Learning Probabilistic Models of Contours

Laure Amate, Maria João Rendas
Laboratoire I3S, CNRS-UNSA, 2000 Rte des Lucioles, 06903 Sophia Antipolis, FRANCE
{amate}{rendas}@i3s.unice.fr

Abstract

We present a methodology for learning spline-based probabilistic models for sets of contours, proposing a new Monte Carlo variant of the EM algorithm to estimate the parameters of a family of distributions defined over the set of spline functions (with fixed complexity). The proposed model effectively captures the major morphological properties of the observed set of contours as well as its variability, as the simulation results presented demonstrate.

1. Introduction

In many situations it is important to be able to find compact representations of the collective morphological properties of a set of objects. In this paper we address this problem, proposing a formal modeling framework to address it, and presenting methods for automatic identification of such models. The morphological properties we consider are the *shape* of planar objects, mathematically described by simple closed curves in the ambient 2D Euclidean space.

Most often, the goal is to capture the intrinsic properties of a given set of objects, independently of the variability induced by viewpoint. The seminal approach of Kendall [8], that formally defines the "shape" of an object as the associated class of equivalence with respect to the action of basic groups of geometric operators, provides an appropriate setting for these studies. One can distinguish two distinct approaches to the definition of "shape spaces": *(i)* the original discrete theory of Kendall [8] and *(ii)* the more recent theory of continuous closed curves presented in [9]. The discrete theory admits the existence of recognizable landmarks on the objects – it thus is intrinsically parsimonious – and the definition of probability distributions can be approached by using the exponential maps, see [12]. It can lead to unstable results when landmark definition is done automatically. The continuous theory provides a more fundamental approach, but leads to shape spaces of infinite dimension. It presents two main drawbacks: it is subject to numerical instabilities, and, more fundamentally, it is still unclear how probability distributions can be defined in them on a convenient manner.

We try in our work to combine the advantages of both approaches, basing our work on *spline* representations of (closed planar) curves: piecewise continuous polynomials defined by sets of knots and control points. Our shapes are thus continuous, benefiting from the plasticity and compactness of spline representations. To model the collective properties of a set of shapes, we propose a generative model for the observed curves that relies on the definition of a parametric family of probability distributions over the spline parameters.

In the context of curve modeling, splines have been widely used and studied [11, 10, 5]. In particular, in [6], the authors address the problem of fitting a fixed knot spline whose complexity is automatically chosen using the MDL principle to the contour between two distinct regions in a digital image. In [2] we addressed the basic problem of fitting free-knots splines of varying complexity to a *single* observed curve, and we concentrate here on the estimation of probability distributions over the spline parameter space.

We start by presenting spline functions and associated spaces in section 2. We then define a parametric model on the space of spline parameters, section 3, and use it to define (section 4) the model that will be used for Maximum Likelihood using a variant of the Expectation-Maximization (EM) algorithm in section 5. Finally, section 6 presents simulation results of the proposed algorithm, and section 7 summarizes our contributions.

2. Model space: Free Knot Splines

We remember in this section the basic definition of splines. The interested reader is referred to [4] for a full presentation.

A *spline* of order m over the interval $I = [\tau_1 \; \tau_{l+1}]$ is a piecewise polynomial (*pp*) function of degree $m-1$ with $l+1$ breakpoints $\tau = (\tau_i)_{i=1}^{l+1}$:

$$f(t) = P_i(t), \qquad \tau_i \leq t < \tau_{i+1}\;, i = 1 \cdots l\;. \quad (1)$$

Each $P_i(t)$, $i = 1 \cdots l$ is a polynomial of degree $m-1$ (we fix $m = 4$, *cubic splines*) on $[\tau_i, \tau_{i+1}]$. Closed curves are conveniently modeled considering an infinite periodic extension of I (we assume a normalized representation with $I = [0, 1]$), obtained by setting $f(t) = f(t+k), k \in \mathbb{N}$. To handle invariance with respect to the definition of origin, we assume that the curve is parametrized such that its origin coincides with the limit of one polynomial piece, i.e., $\tau_i = 0$ and that the polynomial piece of *largest support* is the first one.

It can easily been checked that Π_τ, the set of all *pp* functions $f(t)$ of order m with breaks τ, inherits the vector space structure of polynomials. The knot vector ξ (of dimension $k+m$, $k \geq l$) simultaneously codes information about the breakpoints τ_i and the continuity conditions enforced at each of them. The set of spline functions of order m with knots at ξ, $\mathbb{S}_\xi \subset \Pi_\tau$, is still a vector space. A special basis of this space, the B-spline functions $\{b_j(t;\xi)\}_{j=1}^k$, is a common choice for spline representation, since it satisfies a set of recursive relations that can be exploited for computational purposes. Thus, if $c(t) \in \mathbb{S}_\xi$, it exists a set of coefficients $\beta^k = \{\beta_j\}_{j=1}^k$ such that

$$c(t) = \sum_{j=1}^{k} \beta_j \, b_j(t;\xi) = b(t;\xi)^T \, \beta^k \quad , \tag{2}$$

where we introduced the k-dimensional vector function $b(t;\xi)$ that groups the k B-spline functions. Since ξ is ordered, its elements can be written

$$\xi_j = \sum_{i=1}^{j-1} \Delta_i, \; j > 1 \quad \Delta_i = \xi_{i+1} - \xi_i, \quad \xi_1 = 0 \quad ,$$

where $0 \leq \Delta_j \leq 1$. Vector $\Delta = [\Delta_1 \ldots \Delta_k]$ belongs thus to the $(k-1)$-dimensional simplex $\mathcal{M}^k$. We will consider parametrization of the splines by Δ instead of direct use of ξ. We will stress this using the alternative notation $b(t;\Delta)$ in (2).

Set $\mathbb{S}^k$ of the (cubic) splines with fixed *number of knots k*

$$\mathbb{S}^k = \left\{ \mathbb{S}_\xi ; \xi = (\xi_i)_{i=1}^k \right\} \quad ,$$

plays a central role in our control modeling approach, and in the following section we define a parametric probability distribution over its parameter space Θ^k:

$$w = (\beta, \Delta) \in \Theta^k = \mathbb{C}^k \times \mathcal{M}^k \quad , \tag{3}$$

the product of the coefficient space (the k-dimensional complex space $\mathbb{C}^k$ – considering the complex representation of planar curves) and of the inter-knot spacings space $\mathcal{M}^k$.

3. Parametric Probability Model

We parametrize distributions over Θ^k by

$$\gamma = \left[\mu_0, \sigma^2, \alpha\right]$$

with the following factored structure

$$p_{\mu_0,\sigma^2,\alpha}(\beta, \Delta) = p_{\mu_0,\sigma^2}(\beta|\Delta) p_\alpha(\Delta) \quad , \tag{4}$$

i.e., with dependency structure

$$\begin{array}{ccc} \alpha & \longrightarrow & \Delta \\ & & \downarrow \\ \mu_0, \sigma^2 & \longrightarrow & \beta \end{array}$$

The first factor in (4) is the normal distribution over $\mathbb{C}^k$ with mean $\mu_0 \in \mathbb{C}^k$ and information matrix $\Sigma(\Delta)^{-1} = \frac{1}{\sigma^2} \int_I b(t;\xi) b^T(t;\Delta) \, dt$, $(\sigma^2 > 0)$ proportional to the Gramian matrix of the vector of B-spline basis functions $b(t,\Delta)$ with inter-knot vector Δ. Vector Δ is drawn from a Dirichlet distribution [7] with parameter $\alpha \in \mathbb{R}^{+k}$:

$$\Delta \sim \mathcal{D}(\Delta|\alpha) = \frac{1}{B(\alpha)} \prod_{j=1}^{k} (\Delta_j)^{\alpha_j - 1} \quad , \tag{5}$$

where $B(\alpha) = \frac{\prod_{j=1}^k \Gamma(\alpha_j)}{\Gamma\left(\sum_{j=1}^k \alpha_j\right)}$, and $\Gamma(\cdot)$ is the Gamma function. This distribution has a single mode at $\left(\frac{\alpha_1}{\bar{\alpha}}, \frac{\alpha_2}{\bar{\alpha}}, \cdots, \frac{\alpha_k}{\bar{\alpha}}\right)$, where $\bar{\alpha} = \sum_{j=1}^k \alpha_j$, its dispersion being determined by $\bar{\alpha}$.

The parameter space of the family (4) of distributions over Θ^k is thus

$$\mathcal{G} = \mathbb{C}^k \times \mathbb{R}_*^+ \times \mathbb{R}^{+k} \quad . \tag{6}$$

In the following section we present a generative model for the observed set of contours that is based on this family.

4. Generative model

We now address the problem of learning the value of γ that best fits the set of N observed contours $Z = \{Z_{(i)}\}_{i=1}^N$, $Z_{(i)} = \{z_{(i)}(t_{i,n}), n = 1, \ldots, N_{(i)}\}$. Our observation assumes that each observed contour is a noisy sampled version of a spline curve with complexity (number of knots) k:

$$z_{(i)}(t_{i,n}) = c_{(i)}(t_{i,n}) + \epsilon_{(i)}(t_{i,n}) = B_{(i)}(\Delta_{(i)})\beta_{(i)} + \epsilon(t_{i,n}),$$

where $B_{(i)}(\Delta_{(i)})$ is the design matrix of generic element $\left[B_{(i)}(\Delta_{(i)})\right]_{pq} = b_q(t_{i,p}; \Delta_{(i)})$. We assume that

$$\epsilon(t_{i,n}) \overset{iid}{\sim} \mathcal{N}\left(0, \sigma^2\right)$$

are statistically independent of the parameters w of the spline curve, which are *iid* samples from (4)

$$\begin{aligned} \{\Delta_{(i)}\}_{i=1}^N &\overset{iid}{\sim} \mathcal{D}(\Delta|\alpha) \\ \{\beta_{(i)}|\Delta_{(i)}\}_{i=1}^N &\overset{iid}{\sim} \mathcal{N}\left(\mu_0, \Sigma(\Delta_{(i)})\right) \quad . \end{aligned}$$

Assumptions above allow us to write the conditional density of the observations Z given γ:

$$\begin{aligned} p(Z|\gamma) &= \prod_i p(Z_{(i)}|\gamma) \\ p(Z_{(i)}|\gamma) &= \int p(Z_{(i)}, w_{(i)}|\gamma) dw \quad . \end{aligned}$$

It can be shown that the joint density

$$p(Z_{(i)}, w_{(i)}|\gamma) = p(Z_{(i)}|w_{(i)}, \gamma)p(w_{(i)}|\gamma)$$

belongs to the (curved) exponential family and,

$$p(Z_{(i)}, w_{(i)}|\gamma) = h(Z_{(i)}, w_{(i)})e^{-\Psi(\gamma)+\left\langle S(Z_{(i)},w_{(i)}),\Phi(\gamma)\right\rangle}. \quad (7)$$

See [1] for the detailed definitions of $h(\cdot,\cdot)$, $\Psi(\cdot)$, $\Phi(\cdot)$ and of the sufficient statistic $S(\cdot,\cdot)$.

5. Model identification

We now address estimation of the parameter γ using the Maximum Likelihood criterion:

$$\hat{\gamma} = \arg\max_{\gamma\in\mathcal{G}} p(Z|\gamma), \qquad \gamma = \left(\mu_0, \sigma^2, \alpha\right) \quad . \quad (8)$$

Note that the spline parameters $w_{(i)}$ appear in the observation model as hidden variables with respect to the estimation of γ.We estimate γ using a novel version of the EM algorithm, which is a modification of the Online EM algorithm, proposed in [3] for estimation in exponential families. We name this new algorithm *MC* (Monte Carlo) *Online EM*.

As Online EM, our algorithm processes the observed set of curves sequentially, producing an updated value of $\hat{\gamma}$ at each iteration, and, as the original EM algorithm, each iteration is composed of two basic steps, Expectation and Maximization. Online EM computes, iteratively, stochastic approximations of the complete (batch) Estimation step.

Moreover, for problems where the complete data belongs to the exponential family, as it is our case, this recursive update only requires computation of the expected value with respect to $w = (\beta, \Delta)$ of the sufficient statistic: $\bar{s}(Z_{(i)}, \hat{\gamma}_{(i-1)}) = \mathrm{E}_w\left[S(Z,w)|Z_{(i)}, \hat{\gamma}_{(i-1)}\right]$. In our case the integration over β can be done analytically (see [1] for details) leading to an expression of the general form

$$\bar{s}(Z_{(i)}, \hat{\gamma}_{(i-1)}) = \int_{\mathcal{M}^k} g\left(\Delta\right) q\left(\Delta|Z_{(i)}, \hat{\gamma}_{(i-1)}\right) d\Delta \quad (9)$$

where $q\left(\Delta|Z_{(i)}, \hat{\gamma}_{(i-1)}\right)$ is a tilted version of $p(\Delta|\gamma_{(i-1)})$. The integral over Δ in the above expression is untractable, and we resort to Monte Carlo to approximate it:

$$\bar{s}(Z_{(i)}, \hat{\gamma}_{(i-1)}) \approx \frac{1}{M^i}\sum_{j=1}^{M^i} g\left(\Delta_{(i)}^j\right) \quad , \quad (10)$$

where $\left\{\Delta_{(i)}^j\right\}_{j=1}^{M^i}$ are M^i samples of the normalized version of $q(\Delta|Z_{(i)}, \hat{\gamma}_{(i-1)})$, obtained using a Metropolis Hasting (MH) sampler. The value of (10) is then used to update the log-likelihood

$$\ell(\gamma)_{(i)} = -\Psi(\gamma) + \left\langle\hat{s}_{(i)}, \Phi(\gamma)\right\rangle \quad . \quad (11)$$

In this equation

$$\hat{s}_{(i)} = \hat{s}_{(i-1)} + \eta_i\left(\bar{s}(Z_{(i)}, \hat{\gamma}_{(i-1)}) - \hat{s}_{(i-1)}\right) \quad , \quad (12)$$

with $\eta_i = \eta_0 i^{-\kappa}$, $\kappa \in]1/2, 1[$, $\eta_0 \in [0, 1]$, following [3].

Maximization of $\ell(\cdot)_{(i)}$, (11), can be split into two independent maximizations, over α, and over the parameters (μ_0, σ^2) (see [1] for details). We use numerical optimization methods to solve each problem.

6. Results

In this section, we show results obtained on curves sampled from model (4) for cubic splines $k = 12$ knots. Figure 1 shows 20 of the 200 simulated curves. The red crosses show μ_0, $\sigma^2 = 2.3\,10^{-3}$, and $\alpha = [3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27]$.

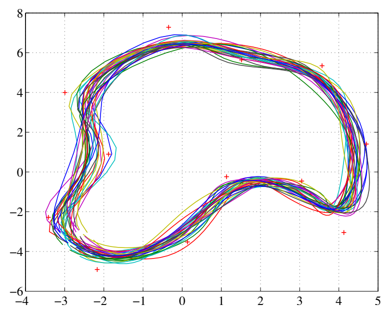

Figure 1. Subset of the simulated curves .

Results obtained with 3 different proposals for the MH sampler are shown in figure 2. We can see that the estimated μ_0 (far left) are similar and close to its true value. While no significant differences can be detected on the estimates of the Dirichlet mode $\frac{\alpha}{\bar{\alpha}}$, which are close to the true value for all the proposals, α is better estimated with one of them (random changes of a knot in its neighborhood, according to a triangular distribution), in magenta, while with the other proposals the Markov chain is not well mixed (second plot). The noise covariance σ^2 shows a small bias in all cases, either positive or negative for these examples (right plot).

The locality of the EM algorithm makes it sensitive to initialization. Figure 3 show results of Monte Carlo runs of MCOnlineEM with different initializations, which clearly show that it can fail to identify the correct mode of the likelihood.

Finally, Figure 4 plots random samples of the models identified. The green curves (left) correspond to a good initialization, while the magenta curves (right) show samples from a model corresponding to a local mode of the likelihood, showing that even if the overall shape is consistent with the true model in both cases, the variability of the local estimate does not reflect that of the true model (higher σ^2, wrong control points mode and dispersion).

7. Conclusions

In this paper, we propose a methodology to learn models of sets of planar closed curves. It is a proba-

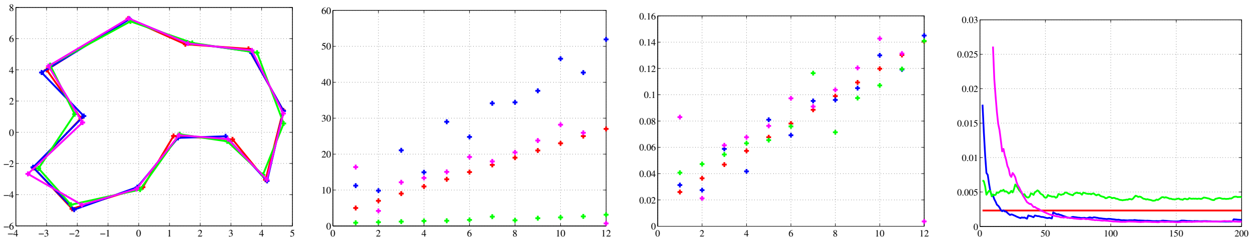

Figure 2. Real (red) and estimated values of, left to right, μ_0, α (vector index in abscissa), $\frac{\alpha}{\bar{\alpha}}$ and σ^2 (iterations in abscissa), with different proposals in magenta (P_1), green (P_2) and blue (P_3).

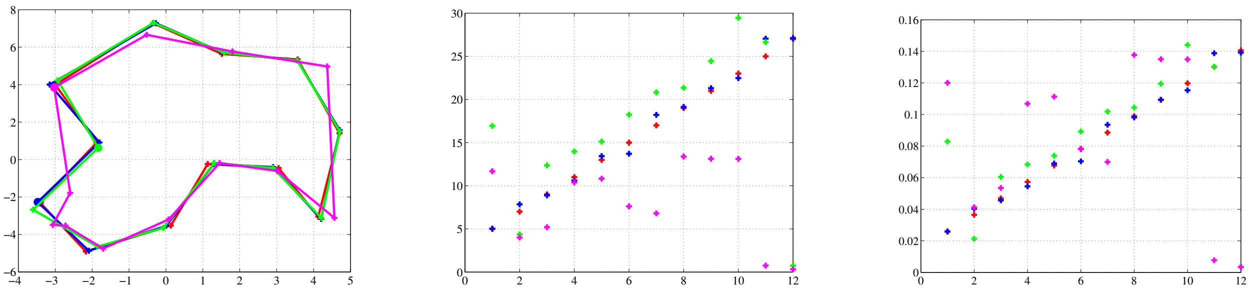

Figure 3. Parameters estimated with different initializations (magenta, green and blue): from left to right, μ_0, α and $\frac{\alpha}{\bar{\alpha}}$. Real model is in red.

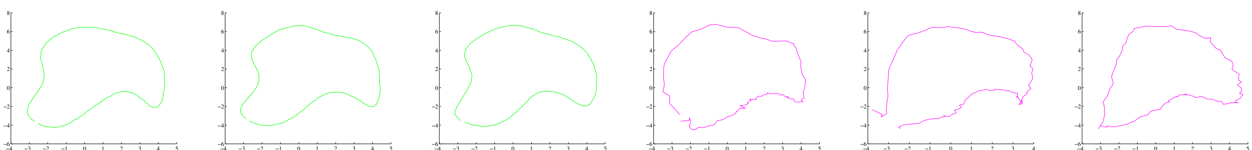

Figure 4. Simulated curves from the model estimated with a good estimation (left, in green) and with a bad estimation (right, in magenta).

bilistic approach, relying on the definition of a parametric family of distributions in the set of spline parameters and a new variant (Monte Carlo version) of the Online EM algorithm proposed in [3] to perform Maximum Likelihood estimation in this model. We present simulation results that illustrate the behavior of the algorithm, in particular the problems of choice of the importance sampling scheme used in the MC approximation and its sensitivity with respect to initialization. Our results show that even when the EM is trapped in a local mode the most important morphological characteristics of the set of contours are still captured, while its variability structure can be poor.

Acknowledgements
This work has been sponsored by French Defense (DGA) and National French Research Council (CNRS).

References

[1] L. Amate. *Learning sparse spline-based shape models*. PhD thesis, Université de Nice-Sophia Antipolis, 2009.

[2] L. Amate and M. Rendas. Sparse model fitting in nested families: Bayesian approach vs. penalized likelihood. In *Proc. Eusipco 2009, Glasgow, UK*, 2009.

[3] O. Cappé and E. Moulines. Online expectation maximization algorithm for latent data models. *J. Roy. Stat. Soc. B*, 71, 2009.

[4] C. DeBoor. *A Practical Guide to Splines*. Springer, 1978.

[5] I. DiMatteo & al. Bayesian curve-fitting with free-knot splines. *Biometrika*, 88:1055–1071, 2001.

[6] M. Figueiredo & al. Unsupervised contour representation and estimation using b-splines and a minimum description length criterion. *IEEE Trans. on Image Processing*, 9(6):1075–1087, June 2000.

[7] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Discrete multivariate distributions*. Wiley - Interscience, 1997.

[8] D. Kendall & al. *Shape and Shape Theory*. John Wiley and sons, 1999.

[9] E. Klassen & al. Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:372–383, 2004.

[10] M. J. Lindstrom. Penalized estimation of free-knot splines. *J. Comp. Graph. Stat.*, 8(2):333–352, 1999.

[11] Z. Luo and G. Wahba. Hybrid adaptive splines. *J. Am. Stat. Assoc.*, 92(437):107–116, 1997.

[12] X. Pennec. Probabilities and statistics on Riemannian manifolds: Basic tools for geometric measurements. In *Proc. of NSIP'99*, vol. 1, 194–198, Turkey, 1999.