Supplemental Material: Fast Two-View Motion Segmentation Using Christoffel Polynomials

B. Ozbay, O. Camps, and M. Sznaier

ECE Dept., Northeastern University, Boston, MA 02115, USA
ozbay.b@northeastern.edu, {camps,msznaier}@coe.neu.edu

Abstract. In this supplemental material we provide the theorerical background in algebraic geometry and Christoffel functions that justifies our approach.

1 Algebraic Geometry Background

In this section we briefly cover the theoretical results that allow for segmenting algebraic varieties by estimating the polynomials that define each one.

Definition 1 *An algebraic variety is the set of solutions of a system of multivariate polynomial equations over the real or complex numbers.*

In this paper we consider second order varieties V_i defined by the real roots of a single multivariate polynomial $p(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x_1}^T \mathbf{F} \mathbf{x}_2$.

Definition 2 *A variety arrangement $\mathscr{A}$ in $\mathbb{R}^d$ is the union of n_v algebraic varieties $V_i \in \mathbb{R}^d$:*

$$\mathscr{A}(V) \doteq V_1 \cup V_2 \cup \ldots \cup V_{n_v} \tag{1}$$

Definition 3 *The vanishing ideal $I(\mathscr{A})$ of a variety arrangement $\mathscr{A} \subseteq R^d$ is the set of all multivariate polynomials in d variables that vanish on all points in $\mathscr{A}$, that is:*

$$I(\mathscr{A}) \doteq \left\{ P \in \mathscr{P}^d : P(\mathbf{z}) = 0 \; \forall \mathbf{z} \in \mathscr{A} \right\} \tag{2}$$

Note that this vanishing ideal of a variety arrangement is indeed the intersection of the vanishing ideals of the individual varieties:

$$I(\mathscr{A}) = I(V_1 \cup V_2 \cup \ldots \cup V_{n_v}) = I(V_1) \cap I(V_2) \cap \ldots I(V_{n_v}) \tag{3}$$

The homogeneous component of this vanishing ideal of degree r, $I_r(\mathscr{A}) \subseteq I(\mathscr{A})$, is formed by homogeneous polynomials of degree r.

Definition 4 *Given a set I of polynomials, $\mathscr{Z}(I)$, the zero set of I is the set of all common roots:*

$$\mathscr{Z}(I) \doteq \{ \mathbf{x} \in \mathbb{R}^d : P(\mathbf{x}) = 0 \, \textit{for all} \; P \in I \} \tag{4}$$

Hence the arrangement $\mathscr{A}$ is completely characterized by its associated homogeneous ideal. Using the definitions above, the following result can be easily established:

Lemma 1. *The variety arrangement $\mathscr{A}$ is the zero set of $I_{n_v \times n}(\mathscr{A})$, the subset of I formed by homogeneous polynomials of degree $n_v \times n$.*

In principle, the results above allow for reducing the algebraic variety segmentation problem to a polynomial estimation one: given the data points, first estimate $I_{n \times n_v}$. The polynomials that generate each variety can then be estimated by factoring the generator of $I_{n \times n_v}$[1] into a product of n_v homogeneous degree n polynomials. In the case of linear subspaces, this is precisely the approach used by GPCA. However, as indicated in the paper, such an approach is fragile to noise and outliers. Thus, in this paper we pursue an alternative approach, using Christoffel function arguments, that identifies the generator of each variety "one-at-a-time".

2 Approximating support sets via Christoffel polynomials

Given a probability measure μ supported on $\mathbb{R}^d$, its associated moments sequence is given by

$$m_\alpha = \mathscr{E}_\mu(\mathbf{x}^\alpha) = \int_{\mathbb{R}^d} \mathbf{x}^\alpha d\mu \tag{5}$$

where $\mathbf{x} \doteq \begin{bmatrix} x_1 \; x_2 \; \dots \; x_d \end{bmatrix}^T$, $\alpha \doteq \begin{bmatrix} \alpha_1 \; \alpha_2 \; \dots \; \alpha_d \end{bmatrix}$ and $\mathbf{x}^\alpha$ stands for $x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_d^{\alpha_d}$. Each sequence m can be associated with a moment matrix $\mathbf{M}_n$ of size $\binom{n+d}{d}$, with entries $\mathbf{M}_{i,j} = m_{\alpha_i + \alpha_j}$, containing moments of order up to $2n$. In this paper we use the submatrix $\mathbf{L}_n$ of $\mathbf{M}_n$, of size $s_{n,d} \doteq \binom{n+d-1}{d-1}$ containing only moments of order $2n$. For instance, for moments of order 4 in two variables, we have

$$\mathbf{L}_2 = \begin{bmatrix} m_{(4,0)} & m_{(3,1)} & m_{(2,2)} \\ m_{(3,1)} & m_{(2,2)} & m_{(1,3)} \\ m_{(2,2)} & m_{(1,3)} & m_{(0,4)} \end{bmatrix}$$

The non-negative function $Q_n^{-1}(\mathbf{x}) \doteq \left[\mathbf{v}_n^T(\mathbf{x}) \mathbf{L}_n^{-1} \mathbf{v}_n(\mathbf{x})\right]^{-1}$ is known as the Christoffel function associated with $\mathbf{L}_n$ [1]. It is related to the measure μ that induces $\mathbf{L}_n$ through the following optimization problem over homogeneous polynomials of degree n [2,1]:

$$p_{\mathbf{y}}^*(.) = \underset{p \in \mathscr{P}_n^d}{argmin} \int_{\mathbb{R}^p} p^2(\xi) d\mu \;\; \text{s.t. } p(\mathbf{y}) = 1 \tag{6}$$

where $\mathbf{y}$ is an arbitrary given data point. An explicit expression for $p_{\mathbf{y}}^*(.)$ in terms of the singular vectors $\mathbf{u}_i$ and singular values σ_i of $\mathbf{L}_n$ is given by [3]:

$$\begin{aligned} & p_{\mathbf{y}}^*(.) = \mathbf{v}_n(.)^T \mathbf{c}_{\mathbf{y}}^* \\ & \text{where } \mathbf{c}_{\mathbf{y}}^* = \frac{1}{\sum_{i=1}^{s_{n,d}} (\frac{1}{\sqrt{\sigma_i}} \mathbf{u}_i^T \mathbf{v}_n(\mathbf{y}))^2} \sum_{i=1}^{s_{n,d}} \frac{1}{\sigma_i} \mathbf{u}_i^T \mathbf{v}_n(\mathbf{y}) \mathbf{u}_i \end{aligned} \tag{7}$$

[1] Under suitable conditions I is a principal ideal and this generator is unique up to a scaling factor.

As shown next, $Q_n^{-1}(\mathbf{y}) = \mathscr{E}_\mu[(p^*_{\mathbf{y}}(.))^2]$. From (6) we have

$$p^*_{\mathbf{y}}(\mathbf{x})^2 = \frac{1}{D^2} \sum_{i,j=1}^{s_{n,d}} \frac{1}{\sigma_i \sigma_j} [\mathbf{u}_i^T \mathbf{v}_n(\mathbf{y})][\mathbf{u}_j^T \mathbf{v}_n(\mathbf{y})][\mathbf{u}_i^T \mathbf{v}_n(\mathbf{x})][\mathbf{v}_n^T(\mathbf{x})\mathbf{u}_j]$$

where $D \doteq \sum_{i=1}^{s_{n,d}} (\frac{1}{\sqrt{\sigma_i}} \mathbf{u}_i^T \mathbf{v}_n(\mathbf{y}))^2 = \mathbf{v}_n^T(\mathbf{y}) \mathbf{L}_n^{-1} \mathbf{v}_n(\mathbf{y})$. Hence

$$\begin{aligned}
\mathscr{E}_\mu[(p^*_{\mathbf{y}}(.))^2] &= \frac{1}{D^2} \sum_{i,j=1}^{s_{n,d}} \frac{1}{\sigma_i \sigma_j} [\mathbf{u}_i^T \mathbf{v}_n(\mathbf{y})][\mathbf{u}_j^T \mathbf{v}_n(\mathbf{y})][\mathbf{u}_i^T \mathscr{E}_\mu \left\{ \mathbf{v}_n(\mathbf{x}) \mathbf{v}_n^T(\mathbf{x}) \right\} \mathbf{u}_j] \\
&= \frac{1}{D^2} \sum_{i,j=1}^{s_{n,d}} \frac{1}{\sigma_i \sigma_j} [\mathbf{u}_i^T \mathbf{v}_n(\mathbf{y})][\mathbf{u}_j^T \mathbf{v}_n(\mathbf{y})] \mathbf{u}_i^T \mathbf{L}_n \mathbf{u}_j \\
&= \frac{1}{D^2} \sum_{i=1}^{s_{n,d}} \frac{1}{\sigma_i} [\mathbf{u}_i^T \mathbf{v}_n(\mathbf{y})][\mathbf{u}_j^T \mathbf{v}_n(\mathbf{y})] \mathbf{u}_i^T \mathbf{u}_j = \frac{1}{D^2} \sum_{i=1}^{s_{n,d}} \frac{1}{\sigma_i} [\mathbf{u}_i^T \mathbf{v}_n(\mathbf{y})]^2 = \frac{1}{D} \\
&= Q_n^{-1}(\mathbf{y})
\end{aligned}$$

where we use the facts that $\mathscr{E}_\mu \left\{ \mathbf{v}_n(\mathbf{x}) \mathbf{v}_n^T(\mathbf{x}) \right\} = \mathbf{L}_n$, $\mathbf{L}_n \mathbf{u}_j = \sigma_j \mathbf{u}_j$ and $\mathbf{u}_i^T \mathbf{u}_j = 0$ if $i \neq j$. Direct application of Markov's inequality yields

$$\text{prob}\left\{ (p^*_{\mathbf{y}}(\mathbf{x}))^2 \geq \frac{1}{t Q_n(\mathbf{y})} \right\} \leq t \tag{8}$$

Next, note that

$$\begin{aligned}
\mathscr{E}_\mu(Q_n) &= \mathscr{E}_\mu \left\{ \mathbf{v}_n^T(\mathbf{x}) \mathbf{L}_n^{-1} \mathbf{v}_n(\mathbf{x}) \right\} = \mathscr{E}_\mu \left\{ \text{Trace} \left(\mathbf{L}_n^{-1} \mathbf{v}_n(\mathbf{x}) \mathbf{v}_n^T(\mathbf{x}) \right\} \right. \\
&= \text{Trace} \left(\mathbf{L}_n^{-1} \mathscr{E}_\mu \left\{ \mathbf{v}_n(\mathbf{x}) \mathbf{v}_n^T(\mathbf{x}) \right\} \right) = \text{Trace} \left(\mathbf{L}_n^{-1} \mathbf{L}_n \right) = s_{n.d} = \binom{n+d-1}{d-1}
\end{aligned}$$

From Markov's inequality we have that

$$\text{prob}\left\{ Q_n(\mathbf{y}) \geq t . s_{n,d} \right\} \leq \frac{1}{t} \tag{9}$$

Thus, values of Q_n much higher than $s_{n,d}$ indicate points with a high probability of being outliers. This property, combined with (8) shows that if $\mathbf{y}$ is chosen to be an outlier to the distribution μ, then the polynomial $p^{*2}_{\mathbf{y}}(.)$ will approximate the complement of the support of μ, in the sense that its value will be large in places where μ is small and viceversa (Fig. 1). This follows from the observation that if $\mathbf{y}$ is an outlier to the distribution μ, then $Q_n(\mathbf{y})$ is large and, from (8), $(p^*_{\mathbf{y}}(\mathbf{x}))^2$ is small if $\mathbf{x}$ is an inlier. Intuitively, if $\mathbf{y}$ is an outlier, a solution to (6) will be a polynomial that is close to one in a neighborhood of $\mathbf{y}$, to satisfy

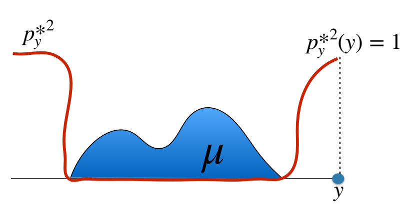

Fig. 1: (a) The square Christoffel polynomial $p^{*2}_{\mathbf{y}}$ for an outlier $\mathbf{y}$ is small at inlier points.

the constraint $p(\mathbf{y}) = 1$, and small in regions where μ is large, to minimize the overall integral (Fig. 1). Since the region around $\mathbf{y}$ has low density, it contributes little to the integral of p^2, while setting p^2 small in regions where μ is large minimizes their cost. Thus, as noted in [2], both $Q_n(.)$ and $p^*_{\mathbf{y}}(.)$ can be used to approximate the support of the distribution μ and to detect outliers.

References

1. Y. Xu. On orthogonal polynomials in several variables. *Special functions, q-series and related topics, The Fields Institute for Research in Mathematical Sciences, Communications Series*, 14:247–270, 1997. 2
2. Edouard Pauwels and Jean B Lasserre. Sorting out typicality with the inverse moment matrix sos polynomial. In *Advances in Neural Information Processing Systems*, pages 190–198, 2016. 2, 4
3. M. Sznaier and O. Camps. Sos-rsc: A sum-of-squares polynomial approach to robustifying subspace clustering algorithms. In *IEEE CVPR*, pages 8033–8041, 2018. 2