Dual Distributions of Multilinear Geometric Entities

Sami S. Brandt[1,2]
[1]University of Oulu, Machine Vision Group
P.O. Box 4500, 90014 University of Oulu, FINLAND
[2]Malmö University, Applied Mathematics Group,
Östra Varvsgatan 11A, 20506 Malmö, SWEDEN

Abstract

In this paper, we propose how the parameter distributions of multilinear geometric entities can be dualised. The dualisation concern, for example, the parameter distributions of conics, multiple view tensors, homographies, or as simple entities as points, lines, and planes. The dual distributions are related to Triggs' joint feature distributions but our approach is different in certain fundamental aspects. Our starting point is in the assumption that the maximum likelihood estimate, or the corresponding robust estimate, and the covariance matrix of the parameters of the geometric entity are available. We then use the asymptotic normality property of the MLE which allows us to transform the parameter uncertainty distribution in a dual form. The dualisation of the parameter distribution allows us, for instance, to look at the uncertainty distributions in feature distributions, which are essentially tied to the distribution of training data, and helps us to derive conditional distributions for point or line transfer and characterise confidence intervals of the estimates. Applications of the proposed approach are thus uncertainty analysis, statistical prediction, probabilistic transfer, etc.

1. Introduction

An essential part of geometric computer vision are multilinear models which include e.g. homographies, multiple view tensors, quadric surfaces, as well as the simple entities of points, lines and planes. There are numerous works related to the estimation of these kinds of multilinear relations. When a statistical approach is selected for the estimation, one should be able to compute the geometric model parameters along with their uncertainty distribution. This paper considers how these parameter distributions of multi-linear geometric entities can be dualised. The simplest form of this dualisation is the transformation of the line-probability-density into a point-probability-density as

proposed in [1]. This paper generalises that approach for general multilinear models.

By the way of an example, consider fitting a conic section to a set of points using maximum likelihood estimation. We would be interested in the confidence intervals of the conic, but the normal distribution assumption can be made only for the MLE in the parameter space. However, we will show that by the dualisation of the parameter distribution we will obtain a distribution of points that can be used to plot the selected confidence intervals for the estimated conic section. As the second example, let us consider the trifocal point transfer. Given a point match in two views, it would be interesting to compute the conditional position distribution of the transferred point in the third view if we had the uncertainty information of the trifocal tensor available. In fact, by dualising the trifocal tensor parameter distribution, an exact form for this conditional distribution can be computed, as will be shown later in this paper.

Our approach is closely related to the branch of *integral geometry* in mathematics. There are two main schools of integral geometry of which the traditional is that of Santaló and Blaschke [7]. The classical example is that the length of a plane curve is the probability of random lines intersecting it. The more recent meaning is the school of Gelfand (see e.g. [3]). It studies integral transforms, modelled with the Radon transform, that relates the underlying geometrical incidence relations by incidence graphs. Our approach seems to be somewhat in between these two schools as we compute Radon and Radon like transforms for the probability densities in such a way that the probability measure is preserved. The dualisation is constructed from the fact that the probability of an element (e.g. a point) is the total probability of all the geometric entities (e.g. planes) that coincide with the element. Then by integrating the distribution of the entity over the affine subspaces corresponding to the selected incidence relation, Radon like transforms follow.

As the principal assumption we use the normal distribution assumption for the multilinear model parameters. This is reasonable due to the asymptotic normality property of

the maximum likelihood estimator, i.e., the fact that, with certain general regularity conditions, the distribution of the maximum likelihood parameter estimator converges in distribution to the normal distribution with the (pseudo)inverse of the Fisher information as the parameter covariance matrix [5]. This makes our approach fundamentally different from [9] where an algebraic linear system and Gaussian approximation in the feature space were used. In contrast to [9], the dual distributions considered here have analytic forms and are exact with the assumptions above.

This paper is organised as follows. In Section 2, we introduce the multilinear models considered in this paper. In Section 3, we derive the dual distributions by first assuming a single constraint equation and then generalise the approach for multiple constraint equations. In Section 4, we show how interesting conditional distributions can be extracted. In Section 5, we compute confidence intervals for conics and compute the point transfer density from two views into the third view. Conclusions are in Section 6.

2. Multilinear Model Definition

In this paper we assume a multilinear model

$$f(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n; \theta) = \mathbf{0}, \tag{1}$$

where θ is the parameter vector. The multilinearity means that the function f is separately linear in each of its $n+1$ arguments; we also allow that any of the n first arguments may be repeated so that, e.g., quadratic forms are included. With a fixed θ, the multilinear model defines a multilinear system of which each equation is equivalent to a linear subspace, with codimension one, in the space of the *joint feature vector*

$$\mathbf{y} \hat{=} \mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \ldots \otimes \mathbf{x}_n, \tag{2}$$

where $\hat{=}$ denotes the correspondence between the expressions, $\otimes$ means the tensor product, and $\mathbf{x}_i \in \mathbb{P}^m$, $i = 1, 2, \ldots, n$. The vector $\mathbf{y}$ contains the elements of the tensor product where the multifold elements have been dropped (see Section 5.1 for an example). In $\mathbb{P}^2$, for instance, the interpretation of $\mathbf{x}_i$ can be either a point or a line depending on the multilinear relation selected.

Without a loss of generality, we assume that each equation l of the multilinear system is in the form of

$$\theta^{\mathrm{T}} \mathbf{T}_l \mathbf{y} \equiv \theta^{\mathrm{T}} \mathbf{y}_l = 0, \tag{3}$$

where $\mathbf{T}_l$ is a matrix defined by the multilinear relation. For instance, the well known point and line incidence relations [4, 2] characterising multiple projective views of a scene can be written in this form.

3. Parameter Distributions and Their Duals

In this section, we dualise the parameter uncertainty distributions starting from the models defined by a single constraint equation and generalise the point–line case [1] to the
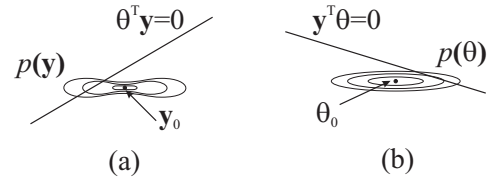

Figure 1. Illustration of the duality. (a) The space of $\mathbf{y}$ is illustrated as the 2D plane. The parameter θ here represents the line $\theta^{\mathrm{T}} \mathbf{y} = 0$. (b) The dual space or the space of θ. In the dual space, $\mathbf{y}$ represents the line $\mathbf{y}^{\mathrm{T}} \theta = 0$. The dual distribution $p(\mathbf{y})$ is constructed from $p(\theta)$ by making Radon like integral transform to it so that the probability measure is preserved.

point–hyperplane duality (Section 3.1). The general case of multiple constraint equations is considered in Section 3.2.

3.1. Single Constraint Equation

Let θ be a parameter vector that defines the multilinear model by a single constraint equation

$$\theta^{\mathrm{T}} \mathbf{y} = 0, \tag{4}$$

with the joint feature vector $\mathbf{y} \in \mathbb{P}^{N-1}$. Let us assume that we have the maximum likelihood estimate θ_0 for the parameter vector and its the covariance matrix $\mathbf{C}_\theta$ available, where θ_0 is constrained to lie on the unit hypersphere $\mathbb{S}^{N-1}$ in $\mathbb{R}^N$. Using the asymptotic normality property of the MLE, we assume that $\theta \sim N(\theta_0, \mathbf{C}_\theta)$. Since $\theta_0 \in \mathrm{Ker}\{\mathbf{C}_\theta\}$, θ has the density function

$$p(\theta) \propto \exp\left(-\frac{1}{2}\theta^{\mathrm{T}} \mathbf{C}_\theta^\dagger \theta\right), \tag{5}$$

where $\cdot^\dagger$ is the Moore–Penrose pseudoinverse. That is, we use the tangential, normal approximation for the variation of θ on the unit hypersphere at θ_0.

Our intention is to compute the dual of $p(\theta)$. The duality between the parameter vector θ and the feature vector $\mathbf{y}$ is illustrated in Fig. 1. As (4) represents the hyperplane θ in the space of $\mathbf{y}$, it can also be seen as the hyperplane $\mathbf{y}$ in the dual space or space of θ. Now, as the dual distribution of θ, which is a distribution for $\mathbf{y}$, we will identify *the total probability (density) of the all the models θ lying on the hyperplane $\mathbf{y}$ in the dual space*. To construct the dual pdf $p(\mathbf{y})$, we first make the whitening transformation for variables. Then we compute the Radon transform, i.e., transform the whitened dual domain by computing the integrals of $p(\theta)$ over all the hyperplanes in the dual space. The subsets of these integrals corresponding to parallel hyperplanes form the conditional probability density (marginal density) conditioned on a fixed normal direction of the hyperplane. Multiplying this conditional density with the pdf of normal directions, which is the uniform density on the half of the
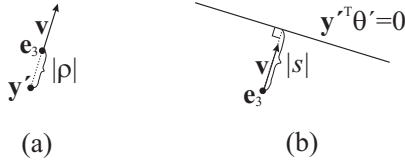
Figure 2. The hyperplane parametrisation when $M = 3$, thus, the hyperplane $\mathbf{y}'$ is a line and the tangent affine subspace π' is the real plane. (a) Reduced joint feature space, where $\mathbf{y}'$, represented by the homogeneous vector $(\rho\mathbf{v}^T\ 1)^T$, is a point; (b) the corresponding dual space, where $\mathbf{y}'$ is the line.

unit hypersphere with the whitened Gaussian model, we obtain a valid probability density which can be interpreted in the parameter space of $\mathbf{y}$. The details are below.

Let us make the whitening transform from the eigenvalue decomposition of $\mathbf{C}_\theta$, which has been constructed so that θ_0 corresponds to the last eigenvector. We may write

$$
\begin{aligned}
\theta^T\mathbf{C}_\theta^\dagger\theta &= \theta^T\mathbf{U}\mathbf{\Lambda}^\dagger\mathbf{U}^T\theta \\
&= \underbrace{\theta^T\tilde{\mathbf{U}}\begin{pmatrix}\tilde{\mathbf{\Lambda}}^{-1/2} & \mathbf{0} \\ \mathbf{0}^T & 1\end{pmatrix}}_{\triangleq\theta'^T}\underbrace{\begin{pmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{0}^T & 0\end{pmatrix}}_{\triangleq\tilde{\mathbf{I}}}\underbrace{\begin{pmatrix}\tilde{\mathbf{\Lambda}}^{-1/2} & \mathbf{0} \\ \mathbf{0}^T & 1\end{pmatrix}\tilde{\mathbf{U}}^T\theta}_{\triangleq\theta'} \\
&= \theta'^T\tilde{\mathbf{I}}\theta',
\end{aligned}
\tag{6}
$$

where the diagonal matrix $\tilde{\mathbf{\Lambda}}$ contains the $M-1$ non-zero eigenvalues of $\mathbf{C}_\theta$, sorted in the descending order, and $\tilde{\mathbf{U}}$ contains the corresponding eigenvectors and the eigenvector representing θ_0. Now, $\theta' \sim N(\mathbf{e}_M, \tilde{\mathbf{I}})$, where $\mathbf{e}_M$ is the standard basis vector $(0, \ldots, 0, 1) \in \mathbb{R}^M$. Furthermore, for all the feature vectors that are consistent with the model θ, lying on the tangent space, we may write

$$
0 = \theta^T\mathbf{y} = \underbrace{\theta^T\tilde{\mathbf{U}}\begin{pmatrix}\tilde{\mathbf{\Lambda}}^{-1/2} & \mathbf{0} \\ \mathbf{0}^T & 1\end{pmatrix}}_{\theta'^T}\underbrace{\begin{pmatrix}\tilde{\mathbf{\Lambda}}^{1/2} & \mathbf{0} \\ \mathbf{0}^T & 1\end{pmatrix}\tilde{\mathbf{U}}^T\mathbf{y}}_{\triangleq\mathbf{y}'} = \theta'^T\mathbf{y}',
\tag{7}
$$

where $\mathbf{y}' \in \mathbb{P}^{M-1}$ represents a *reduced joint feature vector*. For simplicity, we now investigate the reduced, transformed model θ'.

On the basis of the construction above, the variation of θ' occurs only in the $M-1$ dimensional affine subspace π' perpendicular to $\mathbf{e}_M$ in the dual space. Since every point on this tangent hyperplane can be identified with a unique one-dimensional linear subspace of $\mathbb{R}^M$, we may regard the tangent hyperplane as a projective space $\mathbb{P}^{M-1}$. Moreover, the points on the tangent plane can be considered to be already in the homogeneous form. Hence, in the following we assume that $\theta' \in \mathbb{P}^{M-1}$.

In the dual space, the $M-2$ dimensional hyperplanes $\mathbf{y}'^T\theta' = 0$ embedded in the $M-1$ dimensional affine sub-

space $\pi' \triangleq \mathbb{R}^{M-1}$, may be parameterised by the signed distance s from the origin and by the unit vector $\mathbf{v} \in \mathbb{S}^{M-2}$, as Fig 2(b) illustrates. The origin can be represented by $\mathbf{e}_M \in \mathbb{P}^{M-1}$ and we assume that $\mathbf{v} = \mathbf{v}(\phi)$ where ϕ parameterises the normal direction of the $M-2$ dimensional hyperplane $\mathbf{y}'$ in $\mathbb{R}^{M-1}$. We choose the sign of s to be equal to sign of the intercept of the hyperplane in the dual space, hence,

$$
s = -\frac{y_N'}{\text{sign}\,(y_{N-1}')\sqrt{\mathbf{y}'^T\mathbf{y}' - y_N'^2}}.
\tag{8}
$$

On the other hand, in the reduced joint feature space, the homogeneous vector $\mathbf{y}'$ can be parameterised by the signed distance ρ from the origin $\mathbf{e}_M$ and the direction $\mathbf{v}$ (Fig 2(a)). The sign of ρ is identified as the last variable sign of the inhomogeneous representation of $\mathbf{y}'$. Then we have

$$
s = -\frac{1}{\rho}.
\tag{9}
$$

As we assume that a $\theta' \sim N(\mathbf{e}_M, \tilde{\mathbf{I}})$, the probability of the hyper plane $\mathbf{y}_s'^T\theta' = 0$, conditioned on the direction $\mathbf{v}(\phi)$ in the dual space, is simply the marginal probability of the Gaussian over the hyperplane, i.e.,

$$
p(s|\phi) = \int_{\mathbf{y}_s'^T\theta'=0} p_G(\theta'; \mathbf{e}_M, \tilde{\mathbf{I}})dS = p(s)
\tag{10}
$$

which is a mean zero, 1-D Gaussian with unity variance [1], where dS denotes the volume differential in the hyperplane.

By using the modified spherical coordinates $\mathbf{y}' = \mathbf{y}'(\rho, \phi)$ in the reduced joint feature space (see Appendix A) and the fact that $s = s(\rho)$, we get

$$
p(\rho|\phi) = \left|\frac{\partial s}{\partial\rho}\right| p\left(s(\rho)|\phi\right) = \frac{1}{\sqrt{2\pi}\rho^2}\exp\left(-\frac{1}{2}\rho^{-2}\right).
\tag{11}
$$

On the other hand, since we have an isotropic Gaussian distribution, the distribution of normal directions is uniform over a half of the unit sphere. By using the modified spherical coordinates, and marginalising the $M-1$-dimensional Gaussian over the signed radial parameter, we obtain

$$
p(\phi) = \frac{\Gamma(M/2 - 1/2)}{\pi^{(M-1)/2}}\prod_{i=1}^{M-3}\sin^{M-2-i}(\phi_i).
\tag{12}
$$

Hence,

$$
\begin{aligned}
p(\rho, \phi) &= p(\rho|\phi)p(\phi) \\
&= \frac{\Gamma(M/2 - 1/2)e^{-\frac{1}{2}\rho^{-2}}\prod_{i=1}^{M-3}\sin^{M-2-i}(\phi_i)}{\sqrt{2\pi^M}\rho^2}.
\end{aligned}
\tag{13}
$$

This is an analytic form for the probability density of the reduced joint feature vector $\mathbf{y}' = \mathbf{y}'(\rho, \phi)$ assuming the model (4) and the Gaussian distribution $p(\theta)$ for the parameter vector θ.

3.2. Multiple Constraint Equations

Now we generalise the computations above for multiple constraint equations. Let θ be a parameter vector of a multilinear model. We define that the set vectors $\mathbf{y}_l$, $l = 1, \ldots, L$ are consistent with the model if and only if

$$\theta^{\mathrm{T}} \mathbf{y}_l = 0, \quad l = 1, \ldots, L. \tag{14}$$

Hence, $(\mathbf{y}_1 \ \ldots \ \mathbf{y}_L)^{\mathrm{T}} \theta = \mathbf{Y}^{\mathrm{T}} \theta = \mathbf{0}$, i.e., $\theta \in \mathrm{Ker}\,\{\mathbf{Y}^{\mathrm{T}}\}$ (c.f. the nullspace representation of lines in $\mathbb{R}^3$). Similarly as in (7), we define the transformed model θ' and the reduced coefficient matrix $\mathbf{Y}'$ so that

$$\mathbf{0} = \theta^{\mathrm{T}} \mathbf{Y} = \underbrace{\theta^{\mathrm{T}} \tilde{\mathbf{U}} \begin{pmatrix} \tilde{\mathbf{\Lambda}}^{-1/2} & \mathbf{0} \\ \mathbf{0}^{\mathrm{T}} & 1 \end{pmatrix}}_{\theta'^{\mathrm{T}}} \underbrace{\begin{pmatrix} \tilde{\mathbf{\Lambda}}^{1/2} & \mathbf{0} \\ \mathbf{0}^{\mathrm{T}} & 1 \end{pmatrix} \tilde{\mathbf{U}}^{\mathrm{T}} \mathbf{Y}}_{\triangleq \mathbf{Y}'} \tag{15}$$

$$= \theta'^{\mathrm{T}} \mathbf{Y}'.$$

Let W be the affine subspace corresponding to the set $\{\theta' \in \mathbb{P}^{M-1} | \mathbf{Y}'^{\mathrm{T}} \theta' = 0\}$ or $W = \{\tilde{\theta}' \in \mathbb{R}^{M-1} | \mathbf{A} \tilde{\theta}' = \mathbf{b}\}$, where $\theta' \hat{=} (\tilde{\theta}', 1)$ and $\mathbf{Y}'^{\mathrm{T}} = (\mathbf{A} \ -\mathbf{b})$. Assuming that the W contains finite points, it is easy to show that the nearest point to the origin $\mathbf{e}_M$ is

$$\theta'_{\min} \triangleq \begin{pmatrix} \mathbf{w} \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{A}^{\dagger} \mathbf{b} \\ 1 \end{pmatrix}. \tag{16}$$

In other words, the affine subspace $W = \mathbf{w} + V$ where V denotes the linear subspace parallel to W and $\dim(V) \equiv K$.

We are interested in the probability density of $\mathrm{Ker}\,\{\mathbf{Y}'^{\mathrm{T}}\}$, that is, *the total probability (density) of the affine subspace W as it represents the set of all the models θ' that are consistent with $\mathbf{Y}'$.* To construct this dual density, instead of integrating over all hyperplanes as in the previous subsection, we must generalise the Radon transform by integrating $p(\theta')$ over all the K-dimensional affine subspaces. To perform this kind of integral transform and to interpret its result as probability densities, we must create a unique parameterisation for affine subspaces. For projective subspaces, a unique parameterisation could be constructed with the Grasmannian tensor, which can be seen as a generalisation of Plücker line coordinates [8, 9]. Nevertheless, not all the tensors of the same size and shape are Grasmannians since the Grasmannian must additionally satisfy the Grasmannian simplicity relations.

For affine subspaces, however, things are slightly easier in the construction of unique parameterisation since we may represent an affine subspace by the vector $\mathbf{w}$ and parameters of the parallel linear subspace V. To parameterise V, we use the fact that the related *orthogonal projection matrix*

$$\mathbf{P} = \mathbf{I} - \mathbf{A}^{\dagger} \mathbf{A}, \tag{17}$$

projecting onto the subspace is unique. It is well known that a matrix represents an orthogonal projection onto a linear subspace if and only if it is *idempotent*, $\mathbf{P}^2 = \mathbf{P}$, and *self-adjoint*, $\mathbf{P}^{\mathrm{T}} = \mathbf{P}$. Now, to uniquely parameterise the matrix $\mathbf{P}$ we need the following lemma.

Lemma 3.1 *Let $\mathbf{P}$ be an idempotent and symmetric $(M-1) \times (M-1)$ matrix. If $\mathbf{p}_i = \mathbf{P} \mathbf{e}_i$, $i = 1, \ldots, K$ are linearly independent and span the range of $\mathbf{P}$, where $K \leq \frac{M}{2}$, is the dimension of the range, then there is a unique lower triangular $K \times (M-1)$ matrix $\mathbf{L}$ with orthonormal columns and strictly positive diagonal so that $\mathbf{P} = \mathbf{L}\mathbf{L}^{\mathrm{T}}$.*

The proof is in Appendix B.

The lemma suggests that we may parameterise the orthogonal projection matrix $\mathbf{P}$ by parameterising the elements of $\mathbf{L} = \mathbf{L}(\Phi)$ when we may define

$$\mathbf{P}(\Phi) = \begin{cases} \mathbf{L}(\Phi)\mathbf{L}(\Phi)^{\mathrm{T}}, & \text{if } K \leq M/2 \\ \mathbf{I} - \mathbf{L}(\Phi)\mathbf{L}(\Phi)^{\mathrm{T}}, & \text{otherwise.} \end{cases} \tag{18}$$

The matrix $\mathbf{L}$ can be parameterised by the modified spherical coordinates (see Appendix A) with the radial parameter equal to unity. The first column vector is on the half of the unit sphere $\mathbb{S}^{M-2}$, the second column vector is orthogonal to the first and it has one element less and is hence on $\mathbb{S}^{M-4}$. To parameterise $\mathbb{S}^{M-4}$, we form an orthogonal basis in the orthogonal complement of the first column vector, by creating the orthogonal projection matrix that projects onto the orthogonal complement, and employ the Gram–Schmidt orthonormalisation procedure, as in (34), and form the unit sphere in the subspace spanned by the orthonormal vectors. The other columns $l \leq \tilde{K}$ can be similarly parameterised on the half of the unit sphere $\mathbb{S}^{M-2l}$, whereas $\tilde{K} = K$, if $K \leq M/2$, and $\tilde{K} = M - K - 1$, otherwise. Let Φ_k denote the vector of modified spherical coordinates of the column k in $\mathbf{L}$. We collect the parameters in the vector $\Phi = (\Phi_1, \Phi_2, \ldots, \Phi_{\tilde{K}})$. In total, there are $K' = (M-1)\tilde{K} - \tilde{K}^2$ free parameters in $\mathbf{L}$.

Now, we are ready to construct the probability density for the affine subspace W. We create an orthogonal basis $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{M-1}$ for the subspace V and its orthogonal complement $V^{\perp}$ using the Gram-Schmidt orthonormalisation procedure for the projections $\mathbf{p}_i = \mathbf{P} \mathbf{e}_i$, $i = 1, 2, \ldots, K$ onto the subspace V as in (34) and similarly for the projections $\mathbf{p}_i = (\mathbf{I} - \mathbf{P}) \mathbf{e}_i$, $i = K+1, K+2, \ldots, M-1$ on its orthogonal complement $V^{\perp}$. We marginalise the $M-1$ dimensional Gaussian over all the parallel affine subspaces of V, i.e., those which are of the form $W = \mathbf{w} + V$. We obtain the conditional probability density of $\mathbf{w} \hat{=} \mathbf{s} \in \mathbb{R}^{M-K-1}$

$$p(\mathbf{s}|\Phi) = \int_V p_{\mathrm{G}}(\theta'; \mathbf{e}_M, \tilde{\mathbf{I}}) \mathrm{d}S \tag{19}$$
$$= p(\mathbf{s}),$$

which is a mean zero, $M - K - 1$ dimensional Gaussian with the identity matrix as the covariance matrix.

As to the distribution of directions Φ, isotropic Gaussians imply uniform directions on half hyperspheres. We decompose the vector as $\Phi = (\Phi_1, \Phi_2, \ldots, \Phi_{\tilde{K}})$ and further $\Phi_k = (\phi_1^k, \phi_2^k, \ldots, \phi_{M-2k}^k)$. Thus we have

$$
\begin{aligned}
p(\Phi) &= p(\Phi_1, \Phi_2, \ldots, \Phi_{\tilde{K}}) \\
&= p(\Phi_1) p(\Phi_2|\Phi_1) p(\Phi_3|\Phi_1, \Phi_2) \cdots \\
&\quad p(\Phi_{\tilde{K}}|\Phi_1, \Phi_2, \ldots, \Phi_{\tilde{K}-1}),
\end{aligned} \tag{20}
$$

where

$$
\begin{aligned}
&p(\Phi_k|\Phi_1, \Phi_2, \ldots, \Phi_{k-1}) \\
&= \frac{\Gamma\left((M - 2k + 1)/2\right)}{\pi^{(M-2k+1)/2}} \prod_{i=1}^{M-2k-1} \sin^{M-2k-i}(\phi_i^k),
\end{aligned} \tag{21}
$$

$k = 1, 2, \ldots, \tilde{K}$.

So finally,

$$
\begin{aligned}
p(W) &= p(\mathbf{s}, \Phi) = p(\mathbf{s}|\Phi)p(\Phi) \\
&= \frac{\mathrm{e}^{-\frac{1}{2}\mathbf{s}^{\mathrm{T}}\mathbf{s}}}{2^{(M-K-1)/2}\pi^{(M-K+\tilde{K}M-\tilde{K}^2-1)/2}} \\
&\quad \times \prod_{k=1}^{\tilde{K}} \Gamma\left((M - 2k + 1)/2\right) \prod_{i=1}^{M-2k-1} \sin^{M-2k-i}(\phi_i^k).
\end{aligned} \tag{22}
$$

This is an analytic form for the probability density of the affine subspace $W = W(\mathbf{s}, \Phi)$ or, equivalently, the probability density of the left nullspace of the reduced coefficient matrix $\mathbf{Y}'$, assuming the model (14) and Gaussian distributed parameter vector θ.

4. Mappings to Feature Distributions

In the previous section, we derived the general form for the dual distributions in the function of the parameters of the corresponding affine subspace. Now we discuss how we can extract interesting feature distributions from the dual distributions. The interesting feature distribution often has less parameters than the affine subspace or one may be interested only in certain conditional feature distributions, conditioned on some fixed a subset of the features. In these cases the mapping from the feature distribution to the affine subspace W is not necessarily one-to-one and onto. However, we may form the conditional distribution

$$
p(\mathbf{s}, \Phi|R) = \frac{p(\mathbf{s}, \Phi)}{\iint_R p(\mathbf{s}, \Phi)\mathrm{d}\mathbf{s}\mathrm{d}\Phi} \propto p(\mathbf{s}, \Phi). \tag{23}
$$

conditioned on the restriction $R \subset \mathbb{R}^{K'}$ of W, where R is parameterised by the interesting part $\tilde{\mathbf{x}} \in \mathbb{R}^{\tilde{N}}$ of the feature distribution. Let us thus define $\varphi : \mathbb{R}^{\tilde{N}} \to R$ so that

$(\mathbf{s}, \Phi) = \varphi(\tilde{\mathbf{x}})$, where φ is continuous and invertible almost everywhere.

Now, we intend to transform the conditional density (23) into the probability density of $\tilde{\mathbf{x}}$. To find the right magnification factor, corresponding to the absolute value of the Jacobian determinant in the well known substitution rule for integrals, we first make the local linearisation of φ around $\tilde{\mathbf{x}}_0$. That is,

$$
\begin{aligned}
\varphi(\tilde{\mathbf{x}}) &\approx \varphi(\tilde{\mathbf{x}}_0) + \left.\frac{\partial \varphi}{\partial \tilde{\mathbf{x}}}\right|_{\tilde{\mathbf{x}}=\tilde{\mathbf{x}}_0} (\tilde{\mathbf{x}} - \tilde{\mathbf{x}}_0) \\
&= \varphi_0 + \mathbf{J}_0(\tilde{\mathbf{x}} - \tilde{\mathbf{x}}_0).
\end{aligned} \tag{24}
$$

Given a local orthonormal basis $B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{\tilde{N}}\}$ of the tangent space of R at φ_0, we may write

$$
\varphi - \varphi_0 = \left(\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_{\tilde{N}}\right)^{\mathrm{T}} \left(\xi_1\ \xi_2\ \cdots\ \xi_{\tilde{N}}\right)^{\mathrm{T}}. \tag{25}
$$

The Jacobian of the mapping $\tilde{\mathbf{x}} \mapsto \xi$ is the $\tilde{N} \times \tilde{N}$ matrix

$$
\mathbf{J}_\xi = \frac{\partial \xi}{\partial \tilde{x}} = \frac{\partial \xi}{\partial \varphi} \mathbf{J}_0 = \left(\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_{\tilde{N}}\right)^{\mathrm{T}} \mathbf{J}_0, \tag{26}
$$

where we have used the local property $\xi_i = \mathbf{u}_i^{\mathrm{T}}(\varphi - \varphi_0)$, $i = 1, \ldots, \tilde{N}$. The local magnification factor at $\tilde{\mathbf{x}}_0$ now takes the form

$$
\begin{aligned}
|\det \mathbf{J}_\xi| &= \sqrt{|\det \mathbf{J}_\xi|^2} \\
&= \sqrt{\det(\mathbf{J}_0^{\mathrm{T}}\left(\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_{\tilde{N}}\right)\left(\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_{\tilde{N}}\right)^{\mathrm{T}} \mathbf{J}_0} \\
&= \sqrt{\det(\mathbf{J}_0^{\mathrm{T}}\mathbf{J}_0)}.
\end{aligned} \tag{27}
$$

As this identity holds almost everywhere, we may finally drop the subscript from the Jacobian $\mathbf{J}_0$, and the conditional distribution of $\tilde{\mathbf{x}}$ hence takes the form

$$
p(\tilde{\mathbf{x}}|R) \propto \sqrt{\det(\mathbf{J}^{\mathrm{T}}\mathbf{J})} p(\mathbf{s}(\tilde{\mathbf{x}}), \Phi(\tilde{\mathbf{x}})). \tag{28}
$$

As $p(\tilde{\mathbf{x}}|R)$ is known up to a global constant, we may draw samples from it by generating a MCMC chain with the Metropolis–Hastings sampling rule [6].

5. Experiments

In this section we show two application examples of the dual distributions. We create confidence intervals for conics (Section 5.1) and show how probabilistic point transfer can be constructed by using the covariance information of the estimated trifocal tensor (Section 5.2).

5.1. Dual Distributions for Conics

The points on a conic satisfy the homogeneous quadratic equation

$$
\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = 0, \tag{29}
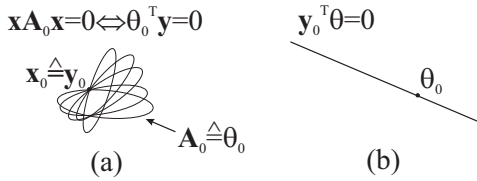$$

Figure 3. (a) The dual distribution as the point distribution is here generated from the total probability of all the conics containing the point. (b) The subset of conics containing a given point is a five-dimensional linear subspace in the six-dimensional dual space, thus the probability of a point can be identified as the total probability of the corresponding linear subspace.

which is a bilinear equation in $\mathbf{x}$ and $\mathbf{A}$ is a symmetric 3×3 matrix. This equation can be written in the form

$$\theta^{\mathrm{T}}\mathbf{y} = 0, \qquad (30)$$

where $\theta \hat{=} (a_{11}, a_{22}, a_{33}, 2a_{12}, 2a_{23}, 2a_{13})$ and $\mathbf{y} \hat{=} (x_1^2, x_2^2, x_3^2, x_1 x_2, x_2 x_3, x_1 x_3)$, i.e., a conic forms a 5-dimensional linear subspace in the six dimensional joint feature space. We assume that the parameter vector estimate $\theta = \theta_0$ and its covariance matrix $\mathbf{C}_\theta$ are available.

If we dualise the relationship (30) above, we see that a fixed point on the image plane determines a 5-dimensional linear subspace in the dual space, i.e., the space of all conics that intersect the point on the image plane, see Fig. 3. Moreover, to characterise the probability of the point on the image we may construct the *total probability (density) of all those conics containing the point* by using the uncertainty distribution of the estimated conic. In other words, the dual distribution characterises the confidence of the estimated conic by illustrating what has been learnt from the locations of the points on the true conic.

By the way of an example, we estimated the maximum likelihood estimate and its covariance matrix for the conic containing 25 points, shown in Fig. 4a, assuming i.i.d. Gaussian noise in the 2D measurements. To evaluate the dual density at the selected location (x_1, x_2) on the 2D plane, we need to evaluate (13) as well as compute the right magnification factor. As we parameterised the reduced joint feature vector $\mathbf{y}'$ by the modified spherical coordinates, we may construct the mapping $(x_1, x_2) \mapsto (\rho, \phi)$ and its Jacobian to evaluate the dual pdf using (28). The dual pdf for the estimated conic is illustrated in Fig. 4b. The contours visualise the fact that the we have a strong belief about the true conic points near the training data but extrapolation beyond the training data contains a substantial risk.

5.2. Probabilistic Point Transfer with an Uncertain Trifocal Tensor

The geometry of three projective views is characterised by several incidence relations or trilinearities, which are
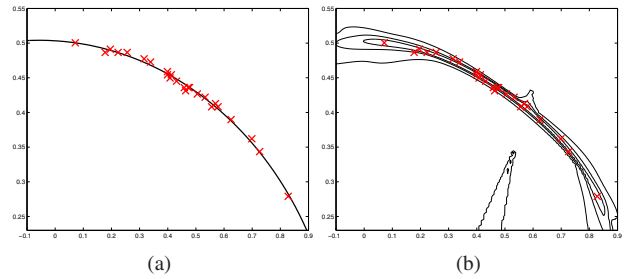

(a)   (b)

Figure 4. (a) The training data and the maximum likelihood estimate for the conic. (b) Contours of the dual pdf (point density) of the estimated conic characterising the current evidence where the points of the true conic locate.

collected into Table 1 from [4]. According to our preferences, we could use any of the trilinearities to create dual distributions or conditional dual distributions. As an example, we now illustrate how the trifocal point transfer can be "probabilised" by constructing a dual distribution for the trifocal tensor given estimates for the tensor and its covariance matrix as well as a novel point match in two views. In this case, the dual distribution is the conditional position distribution of the transferred point in the third view. We use the nine point–point–point constraint equations (the first relation in Table 1) of which only four equations are independent. Therefore the dimension of the affine subspace of this example is $K = (M - 1) - 4 = 14$.

Given the point match $\mathbf{m}' \leftrightarrow \mathbf{m}''$ in the views two and three, the construction of the conditional probability density $p(\mathbf{m}|\mathbf{m}', \mathbf{m}'', \mathcal{T}, \mathbf{C}_\mathcal{T})$ is as follows. The trilinear relations for three points define the elements of $\mathbf{Y}$ in the model (14). After constructing $\mathbf{Y}$, the whitened matrix $\mathbf{Y}'$ is obtained from (15). However, we regularised the whitening transform by replacing $\tilde{\Lambda}$ by $\tilde{\Lambda} + \lambda \mathbf{I}$, where $\lambda = 10^{-8}$ since the number of data points was limited in our experiment and hence, due to overfitting of the covariance matrix, the smallest ones of the eighteen non-zero eigenvalues were (assuming a non-degenerate configuration) relatively small.[1] The probability density for the affine subspace, corresponding to

---

[1] The diagonal Tikhonov regulariser was adequate here but a more elegant way would have been investigating the numerical rank of the covariance matrix and developing an automatic model selection scheme to determine the dimension of the affine subspace. Degenerate configurations could be also handled in this way.

Table 1. The trilinearities, adopted from [4].

Correspondence	Relation	dof
three points	$x^i x'^j x''^k \epsilon_{jqs} \epsilon_{krt} \mathcal{T}_i^{qr} = 0_{st}$	4
two points, one line	$x^i x'^j l''_r \epsilon_{jqs} \mathcal{T}_i^{qr} = 0_s$	2
one point, two lines	$x^i l'_q l''_r \mathcal{T}_i^{qr} = 0$	1
three lines	$l_p l'_q l''_r \epsilon^{piw} \mathcal{T}_i^{qr} = 0^w$	2

Figure 5. Training images for estimating the trifocal tensor.

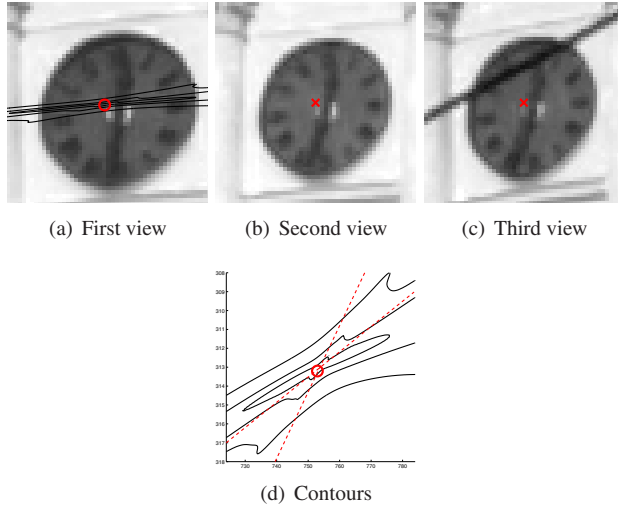(a) First view   (b) Second view   (c) Third view

(d) Contours

Figure 6. Probabilistic point transfer. (a) Three equi-probability contours of the point transfer pdf in the first view, at the levels 10^{-1}, 10^{-2}, and 10^{-3} times the maximum value, conditioned on the two-view point match in the views (b) two and (c) three. The circle indicates the transferred point using the (deterministic) point transfer [4] with only the ML estimate for the trifocal tensor, and the two points in the other views; (d) the same contours shown after scaling of the y-axis and superimposing the ML epipolar lines (dashed) corresponding to the points given in the second and third view. The contours are closed curves surrounding the most likely match locations and illustrate the fact there the trifocal transfer is more than the epipolar transfer from the views two and three.

$\mathbf{Y}'$, is given by (22) and the desired conditional probability density values are finally obtained from (28), up to scale. In short, to compute the conditional probability density value for a location of interest in the first view, the corresponding affine subspace parameters are computed from $\mathbf{Y}'$ and the probability density value of the corresponding affine subspace is weighted by the term containing the Jacobian of the transformation.

To illustrate the conditional point transfer density, we estimated the maximum likelihood trifocal tensor and its covariance matrix from 274 outlier free point correspondences for the image triplet shown in Fig. 5 assuming i.i.d. Gaussian noise in the measurements (see [4]). Then, by using a novel point correspondence in the views two (Fig. 6b) and three (Fig. 6c), we computed the conditional probability density in the first view, as reported above. To visualise the

shape of the pdf, we selected, three pdf values at the levels 10^{-1}, 10^{-2}, and 10^{-3} times the maximum value and show the corresponding contours in Fig. 6a and 6d. It can be seen that the transferred density is non-Gaussian while it indicates the feasible locations for the correspondence. The pdf has its maximum close to the point where the ML epipolar lines meet whereas the local shape of the peak is oriented towards the mean axis of the two epipolar lines. The pdf shape also seems to illustrate the well known fact that the trifocal constraint is more versatile than the mere epipolar geometries between the three views.

6. Conclusions

In this paper we have shown how the parameter distributions of multilinear models can be dualised. The proposed approach provides means for a pure statistical treatment of multiple view relations where the uncertainty of the geometric entity is taken into account. The dual distributions are closely connected to integral geometry in mathematics and have statistically sound, analytic forms with relatively mild assumptions. We demonstrated the applicability of the theory by characterising the confidence of estimated conics and constructed the probabilistic trifocal point transfer by using an uncertain trifocal tensor with its covariance matrix. In future, an interesting research direction is to investigate the dual distributions from the view point of integral geometry to deeply understand the theoretical connections as well as to develop additional numerical methods for vision applications.

References

[1] S. S. Brandt. On the probabilistic epipolar geometry. *Image and Vision Computing*, 26(3):405–414, 2008.

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

[3] I. Gelfand, S. Findkin, and M. Graey. *Selected Topics in Integral Geometry*, volume 220 of *Translations of Mathematical Monographs*. American Mathematical Society, 2003.

[4] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge, 2000.

[5] A. F. Karr. *Probability*. Springer, 1992.

[6] D. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge, 2003.

[7] L. Santaló. *Introduction to integral geometry*. Hermann, Paris, 1953.

[8] B. Triggs. Matching constraints and the joint image. In *Proc. ICCV*, Cambridge, MA, June 1995.

[9] B. Triggs. Joint feature distributions for image correspondence. In *Proc. ICCV*, volume II, pages 201–208, Vancouver, Canada, July 2001.

[10] J. R. Weeks. *The Shape of Space: How to Visualize Surfaces and Three-Dimensional Manifolds*. Marcel Dekker, 1985.

A. Modified Spherical Coordinates in $\mathbb{R}^N$

In this paper, we use modified spherical coordinates in N-dimensional space. That is, we assign a sign for the radius in the conventional N-dimensional spherical coordinates (see e.g. [10]) and parameterise the directions using points only on the other half of the unit hypersphere. This parameterisation is natural for one-dimensional subspaces in $\mathbb{R}^N$ as the directions then uniquely parameterise the linear subspaces almost everywhere and the (signed) radial parameter parameterises the points in the subspace.

Assume that[2] $N \geq 3$ and let $\rho \in \mathbb{R}$ be the radial coordinate and $\phi_1, \phi_2, \ldots, \phi_{N-1}$ be the angular coordinates so that ϕ_1 takes values between 0 and $\pi/2$, $\phi_2, \ldots, \phi_{N-2}$ are between 0 and π and ϕ_{N-1} is between 0 and 2π. The Cartesian coordinates x_i, $i = 1, \ldots, N$ are then defined as

$$x_1 = \rho \cos(\phi_1),$$
$$x_i = \rho \cos(\phi_i) \prod_{k=1}^{i-1} \sin(\phi_k), \quad i = 2, 3, \ldots, N-1,$$
$$x_N = \rho \prod_{i=1}^{N-1} \sin(\phi_i). \tag{31}$$

The inverse transform is

$$\rho = \text{sign}(x_1)\|\mathbf{x}\|, \quad \phi_1 = \arctan\left(\sqrt{\frac{\|\mathbf{x}\|^2}{x_1^2} - 1}\right),$$

$$\phi_i = \arctan\left(\frac{\sqrt{\sum_{k=i+1}^{N} x_k^2}}{\text{sign}(x_1)x_i}\right), \quad i = 2, 3, \ldots N-2,$$

$$\phi_{N-1} = \arctan(\text{sign}(x_1)x_N, \text{sign}(x_1)x_{N-1}), \tag{32}$$

and the volume element is obtained as

$$\left|\det \frac{\partial(x_1, x_2, \ldots, x_N)}{\partial(\rho, \phi_1, \phi_2, \ldots, \phi_{N-1})}\right| d\rho d\phi_1 d\phi_2 \cdots d\phi_{N-1}$$
$$= |\rho|^{N-1} \left(\prod_{k=1}^{N-2} \sin^{N-k-1}(\phi_k)\right) d\rho d\phi_1 d\phi_2 \cdots d\phi_{N-1}. \tag{33}$$

B. Proof of Lemma 1

Proof. It is well known that an orthogonal projection matrix $\mathbf{P}$ may be decomposed as $\mathbf{P} = \mathbf{U}\mathbf{U}^T$ where the columns

---

of $\mathbf{U}$ form an orthogonal basis of the range of $\mathbf{P}$. We hence form an orthonormal basis for the range from the basis $\mathbf{p}_i = \mathbf{P}\mathbf{e}_i$, $i = 1, \ldots, K$ by using the Gram–Schmidt orthonormalisation procedure. Now, let

$$\mathbf{u}_1 = \frac{\mathbf{P}\mathbf{e}_1}{\|\mathbf{P}\mathbf{e}_1\|}, \quad \mathbf{u}_2 = \frac{\mathbf{P}\mathbf{e}_2 - (\mathbf{u}_1^T\mathbf{P}\mathbf{e}_2)\mathbf{u}_1}{\|\mathbf{P}\mathbf{e}_2 - (\mathbf{u}_1^T\mathbf{P}\mathbf{e}_2)\mathbf{u}_1\|}, \quad \ldots,$$

$$\mathbf{u}_K = \frac{\mathbf{P}\mathbf{e}_K - \sum_{i=1}^{K-1}(\mathbf{u}_i^T\mathbf{P}\mathbf{e}_K)\mathbf{u}_i}{\|\mathbf{P}\mathbf{e}_K - \sum_{i=1}^{K-1}(\mathbf{u}_i^T\mathbf{P}\mathbf{e}_K)\mathbf{u}_i\|}. \tag{34}$$

and let $\mathbf{L} = (\mathbf{u}_1 \ \mathbf{u}_2 \ \ldots \ \mathbf{u}_K)$, i.e., the columns of $\mathbf{L}$ form an orthonormal basis for the range of $\mathbf{P}$, and $\mathbf{P} = \mathbf{L}\mathbf{L}^T$.

We need to show that $\mathbf{L}$ is additionally a lower triangular matrix which has a strictly positive diagonal. We note that

$$\mathbf{u}_1 = \frac{\mathbf{P}\mathbf{e}_1}{\|\mathbf{P}\mathbf{e}_1\|} = \frac{\mathbf{L}\mathbf{L}^T\mathbf{e}_1}{\|\mathbf{L}\mathbf{L}^T\mathbf{e}_1\|} = \frac{\sum_{k=1}^{K}\mathbf{u}_k\mathbf{u}_k^T\mathbf{e}_1}{\|\sum_{k=1}^{K}\mathbf{u}_k\mathbf{u}_k^T\mathbf{e}_1\|}$$

$$= \frac{u_{11}\mathbf{u}_1 + \overbrace{\sum_{k=2}^{K} u_{1k}\mathbf{u}_k}^{=0}}{\|\sum_{k=1}^{K} u_{1k}\mathbf{u}_k\|}, \tag{35}$$

where the second term in the nominator is zero because the vectors $\mathbf{u}_k$, $k = 1, 2, \ldots, K$ are linearly independent by assumption. It follows that $u_{1k} = 0$, $k = 2, \ldots, K$. Thus, $u_{11}/|u_{11}| = 1 \implies u_{11} > 0$. Similarly,

$$\mathbf{u}_2 = \frac{u_{22}\mathbf{u}_2 - (\mathbf{u}_1^T\mathbf{P}\mathbf{e}_2)\mathbf{u}_1 + \overbrace{\sum_{k\neq 2}^{K} u_{2k}\mathbf{u}_k}^{=0}}{\|\sum_{k=1}^{K} u_{2k}\mathbf{u}_k\|}, \tag{36}$$

that implies $u_{2k} = 0$, $k = 3, \ldots, K$ and $u_{22} > 0$. Likewise, $u_{lk} = 0$, and $u_{ll} > 0$ when $k = l + 1, \ldots, K$, hence, $\mathbf{L}$ is a lower triangular matrix with strictly positive diagonal.

We still need to show that the representation is unique. Let us assume the contrary, i.e., there are two different lower triangular matrices $\mathbf{L}_1 = (\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_K)$ and $\mathbf{L}_2 = (\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_K)$ which have the the indicated properties. Then $\mathbf{P} = \mathbf{L}_1\mathbf{L}_1^T = \mathbf{L}_2\mathbf{L}_2^T$ and

$$\mathbf{P}\mathbf{e}_1 = u_{11}\mathbf{u}_1 = v_{11}\mathbf{v}_1. \tag{37}$$

Since $\mathbf{u}_1$ and $\mathbf{v}_1$ are unit vectors, it follows that $|u_{11}| = |v_{11}|$. The diagonal elements are additionally strictly positive so $u_{11} = v_{11}$ that implies $\mathbf{u}_1 = \mathbf{v}_1$. Similarly, we see that $\mathbf{u}_l = \mathbf{v}_l$, when $l = 2, \ldots, K$, i.e., $\mathbf{L}_1 \equiv \mathbf{L}_2$ that contradicts the assumption, hence, $\mathbf{L}$ must be unique.

$\square$

---

[2]If $N = 1$ or $N = 2$ the construction is similar but these cases are omitted here for simplicity.