HAL
open science

Coordinate-Free Carlsson-Weinshall Duality and Relative Multi-View Geometry

Matthew Trager, Martial Hebert, Jean Ponce

To cite this version:

Matthew Trager, Martial Hebert, Jean Ponce. Coordinate-Free Carlsson-Weinshall Duality and Relative Multi-View Geometry. CVPR 2019 - IEEE Conference on Computer Vision and Pattern Recognition, Jun 2019, Long Beach, United States. hal-01676732v4

HAL Id: hal-01676732

https://inria.hal.science/hal-01676732v4

Submitted on 15 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Coordinate-Free Carlsson-Weinshall Duality and Relative Multi-View Geometry

Matthew Trager[1], Martial Hebert[2], and Jean Ponce[3,4]

[1]New York University [2]Carnegie Mellon University [3]INRIA, Paris, France
[4]Département dinformatique de lENS, ENS, CNRS, PSL University, Paris, France

Abstract

We present a coordinate-free description of Carlsson-Weinshall duality between scene points and camera pinholes and use it to derive a new characterization of primal/dual multi-view geometry. In the case of three views, a particular set of reduced trilinearities provide a novel parameterization of camera geometry that, unlike existing ones, is subject only to very simple internal constraints. These trilinearities lead to new "quasi-linear" algorithms for primal and dual structure from motion. We include some preliminary experiments with real and synthetic data.

1. Introduction

The idea of picking a few scene features as anchors to simplify the solution of structure-from-motion (SFM) problems dates back to the 1990s, notably with the pioneering work of Koenderink & van Doorn [14] and Faugeras [4], among others [11, 17]. This approach involves fewer parameters than traditional ones [4, 14] and leads to the so-called *Carlsson-Weinshall* (in this presentation, *CW*) *duality* [1], where camera pinholes and scene points play symmetric roles and can easily be swapped in SFM algorithms. However, methods based on this type of "relative" multi-view geometry are reputed to lead to poor-quality reconstructions, in part because the corresponding algorithms do not benefit from traditional data preconditioning methods [9]. We propose to revisit this approach from a geometric perspective, shedding new light on some well-known problems with a string of new results (Props. 2.3, 2.5, 3.4, 4.3), and dispelling through experiments some of its bad reputation.

1.1. Background

As shown in [19, 23] for example, point correspondences across multiple images can be characterized by studying incidence relations among the corresponding visual rays. This approach has the merit of making explicit the geometric constraints defining correspondences, which are often hidden behind algebra in the traditional multilinear approaches to structure from motion [1, 5, 6, 7, 12, 15, 16, 21, 25]. In particular, Ponce, Sturmfels and Trager introduced in [19] the *concurrent lines variety* V_n formed by all n-tuples of lines in $\mathbb{P}^3$ that meet at some point, and showed that constraining the lines in each tuple to pass through n fixed and distinct points yields a three-dimensional sub-variety of V_n isomorphic to Triggs's joint image [24], that can either be seen as the set of all possible images taken by n fixed perspective cameras (Fig. 1 [a]), or as the set of all possible images of n fixed points (Fig. 1 [b]), revealing a profound *geometric duality* between camera pinholes and scene points.

Unfortunately, this duality collapses when one introduces image measurements, since the retinal plane of a camera (or, equivalently, the line bundle of its pinhole) must be equipped with a coordinate system for the measurements to make sense. Contrary to images and the corresponding bundles (Fig. 1 [c]), however, scene points are not associated with coordinate systems. It was shown by Carlsson and Weinshall that this disparity can be addressed by using four fiducial scene points observed by all cameras, and by algebraically manipulating the coordinates of pinholes and scene points before inverting their roles (see [4, 12, 15] for related work). In particular, as argued in [1, 12], this implies that any algorithm for solving the structure-from-motion (SFM) problem from m images of n scene points also provides a (dual) solution to the SFM problem from $n-4$ images and $m+4$ scene points. Carlsson and Weinshall's take on duality is however mainly *analytical*. Our point of departure in this presentation is to bridge the gap between their approach and the geometric viewpoint advocated earlier.

1.2. Objectives and contributions

Our aim in this presentation is threefold:

(1) To explain CW duality [1] which, in its classical textbook form [12], emerges from seemingly accidental algebraic symmetries like Venus from the sea. Concretely, we introduce in Sect. 2 a new, coordinate-free derivation of the duality between scene points and camera pinholes (Prop. 2.3). Our viewpoint hopefully clarifies the geometry that underlies CW duality, and also emphasizes that analyt-

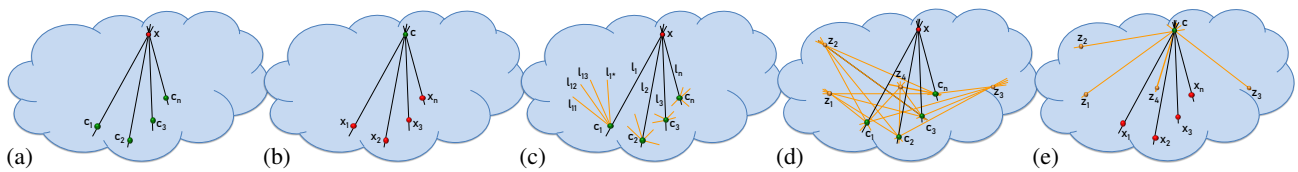

Figure 1. The sub-variety of the concurrent lines variety formed by all concurrent n-tuples of lines passing through n fixed points represents **(a)** the set of all perspective images of these points, as well as **(b)** the set of all images taken by the corresponding pinholes. The introduction of image (or equivalently, bundle) coordinate systems **(c)** breaks this duality, but it can be restored by **(d)-(e)** using four fiducial points observed by all cameras to define the corresponding image coordinate systems.

ical formulations of duality can be given *any* scene and image coordinate systems (Prop. 2.5 and Fig. 1[d,e]) [1, 4, 12].
(2) To characterize reduced multi-view geometry. We present in Sect. 3 a description of multi-view geometry in terms of the *reduced joint image* and its dual (Prop. 3.4). We also introduce a new parametrization of trinocular geometry in terms of both primal and dual *reduced trilinearities*. An interesting feature of these conditions is that, unlike trifocal tensors [10, 21, 25], they are subject to very *simple* internal constraints [6, 7, 12] (Prop. 4.3).
(3) To add to the three-view SFM arsenal. Our reduced trilinearities lead to new algorithms for structure from motion from primal and dual trilinearities, with competitive performance in experiments with real and synthetic data (Sect. 5).

1.3. Notation and elements of line geometry

Much of our presentation will distinguish purely geometric, coordinate-free properties of point configurations from analytical properties established in some coordinate system. To avoid confusion, we will use a teletype font to designate points in $\mathbb{P}^n$, e.g., $\mathtt{x}$, $\mathtt{y}$, and a bold italic font to designate their homogeneous coordinates in some coordinate frame, e.g., $\boldsymbol{x}$, $\boldsymbol{y}$. Whether we speak of points or their homogeneous coordinates should thus be clear, and we will often call both representations *points* for simplicity. We will call the first $n+1$ points of any projective basis $(\mathtt{x}_1, \ldots, \mathtt{x}_{n+1}, \mathtt{x}_{n+2})$, with coordinates $(1, 0, \ldots, 0)^T$ to $(0, \ldots, 0, 1)^T$, the *coordinate points*. The last one, $\mathtt{x}_{n+2}$, with coordinates $(1, \ldots, 1)^T$, is called the *unit point*. Let us also recall here some basic concepts of line geometry. The *join operator* associates with two distinct points $\mathtt{x}$ and $\mathtt{y}$ the unique line $\mathtt{x} \vee \mathtt{y}$ passing through them. Given some coordinate system for $\mathbb{P}^3$, this geometric operator admits an analytical counterpart, and the line $\mathtt{l} = \mathtt{x} \vee \mathtt{y}$ joining two points with coordinates $\boldsymbol{x} = (x_1, \ldots, x_4)^T$ and $\boldsymbol{y} = (y_1, \ldots, y_4)^T$ has homogeneous *Plücker coordinates*

$$\boldsymbol{l} = \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} \text{ with } \boldsymbol{u} = \begin{bmatrix} x_4y_1 - x_1y_4 \\ x_4y_2 - x_2y_4 \\ x_4y_3 - x_3y_4 \end{bmatrix}, \boldsymbol{v} = \begin{bmatrix} x_2y_3 - x_3y_2 \\ x_3y_1 - x_1y_3 \\ x_1y_2 - x_2y_1 \end{bmatrix}. \quad (1)$$

The vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in (1) are orthogonal by construction, and Plücker coordinates identify the four-dimensional set of lines in $\mathbb{P}^3$ with a quadratic hypersurface of $\mathbb{P}^5$, known as the *Klein quadric*. Two lines with Plücker coordinates $\boldsymbol{l} = (\boldsymbol{u}; \boldsymbol{v})$ and $\boldsymbol{l}' = (\boldsymbol{u}'; \boldsymbol{v}')$ intersect (or, equivalently, are coplanar) if and only if $\boldsymbol{u} \cdot \boldsymbol{v}' + \boldsymbol{u}' \cdot \boldsymbol{v} = 0$. The *line bundle* associated with a point $\mathtt{x}$ in $\mathbb{P}^3$ is the set of lines passing through that point. It corresponds to a two-dimensional projective subspace of the Klein quadric, (projectively) isomorphic to any plane π not passing through $\mathtt{x}$, each line in the bundle being associated with the point where it intersects π. Finally, the following result from [18] will be used repeatedly in the sequel.

Proposition 1.1 ([18])**.** *A necessary (and generically sufficient) condition for three lines with Plücker coordinates $\boldsymbol{l}$, $\boldsymbol{l}'$, $\boldsymbol{l}''$ to intersect is that the four minors*

$$T_1 = \begin{vmatrix} l_2 l_2' l_2'' \\ l_3 l_3' l_3'' \\ l_4 l_4' l_4'' \end{vmatrix}, T_2 = \begin{vmatrix} l_3 l_3' l_3'' \\ l_1 l_1' l_1'' \\ l_5 l_5' l_5'' \end{vmatrix}, T_3 = \begin{vmatrix} l_1 l_1' l_1'' \\ l_2 l_2' l_2'' \\ l_6 l_6' l_6'' \end{vmatrix}, T_4 = \begin{vmatrix} l_4 l_4' l_4'' \\ l_5 l_5' l_5'' \\ l_6 l_6' l_6'' \end{vmatrix} \quad (2)$$

of the 6×3 matrix $[\boldsymbol{l}, \boldsymbol{l}', \boldsymbol{l}'']$ all vanish. In addition, the vanishing of a single minor T_k ($k = 1, 2, 3, 4$) is a necessary and sufficient condition for these lines to admit a common transversal through the kth coordinate point.

2. Point configurations and CW duality

2.1. Geometric point of view

Moving together scene points and cameras without changing their relative positions will not change the images of the scene recorded by the cameras. This is sometimes referred to as the projective ambiguity of structure from motion, but we propose here instead to capture the underlying "projective rigidity" in terms of projective *configurations*.

Definition 2.1. *Two k-tuples of points in $\mathbb{P}^n$ are* isomorphic *if they are related by a projective transformation of $\mathbb{P}^n$. Isomorphism is an equivalence relation, and its equivalence classes are called k-*configurations.

The configuration associated with k points $\mathtt{x}_1$ to $\mathtt{x}_k$ is denoted by $\langle \mathtt{x}_1, \ldots, \mathtt{x}_k \rangle$. For $k \leq n+2$, generic point configurations are always isomorphic, so we will assume $k > n+2$ from now on. Given some pinhole $\mathtt{c}$ in $\mathbb{P}^3$ and some retinal plane π not passing through $\mathtt{c}$, the corresponding perspective projection can be defined in a purely geometric manner

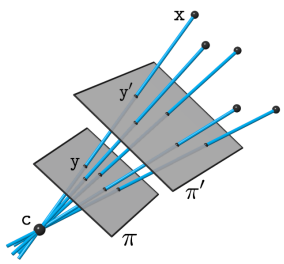

Figure 2. Image point and viewing ray configurations are isomorphic and independent of the retinal plane.

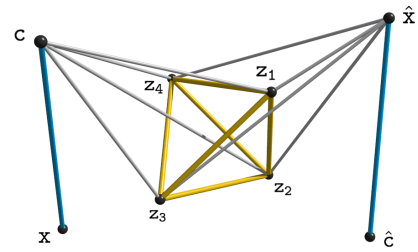

Figure 3. Geometric Carlsson-Weinshall duality between scene point and pinhole configurations.

as the mapping that associates with any point $\mathtt{x} \neq \mathtt{c}$ in $\mathbb{P}^3$ the point $\mathtt{y}$ where the viewing ray joining $\mathtt{c}$ to $\mathtt{x}$ intersects π. In turn, this mapping induces an isomorphism between k-configurations of points in different image planes π and π', and a second isomorphism between these and the corresponding k-configurations of visual rays through $\mathtt{c}$, seen as elements of a line bundle (Fig. 2). This is of course just a retelling of a familiar story in the language of configurations. But it also shows that perspective projection may be viewed as a mapping between the *scene* configurations $\langle \mathtt{x}_1, \ldots, \mathtt{x}_k, \mathtt{c} \rangle$ that determine the visual rays $\mathtt{c} \vee \mathtt{x}_i$ and their *image* counterparts $\langle \mathtt{y}_1, \ldots, \mathtt{y}_k \rangle$. We will sometimes write a scene configuration as $\langle \mathtt{x}_1, \ldots, \mathtt{x}_k \,|\, \mathtt{c} \rangle$ instead of $\langle \mathtt{x}_1, \ldots, \mathtt{x}_k, \mathtt{c} \rangle$ to emphasize that the last point is viewed as a pinhole. In this setting, swapping the roles of pinhole and scene point results in permuting the corresponding elements of a scene configuration. The effect of permutations on point configurations can be described in terms of so-called *Cremona transformations* of $\mathbb{P}^3$, as explained by Coble in a paper from 1915 [2] (see also [3] for a more recent account on this topic). As we will argue in the next section, the analytical map $(x_1, x_2, x_3, x_4)^T \mapsto (x_1^{-1}, x_2^{-1}, x_3^{-1}, x_4^{-1})^T$ that is used in the standard formulation of CW duality is indeed an example of a Cremona transformation. For the moment, we can state the following more general geometric result that follows from Coble's theory (see [2, Sect. 7] or [3, Chap. 6]).

Lemma 2.2. *If* $\mathtt{Z} = (\mathtt{z}_1, \ldots, \mathtt{z}_4)$ *is a quadruple of fixed points of* $\mathbb{P}^3$ *in general position, then there exists a family of birational involutions* $T_{\mathtt{Z}} : \mathtt{x} \mapsto \hat{\mathtt{x}}$ *(*Cremona involutions*), defined on a dense open set of* $\mathbb{P}^3$*, such that for any points* $\mathtt{x}$ *and* $\mathtt{y}$ *in that set* $\langle \mathtt{z}_1, \mathtt{z}_2, \mathtt{z}_3, \mathtt{z}_4, \mathtt{x}, \mathtt{y} \rangle = \langle \mathtt{z}_1, \mathtt{z}_2, \mathtt{z}_3, \mathtt{z}_4, \hat{\mathtt{y}}, \hat{\mathtt{x}} \rangle$ *holds (with equality as configurations). Any two such involutions are related by a projective transformation of* $\mathbb{P}^3$ *that fixes* $\mathtt{Z}$.

Note that this statement does not involve pinholes and scene points. However, we obtain as an immediate corollary a geometric and coordinate-free formulation of Carlsson-Weinshall duality, valid for any Cremona involution associated with a quadruple of points $\mathtt{Z}$.

Proposition 2.3 (Figure 3)**.** *If* $\mathtt{x} \mapsto \hat{\mathtt{x}}$ *is a Cremona involution relative to a quadruple* $\mathtt{Z} = (\mathtt{z}_1, \ldots, \mathtt{z}_4)$*, then any two scene configurations* $\langle \mathtt{z}_1, \mathtt{z}_2, \mathtt{z}_3, \mathtt{z}_4, \mathtt{x} \,|\, \mathtt{c} \rangle$ *and* $\langle \mathtt{z}_1, \mathtt{z}_2, \mathtt{z}_3, \mathtt{z}_4, \hat{\mathtt{c}} \,|\, \hat{\mathtt{x}} \rangle$ *are equal, and thus give rise to the same image configuration* $\langle \mathtt{y}_1, \ldots, \mathtt{y}_4, \mathtt{y} \rangle$*. Here* $\mathtt{y}$ *can be thought of as either the projection of* $\mathtt{x}$ *from* $\mathtt{c}$ *or that of* $\hat{\mathtt{c}}$ *from* $\hat{\mathtt{x}}$.

2.2. Analytical point of view

Let us now introduce a local parameterization of the space of k-configurations in $\mathbb{P}^n$, with $k > n + 2$: we pick $n + 2$ of the points and assign them arbitrary but fixed homogeneous coordinates (it is often convenient, *but by no means necessary*, to choose these points as a basis for $\mathbb{P}^n$). Assuming that the points are in general position and the coordinates assigned to any $n + 1$ of them are linearly independent, this uniquely defines a coordinate system for $\mathbb{P}^n$, dependent on the choice of the $n + 2$ points, but *intrinsic* to the whole configuration. In particular, the coordinates of the $k - n - 2$ remaining points can be used to parameterize the configuration. In our setting, this translates into assigning arbitrary coordinates to the four fixed points $\mathtt{z}_1$ to $\mathtt{z}_4$ in the form of a 4×4 matrix $\boldsymbol{Z} = [\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3, \boldsymbol{z}_4]$, and assigning to the pinhole arbitrary coordinates $\boldsymbol{c}$. This freezes the coordinate system of $\mathbb{P}^3$ and provides a parameterization of the configurations $\langle \mathtt{z}_1, \ldots, \mathtt{z}_4, \mathtt{x} \,|\, \mathtt{c} \rangle$ using the coordinates $\boldsymbol{x}$ of the point $\mathtt{x}$. We also pick the four visual rays $\mathtt{l}_1$ to $\mathtt{l}_4$ joining the pinhole to the points $\mathtt{z}_1$ to $\mathtt{z}_4$ as reference points for the corresponding bundle, and assign them arbitrary coordinates in the form of a 3×4 matrix $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \boldsymbol{u}_4]$. This freezes the coordinate frame for the bundle and provides a parameterization for the configurations $\langle \mathtt{l}_1, \ldots, \mathtt{l}_4, \mathtt{l} \rangle$ of its lines by the coordinates $\boldsymbol{u}$ of the ray $\mathtt{l}$. This also provides, of course, a parameterization of the configurations $\langle \mathtt{y}_1, \ldots, \mathtt{y}_4, \mathtt{y} \rangle$ of the corresponding image points using the coordinates $\boldsymbol{u}$ of the point $\mathtt{y}$. The following result follows from some straightforward computations (see the supplementary material for details).

Proposition 2.4. *Given arbitrary general matrices* $\boldsymbol{U}$ *and* $\boldsymbol{Z}$*, the perspective projection associated with pinhole* $\mathbf{c}$ *can be represented analytically as the projective map* $P_{\mathbf{c}}$ *from*

$\mathbb{P}^3$ *to* $\mathbb{P}^2$ *defined by*[1]

$$P_{\boldsymbol{c}}(\boldsymbol{x}) = [\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \boldsymbol{u}_4] \begin{bmatrix} \frac{|\boldsymbol{x}, \boldsymbol{z}_2, \boldsymbol{z}_3, \boldsymbol{z}_4|}{|\boldsymbol{c}, \boldsymbol{z}_2, \boldsymbol{z}_3, \boldsymbol{z}_4|} \\ \frac{|\boldsymbol{z}_1, \boldsymbol{x}, \boldsymbol{z}_3, \boldsymbol{z}_4|}{|\boldsymbol{z}_1, \boldsymbol{c}, \boldsymbol{z}_3, \boldsymbol{z}_4|} \\ \frac{|\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{x}, \boldsymbol{z}_4|}{|\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{c}, \boldsymbol{z}_4|} \\ \frac{|\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3, \boldsymbol{x}|}{|\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3, \boldsymbol{c}|} \end{bmatrix}, \tag{3}$$

where we assume wlog that the coordinates vectors $\boldsymbol{u}_i$ *have been scaled so* $\boldsymbol{u}_1 + \boldsymbol{u}_2 + \boldsymbol{u}_3 + \boldsymbol{u}_4 = 0$.

For $\boldsymbol{Z} = \mathrm{Id}_4$ and $\boldsymbol{U} = [\mathrm{Id}_3, -\mathbf{1}_3]$, we have that $P_{\boldsymbol{c}}(\boldsymbol{x}) = (x_1/c_1, x_2/c_2, x_3/c_3, x_4/c_4)^T$, and the projection matrix associated with $P_{\boldsymbol{c}}$ is the *reduced camera model* appearing in different guises in [1, 4, 12]:

$$P_{\boldsymbol{c}} = \begin{bmatrix} 1/c_1 & 0 & 0 & -1/c_4 \\ 0 & 1/c_2 & 0 & -1/c_4 \\ 0 & 0 & 1/c_3 & -1/c_4 \end{bmatrix}. \tag{4}$$

In this case, the expression for $P_{\boldsymbol{c}}(\boldsymbol{x})$ is symmetric in $\boldsymbol{x}$ and $\hat{\boldsymbol{c}}$, where $\boldsymbol{y} \mapsto \hat{\boldsymbol{y}} = (y_1^{-1}, y_2^{-1}, y_3^{-1}, y_4^{-1})^T$ is the *standard Cremona involution*. This is a Cremona transformation in the sense of Lemma 2.2 since one easily sees that two sextuples $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_4, \boldsymbol{y}_1, \boldsymbol{y}_2$ and $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_4, \hat{\boldsymbol{y}}_2, \hat{\boldsymbol{y}}_1$ are always related by a projective transformation of $\mathbb{P}^3$. More generally, we have the following result, which can be shown by direct computations (see the supplementary material).

Proposition 2.5 (Analytical Carlsson-Weinshall duality)**.** *The rational map of* $\mathbb{P}^3$ *given by*

$$\hat{\boldsymbol{y}} = \boldsymbol{Z} \left[\frac{1}{|\boldsymbol{y}\boldsymbol{z}_2\boldsymbol{z}_3\boldsymbol{z}_4|}, \frac{1}{|\boldsymbol{z}_1\boldsymbol{y}\boldsymbol{z}_3\boldsymbol{z}_4|}, \frac{1}{|\boldsymbol{z}_1\boldsymbol{z}_2\boldsymbol{y}\boldsymbol{z}_4|}, \frac{1}{|\boldsymbol{z}_1\boldsymbol{z}_2\boldsymbol{z}_3\boldsymbol{y}|} \right]^T \tag{5}$$

is a Cremona involution relative to the points in $\mathbb{P}^3$ *with coordinates* $\boldsymbol{Z}$*. Using this* $\boldsymbol{Z}$ *and arbitrary image coordinates* $\boldsymbol{U}$ *in* (3)*, we have that* $P_{\boldsymbol{c}}(\boldsymbol{x}) = P_{\hat{\boldsymbol{x}}}(\hat{\boldsymbol{c}})$.

3. Reduced multi-view geometry

We now restrict our study to the case where $\boldsymbol{Z} = \mathrm{Id}_4$ and $\boldsymbol{U} = [\mathrm{Id}_3, -\mathbf{1}_3]$ so that all cameras can be represented by projection matrices in the "standard" reduced form of Eq. (4), and thus identify from now on scene points $\mathtt{x}$ and their images $\mathtt{y}$ with their coordinates $\boldsymbol{x}$ and $\boldsymbol{u}$ in $\mathbb{P}^3$ and $\mathbb{P}^2$.

3.1. Reduced joint images

Let S denote the set of triples $(\boldsymbol{c}, \boldsymbol{x}, \boldsymbol{u})$ in $\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^2$ such that $P_{\boldsymbol{c}}(\boldsymbol{x}) = \boldsymbol{u}$ where $P_{\boldsymbol{c}}$ is a reduced camera as in (4). For fixed $\boldsymbol{c}$ and $\boldsymbol{x}$, there is of course a single $\boldsymbol{u}$ such that $(\boldsymbol{c}, \boldsymbol{x}, \boldsymbol{u})$ belongs to S. More generally, we have the following result which describes the set S (see the supplementary material for details).

[1] Although $P_{\boldsymbol{c}}$ obviously depends on $\boldsymbol{U}$ and $\boldsymbol{Z}$, we leave this dependency implicit in the notation to avoid clutter.

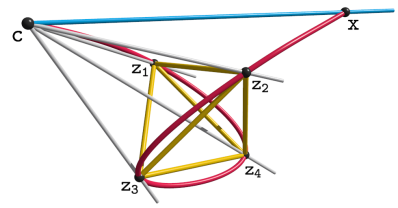

Figure 4. The twisted cubic formed by the points $\mathtt{c}$ such that $\mathtt{c} \vee \mathtt{x}$ has constant coordinates relative to $\mathtt{c} \vee \mathtt{z}_1, \mathtt{c} \vee \mathtt{z}_2, \mathtt{c} \vee \mathtt{z}_3, \mathtt{c} \vee \mathtt{z}_4$.

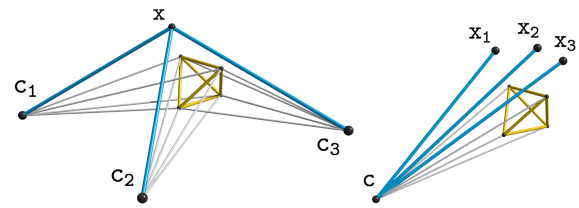

Figure 5. **Left:** A reduced joint image characterizes all converging visual rays from three pinholes. **Right:** A dual reduced joint image characterizes all perspective images of three scene points.

Proposition 3.1. **(1)** *For fixed* $\boldsymbol{c}$ *and* $\boldsymbol{u}$*, the set of points* $\boldsymbol{x}$ *such that* $(\boldsymbol{c}, \boldsymbol{x}, \boldsymbol{u})$ *belongs to* S *is a line with Plücker coordinates*

$$\boldsymbol{\xi} = Q_{\boldsymbol{c}}\boldsymbol{u} \text{ where } Q_{\boldsymbol{c}} = \begin{bmatrix} c_1c_4 & 0 & 0 \\ 0 & c_2c_4 & 0 \\ 0 & 0 & c_3c_4 \\ 0 & -c_2c_3 & c_2c_3 \\ c_1c_3 & 0 & -c_1c_3 \\ -c_1c_2 & c_1c_2 & 0 \end{bmatrix}. \tag{6}$$

(2) *For fixed* $\boldsymbol{x}$ *and* $\boldsymbol{u}$*, the set of points* $\boldsymbol{c}$ *such that* $(\boldsymbol{c}, \boldsymbol{x}, \boldsymbol{u})$ *belongs to* S *is a twisted cubic passing through* $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_4$ *and* $\boldsymbol{x}$ *(Fig. 4).*

Variants of this proposition can be found in [1, 4, 12]. The formula for $Q_{\boldsymbol{c}}$ in Eq. (19) is of course an instance of the classical (transposed) line projection matrix. It will play a key role in the rest of this presentation. We now consider n pinholes $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n$, and the associated reduced cameras $P_{\boldsymbol{c}_1}, \ldots, P_{\boldsymbol{c}_n}$. Following [22, 24], we describe the geometry of these cameras using the *joint image* in $(\mathbb{P}^2)^n$.

Definition 3.2. *The reduced joint image* $V_Z(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n)$ *associated with* n *fixed pinholes* $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n$ *is the set of* n*-tuples* $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)$ *in* $(\mathbb{P}^2)^n$ *such that the corresponding visual rays are concurrent or, equivalently, such that there exists some point* $\boldsymbol{x}$ *such that* $(\boldsymbol{c}_i, \boldsymbol{x}, \boldsymbol{u}_i)$ *belongs to* S.

The reduced joint image is a special case of Triggs's joint image [24] where image (or bundle) basis points are in correspondence. Projective transformations of $\mathbb{P}^3$ do not affect the joint image, so $V_Z(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n)$ is completely determined by $\langle \mathtt{c}_1, \ldots, \mathtt{c}_n, \mathtt{z}_1, \ldots \mathtt{z}_4 \rangle$ (Fig. 5, left). Let us now consider instead n fixed scene points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$.

Definition 3.3. *The* dual *reduced joint image* $\hat{V}_Z(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ *associated with* n *scene points* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$

is the set of n-tuples $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)$ in $(\mathbb{P}^2)^n$ that are image coordinates for the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ for some reduced camera $P_{\boldsymbol{c}}$ with (unknown) pinhole $\boldsymbol{c}$ or, equivalently, such that there exists some point $\boldsymbol{c}$ such that $(\boldsymbol{c}, \boldsymbol{x}_i, \boldsymbol{u}_i)$ belongs to S. Note that this condition imposes that n twisted cubics passing through $\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3, \boldsymbol{z}_4$ intersect at a point $\boldsymbol{c}$.

The set $\hat{V}_Z(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ characterizes all perspective images of n fixed points (Fig. 5, right). It is invariant to projective transformations of $\mathbb{P}^3$, and completely determined by $\langle \mathtt{x}_1, \ldots, \mathtt{x}_n, \mathtt{z}_1, \ldots \mathtt{z}_4 \rangle$ in $\mathbb{P}^3$. Algebraic characterizations of dual multi-view constraints are mostly absent from the literature (see [15] for an exception), but the following result is an immediate corollary of CW duality.

Proposition 3.4. *The dual reduced joint image associated with n scene points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ is the reduced joint image associated with their images $\hat{\boldsymbol{x}}_1, \ldots, \hat{\boldsymbol{x}}_n$ under any Cremona involution relative to $\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3, \boldsymbol{z}_4$.*

In particular, like their primal counterparts [22], dual joint images induce multilinear constraints on point correspondences.

3.2. Reduced multilinearities

Let us now apply the general approach presented so far to the bilinear and trilinear constraints associated with point correspondences for reduced cameras of the form (4). Given two image points $\boldsymbol{u}$ and $\boldsymbol{u}'$ associated with cameras with pinholes $\boldsymbol{c}$ and $\boldsymbol{c}'$, a necessary and sufficient condition for $\boldsymbol{u}$ and $\boldsymbol{u}'$ to form a correspondence is that the visual rays $\boldsymbol{l} = Q_{\boldsymbol{c}}\boldsymbol{u}$ and $\boldsymbol{l}' = Q_{\boldsymbol{c}'}\boldsymbol{u}'$ intersect each other or, equivalently, $(\boldsymbol{l} \,|\, \boldsymbol{l}') = 0$. This immediately yields a bilinear relation $\boldsymbol{u}^T F \boldsymbol{u}' = 0$. When $\boldsymbol{c} = \boldsymbol{1}_4$, F is the *reduced fundamental matrix* of [1, 12] (see also [4]). In turn, substituting $\hat{\boldsymbol{x}}', \hat{\boldsymbol{x}}''$ for $\boldsymbol{c}', \boldsymbol{c}''$ in this equation, we also obtain the expression for the reduced *dual fundamental matrix*, which characterizes the dual reduced joint image $\hat{V}_Z(\boldsymbol{x}, \boldsymbol{x}')$ for two fixed scene points $\boldsymbol{x}, \boldsymbol{x}'$. We can use the same approach to characterize correspondences in three images. Indeed, substituting $Q_{\boldsymbol{c}}\boldsymbol{u}$, $Q_{\boldsymbol{c}}\boldsymbol{u}'$, and $Q_{\boldsymbol{c}}\boldsymbol{u}''$ to $\boldsymbol{l}$, $\boldsymbol{l}'$ and $\boldsymbol{l}''$ in Eq. (2) from Prop. 1.1 immediately yields the following result.

Proposition 3.5. *Taking $\boldsymbol{c} = (1,1,1,1)^T$, a necessary (and generically sufficient) condition for $\boldsymbol{u}$, $\boldsymbol{u}'$ and $\boldsymbol{u}''$ to form a correspondence for the reduced cameras $P_{\boldsymbol{c}}, P_{\boldsymbol{c}'}, P_{\boldsymbol{c}''}$ is that the four determinants T_1 to T_4, respectively given by*

$$\begin{vmatrix} u_2, \hat{c}_3' u_2', \hat{c}_3'' u_2'' \\ u_3, \hat{c}_2' u_3', \hat{c}_2'' u_3'' \\ v_1, \hat{c}_4' v_1', \hat{c}_4'' v_1'' \end{vmatrix}, \begin{vmatrix} u_3, \hat{c}_1' u_3', \hat{c}_1'' u_3'' \\ u_1, \hat{c}_3' u_1', \hat{c}_3'' u_1'' \\ v_2, \hat{c}_4' v_2', \hat{c}_4'' v_2'' \end{vmatrix}, \begin{vmatrix} u_1, \hat{c}_2' u_1', \hat{c}_2'' u_1'' \\ u_2, \hat{c}_1' u_2', \hat{c}_1'' u_2'' \\ v_3, \hat{c}_4' v_3', \hat{c}_4'' v_3'' \end{vmatrix}, \begin{vmatrix} v_1, \hat{c}_1' v_1', \hat{c}_1'' v_1'' \\ v_2, \hat{c}_2' v_2', \hat{c}_2'' v_2'' \\ v_3, \hat{c}_3' v_3', \hat{c}_3'' v_3'' \end{vmatrix} \tag{7}$$

all vanish, with $v_i = u_{i+2} - u_{i+1}$, $v_i' = u_{i+2}' - u_{i+1}'$ and $v_i'' = u_{i+2}'' - u_{i+1}''$, and index addition modulo 3.

The proposition follows immediately from (2) and the form of the matrix $Q_{\boldsymbol{c}}$. Its dual involves three scene points $\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{x}''$ instead of three pinholes, and is obtained by substituting $\hat{\boldsymbol{x}}$ for $\boldsymbol{c}$ in the primal trilinearities from Prop. 3.5.

Proposition 3.6. *Taking $\boldsymbol{x} = (1,1,1,1)^T$, a necessary (and generically sufficient) condition for $\boldsymbol{u}$, $\boldsymbol{u}'$ and $\boldsymbol{u}''$ to be projections of $\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{x}''$ for a reduced camera $P_{\boldsymbol{c}}$ (for some unknown $\boldsymbol{c}$) is that the four determinants $\hat{T}_1$ to $\hat{T}_4$, respectively given by*

$$\begin{vmatrix} u_2, x_3' u_2', x_3'' u_2'' \\ u_3, x_2' u_3', x_2'' u_3'' \\ v_1, x_4' v_1', x_4'' v_1'' \end{vmatrix}, \begin{vmatrix} u_3, x_1' u_3', x_1'' u_3'' \\ u_1, x_3' u_1', x_3'' u_1'' \\ v_2, x_4' v_2', x_4'' v_2'' \end{vmatrix}, \begin{vmatrix} u_1, x_2' u_1', x_2'' u_1'' \\ u_2, x_1' u_2', x_1'' u_2'' \\ v_3, x_4' v_3', x_4'' v_3'' \end{vmatrix}, \begin{vmatrix} v_1, x_1' v_1', x_1'' v_1'' \\ v_2, x_2' v_2', x_2'' v_2'' \\ v_3, x_3' v_3', x_3'' v_3'' \end{vmatrix} \tag{8}$$

all vanish.

4. Algebraic constraints on trilinearities

In this section, we investigate the special primal and dual trilinear conditions (7) and (8). In particular, we show that the coefficients of these polynomial forms are subject to very *simple* algebraic constraints. This contrasts with the classical trifocal tensor and the induced trilinearities [10, 21, 25], which also characterize correspondences among three views, but are known to satisfy very complex internal constraints [6, 7, 12]. This feature suggests that the coefficients of our trilinearities can be estimated easily from image data and used in reconstruction algorithms. This will be confirmed by our experiments in Section 5.

Our first observation is that the conditions (7) and (8) can be seen as polynomials in the "mixed" coordinates u's and v's (as defined in Prop. 3.5), or also in the "pure" image coordinates u's (by replacing the v's with the corresponding expressions). Depending on this choice of variables, the trilinearities have different coefficients. These coefficients are related by a (non-invertible) linear transformation but present some differences, as shown next.

4.1. Trilinearities in mixed coordinates

The four reduced primal trilinearities in "mixed" coordinates u's and v's can be written explicitly as follows

$$\begin{aligned} T_1 &= -\rho_{23} v_1 u_3' u_2'' + \rho_{24} u_2 u_3' v_1'' + \rho_{32} v_1 u_2' u_3'' \\ &\quad -\rho_{34} u_3 u_2' v_1'' - \rho_{42} u_2 v_1' u_3'' + \rho_{43} u_3 v_1' u_2'', \\ T_2 &= \rho_{13} v_2 u_3' u_1'' - \rho_{14} u_1 u_3' v_2'' - \rho_{31} v_2 u_1' u_3'' \\ &\quad +\rho_{34} u_3 u_1' v_2'' + \rho_{41} u_1 v_2' u_3'' - \rho_{43} u_3 v_2' u_1'', \\ T_3 &= -\rho_{12} v_3 u_2' u_1'' + \rho_{14} u_1 u_2' v_3'' + \rho_{21} v_3 u_1' u_2'' \\ &\quad -\rho_{24} u_2 u_1' v_3'' - \rho_{41} u_1 v_3' u_2'' + \rho_{42} u_2 v_3' u_1'', \\ T_4 &= \rho_{12} v_3 v_1' v_2'' - \rho_{13} v_2 v_1' v_3'' - \rho_{21} v_3 v_2' v_1'' \\ &\quad +\rho_{23} v_1 v_2' v_3'' + \rho_{31} v_2 v_3' v_1'' - \rho_{32} v_1 v_3' v_2'', \end{aligned} \tag{9}$$

with $\rho_{ij} = \hat{c}_i' \hat{c}_j''$. The six non-zero coefficients of T_i are ρ_{jk} with j, k distinct in $\{1,2,3,4\} \setminus \{i\}$. Hence, sextuples of

valid coefficients for T_i are vectors in $\mathbb{R}^6$ that have no zero entries and can be written in the form $(a_i b_j)_{i \neq j}$ with $\boldsymbol{a}, \boldsymbol{b}$ in $\mathbb{R}^3$. These vectors are completely characterized by the following result.

Proposition 4.1. *A vector $\boldsymbol{d} = (d_{12}, d_{13}, d_{23}, d_{21}, d_{31}, d_{32})$ in $\mathbb{R}^6$ can be written as $d_{ij} = a_i b_j$ for some vectors $\boldsymbol{a} = (a_1, a_2, a_3)^T, \boldsymbol{b} = (b_1, b_2, b_3)^T$ in $\mathbb{R}^3$ if and only if $d_{12} d_{23} d_{31} = d_{21} d_{32} d_{13}$ holds.*

Proof. By replacing d_{ij} with $a_i b_j$, we see that the condition is necessary. Conversely, if an element (say) d_{12} of $\boldsymbol{d}$ is not zero and $\boldsymbol{d}$ satisfies the condition, we can always set $a_1 = 1$ and solve for all the remaining entries. □

The coefficients of each trilinearity are thus constrained by a single cubic relation. On the other hand, each pair of trilinearities shares two coefficients, so we expect additional constraints for the consistency of $T_1, \ldots, T_4$. This is relevant for the reconstruction method that will be described in Section 5, in which the trilinearities (7) evaluated on point correspondences are treated as linear conditions on a single vector $\boldsymbol{\rho} = (\rho_{ij})$ in $\mathbb{R}^{12}$, such that $\rho_{ij} = \hat{c}'_i \hat{c}''_j$ with distinct i, j in $\{1, 2, 3, 4\}$. In this case, the conditions for $\boldsymbol{\rho}$ in $\mathbb{R}^{12}$ to be a valid solution are given by the following result. The proof is analogous to that of Prop. 4.1. We refer to the supplementary material for details.

Proposition 4.2. *A vector $\boldsymbol{d} = (d_{ij})$ in $\mathbb{R}^{12}$ with no zero entries can be written as $d_{ij} = a_i b_j$ for some vectors $\boldsymbol{a} = (a_1, a_2, a_3, a_4)^T, \boldsymbol{b} = (b_1, b_2, b_3, b_4)^T$ in $\mathbb{R}^3$ if and only if $d_{ij} d_{kl} = d_{il} d_{kj}$ holds for all permutations (i, j, k, l) of $(1, 2, 3, 4)$.*

4.2. Trilinearities in image coordinates

A reduced trilinearity expressed in terms of mixed u's and v's has very simple coefficients, but its variables are not independent. We now describe the trilinearities in pure image coordinates. Remarkably, the internal constraints of each trilinearity are completely *linear* in this setting. To simplify our presentation, we focus on the trilinearity T_1. The trilinearities T_2, T_3 are identical to T_1 up to permutation of indices. The trilinearity T_4 has a different analytical form, but it encodes the same geometric information and also enjoys similar properties. A general treatment of all the trilinearities can be found in the supplementary material.

A simple computation shows that

$$\begin{aligned} T_1 = & (-\hat{c}'_3 \hat{c}''_2 + \hat{c}'_4 \hat{c}''_2) u_2 u'_2 u''_3 + (\hat{c}'_2 \hat{c}''_3 - \hat{c}'_2 \hat{c}''_4) u_2 u'_3 u''_2 + \\ & (-\hat{c}'_4 \hat{c}''_2 + \hat{c}'_2 \hat{c}''_4) u_2 u'_3 u''_3 + (-\hat{c}'_4 \hat{c}''_3 + \hat{c}'_3 \hat{c}''_4) u_3 u'_2 u''_2 + \\ & (\hat{c}'_3 \hat{c}''_2 - \hat{c}'_3 \hat{c}''_4) u_3 u'_2 u''_3 + (-\hat{c}'_2 \hat{c}''_3 + \hat{c}'_4 \hat{c}''_3) u_3 u'_3 u''_2 . \end{aligned} \tag{10}$$

The six non-zero coefficients are linear combinations of the scalars $\rho_{ij} = \hat{c}'_i \hat{c}''_j$. We now address the internal constraints that are satisfied by these coefficients.

Proposition 4.3. *If we allow $\hat{c}', \hat{c}''$ to take complex values, then a vector $\boldsymbol{\tau}$ in $\mathbb{R}^6$ represents a set of feasible coefficients for T_1 in* (10) *if and only if its elements sum to zero. In this case, up to scale factors, there are* two *pairs of (possibly complex) solutions for $(\hat{c}'_2, \hat{c}'_3, \hat{c}'_4)^T$ and $(\hat{c}''_2, \hat{c}''_3, \hat{c}''_4)^T$.*

Proof sketch. We write $\boldsymbol{\tau} = (\tau_{ijk})$ with τ_{ijk} being the coefficient of $u_i u'_j u''_k$ in T_1. It follows from (10) that it is necessary for the elements of τ to sum to zero. For the converse, we note that $\boldsymbol{\tau} = A \boldsymbol{\rho}_1$, where

$$\boldsymbol{\tau} = \begin{bmatrix} \tau_{223} \\ \tau_{232} \\ \tau_{233} \\ \tau_{322} \\ \tau_{323} \\ \tau_{332} \end{bmatrix}, A = \begin{bmatrix} 0 & 0 & -1 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{\rho}_1 = \begin{bmatrix} \rho_{23} \\ \rho_{24} \\ \rho_{32} \\ \rho_{34} \\ \rho_{42} \\ \rho_{43} \end{bmatrix}. \tag{11}$$

The matrix A_1 has rank 5 and its rows and columns sum to zero. In particular, a solution $\boldsymbol{\rho}_1$ of Eq. (11) can always be written as $\boldsymbol{\rho}_1 = \boldsymbol{\rho}_0 + t \mathbf{1}_6$ for some scalar t, where $\boldsymbol{\rho}_0$ is any solution of $\boldsymbol{\tau} = A \boldsymbol{\rho}_0$. Such $\boldsymbol{\rho}_0$ always exists if the elements of $\boldsymbol{\tau}$ sum to zero. In order for $\boldsymbol{\tau}$ to be a valid set of coefficients, the vector $\boldsymbol{\rho}_1$ must factor as $\hat{c}'_i \hat{c}''_j$. According to Prop. 4.1, this corresponds to a single equation in the entries ρ_{ij}, which we can use to solve for t. Moreover, while this equation was cubic in ρ_{ij}, it is easy to see that there is a cancellation of the term t^3, so we are left with a quadratic equation in t. Each of the two (possibly complex) solutions to this equation determines up to scale the vectors $(\hat{c}'_2, \hat{c}'_3, \hat{c}'_4)^T$ and $(\hat{c}''_2, \hat{c}''_3, \hat{c}''_4)^T$. □

Although the solutions in Prop. 4.3 may in principle be complex conjugate, they will be real when the vector $\boldsymbol{\tau}$ is estimated from exact correspondences, and remain real under small perturbations. We also note that the statement of Prop. 4.3 holds without modifications for trilinearities T_2, T_3. In the case of T_4, there are 24 non-zero coefficients rather than six, but these are also only constrained by linear conditions (in all cases, the set of valid coefficients is a vector space of dimension five).

The fact that the coefficients of each trilinearity are not subject to non-linear constraints may seem surprising. One way to justify this property is to point out the close relationship between our trilinearities and *2D trifocal tensors*. These $(2 \times 2 \times 2)$-tensors characterize correspondences for triples of projections from $\mathbb{P}^2$ to $\mathbb{P}^1$. It was shown by Quan [20] that the entries of these tensors are also not bound by any internal constraints. In the supplementary material, we argue that the reduced trilinearities can be obtained by composing a 2D trifocal tensor with linear changes of coordinates. The idea behind this fact is that, according to Prop. 1.1 [23], each trilinarity T_i imposes the condition that three viewing rays admit a common transversal passing through the point $\boldsymbol{z}_i$. If we fix a retinal plane

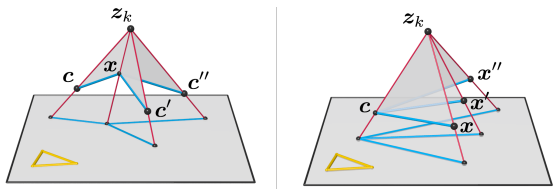

Figure 6. Trilinearities express the condition that three coplanar lines intersect. These are the projections from $\boldsymbol{z}_k$ of three viewing rays converging at some scene point $\boldsymbol{x}$ in the primal case (**left**), and some pinhole $\boldsymbol{c}$ in the dual one (**right**).

with appropriate coordinates, this is the same as requiring three coplanar lines to meet at a point, which is the condition imposed by a 2D trifocal tensor. See Fig. 6 and the supplementary material for details.

5. A quasi-linear three-view SFM algorithm

We now show that the four trilinearities (primal or dual) can be used to solve structure from motion when sufficient correspondences are available. Let us first consider primal trilinearities. As noted in the previous section, each point correspondence imposes four linear conditions on the vector $\boldsymbol{\rho}$ whose entries are $\rho_{ij} = \hat{c}'_i \hat{c}''_j$. Given p correspondences we can stack the corresponding linear equations to form a $4p \times 12$ matrix T and write all the constraints as $T\boldsymbol{\rho} = 0$. However, it is easy to realize that $T\mathbf{1}_{12} = 0$, independently of the image coordinates. In the absence of noise, the kernel of T is thus always at least two-dimensional, since it contains $\mathbf{1}_{12}$ and the "true" solution $\boldsymbol{\rho}$ to our problem. We address this issue by exploiting the special form of the vector $\boldsymbol{\rho}$. In particular, let $\boldsymbol{e}$ be a vector independent of $\mathbf{1}_{12}$ in the kernel, so that we may write $\boldsymbol{\rho} = \boldsymbol{e} + \lambda \mathbf{1}_{12}$ for some scalar λ, and thus $\rho_{ij} = \hat{c}'_i \hat{c}''_j = e_{ij} + \lambda$ for $j \neq i$. We have, for example $\hat{c}''_3 = (e_{13} + \lambda)c'_1 = (e_{23} + \lambda)c'_2$, and $\hat{c}''_4 = (e_{14} + \lambda)c'_1 = (e_{24} + \lambda)c'_2$. This allows us to eliminate λ, obtaining $(e_{23} - e_{24})\hat{c}'_1 + (e_{14} - e_{13})\hat{c}'_2 = 0$. Collecting all similar constraints finally yields the following equations:

$$\begin{bmatrix} e_{23} - e_{24} & e_{14} - e_{13} & 0 & 0 \\ e_{32} - e_{34} & 0 & e_{14} - e_{12} & 0 \\ e_{42} - e_{43} & 0 & 0 & e_{13} - e_{12} \\ 0 & e_{31} - e_{34} & e_{24} - e_{21} & 0 \\ 0 & e_{43} - e_{41} & 0 & e_{21} - e_{23} \\ 0 & 0 & e_{41} - e_{42} & e_{32} - e_{31} \end{bmatrix} \hat{\boldsymbol{c}}' = 0, \qquad (12)$$

$$\begin{bmatrix} e_{32} - e_{42} & e_{41} - e_{31} & 0 & 0 \\ e_{23} - e_{43} & 0 & e_{41} - e_{21} & 0 \\ e_{24} - e_{34} & 0 & 0 & e_{31} - e_{21} \\ 0 & e_{13} - e_{43} & e_{42} - e_{12} & 0 \\ 0 & e_{14} - e_{34} & 0 & e_{32} - e_{12} \\ 0 & 0 & e_{14} - e_{24} & e_{23} - e_{13} \end{bmatrix} \hat{\boldsymbol{c}}'' = 0 \qquad (13)$$

These equations are sufficient to determine $\hat{\boldsymbol{c}}'$ and $\hat{\boldsymbol{c}}''$ and thus $\boldsymbol{c}'$ and $\boldsymbol{c}''$. Three correspondences (in addition to the four "reference" correspondences) are necessary to obtain a unique solution for the 11 unknowns.[2] Note that although the vector $\boldsymbol{\rho}$ must satisfy certain algebraic constraints in order to be of the form $\rho_{ij} = c'_i c''_j$ (see Prop. 4.2), our strategy bypasses this difficulty by directly recovering the vectors $\boldsymbol{c}'$ and $\boldsymbol{c}''$. In other words, this "quasi-linear" method is always guaranteed to return a valid solution (which will approximate the "true" solution in the presence of noise). In practice, we pick four reference points among the (known) correspondences between these pictures, and apply appropriate image coordinate changes so they become basis points. We then use singular value decomposition and the remaining points to find the least-squares solution $\boldsymbol{e}$ of the system of equation in $\boldsymbol{\rho}$ associated with T which is orthogonal to $\mathbf{1}_{12}$, and finally use Eq. (13) to compute the position of the pinholes. Linear least-squares can then be used again to reconstruct all scene points from the known pinholes and image coordinates. We repeat this process for random quadruples of reference points and a fixed number of iterations, and report the results.

The dual algorithm is very similar except that this time we fix three points $\boldsymbol{x}$, $\boldsymbol{x}'$, and $\boldsymbol{x}''$ instead of three images. We again repeatedly pick four random reference correspondences, and use all images (at least three) of $\boldsymbol{x}$, $\boldsymbol{x}'$ and $\boldsymbol{x}''$ to reconstruct them. The procedure is the same as above, replacing $\boldsymbol{c}, \boldsymbol{c}', \boldsymbol{c}''$ with $\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}', \hat{\boldsymbol{x}}''$.

5.1. Preliminary experiments

We have implemented both the primal and dual versions of the proposed algorithm and we present below some preliminary experiments with real and synthetic data. We emphasize that the main thrust of our presentation is theoretical, with the objective of reaching a better understanding of multi-view geometry. We do not make any claim here of outperforming the state of the art, and our experiments are only included as a proof of concept to validate two points: (1) our algorithm gives reasonable reconstructions on real data in a least-squares setting; and (2) its primal version also gives reasonable results on synthetic data with additive Gaussian noise in a setting with only 7 correspondences.

Inria toy house data [17]. This dataset consists of 6 images of the same 38 points. It is small by any standard, but with enough views and correspondences to demonstrate both the primal and dual versions of our algorithm. It also makes it easy to visualize the results since it contains edges linking the data points (this information is of course only used for display). Figure 7 shows the reconstructions obtained for the best choice among 5, 20 and 50 different quadruples of reference points based on the reprojection error. The reconstructions are overlaid on the ground-truth 3D points after projective registration, along with the corresponding mean

[2] The minimum number of correspondences for three-view structure from motion is six [12]. The proposed algorithm is not "minimal" and the solution is over-determined.

reprojection errors in pixels, and the mean relative reconstruction errors in percentage of the radius of the scene.[3] Reprojection errors are quite reasonable in both cases for 20 different choices, and the reconstructions themselves are also quite reasonable after 5 random choices only. For comparison, we have also tested the code for linear estimation of the trifocal tensor by Juliá and Monasse [13][4], dubbed here *3LTFT*. It yields mean reprojection and reconstruction errors of 0.7 pixel and 0.3% respectively but, unlike our method, benefits from Hartley's data preconditioning [9].

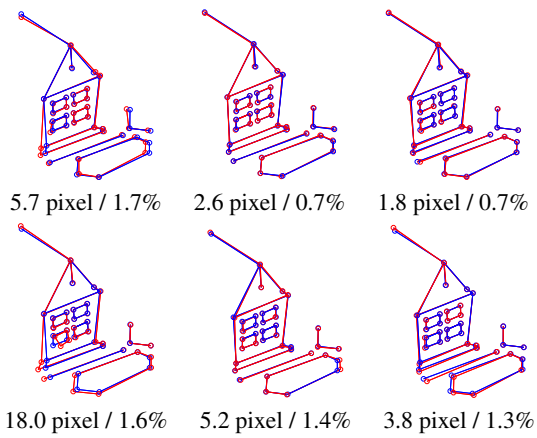

Figure 7. Experiments on the Inria house data [17]. Registered reconstruction (in red) vs. ground truth (in blue) for the primal (**top**) and dual (**bottom**) versions of our algorithm. The results are shown, from left to right, for 5, 20, and 50 different choices of image basis points.

Synthetic noisy data [13]. We have also compared our primal method with 3LTFT on synthetic data with various amounts of additive Gaussian noise. In this setting, the camera parameters are estimated from 7 correspondences from three images (the minimum number for both LTFT and our algorithm). The quality of the reconstruction is evaluated by measuring how well it predicts the reprojection of the remaining points in the dataset as well as their 3D reconstruction, once again registered to the ground truth through a homography. This is the setting where both methods could be used in practice to establish correspondences via RANSAC, for example, before a final bundle adjustment step. Following [13], we have constructed a scene consisting of 100 points randomly distributed in a cube of side 400mm observed by three 1200×1800 35mm cameras with 50mm lenses about 1m away, and added Gaussian noise with a standard deviation σ varying between 0 and 2 pixel to image coordinates. Figure 8 shows the median values of the mean reprojection and reconstruction errors, given respectively in pixel and mm, for 40 random choices of the 7 point correspondences and different values of σ.[5] The 3LTFT plots are shown in black, and the curves for our method are drawn in blue, green and red for the best random draw among 5, 10, and 20 choices of 4 reference points among the original 7. As shown in Fig. 8, 3LTFT does better in general than our primal method but both algorithms give reasonable reprojection and reconstruction errors for low levels of noise.

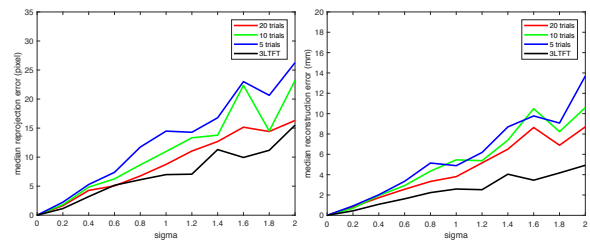

Figure 8. Experiments with synthetic data [13]. See text for details.

6. Conclusion

We have proposed a new coordinate-free approach to multi-view geometry that explains Carlsson-Weinshall duality and leads to new algorithms for primal and dual structure from motion. We believe that this type of work, whose objective is to complete our understanding of the geometric underpinnings of 3D computer vision, must be pursued for a clear and unified picture of multi-view geometry to emerge. Although, we do not claim by any means to establish a new state of the art with the proposed algorithms, we also believe that our preliminary experiments demonstrate that they can serve as yet another useful toolset in the existing arsenal of approaches to SFM.

Acknowledgments. This work was supported in part by the the Inria/NYU collaboration agreement, the Louis Vuitton/ENS chair on artificial intellgence, the Inria-CMU associated team GAYA, ANR Recap, and Samsung Electronics.

[3] Here, the "best" basis is the one minimizing the mean reprojection error over the three images, without of course using 3D ground-truth information.

[4] Available at `https://github.com/LauraFJulia/TFT_vs_Fund`.

[5] We use the median instead of the mean because the latter is occasionally totally off course for certain choices of the 7 correspondences for both methods. See supplemental material for details.

References

[1] S. Carlsson and D. Weinshall. Dual computation of projective shape and camera positions from mutliple images. *IJCV*, 27(3):227–241, 1998.

[2] A. B. Coble. Points sets and allied cremona groups (part i). *Transactions of the American Mathematical Society*, 16(2):155–198, 1915.

[3] I. Dolgachev. Point sets in projective spaces and theta functions. *Astérisque*, 165, 1988.

[4] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *ECCV*, volume 588 of *Lecture Notes in Computer Science*, pages 563–578, Santa Margherita, Italy, 1992. Springer-Verlag.

[5] O. Faugeras, Q.-T. Luong, and T. Papadopoulo. *The Geometry of Multiple Images*. MIT Press, 2001.

[6] O. Faugeras and B. Mourrain. On the geometry and algebra of the point and line correspondences between n images. Technical Report 2665, INRIA Sophia-Antipolis, 1995.

[7] O. Faugeras and T. Papadopoulo. Grassman-Cayley algebra for modeling systems of cameras and the algebraic equations of the manifold of trifocal tensors. Technical Report 3225, INRIA Sophia-Antipolis, 1997.

[8] J. Harris. *Algebraic geometry: a first course*, volume 133. Springer Science & Business Media, 2013.

[9] R. Hartley. In defence of the 8-point algorithm. In *ICCV*, pages 1064–1070, Boston, MA, 1995.

[10] R. Hartley. Lines and points in three views and the trifocal tensor. *IJCV*, 22(2):125–140, 1997.

[11] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *CVPR*, pages 761–764, Champaign, IL, 1992.

[12] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2nd edition, 2004.

[13] L. Julià and P. Monasse. A critical review of the trifocal tensor estimation. In *PSIVT*, 2017.

[14] J. Koenderink and A. van Doorn. Affine structure from motion. *J. Opt. Soc. Am. A*, 8:377–385, 1990.

[15] A. Levin and A. Shashua. Revisiting single-view shape tensors: Theory and applications. In *ECCV*, 2002.

[16] Q.-T. Luong and O. Faugeras. The fundamental matrix: theory, algorithms, and stability analysis. *IJCV*, 17(1):43–76, 1996.

[17] R. Mohr, L. Quan, F. Veillon, and B. Boufama. Relative 3D reconstruction using multiple uncalibrated images. Technical Report RT 84-IMAG 12-LIFIA, LIFIA-IRIMAG, June 1992.

[18] J. Ponce, T. Papadopoulo, M. Teillaud, and B. Triggs. The absolute quadratic complex and its application to camera self calibration. In *CVPR*, 2005.

[19] J. Ponce, B. Sturmfels, and M. Trager. Congruences and concurrent lines in multi-view geometry. *Advances in Applied Mathematics*, 88:62–91, 2017.

[20] L. Quan. Two-way ambiguity in 2d projective reconstruction from three uncalibrated 1d images. *PAMI*, 23(2):212–216, 2001.

[21] A. Shashua. Algebraic functions for recognition. *PAMI*, 17(8):779–789, 1995.

[22] M. Trager, M. Hebert, and J. Ponce. The joint image handbook. In *ICCV*, 2015.

[23] M. Trager, J. Ponce, and M. Hebert. Trinocular geometry revisited. *IJCV*, 120(2):134–152, 2016.

[24] B. Triggs. Matching constraints and the joint image. In *ICCV*, 1995.

[25] J. Weng, T. Huang, and N. Ahuja. Motion and structure from line correspondences: closed-form solution, uniqueness, and optimization. *PAMI*, 14(3):318–336, 1992.

We present some proofs that were not included in the main part of the paper and some additional results of our algorithm on synthetic data.

A. Analytical cameras and duality

The analytical expression for $P_{\boldsymbol{c}}(\boldsymbol{x})$ given in Proposition 2.4 follows easily from the following general result.

Proposition A.1. *The unique projective transformation of $\mathbb{P}^n$ that maps $n+2$ general points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n+2}$ to the standard basis of $\mathbb{P}^n$ can be described by*

$$\boldsymbol{y} \mapsto \begin{bmatrix} \frac{|\boldsymbol{x}_2 \ldots \boldsymbol{x}_{n+1} \boldsymbol{y}|}{|\boldsymbol{x}_2 \ldots \boldsymbol{x}_{n+1} \boldsymbol{x}_{n+2}|} \\ \frac{|\boldsymbol{x}_1 \boldsymbol{x}_3 \ldots \boldsymbol{x}_{n+1} \boldsymbol{y}|}{|\boldsymbol{x}_1 \boldsymbol{x}_3 \ldots \boldsymbol{x}_{n+1} \boldsymbol{x}_{n+2}|} \\ \vdots \\ \frac{|\boldsymbol{x}_1 \ldots \boldsymbol{x}_n \boldsymbol{y}|}{|\boldsymbol{x}_1 \ldots \boldsymbol{x}_n \boldsymbol{x}_{n+2}|} \end{bmatrix}. \tag{14}$$

Proof. It is clear that (14) describes a projective transformation, since all the expressions are linear in the coordinates of $\boldsymbol{y}$. It also follows from elementary properties of determinants that (14) maps $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n+2}$ to the standard basis. □

We now turn to Cremona transformations and justify the general formula (5) given in the paper.

Proposition A.2. *The rational map of $\mathbb{P}^3$ given by*

$$\hat{\boldsymbol{y}} = \boldsymbol{Z} \begin{bmatrix} \frac{1}{|\boldsymbol{y} \boldsymbol{x}_2 \boldsymbol{x}_3 \boldsymbol{x}_4|} \\ \frac{1}{|\boldsymbol{z}_1 \boldsymbol{y} \boldsymbol{z}_3 \boldsymbol{z}_4|} \\ \frac{1}{|\boldsymbol{z}_1 \boldsymbol{z}_2 \boldsymbol{y} \boldsymbol{z}_4|} \\ \frac{1}{|\boldsymbol{z}_1 \boldsymbol{z}_2 \boldsymbol{z}_3 \boldsymbol{y}|} \end{bmatrix}, \tag{15}$$

where $\boldsymbol{Z}$ has columns $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_4$, is a Cremona transformation relative to $\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3, \boldsymbol{z}_4$ (in the sense of Lemma 2.2 from the main part of the paper).

Proof. We need to show that $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_4, \boldsymbol{y}_1, \boldsymbol{y}_2$ are in the same projective configuration as $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_4, \hat{\boldsymbol{y}}_2, \hat{\boldsymbol{y}}_1$. Indeed, a projective transformation relating the two sets of points is given by

$$\boldsymbol{x} \mapsto \boldsymbol{Z} \begin{bmatrix} \frac{|\boldsymbol{x} \boldsymbol{z}_2 \boldsymbol{z}_3 \boldsymbol{z}_4|}{|\boldsymbol{y}_1 \boldsymbol{z}_2 \boldsymbol{z}_3 \boldsymbol{z}_4||\boldsymbol{y}_2 \boldsymbol{z}_2 \boldsymbol{z}_3 \boldsymbol{x}_4|} \\ \frac{|\boldsymbol{z}_1 \boldsymbol{x} \boldsymbol{z}_3 \boldsymbol{z}_4|}{|\boldsymbol{z}_1 \boldsymbol{y}_1 \boldsymbol{z}_3 \boldsymbol{z}_4||\boldsymbol{z}_1 \boldsymbol{y}_2 \boldsymbol{z}_3 \boldsymbol{x}_4|} \\ \frac{|\boldsymbol{z}_1 \boldsymbol{z}_2 \boldsymbol{x} \boldsymbol{z}_4|}{|\boldsymbol{z}_1 \boldsymbol{z}_2 \boldsymbol{y}_1 \boldsymbol{z}_4||\boldsymbol{z}_1 \boldsymbol{z}_2 \boldsymbol{y}_2 \boldsymbol{x}_4|} \\ \frac{|\boldsymbol{z}_1 \boldsymbol{z}_2 \boldsymbol{z}_3 \boldsymbol{x}|}{|\boldsymbol{z}_1 \boldsymbol{z}_2 \boldsymbol{z}_3 \boldsymbol{y}_1||\boldsymbol{z}_1 \boldsymbol{z}_2 \boldsymbol{z}_3 \boldsymbol{y}_2|} \end{bmatrix}, \tag{16}$$

where $\boldsymbol{y}_1, \boldsymbol{y}_2$ are considered fixed. Note that when $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_4$ are basis points then (15) yields the standard Cremona transformation

$$\hat{\boldsymbol{y}} = \left(\frac{1}{y_1}, \frac{1}{y_2}, \frac{1}{y_3}, \frac{1}{y_4} \right)^T, \tag{17}$$

and (16) becomes

$$\boldsymbol{x} \mapsto \left(\frac{x_1}{y_{11} y_{21}}, \frac{x_2}{y_{12} y_{22}}, \frac{x_3}{y_{13} y_{23}}, \frac{x_4}{y_{14} y_{24}} \right)^T, \tag{18}$$

where $\boldsymbol{x} = (x_1, x_2, x_3, x_4)^T$, $\boldsymbol{y}_1 = (y_{11}, y_{12}, y_{13}, y_{14})^T$ and $\boldsymbol{y}_2 = (y_{21}, y_{22}, y_{23}, y_{24})^T$. □

We next prove Proposition 3.1 from the main part of the paper. We recall that the set S in $\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^2$ was defined as the set of triples $(\boldsymbol{c}, \boldsymbol{x}, \boldsymbol{u})$ such that $P_{\boldsymbol{c}}(\boldsymbol{x}) = \boldsymbol{u}$ where $P_{\boldsymbol{c}}$ is a reduced camera.

Proposition 3.1. **(1)** *For fixed $\boldsymbol{c}$ and $\boldsymbol{u}$, the set of points $\boldsymbol{x}$ such that $(\boldsymbol{c}, \boldsymbol{x}, \boldsymbol{u})$ belongs to S is a line with Plücker coordinates*

$$\boldsymbol{\xi} = Q_{\boldsymbol{c}} \boldsymbol{u} \text{ where } Q_{\boldsymbol{c}} = \begin{bmatrix} c_1 c_4 & 0 & 0 \\ 0 & c_2 c_4 & 0 \\ 0 & 0 & c_3 c_4 \\ 0 & -c_2 c_3 & c_2 c_3 \\ c_1 c_3 & 0 & -c_1 c_3 \\ -c_1 c_2 & c_1 c_2 & 0 \end{bmatrix}. \tag{19}$$

(2) *For fixed $\boldsymbol{x}$ and $\boldsymbol{u}$, the set of points $\boldsymbol{c}$ such that $(\boldsymbol{c}, \boldsymbol{x}, \boldsymbol{u})$ belongs to S is a twisted cubic passing through $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_4$ and $\boldsymbol{x}$.*

Proof. To show (1), we define $\boldsymbol{\xi}_j = \boldsymbol{c} \vee \boldsymbol{z}_j$ $(j = 1, \ldots, 4)$. With our choice of coordinate system, we have

$$\boldsymbol{\xi}_1 = \begin{bmatrix} c_4 \\ 0 \\ 0 \\ 0 \\ c_3 \\ -c_2 \end{bmatrix}, \boldsymbol{\xi}_2 = \begin{bmatrix} 0 \\ c_4 \\ 0 \\ -c_3 \\ 0 \\ c_1 \end{bmatrix}, \boldsymbol{\xi}_3 = \begin{bmatrix} 0 \\ 0 \\ c_4 \\ c_2 \\ -c_1 \\ 0 \end{bmatrix}, \boldsymbol{\xi}_4 = -\begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{20}$$

We write $\boldsymbol{\xi} = \rho_1 u_1 \boldsymbol{\xi}_1 + \rho_2 u_2 \boldsymbol{\xi}_2 + \rho_3 u_3 \boldsymbol{\xi}_3$, where the scalars ρ_1, ρ_2, ρ_3 have been chosen so that $\boldsymbol{\xi}_4 = \rho_1 \boldsymbol{\xi}_1 + \rho_2 \boldsymbol{\xi}_2 + \rho_3 \boldsymbol{\xi}_3$. A simple calculation shows that $\rho_1 = -c_1/c_4$, $\rho_2 = -c_2/c_4$, $\rho_3 = -c_3/c_4$, and (19) immediately follows.

For the second statement, we first note that the set S is characterized algebraically by the relation

$$\operatorname{rk} \begin{bmatrix} c_1 x_4 - c_4 x_1 & u_1 c_1 \\ c_2 x_4 - c_4 x_2 & u_2 c_2 \\ c_3 x_4 - c_4 x_3 & u_3 c_3 \end{bmatrix} = 1. \tag{21}$$

This expression follows from

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \sim P_{\boldsymbol{c}}(\boldsymbol{x}) = \begin{bmatrix} x_1/c_1 - x_4/c_4 \\ x_2/c_2 - x_4/c_4 \\ x_3/c_3 - x_4/c_4 \end{bmatrix}, \tag{22}$$

after we clear denominators and eliminate a factor of c_4. We now observe that for fixed $\boldsymbol{x}$ and $\boldsymbol{u}$, the three quadratic equations in $\boldsymbol{c}$ from the minors of (21) define a twisted cubic curve. Indeed, as shown for example in [8, p.14], if L_1, L_2, L_3, M_1, M_2, M_3 are linear forms in c_1, c_2, c_3, c_4, then the projective set defined by

$$\operatorname{rk} \begin{bmatrix} L_1 & L_2 & L_3 \\ M_1 & M_2 & M_3 \end{bmatrix} = 1, \tag{23}$$

is a twisted cubic if and only if for any (λ, μ) the three linear forms $\lambda L_i + \mu M_i$ $(i = 1, 2, 3)$ are independent. It follows that (21) is a twisted cubic in $\boldsymbol{x}$ if and only if

$$\begin{bmatrix} -\lambda x_4 + \mu u_1 & 0 & 0 & \lambda x_1 \\ 0 & -\lambda x_4 + \mu u_2 & 0 & \lambda x_2 \\ 0 & 0 & -\lambda x_4 + \mu u_3 & \lambda x_3 \end{bmatrix} \tag{24}$$

has rank three for all λ, μ. This is true if the coordinates u_1, u_2, u_3 are all distinct and not zero, and the coordinates x_1, x_2, x_3 are not zero, which is indeed the case under our genericity assumptions. □

A.1. Reduced 2D trifocal tensors and trilinearities

We begin by presenting some properties of the SFM problem for projections from $\mathbb{P}^2$ to $\mathbb{P}^1$. We represent an analytical projection of this type using a 2×3 real matrix $\boldsymbol{P}$ defined up to scale. As for traditional cameras, the center of this projection is the point of $\mathbb{P}^2$ associated with the null-space of $\boldsymbol{P}$. Note that if $\boldsymbol{P}, \boldsymbol{P}'$ are two projections with distinct centers, then every pair of points $(\boldsymbol{t}, \boldsymbol{t}')$ in $\mathbb{P}^1 \times \mathbb{P}^1$ will be a "correspondence" for $\boldsymbol{P}, \boldsymbol{P}'$, *i.e.*, there will always exist a point $\boldsymbol{x}$ in $\mathbb{P}^2$ such that $\boldsymbol{P}\boldsymbol{x} \sim \boldsymbol{t}$ and $\boldsymbol{P}'\boldsymbol{x} \sim \boldsymbol{t}'$. This follows from the fact that two lines in $\mathbb{P}^2$ always intersect. Thus, the first interesting case of multi-view geometry is for $n = 3$.

Proposition A.3. *Given three projections $\boldsymbol{P}, \boldsymbol{P}', \boldsymbol{P}''$ with disjoint centers, there exists a $2 \times 2 \times 2$ "trifocal tensor" $\mathcal{T}$ such that $(\boldsymbol{t}, \boldsymbol{t}', \boldsymbol{t}'')$ in $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$ correspond if an only if*

$$\mathcal{T}^{ijk} t_i t'_j t''_k = 0, \tag{25}$$

where we use Einstein notation for summation. Entries for $\mathcal{T}$ are given by

$$\mathcal{T}^{ijk} = (-1)^{ijk} |P_{3-i} P'_{3-j} P''_{3-k}|, \; i, j, k \in \{1, 2\}, \tag{26}$$

where $|P_i P'_j P''_k|$ denotes the determinant of the 3×3 matrix obtained by stacking the i-th row of $\boldsymbol{P}$, the j-th row of $\boldsymbol{P}'$, and the k-th row of $\boldsymbol{P}''$. Finally, the trifocal tensor $\mathcal{T}$ satisfies the following properties:

1. *Any general $2 \times 2 \times 2$ tensor is a "valid" trifocal tensor.*
2. *Given a general trifocal tensor $\mathcal{T}$, there are* two *(possibly complex) projectively distinct sets of parameters for $\boldsymbol{P}, \boldsymbol{P}', \boldsymbol{P}''$.*

Proof. These facts are shown in [20], but we give a short proof here for completeness. We first note that for any 2×3 projection matrix $\boldsymbol{P}$ there is an associated 3×2 "inverse projection" matrix $\boldsymbol{Q}$ that maps points in $\mathbb{P}^1$ to the corresponding viewing lines in $\mathbb{P}^2$ for $\boldsymbol{P}$ (expressed using dual coordinates). The relation between $\boldsymbol{P}$ and $\boldsymbol{Q}$ is simply

$$\boldsymbol{Q} = \boldsymbol{P}^T \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \tag{27}$$

Since three lines in $\mathbb{P}^2$ written in dual coordinates as $\boldsymbol{l}, \boldsymbol{l}', \boldsymbol{l}''$ converge if and only if $|\boldsymbol{l}, \boldsymbol{l}', \boldsymbol{l}''| = 0$, we see that $(\boldsymbol{t}, \boldsymbol{t}', \boldsymbol{t}'')$ in $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$ form a correspondence for $\boldsymbol{P}, \boldsymbol{P}', \boldsymbol{P}''$ if and only if

$$\begin{vmatrix} \boldsymbol{Q}\boldsymbol{t} & \boldsymbol{Q}'\boldsymbol{t}' & \boldsymbol{Q}''\boldsymbol{t}'' \end{vmatrix} = 0, \tag{28}$$

where $\boldsymbol{Q}, \boldsymbol{Q}', \boldsymbol{Q}''$ denote the inverse projection matrices for $\boldsymbol{P}, \boldsymbol{P}', \boldsymbol{P}''$. The expansion of this determinant, together with (27), immediately yields (25) and (26).

The properties (1) and (2) can be shown computationally. Alternatively, one can explicitly a describe method for reconstructing two projectively distinct projection mappings from a general tensor. We do this below in the case of "reduced" tensors. The fact that every tensor is a valid trifocal tensor can also be argued informally by noting that each projection has 5 degrees of freedom, so after removing projective ambiguity (with 8 parameters) we are left with $5 + 5 + 5 - 8 = 7$ which correspond to all $2 \times 2 \times 2$ tensors up to scale.[6] □ □

We now consider "reduced" projection mappings from $\mathbb{P}^2$ to $\mathbb{P}^1$. Similarly to the 3D case discussed in the paper, a reduced projection is determined by a center together with 3 fixed points in $\mathbb{P}^2$ in general position. In the following, we will always assume that the three fixed points are basis points for a projective reference frame. This leads to projection matrices of the form

$$\boldsymbol{P}_{\boldsymbol{c}} = \begin{bmatrix} 1/c_1 & 0 & -1/c_3 \\ 0 & 1/c_2 & -1/c_3 \end{bmatrix} = \begin{bmatrix} \hat{c}_1 & 0 & -\hat{c}_3 \\ 0 & \hat{c}_2 & -\hat{c}_3 \end{bmatrix}, \tag{29}$$

where $\boldsymbol{c} = (c_1, c_2, c_3)^T$ is the center of projection, and we write $\hat{c}_i = 1/c_i$ for convenience. Specializing (26) for three reduced cameras $P_{\boldsymbol{c}}, P_{\boldsymbol{c}'}, P_{\boldsymbol{c}''}$ with $\boldsymbol{c} = (1, 1, 1)^T$, $\boldsymbol{c}' = (c'_1, c'_2, c'_3)^T$, $\boldsymbol{c}'' = (c''_1, c''_2, c''_3)^T$, yields

$$\begin{array}{ll} \mathcal{T}^{111} = 0, & \mathcal{T}^{112} = \hat{c}'_3\hat{c}''_1 - \hat{c}'_2\hat{c}''_1 \\ \mathcal{T}^{121} = \hat{c}'_1\hat{c}''_2 - \hat{c}'_1\hat{c}''_3, & \mathcal{T}^{122} = \hat{c}'_1\hat{c}''_3 - \hat{c}'_3\hat{c}''_1, \\ \mathcal{T}^{211} = \hat{c}'_2\hat{c}''_3 - \hat{c}'_3\hat{c}''_2, & \mathcal{T}^{212} = \hat{c}'_2\hat{c}''_1 - \hat{c}'_2\hat{c}''_3, \\ \mathcal{T}^{221} = \hat{c}'_3\hat{c}''_2 - \hat{c}'_1\hat{c}''_2, & \mathcal{T}^{222} = 0. \end{array} \tag{30}$$

In addition to $\mathcal{T}^{111} = 0$ and $\mathcal{T}^{222} = 0$, we note that the expressions for the remaining six coefficients always sum to zero. Indeed, there are three "synthetic" constraints arising from the fact that the standard basis points in each image provide by construction three correspondences. However, these three are the *only* constraints that such a tensor must satisfy. In fact, we can give the following simple algorithm for 2D reduced SFM.

1. Change coordinates in each image $\mathbb{P}^1$ using three triples of correspondences to restrict to reduced projection matrices as in (29).
2. Estimate the six non-zero coefficients a, b, c, d, e, f of the reduced trifocal tensor (30), under the condition that they sum to zero, using an arbitrary number of correspondences (at least five) in the new image coordinates.
3. Use (30) to recover $\hat{c}'_1, \hat{c}'_2, \hat{c}'_3$ and $\hat{c}''_1, \hat{c}''_2, \hat{c}''_3$, and hence $\boldsymbol{c}', \boldsymbol{c}''$ from the six coefficients a, b, c, d, e, f. There are in general *two* solutions, which can be computed as follows. Writing $\rho_{ij} = \hat{c}'_i\hat{c}''_j$, we have that (30) yields the following linear relation

$$\begin{bmatrix} 0 & 0 & -1 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \rho_{12} \\ \rho_{13} \\ \rho_{21} \\ \rho_{23} \\ \rho_{31} \\ \rho_{32} \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \end{bmatrix}. \tag{31}$$

The matrix on the left has rank 5, with a null-space generated by $(1, 1, 1, 1, 1, 1)^T$. This means that we may write $\rho_{ij} =$

[6] To make this argument more precise, one needs to observe that projective transformations act freely on triples of cameras (*i.e.*, no projective transformation of $\mathbb{P}^2$ fixes $\boldsymbol{P}, \boldsymbol{P}', \boldsymbol{P}''$ simultaneously).

$e_{ij} + t$ where $(e_{12}, e_{13}, e_{21}, e_{23}, e_{31}, e_{32})$ is any vector that satisfies (31) and t is unknown. However, we may solve for t using the fact that ρ_{ij} must satisfy

$$\rho_{12}\rho_{23}\rho_{31} - \rho_{13}\rho_{21}\rho_{32} = 0. \tag{32}$$

This yields a constraint on t, which is actually quadratic rather than cubic, since the cubic term in t cancels out from the two summands. Given a valid set of ρ_{ij}, it is straightforward to recover $\hat{c}'_i$ and $\hat{c}''_j$. Indeed, it is sufficient for example to set

$$\begin{aligned}(\hat{c}'_1, \hat{c}'_2, \hat{c}'_3) &= (\rho_{12}/\rho_{32}, \rho_{21}/\rho_{31}, 1),\\ (\hat{c}''_1, \hat{c}''_2, \hat{c}''_3) &= (\rho_{21}/\rho_{23}, \rho_{12}/\rho_{13}, 1).\end{aligned} \tag{33}$$

Note that this procedure is very similar to the method used in the proof of Proposition 4.3 in the paper. Indeed, the trilinearities T_1, T_2, T_3, T_4 are closely related to the 2D reduced trifocal tensor, and we will now spell out this relation in detail. We first recall the following fact from [23] (see also Proposition 1.1 in the paper).

Proposition A.4. *Three image points $\boldsymbol{u}, \boldsymbol{u}', \boldsymbol{u}''$ satisfy the trilinearity T_i if and only if the three associated viewing lines $\boldsymbol{\xi}, \boldsymbol{\xi}', \boldsymbol{\xi}''$ admit a common transveral through the basis point $\boldsymbol{z}_i$.*

The geometric condition from Proposition A.4 can be expressed by considering an arbitrary projection $P_{\boldsymbol{z}_i}$ with center $\boldsymbol{z}_i$, and imposing that the projections $\boldsymbol{l}, \boldsymbol{l}', \boldsymbol{l}''$ of $\boldsymbol{\xi}, \boldsymbol{\xi}', \boldsymbol{\xi}''$ under $P_{\boldsymbol{z}_i}$ are lines in $\mathbb{P}^2$ that converge at a point. This is why 2D trifocal tensors come into play.

In what follows, we write $\boldsymbol{c}'_r, \boldsymbol{c}''_r$ for the points in $\mathbb{P}^2$ obtained from $\boldsymbol{c}', \boldsymbol{c}''$ by excluding the r-th coordinate (*e.g.*, $\boldsymbol{c}'_2 = (c'_1, c'_3, c'_4)^T$).

Proposition A.5. *The trilinearity T_r ($r = 1, 2, 3, 4$) applied to $\boldsymbol{u}, \boldsymbol{u}', \boldsymbol{u}''$ can be written as*

$$\mathcal{T}_r^{ijk}(\boldsymbol{M}_r\boldsymbol{u})_i(\boldsymbol{M}_r\boldsymbol{u}')_j(\boldsymbol{M}_r\boldsymbol{u}'')_k = 0, \tag{34}$$

where $\mathcal{T}_r^{ijk}$ is the reduced 2D trifocal tensor associated with $(1, 1, 1)^T$, $\boldsymbol{c}'_r$, $\boldsymbol{c}''_r$ (see (30) *for the coefficients) and*

$$\begin{aligned}\boldsymbol{M}_1 &= \begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}, \quad \boldsymbol{M}_2 = \begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 1\end{bmatrix},\\ \boldsymbol{M}_3 &= \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{bmatrix}, \quad \boldsymbol{M}_4 = \begin{bmatrix}1 & 0 & -1\\ 0 & 1 & -1\end{bmatrix}.\end{aligned} \tag{35}$$

Proof. This property can be verified computationally with a computer algebra system using the expressions for T_i (given in Proposition 8 of the paper), and for the reduced trifocal tensor (30). A geometric justification based on Proposition A.4 is as follows.

Let us focus on T_1, for notational simplicity. The argument is identical for the other trilinearities. We consider the following simple projection with pinhole $\boldsymbol{z}_1$

$$\boldsymbol{P}_{\boldsymbol{z}_1} = \begin{bmatrix}0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}. \tag{36}$$

Any triple of image points $\boldsymbol{u}, \boldsymbol{u}', \boldsymbol{u}''$ (in distinct images) determines three lines $\boldsymbol{l}, \boldsymbol{l}', \boldsymbol{l}''$ in the image plane of $P_{\boldsymbol{z}_1}$, which contain the points $(1, 1, 1)^T$, $\boldsymbol{c}'_1 = (c'_2, c'_3, c'_4)^T$, $\boldsymbol{c}''_1 = (c''_2, c''_3, c''_4)^T$. These three points are in fact the images of $\boldsymbol{c}, \boldsymbol{c}', \boldsymbol{c}''$ under $P_{\boldsymbol{z}_1}$ (*i.e.*, they are three epipoles). We can parameterize the three bundles through $\boldsymbol{c}_1, \boldsymbol{c}'_1, \boldsymbol{c}''_1$ using the reference frame induced by the three points $P_{\boldsymbol{z}_1}(\boldsymbol{z}_2)$, $P_{\boldsymbol{z}_1}(\boldsymbol{z}_3)$, $P_{\boldsymbol{z}_1}(\boldsymbol{z}_4)$. This way, converging triplets $\boldsymbol{l}, \boldsymbol{l}', \boldsymbol{l}''$ are described by a reduced 2D trifocal tensor.

Finally, the coordinates in $\mathbb{P}^1$ of the lines associated with $\boldsymbol{u}, \boldsymbol{u}', \boldsymbol{u}''$ are given simply by $\boldsymbol{M}_1\boldsymbol{u}$, $\boldsymbol{M}_1\boldsymbol{u}'$, $\boldsymbol{M}_1\boldsymbol{u}''$. This is because $\boldsymbol{M}_1$ is the unique projection matrix that maps $(0, 1, 0)^T$, $(0, 0, 1)^T$, $(1, 1, 1)^T$ (which are images of $\boldsymbol{z}_2, \boldsymbol{z}_3, \boldsymbol{z}_4$ for all three cameras) to the standard basis of $\mathbb{P}^1$.

In conclusion, a triplet $\boldsymbol{u}, \boldsymbol{u}', \boldsymbol{u}''$ satisfies T_1 if and only if the associated lines $\boldsymbol{l}, \boldsymbol{l}', \boldsymbol{l}''$ in the image plane of $P_{\boldsymbol{z}_1}$ converge, and this in turn is equivalent to the fact that $\boldsymbol{M}_1\boldsymbol{u}, \boldsymbol{M}_1\boldsymbol{u}', \boldsymbol{M}_1\boldsymbol{u}''$ are three points in $\mathbb{P}^1$ which satisfy the constraint from the reduced 2D trifocal tensor. □

Expanding (34) for $r = 1, 2, 3$ we obtain the expressions for T_1, T_2, T_3 given in equation (10) of the main part of the paper. For completeness, we include here the explicit expanded form of all the trilinearities:

$$\begin{aligned}
T_1 =& u_2u'_2u''_3(-\hat{c}'_3\hat{c}''_2 + \hat{c}'_4\hat{c}''_2) + u_2u'_3u''_2(\hat{c}'_2\hat{c}''_3 - \hat{c}'_2\hat{c}''_4) +\\
& u_2u'_3u''_3(-\hat{c}'_4\hat{c}''_2 + \hat{c}'_2\hat{c}''_4) + u_3u'_2u''_2(-\hat{c}'_4\hat{c}''_3 + \hat{c}'_3\hat{c}''_4) +\\
& u_3u'_2u''_3(\hat{c}'_3\hat{c}''_2 - \hat{c}'_3\hat{c}''_4) + u_3u'_3u''_2(-\hat{c}'_2\hat{c}''_3 + \hat{c}'_4\hat{c}''_3),\\[1em]
T_2 =& u_1u'_1u''_3(-\hat{c}'_3\hat{c}''_1 + \hat{c}'_4\hat{c}''_1) + u_1u'_3u''_1(\hat{c}'_1\hat{c}''_3 - \hat{c}'_1\hat{c}''_4) +\\
& u_1u'_3u''_3(-\hat{c}'_4\hat{c}''_1 + \hat{c}'_1\hat{c}''_4) + u_3u'_1u''_1(-\hat{c}'_4\hat{c}''_3 + \hat{c}'_3\hat{c}''_4) +\\
& u_3u'_1u''_3(\hat{c}'_3\hat{c}''_1 - \hat{c}'_3\hat{c}''_4) + u_3u'_3u''_1(-\hat{c}'_1\hat{c}''_3 + \hat{c}'_4\hat{c}''_3),\\[1em]
T_3 =& u_1u'_1u''_2(-\hat{c}'_2\hat{c}''_1 + \hat{c}'_4\hat{c}''_1) + u_1u'_2u''_1(\hat{c}'_1\hat{c}''_2 - \hat{c}'_1\hat{c}''_4) +\\
& u_1u'_2u''_2(-\hat{c}'_4\hat{c}''_1 + \hat{c}'_1\hat{c}''_4) + u_2u'_1u''_1(-\hat{c}'_4\hat{c}''_2 + \hat{c}'_2\hat{c}''_4) +\\
& u_2u'_1u''_2(\hat{c}'_2\hat{c}''_1 - \hat{c}'_2\hat{c}''_4) + u_2u'_2u''_1(-\hat{c}'_1\hat{c}''_2 + \hat{c}'_4\hat{c}''_2),\\[1em]
T_4 =& u_1u'_1u''_2(-\hat{c}'_2\hat{c}''_1 + \hat{c}'_3\hat{c}''_1) + u_1u'_1u''_3(\hat{c}'_2\hat{c}''_1 - \hat{c}'_3\hat{c}''_1) +\\
& u_1u'_2u''_1(\hat{c}'_1\hat{c}''_2 - \hat{c}'_1\hat{c}''_3) + u_1u'_2u''_2(-\hat{c}'_3\hat{c}''_1 + \hat{c}'_1\hat{c}''_3) +\\
& u_1u'_2u''_3(\hat{c}'_3\hat{c}''_1 - \hat{c}'_1\hat{c}''_2) + u_1u'_3u''_1(-\hat{c}'_1\hat{c}''_2 + \hat{c}'_1\hat{c}''_3) +\\
& u_1u'_3u''_2(\hat{c}'_2\hat{c}''_1 - \hat{c}'_1\hat{c}''_3) + u_1u'_3u''_3(-\hat{c}'_2\hat{c}''_1 + \hat{c}'_1\hat{c}''_2) +\\
& u_2u'_1u''_1(-\hat{c}'_3\hat{c}''_2 + \hat{c}'_2\hat{c}''_3) + u_2u'_1u''_2(\hat{c}'_2\hat{c}''_1 - \hat{c}'_2\hat{c}''_3) +\\
& u_2u'_1u''_3(-\hat{c}'_2\hat{c}''_1 + \hat{c}'_3\hat{c}''_2) + u_2u'_2u''_1(-\hat{c}'_1\hat{c}''_2 + \hat{c}'_3\hat{c}''_2) +\\
& u_2u'_2u''_3(\hat{c}'_1\hat{c}''_2 - \hat{c}'_3\hat{c}''_2) + u_2u'_3u''_1(\hat{c}'_1\hat{c}''_2 - \hat{c}'_2\hat{c}''_3) +\\
& u_2u'_3u''_2(-\hat{c}'_2\hat{c}''_1 + \hat{c}'_2\hat{c}''_3) + u_2u'_3u''_3(\hat{c}'_2\hat{c}''_1 - \hat{c}'_1\hat{c}''_2) +\\
& u_3u'_1u''_1(\hat{c}'_3\hat{c}''_2 - \hat{c}'_2\hat{c}''_3) + u_3u'_1u''_2(-\hat{c}'_3\hat{c}''_1 + \hat{c}'_2\hat{c}''_3) +\\
& u_3u'_1u''_3(\hat{c}'_3\hat{c}''_1 - \hat{c}'_3\hat{c}''_2) + u_3u'_2u''_1(-\hat{c}'_3\hat{c}''_2 + \hat{c}'_1\hat{c}''_3) +\\
& u_3u'_2u''_2(\hat{c}'_3\hat{c}''_1 - \hat{c}'_1\hat{c}''_3) + u_3u'_2u''_3(-\hat{c}'_3\hat{c}''_1 + \hat{c}'_3\hat{c}''_2) +\\
& u_3u'_3u''_1(-\hat{c}'_1\hat{c}''_3 + \hat{c}'_2\hat{c}''_3) + u_3u'_3u''_2(\hat{c}'_1\hat{c}''_3 - \hat{c}'_2\hat{c}''_3).
\end{aligned} \tag{37}$$

We can also use Proposition A.5 to prove theoretical properties of all four trilinearities. The following statement is equivalent to Proposition 4.3 in the main part of the paper.

Proposition A.6. *The internal constraints of each trilinearity T_r are* linear. *More precisely, the coefficients in $\mathbb{R}^{27}$ that are entries of T_i for some choice of $\boldsymbol{c}', \boldsymbol{c}''$ form a* linear space*, of dimension*

five. Moreover, the coefficients of T_r characterize $\boldsymbol{c}'_r$ and $\boldsymbol{c}''_r$ up to a two-fold ambiguity.

Proof. It follows from (34) that the coefficients of T_r are images of the coefficients of $\mathcal{T}_r^{ijk}$ under an injective linear map (note that this map is completely described by the four matrices (35)). Since the entries of a reduced 2D trifocal are only constrained by three linear conditions, the coefficients of T_r form linear spaces of dimension five. Moreover, the knowledge of the T_r is equivalent to that of $\mathcal{T}_r^{ijk}$, which means that $\boldsymbol{c}'_r$ and $\boldsymbol{c}''_r$ are determined up to a two-fold ambiguity. □

Finally, we provide some details on the proof of Proposition 4.2 in the paper.

Proposition 4.2 *A vector $\boldsymbol{d} = (d_{ij})$ in $\mathbb{R}^{12}$ with no zero entries can be written as $d_{ij} = a_i b_j$ for some vectors $\boldsymbol{a} = (a_1, a_2, a_3, a_4)^T, \boldsymbol{b} = (b_1, b_2, b_3, b_4)^T$ in $\mathbb{R}^3$ if and only if $d_{ij}d_{kl} = d_{il}d_{kj}$ holds for all permutations (i, j, k, l) of $(1, 2, 3, 4)$.*

Proof. A factorization of $\boldsymbol{d} = (d_{ij})$ exists if and only if there is a rank-1 completion of

$$\begin{pmatrix} * & d_{12} & d_{13} & d_{14} \\ d_{21} & * & d_{23} & d_{24} \\ d_{31} & d_{32} & * & d_{34} \\ d_{41} & d_{42} & d_{43} & * \end{pmatrix}. \tag{38}$$

Indeed, a completion corresponds to $\boldsymbol{a}^T\boldsymbol{b} \in \mathbb{R}^{4\times 4}$. It is clear that the constraints $d_{ij}d_{kl} = d_{il}d_{kj}$, which correspond to certain 2×2 minors, are necessary. Conversely, it is always possible to deduce the diagonal elements, for example:

$$d_{11} = \frac{d_{13}d_{21}}{d_{23}} = \frac{d_{14}d_{21}}{d_{24}} = \frac{d_{12}d_{31}}{d_{32}} = \frac{d_{14}d_{31}}{d_{34}}. \tag{39}$$

The constraints $d_{ij}d_{kl} = d_{il}d_{kj}$ guarantee equality among all these expressions. Indeed, we have that for example $d_{12}d_{23}d_{31} = d_{21}d_{32}d_{13}$ because

$$\frac{d_{12}}{d_{13}}\frac{d_{23}}{d_{21}}\frac{d_{31}}{d_{32}} = \frac{d_{42}}{d_{43}}\frac{d_{43}}{d_{41}}\frac{d_{31}}{d_{32}} = \frac{d_{42}d_{31}}{d_{43}d_{32}} = 1. \tag{40}$$

This shows that if the quadratic constraints hold, we can solve for the diagonal elements in (38), and the resulting matrix will have rank one. □

B. Synthetic data experiment

For completeness, we show in Figure 9 the *mean* value of the mean reprojection and reconstruction errors for the synthetic data used in Section 5.2 of our submission. Recall that the quality of the reconstruction was evaluated in that section by measuring how well it predicts the reprojection of the remaining points in the dataset as well as their 3D reconstruction, once again registered to the ground truth through a homography. We have used the same data as in Section 5.2 to construct the curves shown in Figure 9. They show the mean values of the mean reprojection and reconstruction errors, given respectively in pixel and mm, for 40 random choices of the 7 point correspondences and different values of the standard deviation σ (in pixel units) of Gaussian noise added to the image coordinates. As noted in our submission, Figure 9 shows that both the linear trifocal tensor estimation are occasionally thrown completely off course for "bad" choices of the 7 correspondences, without a clear winner in this case.

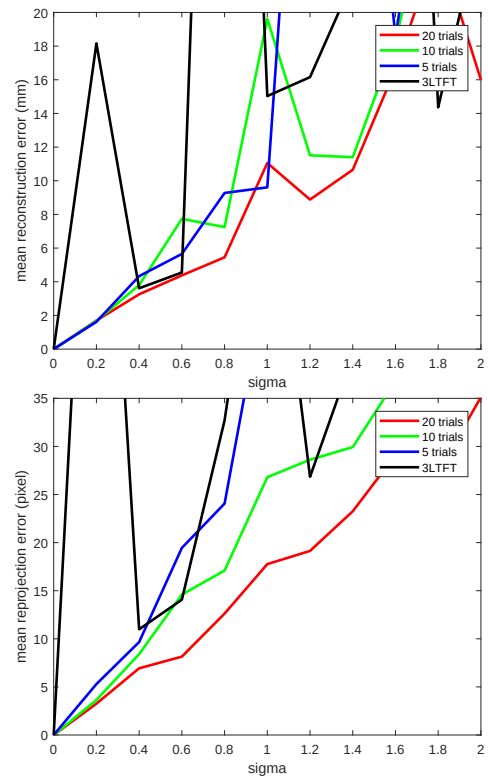

Figure 9. Experiments with synthetic data using the mean of the mean reprojection and reconstruction errors instead of their median. Compare to Figure 8 in our submission.