when $\alpha = 1$, where the DFS seem to depend linearly on $\ln(N)$. The expansion used earlier is not suited for the diagnostic of such behaviour. One can show that, at least for the toy model with a power law, the next leading-order term is diverging for $\alpha > 0$ when considering regular grids (ρ uniform). Even though this divergence can be cured by taking into account the finest regular grid acting as a cut-off, there is a simpler heuristic argument that helps understanding the dependence of regular grids with the scale.

In order to compare with the numerical results, at least qualitatively, we make several assumptions. A single influence function $u(\mathbf{x}) \propto \|\mathbf{x}\|^{-\alpha}$, corresponding to a single

annual-mean observation, is considered in dimension $d = 2$. The global domain is replaced with a finite planar domain surrounding the observation site. Spectral analysis with spherical harmonics is replaced with discrete Fourier analysis. In turn, assuming the finite domain is large enough, continuous Fourier analysis will be preferred to discrete analysis for its simplicity. Hence the following argument is only qualitative.

Transposing Eq. (15) in dimension $d = 2$, the DFS criterion reads

$$\begin{aligned} \mathcal{J}_N &= \sum_{k=1}^N \frac{1}{|S_k|} \left(\int_{S_k} d^2\mathbf{x} u(\mathbf{x}) \right)^2 \\ &= \sum_{k=1}^N |S_k| \left(\frac{1}{|S_k|} \int_{S_k} d^2\mathbf{x} u(\mathbf{x}) \right)^2. \end{aligned} \quad (51)$$

In the second line, it is made clear that u is coarse-grained in cell S_k by taking its average. A heuristically similar filter would be achieved by applying a cut-off Λ in Fourier space. For regular grids and $d = 2$, Λ would be related to N by $\Lambda \sim \sqrt{N}$. If $\widehat{u}(\mathbf{p})$ is the Fourier transform of $u(\mathbf{x})$, the filtering operation is defined by

$$F_\Lambda u(\mathbf{x}) = \int_{\Lambda_0 \leq \|\mathbf{p}\| \leq \Lambda} d^2\mathbf{p} \widehat{u}(\mathbf{p}) e^{-i\mathbf{p}\cdot\mathbf{x}}, \quad (52)$$

where Λ_0 is the large-scale cut-off due to the finite domain (which corresponds to $N_0 = N_{\text{cg}} = 8$ in the simplified CTBTO example).

For grid-cell number N related to the cut-off $\Lambda = \sqrt{N}$, one gets

$$\begin{aligned} \mathcal{J}_N &\simeq \int_{\Omega} d^2\mathbf{x} (F_\Lambda u)^2 \\ &\simeq \int d^2\mathbf{p} \widehat{F_\Lambda u}(\mathbf{p}) \widehat{F_\Lambda u}(-\mathbf{p}) \\ &\simeq \int_{\Lambda_0 \leq \|\mathbf{p}\| \leq \Lambda} d^2\mathbf{p} \widehat{u}(\mathbf{p}) \widehat{u}(-\mathbf{p}). \end{aligned} \quad (53)$$

In the case $u(\mathbf{x}) \propto \|\mathbf{x}\|^{-\alpha}$, one obtains $\widehat{u}(\mathbf{p}) \propto \|\mathbf{p}\|^{\alpha-2}$. As a consequence

$$\mathcal{J}_N \propto \int_{\Lambda_0 \leq \|\mathbf{p}\| \leq \Lambda} d^2\mathbf{p} \|\mathbf{p}\|^{2(\alpha-2)} \propto \int_{\Lambda_0}^{\Lambda} dp p^{2\alpha-3}, \quad (54)$$

where $p = \|\mathbf{p}\|$. The normalization of u allows us to determine the proportionality constant for \mathcal{J}_N :

$$1 = \int_{\Omega} d^2\mathbf{x} u^2(\mathbf{x}) = \int_{\Lambda_0 \leq \|\mathbf{p}\| \leq \Lambda_{\text{fg}}} d^2\mathbf{p} \|\mathbf{p}\|^{2(\alpha-2)}, \quad (55)$$

where $\Lambda_{\text{fg}} = \sqrt{N_{\text{fg}}}$ is the cut-off imposed by the finest regular grid. As a consequence, one obtains for $\alpha \neq 1$

$$\mathcal{J}_N \simeq \frac{N^{\alpha-1} - N_0^{\alpha-1}}{N_{\text{fg}}^{\alpha-1} - N_0^{\alpha-1}}, \quad (56)$$

and for $\alpha = 1$

$$\mathcal{J}_N \simeq \frac{\ln(N/N_0)}{\ln(N_{\text{fg}}/N_0)}. \quad (57)$$

Qualitatively, the predicted behaviour matches well the one reported in the numerical experiments, for $\alpha < 1$, $\alpha = 1$, and $\alpha > 1$.

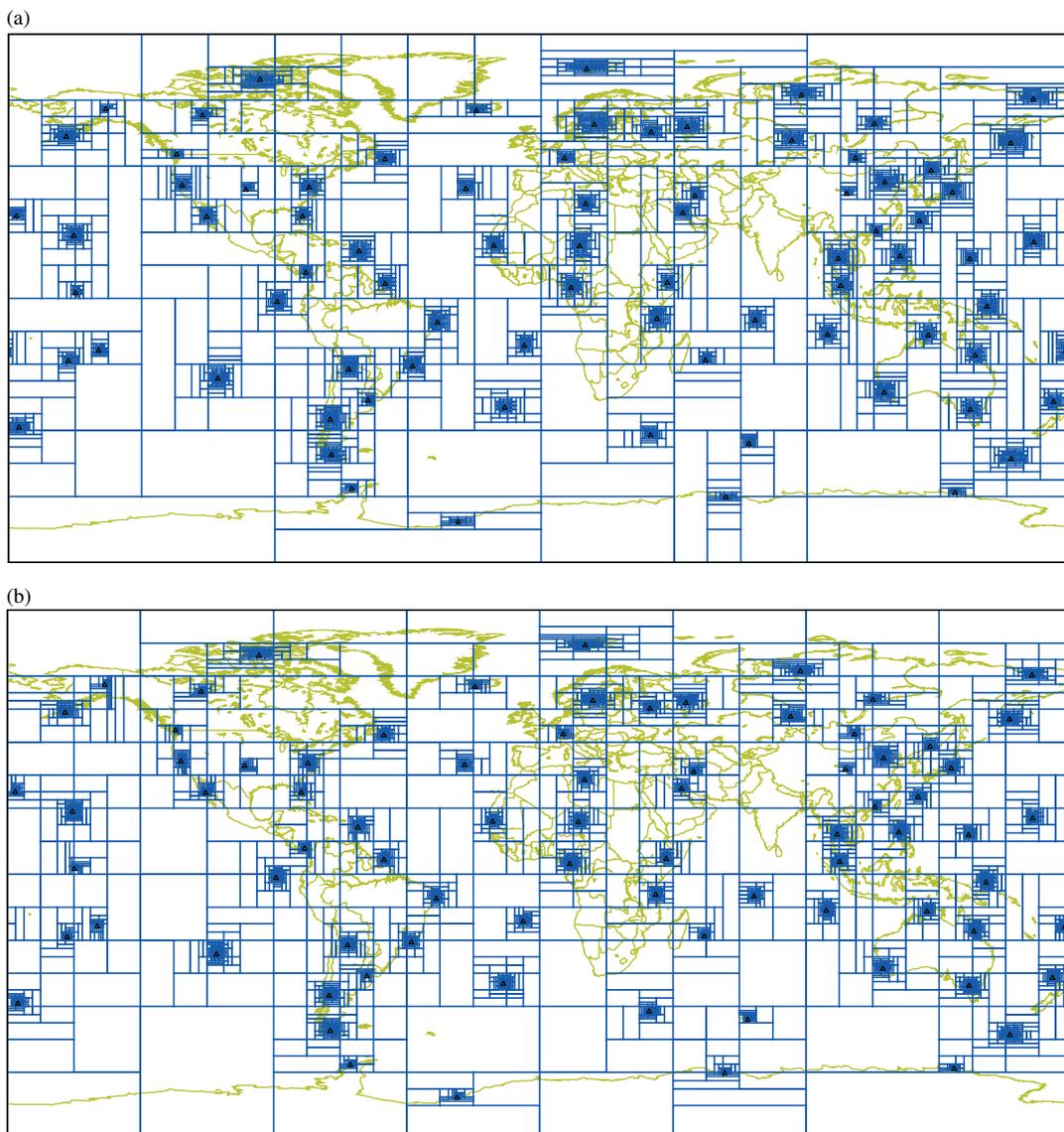


Figure 4. As Figure 2, but for general tilings with $N = 4096$.

5. Summary and conclusion

In Part II, analytical solutions for the representations of control space with a view to optimal data assimilation have been proposed. These densities have been derived in the continuum limit. They are asymptotic, in the sense that they are approximations closer to the optimal solution in the large N limit. The 1D case was first treated, since the solution is expected to be exact in the asymptotic limit. For the 2D cases and beyond, the continuum limit does depend on the dictionary of representations, such as the set of frees, qtrees, or general tilings. The qtree asymptotic limit leads to one single rotationally invariant density. On the contrary, ftree and tilings leads to directional densities, one for each direction, and the total tile density is not rotationally invariant. If this is quite trivial in the ftree case, it is more surprising but understandable in the general tiling case.

Since these densities are continuum limit objects, an algorithm is needed to build discrete adaptive grids from them. Algorithms have been proposed for most efficient multiscale structures: the dictionary of qtrees and the dictionary of general tilings.

All the methodological results have then been tested on an atmospheric dispersion toy model, pertaining to the International Monitoring System network of the CTBTO. The case corresponds to a 2D configuration, but the methodology can be generalized to other dimensional configurations such as the 2D+T ETEX-I example used in Part I.

It was shown that the asymptotic solutions yield suboptimal representations that are quite close to the optimal representations, even for a small fraction of the maximum number of tiles achievable ($N/N_{fg} \ll 1$). They even share a substantial number of tiles with the optimal solutions. Note that there are at least two chained approximations in the process. The first one is the fact that the asymptotic expansion stops at the next leading order. The second one is the heuristic algorithm that we have devised to retrieve discrete representations and which may be sub-optimal. Still, these solutions turned out to be satisfying on the toy model and may well be used for practical problems.

In addition, it was found that the quality of the approximations of these asymptotic solutions depends on the physics. This was illustrated by the α exponent which scales the magnitude of the physical diffusion. In particular,

when $\alpha \geq d/2$, the density is improper which is related to the findings of Bocquet (2005). Discrete representations can nonetheless be built on these improper densities, because of saturation zones that shield the singularities of the density. An algorithm to deal with this problem is also proposed.

The interest of these optimal representations versus the regular grids was shown to depend also on the physics. In particular, the more diffusive the regime is (typically $\alpha > 1$), the less efficient the regular grids are, and the more beneficial the optimal representations are. Note that the regime of the ETEX experiment discussed in Part I, with a real physics Jacobian, corresponds rather to a convex DFS curve where regular grids are particularly inefficient.

An unexplored idea is to use these asymptotic solutions to precondition the discrete optimization presented in Part I, but it would require the dual optimization of Part I to be exchanged for a primal optimization scheme where the filling factors of each tile are explicit. This can be done but is outside the present scope.

The first aim of this two-part article was to establish a set of theoretical results on the topic. A secondary goal is to prepare the application of the methodology to atmospheric chemistry problems with large datasets and real chemistry and transport models (hence large and realistic Jacobians). In that respect, those asymptotic solutions and the associated algorithms are key tools towards that goal. We are currently applying the methodology on a realistic CO₂ inversion experiment, and on the IMS network of the CTBTO preparatory commission, in collaboration with experts in these two fields.

Acknowledgements

The authors would like to thank the associate editor, Dan Cornford, and two anonymous reviewers for their constructive and extensive comments and suggestions. The authors thank Monika Krysta for discussing the CTBO radionuclides monitoring issues, and for a careful reading of the manuscript. This paper is a contribution to the MultiScale Data Assimilation in Geophysics (MSDAG) project supported by the Agence Nationale de la Recherche, grant ANR-08-SYSC-014.

References

- Bocquet M. 2005. Grid-resolution dependence in the reconstruction of an atmospheric tracer source. *Nonlin. Proc. Geophys.* **12**: 219–233.
- Bocquet M. 2009. Towards optimal choices of control space representation for geophysical data assimilation. *Mon. Weather Rev.* **137**: 2331–2348.
- Bocquet M, Wu L, Chevallier F. Bayesian design of control space for optimal assimilation of observations. Part I: Consistent multiscale formalism. 2011. *Q. J. R. Meteorol. Soc.* **137**: DOI: 10.1002/qj.837.
- Byrd RH, Lu P, Nocedal J. 1995. A limited-memory algorithm for bound constrained optimization. *SIAM J. Sci. Statist. Comput.* **16**: 1190–1208.
- Rodgers CD. 2000. *Inverse methods for atmospheric sounding. Series on Atmospheric, Oceanic and Planetary Physics*, Vol. 2. World Scientific.
- Roustan Y, Bocquet M. 2006. Sensitivity analysis for mercury over Europe. *J. Geophys. Res.* **111**: D14304.
- Saïde P, Bocquet M, Osses A, Gallardo L. 2011. Constraining surface emissions of air pollutants using inverse modeling: Method intercomparison and a new two-step multiscale approach. *Tellus B* in press.
- United Nations. 1996. *Protocol on the Comprehensive Nuclear Test Ban Treaty*. Preparatory Commission for the Comprehensive Nuclear Test Ban Treaty: Vienna.