A Primary Decomposition in Computer Vision

Justin Katz & Irena Swanson
Reed College Mathematics Department

Contact Information:
Mathematics
Reed College
3203 SE Woodstock, Portland Oregon, USA

Phone: +1 (219) 210 9452
Email: jukatz@reed.edu

Introduction

A projective pinhole camera is modeled by rank 3 projective transformation $A : \mathbb{P}^3 \setminus \{f_A\} \to \mathbb{P}^2$ where f_A is the generator of the kernel of A. Given a family of m rank-3 projective transformations $A_i : \mathbb{P}^3 - \{f_{A_i}\} \to \mathbb{P}^2$, the m-view scenario is modeled by the rational map

$$\Phi_A : \mathbb{P}^3 - \{f_1, ..., f_m\} \to \underbrace{\mathbb{P}^2 \times \cdots \times \mathbb{P}^2}_{\text{m times}}.$$

To study the image of $\mathbb{P}^3 - \{f_1, .., f_m\}$ under Φ_A is to study the possible configurations of points on photographs that simultaneously arise as projections of the same point in physical space. This is a problem of fundamental interest in computer vision. The means by which we study $\Phi_A(\mathbb{P}^3 - \{f_1, .., f_m\})$ is through *algebraic geometry.* Despite $\Phi_A(\mathbb{P}^3 - \{f_1, .., f_m\})$ not being an projective variety (i.e. is not the zero set of a set of polynomials), its Zariski closure is determined by a family of polynomials that arise from independent interests.

On the other side, the aforementioned family of polynomials spltis into two distinct types: the bilinearities and the trilinearities. The ideal generated by the bilinearities can be seen as the totality of information that one can obtain about the m-view scenario by looking at pairs of cameras, and the ideal generated by the trilinearities gives the information generated by looking at triples of cameras. It has been known for some time that the bilinearities and trilinearities are related, in the sense that they can be computed from one another if one has access to their coefficients. Restricting only to ring operations, however, makes the situation more difficult. That said, knowing (as discussed above) that both the bilinearities and trilinearities are related to $\Phi_A(\mathbb{P}^3 - \{f_1, .., f_m\})$ allows for interesting connections to be made between the two.

In general, for m cameras there are $\binom{m}{2}$ bilinearities. For large values of m, computing all of them is too computationally expensive for practical applications. The codimension of the ideal generated by the bilinearities is $2m-3$, so one would hope that it would suffice to choose just $2m-3$ bilinearities to generate the whole bilinear ideal. Unfortunately this is not the case, as was shown in 1996 by Heyden and Astrom. That said, they conjectured that for a particular choice of $2m-3$ bilinearities, the resulting ideal will contain (effectively) the same information as the whole. That is to say that with this choice of bilinearities, the whole trilinear ideal is a primary component. This is what we proved.

The Objects

Let $X \in \mathbb{P}^3$ and $A_1, \ldots, A_m$ be 3×4 camera matrices such that $A_i X = x_i \in \mathbb{P}^2$ for each $i \in \{1, .., m\}$. Corespondingly there exist representing $\vec{X} \in \mathbb{R}^4 \setminus \{0\}$, $\vec{x}_i \in \mathbb{R}^3 \setminus \{0\}$ and $\lambda_i \in \mathbb{R} \setminus \{0\}$ such that

$$A_i \vec{X} = \lambda_i \vec{x}_i.$$

We can compile this data into the equation

$$Mu = 0,$$

where

$$M = \begin{bmatrix} A_1 & \vec{x}_1 & 0_{3\times 1} & \cdots & 0_{3\times 1} \\ A_2 & 0_{3\times 1} & \vec{x}_2 & \cdots & 0_{3\times 1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_m & 0_{3\times 1} & 0_{3\times 1} & \cdots & \vec{x}_m \end{bmatrix} \qquad u = \begin{bmatrix} \vec{X} \\ \lambda_1 \\ \vdots \\ \lambda_m \end{bmatrix}.$$

We call M the *master matrix of the* m*-view scenario.*

Now because u is nonzero, M has a nontrivial kernel, and the strict inequality

$$\text{rank}\, M < m+4$$

holds. Consequently, the determinant of any $(m+4) \times (m+4)$ submatrix of M will be zero. This suggests that if one thinks of the $\vec{x}_i$ as indeterminates in the polynomial ring $R := \mathbb{R}[x_1, ..., x_m, y_1, .., y_m, z_1, ..., z_m]$, then insisting that all $(m+4) \times (m+4)$ minors of M be zero will impose genuine polynomial constraints on the coordinates in the m-fold product of projective 2-spaces. There are three distinct types of choices of $(m+4) \times (m+4)$ minors, distinguished by how many rows are taken from how many cameras.

To construct the first type, pick two cameras A_i, A_j, and consider the $(m+4) \times (m+4)$ submatrix of M

$$D = \begin{bmatrix} A_1^{q_1} & \vec{x}_1^{q_1} & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ A_i & 0_{3\times 1} & \cdots & \vec{x}_i & \cdots & 0_{3\times 1} & \cdots & 0_{3\times 1} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ A_j & 0_{3\times 1} & \cdots & 0_{3\times 1} & \cdots & \vec{x}_j & \cdots & 0_{3\times 1} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ A_m^{q_m} & 0 & \cdots & 0 & \cdots & 0 & \cdots & \vec{x}_m^{q_m} \end{bmatrix}$$

where each $q_k \in \{1, 2, 3\}$ is picked arbitrarily.Taking the determinant of D, expanding along the columns containing single indeterminates, we discover

$$\det D = \pm \left(\prod_{\substack{k=1 \\ k \neq i,j}}^{m} \vec{x}_k^{q_k} \right) \det \begin{bmatrix} A_i & \vec{x}_i & 0 \\ A_j & 0 & \vec{x}_j \end{bmatrix} = 0.$$

Because we are in the projective setting and are viewing the indeterminates as modeling coordinates, there is some choice of each p_k such that $x_k^{p_k}$ will be nonzero. Thus, it is clear that the genuine constraint that arises in this way is the bilinear form

$$b_{ij}(\vec{x}_i, \vec{x}_j) := \det \begin{bmatrix} A_i & \vec{x}_i & 0 \\ A_j & 0 & \vec{x}_j \end{bmatrix} = 0.$$

We denote the ideal in R generated by all such bilinear forms as J_b, and call it the *unprojectivized bilinear ideal*. When we want to consider the unprojectivized bilinear ideal generated by some subset L of the cameras, we write $J_b(L)$.

The second type of $(m+4) \times (m+4)$ minor arises if one picks one distinguished camera A_i from which all three rows will be taken, two auxilliary cameras A_j and A_k from which two rows will be taken and one row from each of the other $m-3$ cameras. By expanding down each of the columns containing a single indeterminate, one finds the resulting *trilinear constraint* takes the form

$$T_{ij^{pq_j}k^{pq_k}}(\vec{x}_i, \vec{x}_j, \vec{x}_k) := \det \begin{bmatrix} A_i & \vec{x}_i & 0 & 0 \\ A_j^p & 0 & \vec{x}_j^p & 0 \\ A_j^{q_j} & 0 & \vec{x}_j^{q_j} & 0 \\ A_k^p & 0 & 0 & \vec{x}_k^p \\ A_k^{q_k} & 0 & 0 & \vec{x}_k^{q_k} \end{bmatrix} = 0,$$

where $p, q_j, q_k \in \{1, 2, 3\}$ are the rows that have been chosen from the two auxilliary cameras. We denote the ideal in R generated by all trilinear forms by J_t, and call it the *unprojectivized trilinear ideal*. As with the bilinear ideal, we denote the unprojectivized trilinear ideal generated by some subset of cameras L by $J(L)$.

Theorem and Essence of Proof

To motivate the central theorem, we begin with a dimension count. Let $m \geq 4$ and consider the m-view scenario. Because the bilinear $\mathcal{V}_b$ is the Zariski closure of the the image of a nondegenerate rational map from $\mathbb{P}^3$ into the m-fold product of $\mathbb{P}^2$, we know that its (projective) dimension is 3. One would hope that, because the codimension of a (projective) variety is the cardinality of the maximum number of algebraically independent functions in its coordinate ring, I_b can be generated by $2m-3 = \dim\left(\mathbb{P}^2\right)^m - \dim I_b$. Unfortunately this cannot always be the case, as Heyden and Astrom demonstrated in the case $m = 4, 5$, the variety generated by 5 and 7 bilinear forms (respectively) is always strictly larger than that generated by I_b.

This discrepancy leads to the study of a particular subset of the bilinearities that uniformly uses all cameras. We define

$$\tilde{J}_b = \sum_{i=1}^{m-1} (b_{i,i+1}) + \sum_{i=1}^{m-2} (b_{i,i+2})$$

and set

$$\tilde{I}_b = \tilde{J}_b : \left(\prod_{i=1}^{m} (x_i, y_i, z_i) \right)^{\infty}$$

for $i \in \{1, ..., m-1\}$. When we want to be explicit about which cameras we are using to generate $\tilde{I}_b$, we write $\tilde{I}_b(L)$ for a list of cameras L. Despite $\tilde{I}_b$ proveably not being the whole of I_b, Heyden and Astrom conjectured that it captures much of the complete m view scenario, in the sense of the following theorem.

Theorem 1. *I_t is a primary component of $\tilde{I}_b$ for all $m \geq 4$.*

Equivalently, we have

Corollary 1. *$\mathcal{V}_t = \overline{\mathcal{V}_n}$ is an irreducible subvariety of $\mathcal{V}(\tilde{I}_b)$ for all $m \geq 4$.*

To prove the theorem, we use the tool of colon ideals. In particular, if we can find an element w of R such that $\tilde{I}_b : w = I_t$ then I_t will be a primary component of $\tilde{I}_b : w = I_t$. This, in particular, is because I_t is (provably) prime. After much computer experimentation, the candidate w is

$$w = \prod_{i=1}^{m-2} \det \left[\, \vec{x}_i \;\; \vec{x}_{i+1} \;\; \vec{x}_{i+2} \,\right] \tag{1}$$

To prove that $\tilde{I}_b : w = I_t$ it suffices to show $w \cdot I_t \subset \tilde{I}_b$ because I_t is prime. This containment is proven by induction, with two base cases: $m = 3$ and $m = 4$. In the case of $m = 3$, were able to provide an explicit caluculation, listing a generic trilinearity times w as an R linear combination of bilinearities. Indeed

$$T_{12^{pq_2}3^{pq_3}} = w_1^{q_3} w_2^{q_2} b_{12} + w_2^{q_2} w_3^{q_3} b_{23} - w_1^{q_2} w_3^{q_3} b_{13}. \tag{2}$$

where w_j^i is the cofactor of w with the i^{th} row and j^{th} column removed. Further, in the case of $m = 4$ the computations became intractable to do by hand, but symmetry reduces the problem from checking 27 cases to just 4. With the aid of Macaulay2, a Groebner base program, we find

$$\begin{aligned} w_{123} \cdot w_{234} T_{12^{12}4^{12}} = {} & \begin{vmatrix} x_2 & x_2 \\ y_2 & y_4 \end{vmatrix} \cdot \begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix} \cdot w_{234} \cdot b_{12} + \begin{vmatrix} x_2 & x_4 \\ y_2 & y_4 \end{vmatrix} \cdot \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \cdot w_{134} \cdot b_{23} \\ & + \begin{vmatrix} x_2 & x_4 \\ y_2 & y_4 \end{vmatrix} \cdot \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \cdot w_{123} b_{34} - \begin{vmatrix} x_2 & x_4 \\ y_2 & y_4 \end{vmatrix} \cdot \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \cdot w_{234} \cdot b_{13} \\ & - \begin{vmatrix} x_3 & x_4 \\ y_3 & y_4 \end{vmatrix} \cdot \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \cdot w_{123} \cdot b_{14} \end{aligned}$$

$$\begin{aligned} w_{123} \cdot w_{234} T_{12^{12}4^{23}} = {} & \left(x_1 y_2 \begin{vmatrix} y_3 & y_4 \\ z_3 & z_4 \end{vmatrix} - x_3 y_2 \begin{vmatrix} y_1 & y_4 \\ z_1 & z_4 \end{vmatrix} + x_2 y_4 \begin{vmatrix} y_1 & y_3 \\ z_1 & z_3 \end{vmatrix} \right) w_{234} b_{12} \\ & + \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \cdot \begin{vmatrix} y_2 & y_4 \\ z_2 & z_4 \end{vmatrix} w_{134} b_{23} + \begin{vmatrix} x_1 & x_2 \\ z_1 & z_2 \end{vmatrix} \cdot \begin{vmatrix} y_2 & y_4 \\ z_2 & z_4 \end{vmatrix} w_{123} b_{34} \\ & - \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \cdot \begin{vmatrix} y_2 & y_4 \\ z_2 & z_4 \end{vmatrix} w_{234} b_{13} - \begin{vmatrix} y_3 & y_4 \\ z_3 & z_4 \end{vmatrix} \cdot \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} w_{123} b_{24} \end{aligned}$$

$$\begin{aligned} w_{123} \cdot w_{234} T_{21^{23}4^{13}} = {} & \begin{vmatrix} y_1 & y_3 \\ z_1 & z_3 \end{vmatrix} \cdot \begin{vmatrix} x_2 & x_4 \\ y_2 & y_4 \end{vmatrix} w_{234} b_{12} + \begin{vmatrix} y_1 & y_2 \\ z_1 & z_2 \end{vmatrix} \cdot \begin{vmatrix} x_2 & x_4 \\ z_2 & z_4 \end{vmatrix} w_{134} b_{23} \\ & + \begin{vmatrix} y_1 & y_2 \\ z_1 & z_2 \end{vmatrix} \cdot \begin{vmatrix} x_2 & x_4 \\ z_2 & z_4 \end{vmatrix} w_{123} b_{34} - \begin{vmatrix} y_1 & y_2 \\ z_1 & z_2 \end{vmatrix} \cdot \begin{vmatrix} x_2 & x_4 \\ z_2 & z_4 \end{vmatrix} w_{234} b_{13} \\ & - \begin{vmatrix} y_1 & y_2 \\ z_1 & z_2 \end{vmatrix} \cdot \begin{vmatrix} x_3 & x_4 \\ z_3 & z_4 \end{vmatrix} w_{123} b_{24} \end{aligned}$$

$$\begin{aligned} w_{123} \cdot w_{234} T_{41^{12}2^{23}} = {} & \begin{vmatrix} y_2 & y_4 \\ z_2 & z_4 \end{vmatrix} \cdot \begin{vmatrix} x_1 & x_3 \\ z_1 & z_3 \end{vmatrix} \cdot w_{234} \cdot b_{12} + \begin{vmatrix} y_2 & y_4 \\ z_2 & z_4 \end{vmatrix} \cdot \begin{vmatrix} x_1 & x_2 \\ z_1 & z_2 \end{vmatrix} \cdot w_{134} b_{23} \\ & + \begin{vmatrix} y_2 & y_4 \\ z_2 & z_4 \end{vmatrix} \cdot \begin{vmatrix} x_1 & x_2 \\ z_1 & z_2 \end{vmatrix} \cdot w_{123} b_{34} + \begin{vmatrix} y_2 & y_4 \\ z_2 & z_4 \end{vmatrix} \cdot \begin{vmatrix} x_1 & x_2 \\ z_1 & z_2 \end{vmatrix} \cdot w_{234} \cdot b_{13} \\ & + \left(-x_4 z_1 \begin{vmatrix} y_2 & y_3 \\ z_2 & z_3 \end{vmatrix} + x_3 z_1 \begin{vmatrix} y_2 & y_4 \\ z_2 & z_4 \end{vmatrix} - x_1 z_2 \begin{vmatrix} y_3 & y_4 \\ z_3 & z_4 \end{vmatrix} \right) w_{123} b_{14}. \end{aligned}$$

With the base cases taken care of, the induction is easy and uninformative.

Conclusions

The bilinear ideal cannot be generated by $2m-3$ bilinear forms, but a particular selection of $2m-3$ that uniformly use all cameras allows for a slightly rarefied ideal, that still contains the vital information of the ideal of the Zariski closure of the natural descriptor.

Acknowledgements

The first author would like to thank, first and foremost, the second author. Without her extreme generosity, with time and effort, this project could not have happened. This project was funded by the Reed College Science Research Fellowship, an award for which the first author is eternally grateful.