Statistical Validation of Computer Vision Software

Xufei Liu, Tapas Kanungo, R. M. Haralick *

Intelligent Systems Laboratory
Department of Electrical Engineering, FT-10
University of Washington
Seattle, WA 98195

Abstract

Computer vision software is complex involving many tens of thousands of lines of code. Coding mistakes are not uncommon. When the vision algorithms are run on controlled data which meet all the algorithm assumptions, the results are often statistically predictable. This renders it possible to statistically validate the computer vision software and its associated theoretical derivations. In this paper we review the general theory for some relevant kinds of statistical testing and then illustrate this experimental methodology to validate our building parameter estimation software. This software estimates the 3D positions of buildings vertices based on the input data obtained from multi-image photogrammetric resection calculations and 3D geometric information relating some of the points, lines and planes of the buildings to each other.

KEYWORDS: Statistical analysis, multivariate hypothesis testing, 3D parameter estimation, error propagation, software validation.

1 Introduction

Many computer vision problems can be posed as either parameter estimation problems (for example, estimate the pose of the object), or hypothesis testing problems (for example, which of the N objects in a database occurs on a given image.) Since the input data (such as, images, or feature points) to these algorithms is noisy, the estimates produced by the algorithm are noisy. In other words, there is an inherent uncertainty associated with the results produced by any computer vision algorithm. These uncertainties are best expressed in terms of statistical distributions, and the distributions' means and covariances. Details of the theory and application of covariance propagation can be found in [6] [9], and the references cited in [9].

Usually, implementations of vision algorithms run into thousands of lines of code. Furthermore, the algorithms are based on many approximations, and numerous mathematical calculations. One way to check whether the software implementation and the theoretical calculations are correct is by providing the algorithm input data with known (controlled) statistical characteristics, which is possible since the input data can be artificially generated, and then checking if the estimated output is actually distributed as what was predicted by theoretical calculations.

Since many of the estimation problems are multidimensional, testing whether the means and covariances of the empirical distribution and predicted distribution are same is easier than testing whether or not the shapes of the two distributions are same. In this paper, we summarize statistical tests for the case when the random estimates can be assumed to be multivariate Gaussian. We also describe the function interfaces to software we have implemented for conducting these tests. Although the software libraries and environments (e.g. Splus, numerical recipes) are available for conducting the tests for one-dimensional samples, we are un-aware of similar software libraries for multivariate case. In fact, most of the statistics books do not give all the five tests we have give (e.g. Koch [13] does not address the fifth testing problem). The hypothesis testing theory and software are described in [12] and the software can be obtained at no cost from Kanungo (tapas@george.ee.washington.edu). A description of how the software and the theory was tested using statistical techniques is also included.

2 The Kinds of Statistical Hypotheses

Let $x_1, x_2, \ldots, x_n$ be a sample from a multivariate Gaussian distribution with population mean μ and population covariance Σ. That is, $x_i \in R^p$ and $x_i \sim N(\mu, \Sigma)$, where p is the dimension of the vectors x_i.

We can make various hypotheses regarding the population mean and covariance depending on what is known and what is unknown. The data x_i are then used to test whether or not the hypothesis is false. Notice that each population parameter (here we have two – μ and Σ) can be either (i) tested, or (ii) unknown and untested, (iii) or known. If a parameter is being tested, then a claim regarding its value is being made. If a parameter is unknown and untested, no claim is being made about the value of that parameter; its value is not known and therefore we cannot use it in any computation. If the value of a parameter

*Funding from DARPA contract 92-F1428000-000 is gratefully acknowledged.

assumed to be known, then its value is known without error and cannot be questioned or tested, just like the normality assumption is not questioned. Furthermore, when a parameter value is known, the value itself can be used in computation of test statistics for other parameters.

In general, if the distribution has q parameters, then there can be $3^q - 2^q$ tests. The reasoning is as follows. Since each parameter can be either tested, or unknown and untested, or known, the number of possibilities are 3^q. But, of these the number of combinations in which none of the parameters are tested (that is, they are either known, or unknown and untested – and so do not represent a test) is 2^q. Thus, the total number of distinct hypotheses that can be made about a sample from a q-parameter distribution is $3^q - 2^q$.

In the case when the data comes from multivariate normal distribution, $N(\mu, \Sigma)$, we have $q = 2$ and thus can have $3^2 - 2^2 = 5$ possible hypotheses. Now we describe each of the five tests when the data comes from a multivariate normal population.

H_1 : $\mu = \mu_0$, **($\Sigma = \Sigma_1$ known.)** In this test, the question is whether or not the sample is from a Gaussian population whose mean is μ_0. The population covariance Σ is assumed to be known and equal to Σ_1.

H_2 : $\mu = \mu_0$, **(Σ unknown, untested.)** In this test, the question is whether or not the sample is from a Gaussian population whose mean is μ_0. No statement is made regarding the population covariance Σ.

H_3 : $\Sigma = \Sigma_0$, **($\mu = \mu_1$ known.)** In this test, the question is whether or not the sample is from a Gaussian population whose covariance is Σ_0. The population mean μ is assumed to be known equal to μ_1.

H_4 : $\Sigma = \Sigma_0$, **(μ unknown, untested.)** In this test, the question is whether or not the sample is from a Gaussian population whose covariance is Σ_0. No statement is made regarding the mean μ.

H_5 : $\mu = \mu_0$, $\Sigma = \Sigma_0$ In this test, the question is whether or not the sample is from a Gaussian population whose mean is μ_0 and covariance is Σ_0. It is this test that is the principal test we use for the software validation.

3 Definitions

In this section we briefly describe the terms used in the rest of the paper. If the reader is familiar with statistics, he/she should skip this section. For a lucid explanation of the basic univariate concepts please see [4]. A slightly more rigorous treatment of the univariate and multivariate test is given in [2]. Multivariate tests are treated in great detail in [13]. The most authoritative reference on multivariate statistics is [1]. Although this book has most of the results, it is not very readable, and the results are scattered all over the book.

A *statistic* of the data $x_1, \ldots, x_n$ is any function of the data. For example, sample mean, $\bar{x}$, is a statistic, and so is the sample covariance matrix, S. The statistic need not be one-dimensional – $(\bar{x}, S)^t$ together form another statistic of the same data. A *sufficient statistic* is a statistic that contains all the information about the data; any inference regarding the underlying population can be made using just the sufficient statistic – the individual data points do not add any more information to the inference process. For example, the vector of original data $(x_1, \ldots, x_n)^t$ is a sufficient statistic – it contains all the information regarding the data. Another sufficient statistic is $(\bar{x}, S)^t$. Sufficient statistic is not unique. A *minimal sufficient statistic* is a sufficient statistic that has smallest number of entries. For example, for Gaussian data, $(\bar{x}, S)$ is the minimal sufficient statistic.

A *hypothesis* is any statement about a population parameter that is either true or false. The *null hypothesis*, H_0, and the *alternate hypothesis*, H_A, form the two complementary hypothesis in a statistical hypothesis testing problem.

A *test statistic* is just another statistic of the data that is used for testing a hypothesis. The *null distribution* is the distribution of the test statistic when the null hypothesis is true. The *alternate distribution* is the distribution of the test statistic when the alternate hypothesis is true.

There are two types of errors - mis-detection and false alarm. If the null hypothesis is true but the test procedure decides the null hypothesis to be false, it is called a *misdetection*. When the alternate hypothesis is true but the test procedure accepts the *null hypothesis*, it is called a *false alarm*. The misdetection probability of a test procedure is usually fixed by the user also referred to as the *significance level*, α, of the test. Typical value for α is 0.05.

The *power function* of a hypothesis test is a function of the population parameter θ, and value of the function $\beta(\theta)$ is equal to 1 minus the probability of false alarm. Ideally, the power function should be zero for θ where the null hypothesis is true and one for all θ where the alternate hypothesis is true. For most realistic testing problems one cannot create a test procedure with such an ideal power function. Power functions are very useful for evaluating hypothesis testing procedures. For an example where it is used for computer vision problems, see[11]. A *uniformly most powerful test* is a test procedure whose power function is higher than all other test procedures.

There are many methods for creating tests and corresponding test statistics. The test statistics given in this paper were derived by maximizing the likelihood ratio. Please refer to the cited literature for the derivation.

4 Test statistics

In this section we define the test statistic and their distribution under the true null hypothesis [1]. We use the following definitions of $\bar{x}$ and S.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

and

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^t,$$

where we have assumed that the data vectors x_i are p-dimensional and the sample size is n.

4.1 Test 1: $\mu = \mu_0$ with known $\Sigma = \Sigma_1$

Test statistic:

$$T = n(\bar{x} - \mu_0)^t\Sigma_1^{-1}(\bar{x} - \mu_0). \tag{1}$$

Distribution under null hypothesis is Chi-squared (Anderson, page 73):

$$T \sim \chi^2_p.$$

The alternate hypothesis is $H_A : \mu \neq \mu_0$; the distribution under the alternate hypothesis is non-central Chi-squared (Anderson, page 77):

$$T \sim \chi^2_{p,d}$$

where $d = n(\mu - \mu_0)^t\Sigma_1^{-1}(\mu - \mu_0)$ is the non-centrality parameter.
Reference: Anderson, pages 73, 77.

4.2 Test 2: $\mu = \mu_0$ with unknown Σ

Hotelling's Test statistic:

$$T = \frac{n(n-p)}{p(n-1)}(\bar{x} - \mu_0)^t S^{-1}(\bar{x} - \mu_0). \tag{2}$$

Distribution under null hypothesis (F):

$$T \sim F_{p,n-p}.$$

The alternate hypothesis is $H_A : \mu \neq \mu_0$; the distribution under the alternate hypothesis is non-central F:

$$T \sim F_{p,n-p,d}$$

where $d = n(\mu - \mu_0)^t\Sigma^{-1}(\mu - \mu_0)$ is the non-centrality parameter.
Reference: Anderson page 163.

4.3 Test 3: $\Sigma = \Sigma_0$ with known $\mu = \mu_1$

Let

$$C = \sum_{i=1}^{n}(x_i - \mu_1)(x_i - \mu_1)^t = (n-1)S + (\bar{x} - \mu_1)(\bar{x} - \mu_1)^t \; .$$

and

$$\lambda = (e/n)^{pn/2}|C\Sigma_0^{-1}|^{n/2}\exp(-tr(C\Sigma_0^{-1})/2) \; .$$

Test statistic:

$$T = -2\log\lambda. \tag{3}$$

Distribution under null hypothesis is Chi-squared:

$$T \sim \chi^2_{p(p+1)/2}.$$

The alternate hypothesis is $H_A : \Sigma \neq \Sigma_0$; the distribution under the alternate hypothesis is unknown.
Reference: Anderson page 249, 434, 436.

4.4 Test 4: $\Sigma = \Sigma_0$ with unknown μ

Let

$$B = (n-1)S,$$

and

$$\lambda = (e/(n-1))^{p(n-1)/2}|B\Sigma_0^{-1}|^{(n-1)/2}\exp(-tr(B\Sigma_0^{-1})/2)$$

Test statistic:

$$T = -2\log\lambda. \tag{4}$$

Distribution under null hypothesis is Chi-squared:

$$T \sim \chi^2_{p(p+1)/2}.$$

The alternate hypothesis is $H_A : \Sigma \neq \Sigma_0$; the distribution under the alternate hypothesis is unknown.
Reference: Anderson page 249, 434, 436.

4.5 Test 5: $\Sigma = \Sigma_0$ and $\mu = \mu_0$

Define

$$B = (n-1)S$$

and

$$\begin{aligned}\lambda \;=\;& (e/n)^{pn/2}|B\Sigma_0^{-1}|^{n/2} \\ & \exp\left(-[tr(B\Sigma_0^{-1}) + n(\bar{x} - \mu_0)^t\Sigma_0^{-1}(\bar{x} - \mu_0)]/2\right)\end{aligned}$$

Test statistic:

$$T = -2\log\lambda \tag{5}$$

Distribution under true null hypothesis is Chi-squared:

$$T \sim \chi^2_{p(p+1)/2+p}$$

The alternate hypothesis is $H_A : \Sigma \neq \Sigma_0$, and $\mu \neq \mu_0$; the distribution under the alternate hypothesis is unknown.
Reference: Anderson page 442.

5 Validating theory and software

To validate computer vision software two checks have to be performed. The first check is that the theory is correct: the theoretically derived null distributions of the test statistics are actually correct. The second check is that the software is correct: the implementation is exactly what the theory dictates. Both the checks can be done by computing the empirical distributions and comparing them with the theoretically derived distributions. In the next subsection we describe how we empirically compute the null distributions of the five test statistics, and in the following section we describe how we use the Kolmogorov-Smirnov test to check if the empirical distribution and the theoretically-derived distributions are same.

5.1 Empirical null distributions

In order to generate the empirical null distributions we proceed as follows.

1. Choose some values for the multivariate Gaussian population parameters p, μ and Σ.

2. Generate n samples from the population.

3. Compute the value of the statistic, T, for the test you are verifying.

4. Repeat steps 2 and 3 M times to get T_i, $i = 1, \ldots, M$.

5. The empirical distribution T can be computed by computing the histogram of T_i.

5.2 Kolmogorov-Smirnov tests

The Kolmogorov-Smirnov (KS) test tests whether two distributions are alike. The KS test uses the fact that the maximum absolute difference between the empirical cumulative distribution (the KS test statistic) and the theoretical cumulative distribution has a known distribution (the null distribution). For a more detailed discussion on the KS test see [16].

The Kolmogorov-Smirnov test was performed to check if the empirical distributions and the theoretical distributions were close enough. All the empirically computed null distributions passed the KS test. Thus we have confirmed that the theoretical derivations of the null distributions is correct and the software implementing the theory is also correct.

6 Application: 3D Parameter Estimation

We applied the hypothesis testing methodology to validate the 3D parameter estimation software used for constructing the ground truth model from the RADIUS model board data set. In this section we describe the problem and the optimization approach.

6.1 Site Model Construction

The task is to construct 3D object models from the detected 2D image features and the known geometric constraints of the observed perspective projections of the 3D objects. The data set consists of the 78 images from the two RADIUS model boards and the 3D coordinates of some building vertices. Since the purpose of this site model construction was to establish ground truth for automatic site model construction algorithms, the corresponding points of the building vertices that were observable on the images were identified and located manually. Also 3D positions of a few of the building vertices are known. A simultaneous estimation of the interior parameters and exterior orientation parameters of the cameras was done by setting up and solving a very large photogrammetric resection problem. Then using these camera parameters a multi-image triangulation was performed. This yielded the noisy estimates for the building vertices that was the input to the site model construction software whose testing we describe.

The geometric constraint procedure takes the photogrammetrically estimated 3D point positions and their covariance matrices as observations. It use the partial models of the buildings to generate constraints on the building parameters. To estimate the optimal 3D parameters that satisfy the relations in the partial models, a constrained optimization model is solved. By error propagation we derived the covariance matrix of the estimated building vertices which are now guaranteed to satisfy the given constraints. See the paper *Site Model Construction Using Geometric Constrained Optimization* in this IUW proceedings for details on this problem.

6.2 Constrained Optimization

The observed 3D points and the associated covariance matrix Σ are obtained from triangulation. Having the partial object model and the perturbation model, we can define the estimation problem. Let $\Theta \in \mathbb{R}^m$ denote the parameters, $X' \in \mathbb{R}^m$ denote the observations, and $p(X' \mid \Theta)$ denote the likelihood function. In the building estimation problem, the parameters are the coordinates of the points, the normal vectors and distance constants of the planes, and the direction cosines and reference points of the lines.

Assume that the given optimality criterion is the maximum posterior probability, a Bayesian approach can be used to transform the problem into a maximum likelihood problem with constraints. Let the constraints be denoted by $\Theta \in C_\Theta \subset \mathbb{R}^m$. The problem can be expressed as a constrained optimization problem.

$$\min\{-p(X' \mid \Theta) \mid \Theta \in C_\Theta\}$$

The problem can be reformulated by taking logarithm of the probability function. Under the assumption of Gaussian noise, we obtain a least squares model. The objective function is the sum of squared errors between the estimated point positions and the observed points.

$$\begin{array}{ll} \min\limits_{\Theta} & \{f(\Theta) := (X' - X)^T \Sigma^{-1} (X' - X)\} \quad (6) \\ \text{subject to} & \Theta \in C_\Theta \end{array}$$

where X denotes the unknown 3D points, and the feasible set C_Θ is determined by the partial model and the unit length constraint for the directional vectors.

If the noise effecting different 3D points are independent, the objective function can be rewritten as

$$f(\Theta) = \sum_{i=1}^{K} (\mathbf{x}_i' - \mathbf{x}_i)^T \Sigma_i^{-1} (\mathbf{x}_i' - \mathbf{x}_i)$$

where Σ_i is the covariance matrix of the ith point, and K is the number of observed points.

7 Error Propagation

Once the constrained optimization produces a result, we use the error propagation approach [10] [14] to transform the input error covariance matrix to the output covariance matrix. In the building estimation problem, we have the optimization model

$$\begin{array}{ll} \min\limits_{\theta} & f(\Theta) \\ \text{subject to} & \mathbf{h}(\Theta) = 0 \end{array}$$

where f is the sum of squared errors between the estimated 3D points and the observed 3D points.

The Lagrangian function is

$$L(X', \Theta, \Lambda) = f(X', \Theta) + \Lambda^T \mathbf{h}(\Theta)$$

Suppose that $(\tilde{X}, \tilde{\Theta}, \tilde{\Lambda})$ is a optimal point. From the necessary conditions of a local minimum point, the linearized model at the optimal point can be obtained by solving [7] [15]

$$\begin{pmatrix} \tilde{\mathbf{Q}} & \tilde{\mathbf{H}}^T \\ \tilde{\mathbf{H}} & 0 \end{pmatrix} \begin{pmatrix} \Delta\Theta \\ \Delta\Lambda \end{pmatrix} = \begin{pmatrix} -\tilde{\mathbf{B}}\Delta X \\ 0 \end{pmatrix} \quad (7)$$

The Lagrangian matrix at the point of $(\tilde{X}, \tilde{\Theta}, \tilde{\Lambda})$ can be approximated by the Lagrangian matrix at the minimum if the error is small. Hence the linear model can be approximated by

$$\begin{pmatrix} \mathbf{Q}^* & (\mathbf{H}^*)^T \\ \mathbf{H}^* & 0 \end{pmatrix} \begin{pmatrix} \Delta\Theta \\ \Delta\Lambda \end{pmatrix} = \begin{pmatrix} -\mathbf{B}^*\Delta X \\ 0 \end{pmatrix}$$

where

$$\begin{aligned} \mathbf{Q}^* &= \nabla^2_{\theta\theta}\mathcal{L}(X', \hat{\Theta}, \hat{\Lambda}) \\ &= \nabla^2 f(X', \hat{\Theta}) + \sum_{j=1}^{r} \hat{\lambda}_j \nabla^2 h_j(\hat{\Theta}) \\ \mathbf{B}^* &= \nabla^2_{\theta\mathbf{x}}\mathcal{L}(X', \hat{\Theta}, \hat{\Lambda}) \;=\; \nabla^2_{\theta\mathbf{x}} f(X', \hat{\Theta}) \\ \mathbf{H}^* &= \nabla\mathbf{h}(\hat{\Theta}) \end{aligned}$$

Assume that the constraints are linearly independent. Then the row vectors in matrix H^* are linearly independent. We can use the null space method to compute the error propagation matrix J [7] [15].

Once the error propagation matrix is obtained, we can propagate the covariance matrix of the observations Σ to the output. The covariance matrix of the estimated parameters, Σ_Θ, can be approximated by

$$\Sigma_\Theta = J\Sigma J^T \quad (8)$$

8 Experimental Methodology

To validate the optimization algorithm and the error propagation model, an experiment is needed. This section describes the experimental methodology for this validation.

8.1 Ideal Data Generators and Noise Model

Three building types, the cubic box, the peak roof house and the hip roof house, appear frequently in the given sites. They are chosen as our prototype models with unknown length, location and orientation parameters.

In the experiment, ideal data generators randomly generate the ideal parameters for the prototype models and produce the ideal 3D points.

Assume that a 3D coordinate system x-y-z is used. To simulate the site model situation, the ground is assigned as the plane $z = 0$. Without losing generality we assume that the ideal model parameters determining the 3D positions of the building vertices have uniform distributions. The center of the bottom plane of a basic model is in a region defined by $[-x_0, x_0), [-y_0, y_0), [-z_0, z_0)$. The basic model is rotated on the ground with a random angle $\phi \in [\phi_0, \phi_1)$.

The three length parameters for the cubic box model are denoted by a, b, c, with $a_0 \leq a < a_1, b_0 \leq b < b_1$, and $c_0 \leq c < c_1$.

In this experiment the ranges of the parameters for the cubic model are set as follows.

x_0	y_0	z_0	ϕ_0	ϕ_1	a_0	a_1	b_0	b_1	c_0	c_1
50	50	0	0	2π	30	60	30	60	30	60

The peak roof model uses four length parameters, a, b, c, d, where a, b, c are same as those in cubic model and their ranges are the same. The height of the peak roof has parameter d, with $10 \leq d \leq 20$. The hip roof model requires one more parameter e than the peak roof model. The length of the roof edge is $a - 2e$, with $5 \leq e \leq 10$.

For each building type, K ideal buildings are randomly generated. Each one of these K buildings will be associated with n experiments where Gaussian random noise is added to each of the 3D coordinates of the building and the constrained optimization is used to estimate the building vertices that satisfy the various geometric constraints. As a result of these n experiments, n estimates of the building parameters are produced. It is these n estimates on which the hypothesis test statistics will be computed. We call the procedure to determine these n test statistics a trial. Since there are K ideal buildings for each ideal building type, we can compute K test statistics. These K statistics can then be used to test the hypothesis that their distribution is as the statistical theory of the test says it should be.

The noise values are independently sampled from a Gaussian distribution $\mathcal{N}(0, \sigma^2 I)$, where σ is the standard deviations of the random variables $\delta x, \delta y, \delta z$. We repeated each experiment σ set to 1.0, 2.0 or 3.0. The validation results for all three different standard deviations is similar. So here we just discuss the validation for the standard deviation begin equal to 3.0. K is set at 100. n is set at 500 for cubic model and 700 for other models.

8.2 Statistic Test

At each trial a sample of model parameters and corresponding ideal 3D points are produced by the ideal data generator. Let the ideal parameters be denoted by $\tilde{\Theta}$. For each ideal building instance having parameters $\tilde{\Theta}$, n independent perturbations $\{\Delta X_i,\ i = 1, ..., n\}$ are generated from the noise model with distribution $\mathcal{N}(0, \Sigma)$. By adding the perturbations to the ideal points, the perturbed data set $\{X'_1, X'_2, ..., X'_n\}$ is generated. For each of the perturbed data $\{X'_1, X'_2, ..., X'_n\}$, an optimal solution $\hat{\Theta}_i$ is computed by solving

$$\begin{aligned} \min_{\Theta_i} \quad & f(X'_i, \Theta_i) \qquad i = 1, ..., n \\ \text{subject to} \quad & \mathbf{h}(\Theta_i) = 0 \end{aligned}$$

Thus, we have n estimates $\{\hat{\Theta}_i,\ i = 1, ..., n\}$.

Using equation (8), we can transform the input covariance matrix through the error propagation matrix to the output. If the linear model is valid, the estimated parameters $\{\hat{\Theta},\ i = 1, ..., n\}$ should be approximately distributed as $\mathcal{N}(\tilde{\Theta}, J\Sigma J^T)$.

Let $\Delta\hat{\Theta}_i$ denote $\hat{\Theta}_i - \tilde{\Theta}$, $i = 1, ..., n$. Let $\mu_0 = 0$ and $\Sigma_0 = J\Sigma J^T$. Under the linearized model, $\{\Delta\hat{\Theta}_i,\ i = 1, ..., n\}$ have distribution $\mathcal{N}(\mu_0, \Sigma_0)$. Consider $\{\Delta\hat{\Theta}_i,\ i = 1, ..., n\}$ as a random sample from a Gaussian distribution $\mathcal{N}(\mu, \Sigma)$, we can perform any one of the five hypothesis test. Here we just discuss our results on hypothesis H_5: $\mu = \mu_0$ and $\Sigma = \Sigma_0$. Results on the other hypothesis tests are similar.

The significance level α is selected to be 0.05. Under the null hypothesis, the computed statistics of the mean and covariance tests have the null distributions. This can be verified by using Kolmogorov-Smirnov test (K-S test) on the K test statistics generated from the K trials.

8.3 Range Space Analysis

The standard hypothesis test methods require that the covariance matrix be positive definite. However the output covariance of a constrained optimization is generally semi-positive definite, precisely because of the constraints.

Theorem 1 *Suppose that not all of the derivatives of the constraint equations are equal to zero at the local minimum point, then the propagated error covariance $J\Sigma J^T$ is singular.*

Proof : From the given condition we know that the derivative matrix H is not a zero matrix, i.e.,

$$H = \begin{pmatrix} \frac{\partial h_1}{\partial \Theta} \\ \cdot \\ \cdot \\ \frac{\partial h_r}{\partial \Theta} \end{pmatrix} \neq 0$$

Left multiply equation $\Delta\Theta = J\Delta X$ with H.

$$H\Delta\Theta = HJ\Delta X$$

Because both (X, Θ^*, Λ^*) and $(X + \Delta X, \Theta^* + \Delta, \Lambda^* + \Delta\Lambda)$ are local minimum points of the optimization, the following equation is satisfied,

$$H\Delta\Theta = 0.$$

Hence

$$0 = HJ\Delta X$$

Since the formula holds for any ΔX, it implies that

$$0 = HJ \tag{9}$$

Now we use this result to prove that $J\Sigma J^T$ is singular. Left multiply $J\Sigma J^T$ with H and right multiply it with H^T. From (9) we have

$$HJ\Sigma J^T H^T = 0\Sigma 0$$

Since H is not a zero matrix, $J\Sigma J^T$ must be singular.
□

To utilize the standard hypothesis technology, we project a semi-positive definite matrix onto its range space. Suppose that a $n \times n$ covariance matrix Σ_0 has k nonzero eigenvalues $w_1, ..., w_k$ and the associated unit eigenvectors $v_1, ..., v_k$. A basis of the range space of Σ_0 can be composed by

$$B = (v_1, ..., v_n)$$

Using B to perform matrix transform as follows.

$$B^T\Sigma_0 B = \Sigma_B = \begin{pmatrix} w_1 & 0 & . & . & . \\ 0 & w_2 & 0 & . & . \\ . & . & . & & \\ . & . & . & & \\ 0 & . & . & . & w_k \end{pmatrix}$$

Let $B^\perp$ be a basis matrix of the null space of matrix Σ_0. It is obvious that $(B, B^\perp)$ is orthonormal. In the experiment, we check whether $(B^\perp)^T\Delta\Theta$ has very small variances (caused by round off errors and nonlinear items). If it is true, we conduct the hypothesis test on variables $B^T\Delta\Theta$ with covariance matrix Σ_B.

Due to the round-off error and nonlinear items, the zero eigenvalues of matrix Σ_0 may not be exactly zero. We use a small threshold to distinguish the zero eigenvalues and the and nonzero eigenvalues. In all our experiments the threshold is set to 10^{-6} times the maximum eigenvalue.

For the cubic model, the range space of the output error covariance matrix has 7 dimensions. This can be understand as follows. Consider a cubic house whose faces are all at right angles to each other. Count the number of degrees of freedom. The size of a cubic model is defined by 3 independent parameters. The location of the model is specified by 3 translation parameters in 3D space. In the experiments, the normal vector of the cube roof is fix to the vertical direction. The only possible rotation is around the vertical axis of the model. Thus the total number of independent parameters is seven. For the peak roof model the above analysis is similar, except that two more parameters are need to determine the roof height and the ridge position. (In the partial model we do not fix the horizontal position of the roof ridge to the center of the building.) Thus the range space of the output covariance matrix for the peak roof model has 9 dimensions. The hip roof model inherits all the parameters of the peak roof model. It requires two more parameters to determine how much of the ridge being cut from each of the two ends (they are assumed to be independent). These parameters can be thought of as the relative starting and ending points for the the ridge. Thus the range space of the output covariance matrix for the hip roof model has 11 dimensions.

9 Experimental Results

In this section we will show the experimental results on the cubic model, the peak roof model and the hip roof model.

test method	# trials	sample size	signif. level	reject rate
test μ, known Σ	100	500	0.05	0.06
test μ, unknown Σ	100	500	0.05	0.05
test Σ, known μ	100	500	0.05	0.07
test Σ, unknown μ	100	500	0.05	0.06
test μ and Σ	100	500	0.05	0.05

9.1 Test of Cubic Model with $\sigma = 3.0$

The experimental results for the cubic model are summaryized in the table below. The null hypothesis is not rejected at a .05 significance level.

The experimental results of hypothesis test for cubic model with $\sigma = 3.0$ is shown in figure 1, the x axis is the statistic used in the test and the y axis is $1-\alpha$, where α is the significance level.

Figure 1 shows the result of test mean and covariance simultaneously.

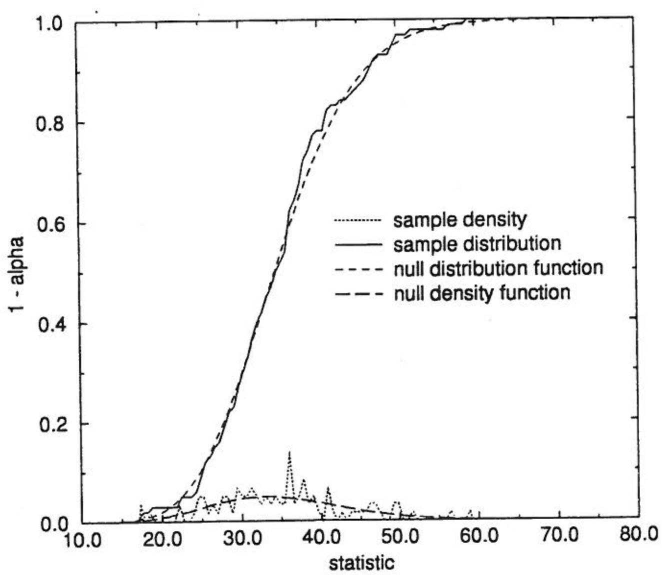

Figure 1: Test mean and covariance : cubic model, $\sigma = 3$

9.1.1 K-S test

For the K-S test result the value of p is 7. All statistics pass the K-S test at a significance level of greater than .2. Thus, the optimization model and the error propagation model are validated.

9.2 Test of Peak Roof Model with $\sigma = 3.0$

The table below summarizes the results for the peak roof model. The null hypothesis is not rejected at the .05 significance level.

The experimental results of hypothesis test for the peak roof model with $\sigma = 3.0$ is shown in figure 2, the x axis is the statistic used in the test and the y axis is $1-\alpha$, where α is the significance level.

Figure 2 shows the result of test mean and covariance simultaneously.

test method	null distribution	sample size	signif. level	K-S test
test μ, known Σ	χ^2_p	100	0.391	pass
test μ, unk. Σ	$F_{p,n-p}$	100	0.217	pass
test Σ, known μ	$\chi^2_{p(p+1)/2}$	100	0.866	pass
test Σ, unk. μ	$\chi^2_{p(p+1)/2}$	100	0.881	pass
test μ and Σ	$\chi^2_{p(p+1)/2+p}$	100	0.700	pass

test method	# trials	sample size	signif. level	reject rate
test μ, known Σ	100	700	0.05	0.03
test μ, unk. Σ	100	700	0.05	0.03
test Σ, known μ	100	700	0.05	0.04
test Σ, unk. μ	100	700	0.05	0.04
test μ and Σ	100	700	0.05	0.02

9.2.1 K-S test

For the K-S test result, the value of p is 9. Using significance level All statistics pass the K-S test at a significance level of greater than .2. Thus, the optimization model and the error propagation model for the peak roof model are validated.

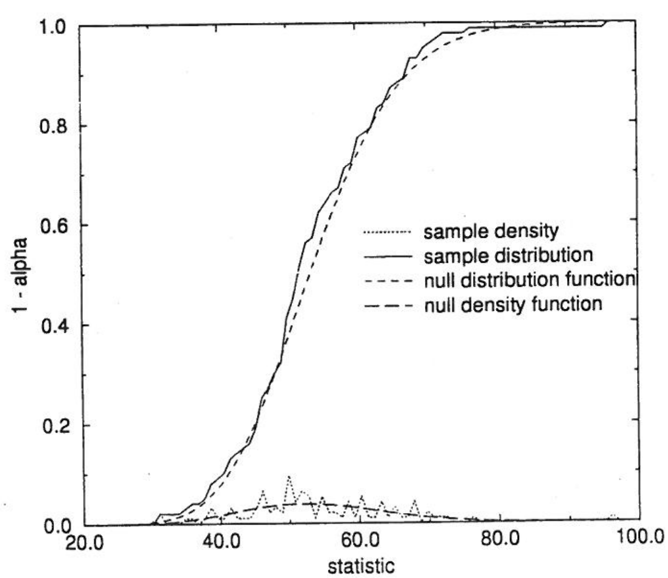

Figure 2: Test mean and covariance : peak roof model, $\sigma = 3$

9.3 Test of Hip Roof Model with $\sigma = 3.0$

9.3.1 Statistic distribution

The experimental result of hypothesis test for hip roof model with $\sigma = 3.0$ is shown in figure 3,where the x axis represents the statistic used in the test and the y axis represents $1-\alpha$, where α is the significance level.

Figure 3 shows the result of test mean and covariance simultaneously.

9.3.2 K-S test

The value of p for the K-S test is 11. All statistics pass the K-S test at a significance level of greater than .09.

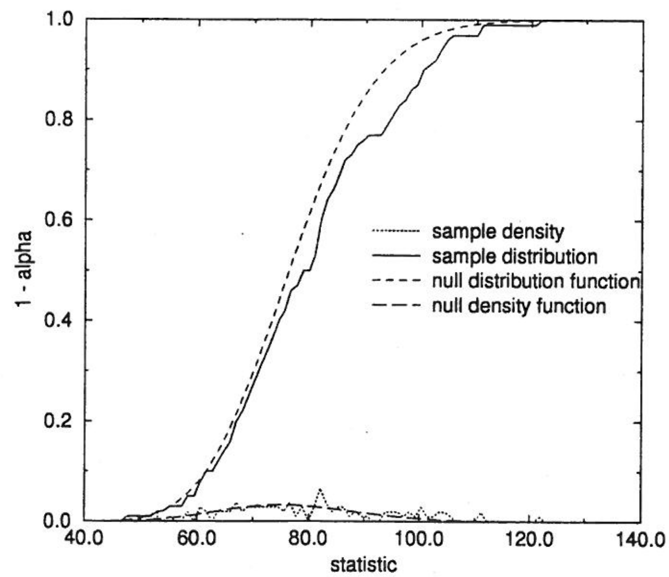

Figure 3: Test mean and covariance : hip roof model, $\sigma = 3$

Acknowledgment

We would like to thank Professors Michael Perlman and David Madigan from the statistics department, Professors Terry Rockafellar and Jim Burke from the mathematics department providing relevant literature and discussion.

References

[1] T. W. Anderson. *Introduction to Multivariate Statistical Analysis.* John Wiley & Sons, Inc., New York, NY, 1984.

[2] S. F. Arnold. *Mathematical Statistics.* Prentice-Hall, New Jersey, 1990.

[3] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.

[4] G. Casella and R. L. Berger. *Statistical Inference.* Wadsworth, CA, 1990.

[5] J. E. Dennis and Robert B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations* Englewood Cliffs, N. J. : Prentice-Hall, 1983.

[6] O. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*, section 5.6, MIT Press, Cambridge, MA, 1993.

[7] R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, New York, 1987.

[8] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*, Academic Press, New York, 1981.

[9] R. M. Haralick. "Propagating covariance in computer vision." In *Proc. of IAPR Int. Conf. on Pattern Recognition*, pages 493–498, Israel, October 1994.

[10] Dean A. Hudson, "Inverse Perspective Projection Algorithm: Error Estimation", Master Thesis, University of Washington, 1990.

[11] T. Kanungo, R. M. Haralick, and H. S. Baird. "Power functions and their use in selecting distance functions for document degradation model validation." In *Proc. of IAPR Int. Conf. on Document Analysis and Recognition*, Montreal, Canada, August 1995.

[12] T. Kanungo and R. M. Haralick. "Multivariate Hypothesis Testing for Gaussian Data: Theory and Software", ISL Tech. Report: ISL-TR-95-05, Univ. Washington, Seattle, WA, Oct., 1995

[13] K. Koch. *Parameter Estimation and Hypothesis Testing in Linear Models.* Springer-Verlag, New York, NY, 1987.

[14] Xufei Liu, R. M. Haralick and Anand Bedekar, "Optimization Methods For Estimating 3D Object Parameters", *Proceedings, 1994 Image Understanding Workshop*, Monterey, Ca, 1994, pp. 1017-1025.

[15] Xufei Liu "Modeling and performance characterization of 3D parameter estimation using perspective geometry", Ph.D. Dissertation, University of Washington, 1995.

[16] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C.* Cambridge University Press, New York, NY, 1990.