Entropy minimizing curves with application to flight path design and clustering *

Stéphane Puechmorel, Florence Nicol

MDPI

Article

Entropy minimizing curves with application to flight path design and clustering*

Stephane Puechmorel [1] and Florence Nicol [2]

[1] Ecole Nationale de l'Aviation Civile (ENAC)
Dept. SINA/MAIAA
7, Avenue Edouard Belin
F31500 TOULOUSE FRANCE
stephane.puechmorel@enac.fr
[2] Ecole Nationale de l'Aviation Civile (ENAC)
Dept. SINA/MAIAA
7, Avenue Edouard Belin
F31500 TOULOUSE FRANCE
nicol@recherche.enac.fr
* Correspondence: stephane.puechmorel@enac.fr; Tel.: +33-5-62 17 95 03

Academic Editor: name
Version July 28, 2016 submitted to Entropy; Typeset by LaTeX using class file mdpi.cls

Air traffic management (ATM) aims at providing companies with a safe and ideally optimal aircraft trajectory planning. Air traffic controllers act on flight paths in such a way that no pair of aircraft come closer than the regulatory separation norms. With the increase of traffic, it is expected that the system will reach its limits in a near future: a paradigm change in ATM is planned with the introduction of trajectory based operations. In this context, sets of well separated flight paths are computed in advance, tremendously reducing the number of unsafe situations that must be dealt with by controllers. Unfortunately, automated tools used to generate such plannings generally issue trajectories not complying with operational practices or even flight dynamics. In this paper, a mean of producing realistic air routes from the output of an automated trajectory design tool is investigated. For that purpose, the entropy of a system of curves is first defined and a mean of iteratively minimizing it is presented. The resulting curves form a route network that is suitable for use in a semi-automated ATM system with human in the loop. The tool introduced in this work is quite versatile and may be applied also to unsupervised classification of curves: an example is given for the french traffic.

1. Introduction

Based on recent studies [3], traffic in Europe is expected to grow on an average yearly rate of 2.6%, yielding a net increase of 2 million flights per year at the 2020 horizon. Long term forecast gives a two fold increase in 2050 over the current traffic, pointing out the need for a paradigm change in the way flights are managed. Two major framework programs, SESAR (Single European Sky Air traffic management Research) in Europe and Nextgen in the US have been launched in order to first investigate potential solutions and to deploy them in a second phase. One of the main changes that the air traffic management (ATM) system will undergo is a switch from airspace based to trajectory based operations with a delegation of the separation task to the crews. Within this framework, trajectories become the basic object of ATM, changing the way air traffic controllers will

*This paper is an extended version of "Entropy minimizing curves with application to automated flight path design published in GSI15

be working. In order to alleviate the workload of controllers, trajectories will be planned several weeks in advance in such a way that close encounters are minimized and ideally removed. For that purpose, several tools are currently being developed, most of them coming from the field of robotics [10]. Unfortunately, flight path issued by these algorithms are not tractable for a human controller and need to be simplified. The purpose of the present work is to introduce an automated procedure that takes as input a set of trajectories and outputs a simplified one that can be used in an operational context. Using an entropy associated with a curves system, a gradient descent is performed in order to reduce it so as to straighten trajectories while avoiding areas with low aircraft density, thus enforcing route-like behavior. This effect is related to the fact that entropy minimizing distributions favor concentration.

2. Entropy minimizing curves

2.1. Motivation

As previously mentioned, air traffic management of the future will make an intensive use of 4D trajectories as a basic object. Full automation is a far reaching concept that will probably not implemented before 2040-2050 and even in such a situation, it will be needed to keep humans in the loop so as to gain a wide societal acceptance of the concept. Starting from SESAR or Nextgen initial deployment and aiming towards this ultimate objective, a transition phase with human-system cooperation will take place. Since ATC controllers are used to a well structured network of routes, it is advisable to post-process the 4D trajectories issued by automated systems in order to make them as close as possible to line segments connecting beacons. To perform this task, in an automatic way, flight paths will be deformed so as to minimize an entropy criterion, that enforces avoidance of low density area and at the same time penalizes length. Compared to already available bundling algorithms [5] that tend to move curves to high density areas, this new procedure generates geometrically correct curves, without excess curvature.

Let a set $\gamma_1, \dots \gamma_N$ of smooth curves be given, that will be aircraft flight paths for the air traffic application. It will be assumed in the sequel that all curves are smooth mappings from $[0,1]$ to a domain Ω of $\mathbb{R}^q$ with everywhere non vanishing derivatives in $]0,1[$. This last condition allows to view them as smooth immersions with boundaries and is sound from the application point of view as aircraft velocities are bounding below by efficiency consideration and ultimately by the stall, and therefore cannot be vanish expect at the endpoints. In air traffic applications, the dimension of the state space is generally 2 and sometimes 3 when the evolution of aircraft in the vertical plane is of interest.

The approach taken in this work is first to get a sound definition of spatial density associated with a curve system, then to derive from it an entropy that will be minimized.

2.2. Spatial density of a system of curves

Due to the fact that aircraft positions are acquired through radar measurements, a trajectory is known only at discrete sampling times. In the operational context, the sampling period ranges from 4 to 10 seconds which corresponds roughly to 100-250m traveling distance. Derived from that, a classical performance indicator used in ATM is the aircraft density [4], obtained from the sampled positions $\gamma_i(t_j)$, $j = 1, \dots, n_i$ on each flight path γ_i, $i = 1, \dots, N$. It is constructed from a partition U_k, $k = 1, \dots, P$ of Ω by counting the number of samples occurring in a given U_k then dividing out by the total number of samples $n = \sum_{i=1}^{N} n_i$. More formally, the density d_k in the subset U_k of Ω is expressed as:

$$d_k = n^{-1} \sum_{i=1}^{N} \sum_{j=1}^{n_i} 1_{U_k} \left(\gamma_i(t_j) \right) \tag{1}$$

with 1_{U_k} the characteristic function of the set U_k. It seems natural to extend the density obtained from samples to another one based on the trajectories themselves using an integral form:

$$d_k = \lambda^{-1} \sum_{i=1}^{N} \int_0^1 1_{U_k}(\gamma_i(t))\, dt \tag{2}$$

where the normalizing constant λ is chosen so as that d_k is a discrete probability distribution:

$$\lambda = \sum_{k=1}^{P} \sum_{i=1}^{N} \int_0^1 1_{U_k}(\gamma_i(t))\, dt = \sum_{i=1}^{N} \int_0^1 \sum_{k=1}^{P} 1_{U_k}(\gamma_i(t))\, dt$$

and since $U_k, k = 1 \dots P$ is a partition:

$$\sum_{k=1}^{P} 1_{U_k}(\gamma_i(t)) = 1 \tag{3}$$

so that $\lambda = N$.

Density can be viewed as an empirical probability distribution with the U_k considered as bins in an histogram. It is thus natural to extend the above computation so as to give rise to a continuous distribution on Ω. For that purpose, local weighting techniques such as kernel density estimation methods are well-known in nonparametric statistics because they are a useful data driven way to yield continuous density estimation. Many references may be found in the literature as in [11,12]. Given observations, the resulting estimation will be the sum of weights taking into account the distance between the observations and the location x at which the density has to be estimated; the more an observation is close to x, the greater is the weighting. The weights are defined by selecting a summable function centered on the observations, called a kernel, usually denoted by $K\colon \mathbb{R} \to \mathbb{R}^+$ in the univariate case, and a smoothed version of the Parzen-Rosenblatt density estimator [8,9] is used. Standard choices for the K function are the ones used for nonparametric kernel estimation like the Epanechnikov function [2]:

$$K\colon x \mapsto \left(1 - x^2\right) 1_{[-1,1]}(x).$$

There exists a large variety of kernel functions and any density function satisfying the normalization condition can be considered so that the estimation is a probability density. Moreover, the kernel function is a symmetric positive function, with a first moment equals to zero and a finite second order moment. In the multivariate case, a multivariate kernel function $\mathcal{K}\colon \mathbb{R}^q \to \mathbb{R}^+$ is selected, that can be expressed by means of a real kernel K associated with a norm, denoted by $\|.\|$, in $\mathbb{R}^q$ as follows

$$\mathcal{K}(x) = K(\|x\|), \qquad x \in \mathbb{R}^q.$$

The normalization condition becomes

$$\int_{\mathbb{R}^q} \mathcal{K}(x)\, dx = \int_{\mathbb{R}^q} K(\|x\|)\, dx = 1.$$

A kernel version of the density is then defined as a mapping d from Ω to $[0,1]$:

$$d\colon x \mapsto \frac{\sum_{i=1}^{N} \int_0^1 K(\|x - \gamma_i(t)\|)\, dt}{\sum_{i=1}^{N} \int_\Omega \int_0^1 K(\|x - \gamma_i(t)\|)\, dt dx}. \tag{4}$$

Normalizing the kernel is not mandatory as the normalization occurs with the definition of d. It is nevertheless easier to consider these kind of kernels, as it is done in nonparametric density estimation.

Note that when K is compactly supported, which is the case of the Epanechnikov function and all its relatives, it comes:

$$\int_\Omega K\left(\|x-\gamma_i(t)\|\right)dx = \int_{\mathbb{R}^q} K\left(\|x\|\right)dx$$

provided that Ω contains the set:

$$\{x\in\mathbb{R}^q, \inf_{i=1\ldots N, t\in[0,1]} \|x-\gamma_i(t)\| \leq A\}$$

where the interval $[-A, A]$ contains the support of K. The case of kernels with unbounded support, like Gaussian functions, may be dealt with provided $\Omega = \mathbb{R}^q$. In the application considered, only compactly supported kernels are used, mainly to allow fast machine implementation of the density computation.

Using the polar coordinates (ρ, θ) and the rotation invariance of the integrand, the relation becomes:

$$Vol\left(\mathbb{S}^{q-1}\right)\int_{\mathbb{R}^+} K(\rho)\rho^{q-1}d\rho = 1$$

which yields a normalizing constant of $2/\pi$ for the Epanechnikov function in dimension 2, instead of the usual $3/4$ in the real case. When the normalization condition is fulfilled, the expression of the density simplifies to:

$$d\colon x \mapsto N^{-1}\sum_{i=1}^{N}\int_0^1 K\left(\|x-\gamma_i(t)\|\right)dt. \tag{5}$$

The normalizing constant is the same as in 2.

As an example, one day of traffic over France is considered and pictured on Figure 1 with the corresponding density map, computed on an evenly spaced grid with a normalized Epanechnikov kernel:

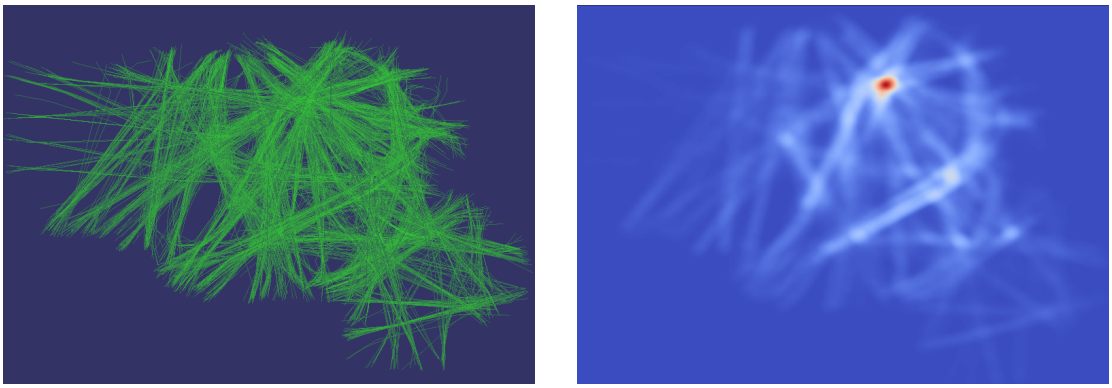

Figure 1. Traffic over France and associated density.

Unfortunately, density computed this way suffers a severe flaw for the ATM application: it is not related to the shape of trajectories but also to the time behavior. Formally, it is defined on the set $\mathbf{Imm}\left([0,1],\mathbb{R}^q\right)$ of smooth immersions from $[0,1]$ to $\mathbb{R}^q$ while the space of primary interest will be the quotient by smooth diffeomorphisms of the interval $[0,1]$, $\mathbf{Imm}\left([0,1],\mathbb{R}^q\right)/\mathbf{Diff}([0,1])$. Invariance of the density under the action of $\mathbf{Diff}([0,1])$ is obtained as in [6] by adding a term related to velocity in the integrals. The new definition of d becomes:

$$\tilde{d}\colon x \mapsto \frac{\sum_{i=1}^N \int_0^1 K\left(\|x-\gamma_i(t)\|\right)\|\gamma_i'(t)\|dt}{\sum_{i=1}^N \int_\Omega \int_0^1 K\left(\|x-\gamma_i(t)\|\right)\|\gamma_i'(t)\|dtdx}. \tag{6}$$

Assuming again a normalized kernel and letting l_i be the length of the curve γ_i, the expression of the density simplifies to:

$$\tilde{d}\colon x \mapsto \frac{\sum_{i=1}^{N} \int_0^1 K\left(\|x-\gamma_i(t)\|\right) \|\gamma_i'(t)\| dt}{\sum_{i=1}^{N} l_i}. \tag{7}$$

The new **Diff**-invariant density is pictured on Figure 2 along with the standard density. While the overall aspect of the plot is similar, one can observe that routes are more apparent the right picture and that the density peak located above Paris area is of less importance and less symmetric is due to the fact that near airports, aircraft are slowing down and this effect exaggerates the density with the non-invariant definition.

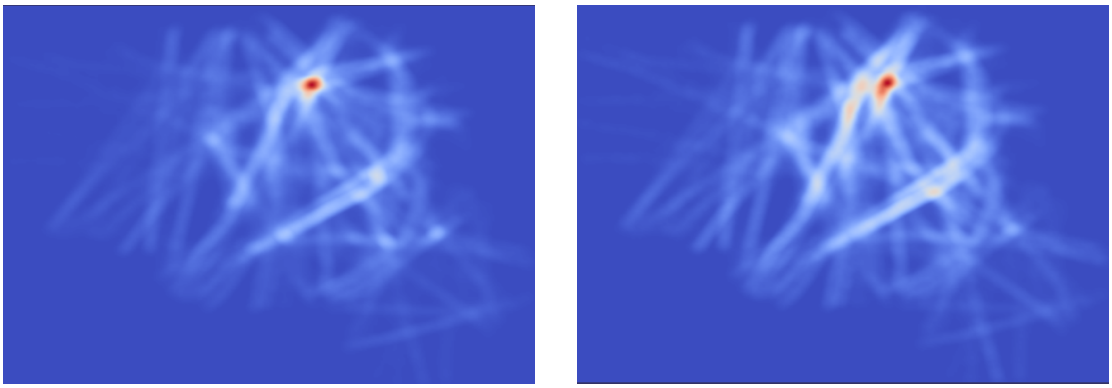

Figure 2. Density (left) and **Diff** invariant density (right) for the 12th February 2013 traffic.

The extension of the two dimensional defined that way to the general case of curves in an arbitrary space $\mathbb{R}^q$ is straightforward.

2.3. Further properties of the density

In this section, curves considered are smooth mapping from the closed interval $[0,1]$ to $\mathbb{R}^q$, with a non-vanishing derivative in $]0,1[$. All multivariate kernels $\mathcal{K}$ will be assumed smooth, positive, with a unit integral, and of the form $x \mapsto K\left(\|x\|\right)$. However, it is not required that they are compactly supported unless explicitly stated. All results are presented for the whole space $\mathbb{R}^q$, but apply almost verbatim to an open subset.

Definition 1. Let f be a smooth summable mapping from $\mathbb{R}$ to $\mathbb{R}$. The scaling f_ν of f is defined, for each $\nu > 0$ to be the mapping:

$$f_\nu\colon x \in \mathbb{R} \mapsto \frac{1}{\nu} f\left(\frac{x}{\nu}\right).$$

It is clear that the L^1-norm of the original mapping is preserved by the scaling. Given a summable kernel function K from $\mathbb{R}$ to $\mathbb{R}^+$, it defines a multivariate kernel $\mathcal{K}$ on $\mathbb{R}^q$ that maps x to $K(\|x\|)$. One may derive from it a parametrized family of kernels in $\mathbb{R}$ by mapping each ν in $]0,1]$ to the scaled kernel K_ν. If the original K is of unit integral, so are all the K_ν.

Proposition 2. *Let $\gamma\colon [0,1] \to \mathbb{R}^q$ be a smooth path with non-vanishing derivative in $]0,1[$. Let $K_\nu, \nu > 0$ be a parametrized family of unit integral kernels. The family of Borel measures μ_ν defined for any Borel set A by:*

$$\mu_\nu(A) = \int_A \int_0^1 K_\nu\left(\|x-\gamma(t)\|\right) \|\gamma'(t)\| dt dx$$

is tight and converges narrowly to the Lebesgue measure on $\gamma\left([0,1]\right)$.

Proof. Let $\epsilon > 0$ be given. By the summability of K, it exists a positive real number r such that:

$$\int_{\mathbb{R}^q - B(0,r)} K(\|x\|)dx < \epsilon$$

with $B(0,r)$ the open ball of radius r centered at the origin. Since $B(0,r) \subset B(0,r\nu^{-1})$ for $\nu > 0$, the same holds for all the family K_ν. Let $B(0,M)$ be an open ball containing $\gamma([0,1])$. Then:

$$\mu_\nu\left(\mathbb{R}^q - \overline{B(0,M+r)}\right) = \int_{\mathbb{R}^q - B(0,M+r)} \int_0^1 K_\nu\left(\|x-\gamma(t)\|\right)\|\gamma'(t)\|dtdx \tag{8}$$

$$= \int_0^1 \int_{\mathbb{R}^q - B(0,M+r)} K_\nu\left(\|x-\gamma(t)\|\right)\|\gamma'(t)\|dxdt \tag{9}$$

$$\leq \epsilon \int_0^1 \|\gamma'(t)\|dt = \epsilon\, l(\gamma) \tag{10}$$

where $l(\gamma)$ denotes the length of γ. This proves the tightness of the family K_ν.

Let $f\colon \mathbb{R}^q \to \mathbb{R}$ be a bounded continuous mapping. It comes:

$$I_\nu(f) = \int_{\mathbb{R}^q} \int_0^1 K_\nu\left(\|x-\gamma(t)\|\right) f(x)\|\gamma'(t)\|dtdx = \int_0^1 \int_{\mathbb{R}^q} K_\nu\left(\|x-\gamma(t)\|\right) f(x)\|\gamma'(t)\|dxdt \tag{11}$$

$$= \int_0^1 \int_{\mathbb{R}^q} K\left(\|x\|\right) f\left(x\nu + \gamma(t)\right)\|\gamma'(t)\|dxdt \tag{12}$$

and since f is bounded, the dominated convergence theorem shows that:

$$\lim_{\nu\to 0} I_\nu(f) = \int_0^1 f\left(\gamma(t)\right)\|\gamma'(t)\|dt$$

proving the second part of the claim. □

The density in 7 is for a single curve of the form $d(x) = l(\gamma)^{-1}\int_0^1 K\left(\|x-\gamma(t)\|\right)\|\gamma'(t)\|dt$ with $l(\gamma)$ the length of the curve γ. It is invariant under change of parameter, and can be written in a more concise way as:

$$\int_0^1 K\left(\|x-\gamma(\eta)\|\right)\|d\eta \tag{13}$$

where η is the arclength times $l(\gamma)^{-1}$.

This form allows a simple probabilistic interpretation of the density d: if a point u is drawn on the curve γ according to a uniform distribution and independently a vector v in $\mathbb{R}^q$ with a density $\mathcal{K}$ (the multivariate kernel corresponding to K), then the density of $x = u + v$ is given by the equation 13.

Proposition 3. *If the multivariate kernel $\mathcal{K}$ has a finite second moment, that is the univariate kernel K is such that:*

$$M = \int_{\mathbb{R}^+} r^{q+1}K(r)\,dr < +\infty$$

then the Wasserstein distance between the densities d_1, d_2 associated to smooth curves γ_1, γ_2 is bounded by:

$$2Vol(\mathbb{S}^{q-1})M + D(\gamma_1,\gamma_2)$$

with :

$$D(\gamma_1,\gamma_2) = \int_0^1 \|\gamma_1(\eta) - \gamma_2(\eta)\|^2 d\eta$$

where each curve is parametrized by the scaled arclength as in 13.

Proof. Let us consider the plan [1] given by the density:

$$d\colon (x,y) \mapsto \int_0^1 K(\|x-\gamma_1(\eta)\|)K(\|y-\gamma_2(\eta)\|)d\eta$$

where each curve is parametrized by the scaled arclength. The associated transport cost is given by:

$$C = \int_{\mathbb{R}^q\times\mathbb{R}^q} \|x-y\|^2 \int_0^1 K\left(\|x-\gamma_1(\eta)\|\right) K\left(\|y-\gamma_2(\eta)\|\right) d\eta dx dy$$

letting $u = y - x$ and using Fubini gives:

$$C = \int_0^1 \int_{\mathbb{R}^q} K\left(\|x-\gamma_1(\eta)\|\right) \int_{R^q} \|u\|^2 K\left(\|u+x-\gamma_2(\eta)\|\right) du dx d\eta.$$

The inner term can be written as:

$$\int_{R^q} \|u\|^2 K\left(\|u+x-\gamma_2(\eta)\|\right) du = \int_{R^q} \|u+\gamma_2(\eta)-x\|^2 K\left(\|u\|\right) du \tag{14}$$

$$= \int_{\mathbb{R}^q} \|u\|^2 K(\|u\|)du + 2\langle \gamma_2(\eta)-x, \int_{\mathbb{R}^q} uK(\|u\|)du\rangle + \|\gamma_2(\eta)-x\|^2. \tag{15}$$

The integral:

$$\int_{\mathbb{R}^q} uK(\|u\|)du$$

is 0 and, using spherical coordinates:

$$\int_{\mathbb{R}^q} \|u\|^2 K(\|u\|)du = \int_{\mathbb{R}^+} r^{q+1}K(r) \int_{\mathbb{S}^{q-1}} d\sigma\, dr = Vol(\mathbb{S}^{q-1})M$$

with $M = \int_{\mathbb{R}^+} r^{q+1}K(r)$. Putting back this value in the expression of the cost gives:

$$C = Vol(\mathbb{S}^{q-1})M \int_0^1 \int_{\mathbb{R}^q} K\left(\|x-\gamma_1(\eta)\|\right) dx d\eta + \int_0^1 \int_{\mathbb{R}^q} K\left(\|x-\gamma_1(\eta)\|\right) \|\gamma_2(\eta)-x\|^2 dx d\eta \tag{16}$$

$$= Vol(\mathbb{S}^{q-1})M + \int_0^1 \int_{\mathbb{R}^q} K\left(\|x-\gamma_1(\eta)\|\right) \|\gamma_2(\eta)-x\|^2 dx d\eta \tag{17}$$

$$= Vol(\mathbb{S}^{q-1})M + \int_0^1 \int_{\mathbb{R}^q} K\left(\|x\|\right) \|\gamma_2(\eta)-\gamma_1(\eta)+x\|^2 dx d\eta. \tag{18}$$

Finally:

$$\int_{\mathbb{R}^q} K\left(\|x\|\right) \|\gamma_2(\eta)-\gamma_1(\eta)+x\|^2 dx = \int_{\mathbb{R}^q} K\left(\|x\|\right) \|\gamma_2(\eta)-\gamma_1(\eta)\|^2 dx d\eta \tag{19}$$

$$+ 2\langle \gamma_2(\eta)-\gamma_1(\eta), \int_{\mathbb{R}^q} xK\left(\|x\|\right) dx \tag{20}$$

$$+ Vol(\mathbb{S}^{q-1})M. \tag{21}$$

As before the middle term vanishes and the first one integrates to:

$$\int_0^1 \|\gamma_2(\eta)-\gamma_1(\eta)\|^2 d\eta$$

so that:

$$C = 2Vol(\mathbb{S}^{q-1})M + \int_0^1 \|\gamma_2(\eta)-\gamma_1(\eta)\|^2 d\eta.$$

95 □

This result indicates that the densities associated to curves γ_1, γ_2 using the smoothing process described above cannot be too far (with respect to the Wasserstein distance) one from each other if the geometric L^2 distance $D(\gamma_1, \gamma_2)$ is small. In fact, the upper bound in Proposition 3 can be interpreted as the cost of moving the smoothed density around γ_1 to the uniform distribution on the curve, then move γ_1 to γ_2 keeping points with equal scaled arclength in correspondence, and finally move the uniform distribution on γ_2 to the smoothed density.

Having a density at hand, the entropy of the system of curves $\gamma_1, \dots, \gamma_N$ is defined the usual way as:

$$E(\gamma_1, \dots, \gamma_N) = -\int_\Omega \tilde{d}(x) \log\left(\tilde{d}(x)\right) dx.$$

The entropy is dependent on the particular choice of the kernel K. As mentioned before, it is a common practice in the field of non-parametric statistics to introduce a tuning parameter $\nu > 0$ in the kernel, called bandwidth, so that is it expressed as a scaled version $K = f_\nu$ of a given function $f\colon \mathbb{R}^+ \to \mathbb{R}^+$. The value of ν is the most influential parameter in the estimation of the density and must be selected carefully. For curves clustering applications, it is defined by the desired interaction length: if ν tends to 0, the curves will behave as independent objects while on the other end of the scale very high bandwidth will tend to remove the influence of the curves themselves. For the moment, no automated mean of finding an optimal ν was used, although it will be part of a future work.

2.4. Minimizing the entropy

In order to fulfill the initial requirement of finding bundles of curve segments as straight as possible, one seeks after the system of curves minimizing the entropy $E(\gamma_1, \dots, \gamma_N)$, or equivalently maximizing:

$$\int_\Omega \tilde{d}(x) \log\left(\tilde{d}(x)\right) dx.$$

The reason why this criterion gives the expected behavior will become more apparent after derivation of its gradient at the end of this part. Nevertheless, when considering a single trajectory its is intuitive that the most concentrated density distribution is obtained with a straight segment connecting the endpoints: this point will be made rigorous later.

Letting ϵ be a perturbation of the curve γ_j such that $\epsilon(0) = \epsilon(1) = 0$, the first order expansion of $-E(\gamma_1, \dots, \gamma_N)$ will be computed in order to get a maximizing displacement field, analogous to a gradient ascent[1] in the finite dimensional setting. The notation:

$$\frac{\partial F}{\partial \gamma_j}$$

will be used in the sequel to denote the derivative of a function F of the curve γ_j in the sense that for a perturbation ϵ:

$$F(\gamma_j + \epsilon) = F(\gamma_j) + \frac{\partial F}{\partial \gamma_j}(\epsilon) + o(\|\epsilon\|_2).$$

First of all, please note that since $\tilde{d}$ has integral 1 over the domain Ω:

$$\int_\Omega \frac{\partial \tilde{d}(x)}{\partial \gamma_j}(\epsilon) dx = 0$$

so that:

$$-\frac{\partial}{\partial \gamma_j} E(\gamma_1, \dots, \gamma_N)(\epsilon) = \int_\Omega \frac{\partial \tilde{d}(x)}{\partial \gamma_j}(\epsilon) \log\left(\tilde{d}(x)\right) dx. \tag{22}$$

---

[1] Choice has been made to maximize the opposite of the entropy, so that the algorithm will be a gradient ascent one.

Starting from the expression of $\tilde{d}$ given in equation 7, the first order expansion of $\tilde{d}$ with respect to the perturbation ϵ of γ_j is obtained as a sum of a term coming from the numerator:

$$\int_0^1 K\left(\|x-\gamma_j(t)\|\right)\|\gamma_j'(t)\|dt. \tag{23}$$

and a second one coming from the length of γ_j in the denominator. This last term is obtained from the usual first order variation formula of a curve length:

$$\int_{[0,1]}\left\|\gamma_j'(t)+\epsilon'(t)\right\|dt = \int_{[0,1]}\left\|\gamma_j'(t)\right\|dt+\int_{[0,1]}\left\langle\frac{\gamma_j'(t)}{\|\gamma_j'(t)\|},\epsilon'(t)\right\rangle dt+o(\|\epsilon\|_2).$$

Using an integration by part, the first order term can be written as:

$$\int_{[0,1]}\left\langle\frac{\gamma_j'(t)}{\|\gamma_j'(t)\|},\epsilon'(t)\right\rangle dt = \tag{24}$$

$$-\int_{[0,1]}\left\langle\left(\frac{\gamma_j''(t)}{\|\gamma_j'(t)\|}\right)_{\mathcal{N}},\epsilon(t)\right\rangle dt \tag{25}$$

with:

$$\left(\frac{\gamma_j''(t)}{\|\gamma_j'(t)\|}\right)_{\mathcal{N}}=\frac{\gamma_j''(t)}{\|\gamma_j'(t)\|}-\frac{\gamma_j'(t)}{\|\gamma_j'(t)\|}\left\langle\frac{\gamma_j'(t)}{\|\gamma_j'(t)\|},\frac{\gamma_j''(t)}{\|\gamma_j'(t)\|}\right\rangle$$

the normal component of:

$$\frac{\gamma_j''(t)}{\|\gamma_j'(t)\|}.$$

Please note that when dealing with planar curves (i.e. with values in $\mathbb{R}^2$), it is $\kappa_j(t)N_j(t)$ with κ_j (resp N_j) the curvature (resp. the unit normal vector) of γ_j.

The integral in 23 can be expanded in a similar fashion. Using as above the notation $()_{\mathcal{N}}$ for normal components, the first order term is obtained as:

$$\int_{[0,1]}\left\langle\left(\frac{\gamma_j(t)-x}{\|\gamma_j(t)-x\|}\right)_{\mathcal{N}},\epsilon(t)\right\rangle K'\left(\|\gamma_j(t)-x\|\right)\|\gamma_j'(t)\|dt \tag{26}$$

$$-\int_{[0,1]}\left\langle\left(\frac{\gamma_j''(t)}{\|\gamma_j'(t)\|}\right)_{\mathcal{N}},\epsilon(t)\right\rangle K\left(\|\gamma_j(t)-x\|\right)dt. \tag{27}$$

From the expressions in 26 and 24, the first order variation of the entropy is:

$$\frac{1}{\sum_{i=1}^N l_i}\left(\int_{[0,1]}\left\langle\int_\Omega\left(\frac{\gamma_j(t)-x}{\|\gamma_j(t)-x\|}\right)_{\mathcal{N}}K'\left(\|\gamma_j(t)-x\|\right)\log(\tilde{d}(x))dx,\epsilon(t)\right\rangle\|\gamma_j'(t)\|dt\right. \tag{28}$$

$$-\int_{[0,1]}\left(\int_\Omega K\left(\|\gamma_j(t)-x\|\right)\log(\tilde{d}(x))dx\right)\left\langle\left(\frac{\gamma_j''(t)}{\|\gamma_j'(t)\|}\right)_{\mathcal{N}},\epsilon(t)\right\rangle dt \tag{29}$$

$$\left.+\left(\int_\Omega\tilde{d}(x)\log(\tilde{d}(x))dx\right)\int_{[0,1]}\left\langle\left(\frac{\gamma_j''(t)}{\|\gamma_j'(t)\|}\right)_{\mathcal{N}},\epsilon(t)\right\rangle dt\right). \tag{30}$$

As expected, only moves normal to the trajectory will change at first order the value of the criterion: the displacement of the curve γ_j will thus be performed at t in the normal bundle to γ_j and is given, up to the $(\sum_{i=1}^{N} l_i)^{-1}$ term by:

$$\int_\Omega \left(\frac{\gamma_j(t)-x}{\|\gamma_j(t)-x\|}\right)_{\mathcal{N}} K'\left(\|\gamma_j(t)-x\|\right)\log(\tilde{d}(x))dx\|\gamma_j'(t)\| \tag{31}$$

$$-\left(\int_\Omega K\left(\|\gamma_j(t)-x\|\right)\log(\tilde{d}(x))dx\right)\left(\frac{\gamma_j''(t)}{\|\gamma_j'(t)\|}\right)_{\mathcal{N}} \tag{32}$$

$$+\left(\int_\Omega \tilde{d}(x)\log(\tilde{d}(x))dx\right)\left(\frac{\gamma_j''(t)}{\|\gamma_j'(t)\|}\right)_{\mathcal{N}}. \tag{33}$$

The first term in the expression will favor moves towards areas of high density, while the second and third one are moving along normal vector and will straighten trajectory. This last point can be enlightened by considering the case of a single planar curve with fixed endpoints.

Proposition 4. *Let a, b be fixed points in $\mathbb{R}^2$ and K be a kernel as in 7. The segment $[a, b]$ is a critical point for the entropy associated with the curve system in $\mathbb{R}^2$ consisting of single smooth paths with endpoints a, b.*

Proof. Let the segment $[a, b]$ be parametrized as $\gamma\colon t \in [0,1] \mapsto a + tv$ with v the vector $(b - a)$. Starting with the expression 31, it is clear that the second and third terms occurring in the formula will vanish as the second derivative of γ is 0. Let u be the unit normal vector to γ. Any point x in $\mathbb{R}^2$ can be written as $x = a + \theta v + \xi u$, $\theta, \xi \in \mathbb{R}$, , so that $\gamma(t) - x = (t - \theta)v - \xi u$ and $\|\gamma(t) - x\| = \sqrt{(t-\theta)^2\|b-a\|^2 + \xi^2}$. The change of variables $x \to (\theta, \xi)$ has Jacobian $\|v\| = \|b - a\|$. For a fixed $t \in [0, 1]$, it comes:

$$\int_{\mathbb{R}^2} \left(\frac{\gamma(t)-x}{\|\gamma(t)-x\|}\right)_{\mathcal{N}} K'\left(\|\gamma(t)-x\|\right)\log(\tilde{d}(x))dx\|\gamma'(t)\| = \tag{34}$$

$$\|b-a\|^2 \int_{\mathbb{R}}\int_{\mathbb{R}} \frac{-\xi}{\sqrt{(t-\theta)^2\|b-a\|^2+\xi^2}} K'\left(\sqrt{(t-\theta)^2\|b-a\|^2+\xi^2}\right)\log(\tilde{d}(\theta,\xi))d\xi d\theta. \tag{35}$$

The density $\tilde{d}$ for the γ curve is expressed in ξ, θ coordinates as:

$$\int_{[0,1]} K\left(\sqrt{(t-\theta)^2\|b-a\|^2+\xi^2}\right)dt$$

and is an even function in ξ. The same is true for $K'\left(\|\gamma(t) - x\|\right)$. Finally, the mapping:

$$\xi \mapsto \frac{-\xi}{\sqrt{(t-\theta)^2\|b-a\|^2+\xi^2}}$$

is odd for a fixed θ so that the whole integrand is odd as a function of ξ. By the Fubini theorem, integrating first in ξ will therefore yield a vanishing integral, proving the assertion. □

The result still holds in $\mathbb{R}^q$, the only different aspect being that x is now expanded as $x = a + \theta v + \sum_{i=1}^{q-1} \xi_i u_i$ with $u_i, i = 1, \ldots, q-1$ an orthonormal basis of the orthogonal complement of $\mathbb{R}v$ in $\mathbb{R}^q$. Rewriting $\gamma(t) - x = (t - \theta)v - \sum_{i=1}^{q-1} \xi_i u_i$ and $\|\gamma(t) - x\| = \sqrt{(t-\theta)^2\|b-a\|^2 + \sum_{i=1}^{q-1}\xi_i^2}$, the same parity argument can be applied on any of the components $\xi_i, i = 1, \ldots, q-1$, showing that the integral is vanishing.

The effect of curve straightening is present when minimizing the entropy of a whole curve system, but is counterbalanced by the gathering effect. Depending on the choice of the kernel bandwidth, one or the other effect is dominant: straightening is preeminent for low values, being

the only remaining effect in the limit, while gathering dominates at high bandwidths. For the air traffic application, a rule of the thumb is to take 2-3 times the separation norm as an effective support for the kernel. Using an adaptive bandwidth may be of some interest also: starting with medium to high values favors curves gathering, then gradually reducing it will straighten the trajectories.

Using the scaled arclength in the entropy gives an equivalent but somewhat easier to interpret result. Starting with the expression 7, that takes in this case the form:

$$\tilde{d}\colon x \mapsto \frac{\sum_{i=1}^{N} l_i \int_0^1 K\left(\|x-\gamma_i(\eta)\|\right) d\eta}{\sum_{i=1}^{N} l_i}. \tag{36}$$

Let $i \in \{1,\dots,N\}$ be fixed. An admissible variation of the curve γ_i is a smooth mapping from $]-a,a[\times[0,1]$ to $\mathbb{R}^q$, with $a>0$ satisfying the following properties:

(a) $\forall \eta \in [0,1], \phi(0,\eta) = \gamma_i(\eta)$.
(b) $\forall (t,\eta) \in]-a,a[\times]0,1[, \|\partial_\eta \phi(t,\eta)\| = l_\phi(t)$ with $l_\phi(t)$ the length of the curve $\eta \mapsto \phi(t,\eta)$.
(c) $\forall t \in]-a,a[, \phi(t,0) = \gamma_i(0),\ \phi(t,1) = \gamma_i(1)$.

Taking the derivative with respect to t at 0 of equation (b) yields:

$$\left\langle \partial_t \partial_\eta \phi(0,\eta), \partial_\eta \phi(0,\eta) \right\rangle = \partial_t l_\phi(0) l_i.$$

Letting $T(\eta)$ be the unit tangent vector to γ_i at η, and noting that $\partial_\eta \phi(0,\eta) = l_i T(\eta)$, it comes:

$$\left\langle \partial_t \partial_\eta \phi(0,\eta), T(\eta) \right\rangle = \partial_t l_\phi(t). \tag{37}$$

This relation puts a constraint on the variation of the tangential component of the curve derivative and shows that it has to be constant in η.

Proposition 5. *Let D be the mapping from $]-a,a[\times\mathbb{R}^q$ to $\mathbb{R}^+$ defined by:*

$$D\colon (t,x) \mapsto \frac{\sum_{j=1,j\neq i}^{N} l_j \int_0^1 K\left(\|x-\gamma_j(\eta)\|\right) d\eta + \int_0^1 K\left(\|x-\phi(t,\eta)\|\right) d\eta}{\sum_{j=1}^{N} l_j}.$$

where η refers collectively to the scaled arclength parameter for each curve. The partial derivative $\partial_t D(0,x)$ is given by :

$$\partial_t D(0,x) = \frac{l_i}{\sum_{j=1}^{N} l_j} \int_0^1 \left\langle \frac{\gamma_i(\eta)-x}{\|\gamma_i(\eta)-x\|}, \partial_t \phi(0,\eta) \right\rangle K'\left(\|\gamma_i(\eta)-x\|\right) d\eta.$$

The proof is straightforward and is omitted. From Proposition 5 the variation of the entropy is derived:

$$\partial_t E = -\int_{\mathbb{R}^q} \frac{l_i}{\sum_{j=1}^{N} l_j} \int_0^1 \left\langle \frac{\gamma_i(\eta)-x}{\|\gamma_i(\eta)-x\|}, \partial_t \phi(0,\eta) \right\rangle K'\left(\|\gamma_i(\eta)-x\|\right) d\eta dx. \tag{38}$$

This relation is equivalent to 28: it can be seen by splitting the terms into a normal and a tangential component. The first one yields:

$$-\int_{\mathbb{R}^q} \frac{l_i}{\sum_{j=1}^{N} l_j} \int_0^1 \left\langle \left(\frac{\gamma_i(\eta)-x}{\|\gamma_i(\eta)-x\|}\right)_{\mathcal{N}}, \left(\partial_t \phi(0,\eta)\right)_{\mathcal{N}} \right\rangle K'\left(\|\gamma_i(\eta)-x\|\right) d\eta dx.$$

For the tangential part, the starting point is the relation:

$$\begin{aligned} \partial_\eta \left(K\left(\|\phi(0,\eta)-x\|\right) T(\eta)\right) =& l_i K'\left(\|\phi(0,\eta)-x\|\right) \left\langle \frac{\phi(0,\eta)-x}{\|\phi(0,\eta)-x\|}, T(\eta) \right\rangle T(\eta) && (39)\\ &+ K\left(\|\phi(0,\eta)-x\|\right) \partial_\eta T(\eta). && (40) \end{aligned}$$

where the subscript $\mathcal{T}$ stands for tangential component. It comes:

$$\frac{l_i}{\sum_{j=1}^N l_j} \int_0^1 \left\langle \left(\frac{\gamma_i(\eta) - x}{\|\gamma_i(\eta) - x\|} \right)_{\mathcal{T}}, (\partial_t \phi(0,\eta))_{\mathcal{T}} \right\rangle K'\left(\|\gamma_i(\eta) - x\|\right) d\eta dx = \tag{41}$$

$$\frac{l_i}{\sum_{j=1}^N l_j} \int_0^1 \left\langle \partial_\eta \left(K\left(\|\phi(0,\eta) - x\|\right) T(\eta) \right), (\phi(0,\eta))_{\mathcal{T}} \right\rangle d\eta dx \tag{42}$$

$$- \frac{l_i}{\sum_{j=1}^N l_j} \int_0^1 \left\langle K\left(\|\phi(0,\eta) - x\|\right) \partial_\eta T(\eta), (\phi(0,\eta))_{\mathcal{T}} \right\rangle d\eta dx. \tag{43}$$

With an integration by parts, the first integral in the right hand side becomes:

$$- \frac{l_i}{\sum_{j=1}^N l_j} \int_0^1 \left\langle K\left(\|\phi(0,\eta) - x\|\right) T(\eta), \partial_\eta \left(\partial_t \phi(0,\eta)\right)_{\mathcal{T}} \right\rangle d\eta dx = \tag{44}$$

$$- \frac{l_i}{\sum_{j=1}^N l_j} \int_0^1 K\left(\|\phi(0,\eta) - x\|\right) \partial_t l_\phi(0) d\eta dx. \tag{45}$$

Gathering terms, the expression 28 is recovered. As expected, only the normal components enter the relation, but it has to be noted that the tangential component of $\partial_t \phi(0,\eta)$ is not arbitrary and can be deduced from 37. The gradient with respect to the i-th curve is obtained from the expression of the entropy variation and can be written in its simplest form as:

$$\frac{l_i}{\sum_{j=1}^N l_j} \int_{\mathbb{R}^q} \int_0^1 \frac{\gamma_i(\eta) - x}{\|\gamma_i(\eta) - x\|} K'\left(\|\gamma_i(\eta) - x\|\right) d\eta \log \tilde{d}(x) dx. \tag{46}$$

where $\tilde{d}$ is the estimated spatial density. One must keep in mind the constraint on $\partial_t \phi(0,\eta)$ that is hidden within the apparent simplicity of the expression.

3. Numerical implementation

The two formulations 31, 46 of the gradient may be used. The first one is more complicated but does not require any additional constraint to be taken into account. The second one cannot be applied readily as the tangential component must comply with relation 37. In both cases, F: it is needed to evaluate a spatial integral, which may yield to prohibitive computational time, especially in high dimensions. In the air traffic application, only planar of 3D curves are considered, greatly simplifying the problem. Nevertheless, the performance of the algorithms is still a concern and the choice made was to replace the spatial integral by a discrete sum over a evenly spaced grid. From now, it is assumed that all curves are planar, so that the ambient space for the spatial density $\tilde{d}$ is $\mathbb{R}^2$. Going back to the expression of $\tilde{d}$ given by 7, a first step is to replace the integral over t by a discrete sum. In practice, curves are described by a sequence of sampled points $\gamma_i(t_{ij})$ where the sampling times t_{ij} will be assumed to be identical for all curves. This assumption is not satisfied in the air traffic application so that a pre-processing step must be taken before the actual entropy minimization stage. It will not be described here as any standard interpolation procedure can be applied with negligible differences on the final result. To obtain the results presented here, a cubic spline smoother was used. Since the sampling times are assumed to be the same for all trajectories, the double subscript will be dropped, so that the samples on each trajectory will be denoted as $\gamma_{ij} = \gamma_i(t_j)$. It is further assumed that the derivative $\gamma'_{ij} = \gamma'_i(t_i)$ is available, most of the time through a numerical approximation. Given a quadrature formula on $[0,1]$ with points $t_j, j = 1,\dots,m$ and weights w_j, $j = 1,\dots,m$, the density may be approximated at $x \in \mathbb{R}^2$ by:

$$\tilde{d}(x) = \frac{1}{\sum_{i=1}^N l_i} \sum_{i=1}^N \sum_{j=1}^m w_j K\left(\|x - \gamma_{ij}\|\right) \|\gamma'_{ij}\|. \tag{47}$$

where the lengths $l_i, i = 1, \ldots, N$ are also obtained with the same quadrature rule:

$$l_i = \sum_{j=1}^{m} w_j \|\gamma'_{ij}\|.$$

When γ'_{ij} is computed in a numerical way, it may be expressed as:

$$\gamma'_{ij} = \sum_{k=1}^{m} \tilde{w}_{jk} \gamma_{i,k}.$$

where the weights $\tilde{w}_{jk}$ are often obtained through the application of the Lagrange interpolation formula to ensure exactness on polynomials up to a given degree. In a more compact form, it can be written in matrix form as:

$$\begin{pmatrix} \gamma'_{i1} \\ \vdots \\ \gamma'_{iq} \end{pmatrix} = \tilde{W} \begin{pmatrix} \gamma_{i1} \\ \vdots \\ \gamma_{iq} \end{pmatrix}$$

where the matrix $\tilde{W}$ has entries the weights $\tilde{w}_{jk}$. The cost of evaluating $\tilde{d}$ at a single point is in $o(Nm)$, with the kernel evaluation being dominant. When dealing with points in $\mathbb{R}^2$ or $\mathbb{R}^3$ and compactly supported kernels a simple trick greatly reduces the time needed to compute $\tilde{d}$. First of all, the domain of interest is discretized on a evenly spaced grid, so that points of evaluation of the density $\tilde{d}$ are its vertices $x_{ij}, i = 1, \ldots, n_x, j = 1, \ldots, n_y$. The grid step δ_x (resp. δ_y) in the first (resp. second) coordinate is the difference between any two adjacent vertices $\delta_x = x_{i+1,j} - x_{i,j}$ (resp. $\delta_y = x_{i,j+1} - x_{i,j}$ (most of the time, $\delta_x = \delta_y$). Since the expression 47 is linear, the computation can be performed by accumulating values $K(\|x_{kl} - \gamma_{ij}\|)\|\gamma'_{ij}\|$ for a fixed couple (i, j), where only the points x_{kl} close enough to γ_{ij} are considered. In fact, the evaluation can be written as a 2D discrete convolution:

$$\tilde{d}(x_{kl}) = \sum_{i=1\ldots,N, j=1\ldots m} w_j K(\|x_{kl} - \gamma_{ij}\|)\|\gamma'_{ij}\|. \tag{48}$$

When the support of K is small compared to the overall spatial domain, a lot of computation is saved using this procedure. Furthermore, it can be thought as 2D filtering, so that highly efficient algorithms coming from the field of image processing can be applied: in particular, computing the density on a graphics processing unit (GPU) is straightforward and allows to decrease the computational time by at least a factor ten. When dealing with the scaled arclength, the derivative term is not present, an a factor of l_i appears in from of the integral. The discrete version becomes:

$$\tilde{d}(x_{kl}) = \sum_{i=1\ldots,N, j=1\ldots m} l_i w_j K(\|x_{kl} - \gamma_{ij}\|) \tag{49}$$

where $\gamma_{ij} = \gamma_i(\eta_j)$, η_j being in correspondence with t_j. Please note that the quadrature weights must be adapted to the abscissa $\eta_j, j = 1 \ldots m$ and not to the $t_j, j = 1 \ldots m$. Therefore, it is advisable to resample the curves so that the points $\eta_j, j = 1 \ldots m$ are for example evenly spaced or of the Gauss-Lobatto form. The former was chosen for the experiments due to its ease of implementation, although the second form is probably more efficient from a numerical point of view and will be investigated in a second stage.

Having the density at hand, the gradient of the entropy with respect to the points $\gamma_{ij}, i = 1 \ldots N, j = 1 \ldots m$ can be easily computed using a straightforward application of the formula 31. When dealing with planar curves, a simplification occurs for the second derivative term since for a smooth curve γ_j:

$$\left(\frac{\gamma''_j(t)}{\|\gamma'_j(t)\|} \right)_{\mathcal{N}} = \kappa(t) N(t).$$

where κ is the curvature and N the unit normal vector. These quantities may be computed using numerical differentiation, but a coarse approximation based on the rotation rate of the vectors $\gamma_{i,j+1} - \gamma_{i,j}$, $\gamma_{i,j+2} - \gamma_{i,j+1}$ works well in many cases.

The case of scaled arclength parametrization needs some extra attention, due to the condition on the tangential component. The simplest approach is to move the points γ_{ij} according to an unconstrained gradient, then to re-sample the obtained curve so as get adjusted γ_{ij} that correspond to the abscissa $\eta_j, j = 1 \dots m$.

In a numerical implementation, the scaling factor in front of the whole expression may be dropped due to the fact that all gradient-based algorithms will use an automatically tuned step length. As usual with gradient algorithms, one must carefully select the step taking in the maximizing direction in order to avoid divergence. A simple fixed step strategy was first applied and gives satisfactory results on small datasets. A safer approach is to adapt the step size so as to ensure a sufficient decrease of the entropy. Due to the potentially huge dimension of the search space, this procedure has to be simple enough. An approximate quadratic search [13] was used in final implementation.

The procedure applied to one day of traffic over France yields the picture of Figure 3. As

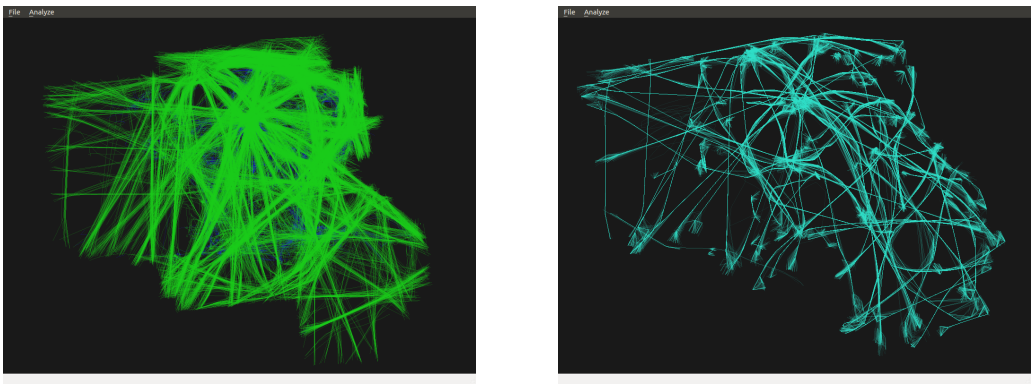

Figure 3. Initial and Bundled traffic of the 24/02/2013.

expected, a route-like network emerges. In such a case, since the traffic comes from an already organized situation, the recovered network is indeed a subset of the route network in the french airspace. Please note that there is a trade-off between density concentration and minimal curvature of the recovered trajectories, as already mentioned. The kernel bandwidth was chosen empirically in the example presented, with the aid of visual interaction.

In the second example of Figure 4, the problem of automatic conflict solving is addressed. In the initial situation, aircraft are converging to a single point, which is unsafe. Air traffic controllers will proceed in such a case by diverting aircraft from their initial flight path so as to avoid each other, but only using very simple maneuvers. An automated tool will make a full use of the available airspace and the resulting set of trajectories may fail to be manageable by a human: in the event of a system failure, no backup can be provided by controllers. The entropy minimization procedure was added to an automated conflict solver in order to end up with flight paths still tractable by humans. The final result is shown on the right part of Figure 4, where encounters no longer exists but aircraft are bound to simple trajectories, with a merging and a splitting point. Note that since the automated planner acts on velocity, all aircraft are separated in time on the inner part.

4. Conclusion and future work

Algorithms coming from the field of shape spaces emerge as a valuable tool for applications in ATM. In this work, the foundations of a post processing procedure that may be applied after an automated flight path planner are presented. Entropy minimization makes straight segments bundle

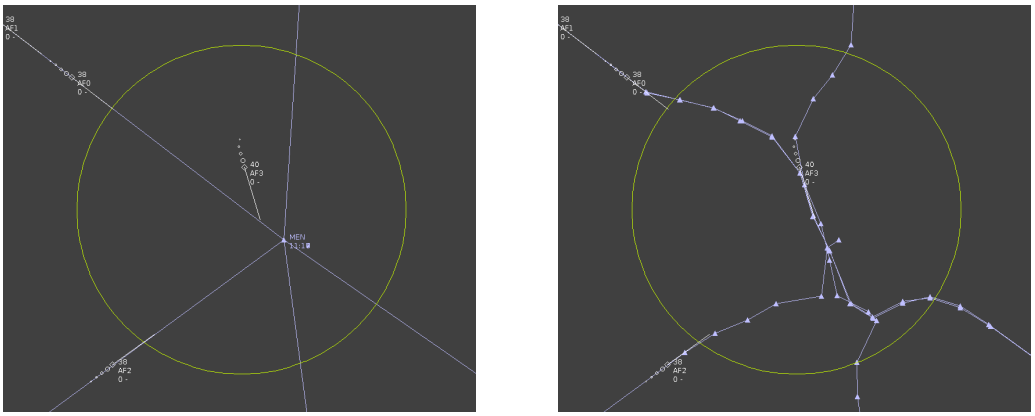

Figure 4. Initial flight plans and final ones.

emerge, which fulfills the operational requirements. Computational efficiency has to be improved in order to release an usable building block for future ATM systems. One way to address this issue is to compute kernel density estimators using GPUs which excel in this kind of task, very similar texture manipulations. Furthermore, statistical properties such as the optimal choice of the bandwidth parameter in the kernel estimation should be explored in more details in the next step of this work.

Another important point that must be addressed in future works deals with the flight paths that are very similar shape but are oriented in opposite direction. As the spatial density is not sensitive to the directional information, the entropy based procedure presented in this paper will tend to aggregate flight paths that should be sufficiently separated in order to prevent hazardous encounters. In [7], a notion of density based on position and velocity is developed. This work relies on a Lie group modeling as an unifying state representation that take into account the direction and the position of the curves. The curve system entropy has been extended to this setting.

Bibliography

1. L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows: In Metric Spaces And In The Space Of Probability Measures*. Lectures in math. Birkhäuser, 2005.
2. V. A. Epanechnikov. Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications*, 14(1):153–158, 1969.
3. EUROCONTROL/NMD/STATFOR. Eurocontrol seven-year forecast. https://www.eurocontrol.int/sites/default/files/content/documents/official-documents/forecasts/seven-year-flights-service-units-forecast-2014-2020-sep2014.pdf, 2014.
4. W.H. Harman. Air traffic density and distribution measurements. Lincoln Laboratory, MIT, Report ATC-80, May 1979.
5. C. Hurter, O. Ersoy, and A. Telea. Smooth bundling of large streaming and sequence graphs. In *Visualization Symposium (PacificVis), 2013 IEEE Pacific*, pages 41–48, Feb 2013.
6. Peter W Michor and David Mumford. Riemannian geometries on spaces of plane curves. *J. Eur. Math. Soc. (JEMS)*, 8:1–48, 2006.
7. Florence Nicol and Stéphane Puechmorel. Unsupervised aircraft trajectories clustering: a minimum entropy approach. In *ALLDATA 2016, The Second International Conference on Big Data, Small Data, Linked Data and Open Data*, Lisbon, Portugal, 2016. IARIA.
8. Emanuel Parzen. On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33(3):1065–1076, 09 1962.
9. Murray Rosenblatt. Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, 27(3):832–837, 09 1956.

10. G.P. Roussos, D.V. Dimarogonas, and K.J. Kyriakopoulos. Distributed 3d navigation and collision avoidance for nonholonomic aircraft-like vehicles. In *Proceedings of the 2009 European Control Conference*, 2009.
11. D.W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. A Wiley-interscience publication. Wiley, 1992.
12. B.W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1986.
13. W. Sun and Y.X. Yuan. *Optimization Theory and Methods: Nonlinear Programming*. Springer Optimization and Its Applications. Springer US, 2006.