Robotica (1997) volume 15, pp 461–464. Printed in the United Kingdom © 1997 Cambridge University Press

The least squares fit of a hyperplane to uncertain data*

David B. Reister

Center for Engineering Systems Advanced Research, Oak Ridge National Laboratory, P.O. Box 2008, Building 6025, MS-6364, Oak Ridge, TN 37831-6364 (USA)

(Received in Final Form: August 15, 1996)

SUMMARY

Sensor based robotic systems are an important emerging technology. When robots are working in unknown or partially known environments, they need range sensors that will measure the Cartesian coordinates of surfaces of objects in their environment. Like any sensor, range sensors must be calibrated. The range sensors can be calibrated by comparing a measured surface shape to a known surface shape. The most simple surface is a plane and many physical objects have planar surfaces. Thus, an important problem in the calibration of range sensors is to find the best (least squares) fit of a plane to a set of 3D points.

We have formulated a constrained optimization problem to determine the least squares fit of a hyperplane to uncertain data. The first order necessary conditions require the solution of an eigenvalue problem. We have shown that the solution satisfies the second order conditions (the Hessian matrix is positive definite). Thus, our solution satisfies the sufficient conditions for a local minimum. We have performed numerical experiments that demonstrate that our solution is superior to alternative methods.

KEYWORDS: Hyperplane; Least squares fit; Sensors; Robotic systems.

1. INTRODUCTION

Despite twenty years of research, many assembly tasks are currently impossible to automate. Several automobile manufacturers have identified robotic part assembly as an emerging technology and are organizing a basic research program to develop next generation robotic technologies for assembly. They would like to develop robotic technologies with human-like dexterity and perception that could work in confined environments (for example, under the dash of a car). Their list of desired technologies includes: real-time three dimensional (3D) perception, real-time force and tactile feedback for use in automated assembly, design of a family of advanced hands, dual arm robotic assembly, and dexterous manipulators.

The primary motivation for 3D vision is to make the robotic assembly more flexible. If sensors are not used, all parts and the robots must be precisely located. If either parts or robot are out of alignment, both the parts and the robot can be damaged. Furthermore, placing parts in precise fixtures is expensive. Sensors allow the assembly to be performed with parts in uncertain locations and provide an imbedded inspection of the process.

A laser range camera is an excellent 3D vision sensor. The camera defines the Cartesian coordinates of surfaces of objects in an environment by scanning a laser beam over a scene and determines the distance (r) by the time required for the beam to reach an object and return. The beam is directed back and forth (ϕ) and up and down (θ) by two rotating mirrors. Thus, the data measured by the camera has the form (r, ϕ, θ). The geometry of the camera can be used to map the measured data to Cartesian coordinates (x, y, z).

Like any sensor, laser range cameras must be calibrated. Since the geometry of the camera may not be known precisely, the cameras can be calibrated by comparing the calculated surface shapes to the known surface shapes. The most simple surface is a plane and many physical objects have planar surfaces. Thus, an important problem in the calibration of range cameras is to find the best (least squares) fit of a plane to a set of 3D points.

For many least squares problems, the uncertainty is in one of the variables [for example, $y = f(x)$ or $z = f(x, y)$]. However, for this problem, the uncertainty is in the geometric transformation from measured data to Cartesian coordinates and all three of the calculated variables are uncertain.

We are all familiar with the linear regression problem of determining the least squares fit of data to a straight line ($y = a + bx$). Two linear equations (the normal equations) are solved to determine the parameters (a and b). For the general case with M basis functions, the M parameters are determined by solving the M normal equations. The book by Lawson[1] provides a good introduction to the methods for solving least squares problems.

When both x and y are uncertain, a line is determined by the three parameters ($a + bx + cy = 0$). To minimize the sum of the squares of the errors, the three parameters satisfy a system of three homogeneous equations [$\mathbf{Ds} = 0$ where $\mathbf{s}^{\mathrm{T}} = (a, b, c)$]. For the general case of a hyperplane in an N dimensional space, the

* Research sponsored by the Engineering Research Program, Office of Basic Energy Sciences, Oak Ridge National Laboratory managed by Lockheed Martin Energy Research Corp. for the U.S. Department of Energy under contract number DE-AC05-96OR22464.

parameters satisfy a system of $N+1$ homogeneous equations. The homogeneous equations will have a zero solution unless the matrix **D** is singular.

In a recent paper,[2] we estimated an approximate solution of the homogeneous equations by adding a small perturbation to make the matrix **D** singular and by solving the resultant eigenvalue problem [$\mathbf{Ds} = \gamma\mathbf{s}$]. The approximate solution was compared to the set of solutions obtained by neglecting one of the homogeneous equations at a time. Pratt[3] estimates the solution of the homogeneous equations by neglecting the last row of the Cholesky decomposition of the matrix **D**. Nievergelt[4] has followed a different path to derive the same algorithm that we shall develop in this paper.

In the next section, we will formulate a constrained optimization problem to determine the least squares fit of a hyperplane to uncertain data. The first order necessary conditions require the solution of an eigenvalue problem. We will show that the solution satisfies the second order conditions (the Hessian matrix is positive definite). Thus, our solution satisfies the sufficient conditions for a local minimum. In the third section, we will discuss the results of numerical experiments that demonstrate that our solution is superior to both our previous method and to Pratt's method.

2. LEAST SQUARES ESTIMATION OF A HYPERPLANE

Our motivation was to fit a plane to three dimensional range data. We will generalize our problem from three dimensions to N dimensions. We assume that we receive noisy measurements of points (**p**) in an N dimensional space:

$$\mathbf{p} = (x_1, x_2, \ldots, x_N). \tag{1}$$

The data will be a set of points ($\mathbf{p}_j$), where j ranges from 1 to M:

$$\mathbf{p}_j = (x_{1j}, x_{2j}, \ldots, x_{Nj}). \tag{2}$$

A hyperplane is defined by the following equation:

$$d + \sum_{i=1}^{N} a_i x_i = 0 \tag{3}$$

If the parameter vector **a** is a unit vector, d is the perpendicular distance from the origin to the plane. We will require that the parameter vector **a** be a unit vector:

$$\sum_{i=1}^{N} (a_i)^2 = 1 \tag{4}$$

Since the data will not lie on the hyperplane, we can define an error (e_j) for each point:

$$e_j = d + \sum_{i=1}^{N} a_i x_{ij} \tag{5}$$

We will choose the parameters (the N vector **a** and the scalar d) to minimize the weighted sum of the squares of the errors subject to the constraint that **a** is a unit vector. We use a Lagrange multiplier (λ) for the constraint and define the Lagrangian function (L) by:

$$L = \sum_{j=1}^{M} w_j (e_j)^2 + \lambda\left[1 - \sum_{i=1}^{N} (a_i)^2\right] \tag{6}$$

where the weights (w_j) are nonnegative and sum to 1.

Using equation (5), L may be written:

$$L = d^2 + 2d \sum_{i=1}^{N} a_i B_i + \sum_{i=1}^{N} \sum_{k=1}^{N} a_i a_k A_{ik} + \lambda\left[1 - \sum_{i=1}^{N} (a_i)^2\right] \tag{7}$$

where:

$$A_{ik} = \sum_{j=1}^{M} w_j x_{ij} x_{kj} \tag{8}$$

$$B_i = \sum_{j=1}^{M} w_j x_{ij} \tag{9}$$

The **A** matrix is symmetric ($A_{ij} = A_{ki}$).

The first order necessary conditions for the parameters that minimize L is that the partial derivatives of L with respect to each parameter are zero:

$$d + \sum_{i=1}^{N} a_i B_i = 0 \tag{10}$$

$$dB_k + \sum_{i=1}^{N} a_i A_{ik} - \lambda a_k = 0, \quad \text{for } k = 1 \text{ to } N. \tag{11}$$

Nievergelt [4] shows that the hyperplane must pass through the centroid of the data. Equation (10) requires that the vector **B** be a point on the hyperplane, while equation (9) defines the vector **B** to be the centroid of the data. Thus, equation (10) supports Nievergelt's conclusion.

We can use equation (10) to eliminate the parameter d from equation (11):

$$-\sum_{i=1}^{N} a_i B_i B_k + \sum_{i=1}^{N} a_i A_{ik} - \lambda a_k = 0, \quad \text{for } k = 1 \text{ to } N. \tag{12}$$

Thus, the first order necessary conditions for the parameters that minimize L [equation (12)] require the solution of an eigenvalue problem:

$$(\mathbf{G} - \lambda\mathbf{I})\mathbf{a} = 0 \tag{13}$$

where the matrix **G** is defined by:

$$G_{ik} = A_{ik} - B_i B_k \tag{14}$$

In reference 2, we did not include the constraint in the Lagrangian function (L). When the Lagrange multiplier (λ) is zero, first order necessary conditions for the

https://doi.org/10.1017/S0263574797000556 Published online by Cambridge University Press

parameters that minimize L [equations (10) and (11)] are a set of $N+1$ homogeneous equations:

$$\mathbf{Ds} = 0 \tag{15}$$

where the column vector s is defined by:

$$\mathbf{s} = \begin{bmatrix} d \\ \mathbf{a} \end{bmatrix} \tag{16}$$

and $\mathbf{D}$ is a real symmetric matrix:

$$\mathbf{D} = \begin{bmatrix} 1 & \mathbf{B} \\ \mathbf{B^T} & \mathbf{A} \end{bmatrix} \tag{17}$$

Equation (15) will have a zero solution unless the matrix $\mathbf{D}$ is singular and in general the matrix $\mathbf{D}$ will not be singular. In reference 2, we estimated an approximate solution of equation (15) by adding a small perturbation to make the matrix $\mathbf{D}$ singular and by solving the resultant eigenvalue problem:

$$\mathbf{Ds} = \gamma\mathbf{s} \tag{18}$$

The solution of equation (18) that corresponds to the eigenvalue [γ] with the smallest magnitude is our best approximate solution to equation (15).

Both our previous method and our current method require the solution of eigenvalue problems. However, the justification for the two methods are quite different. The first order necessary conditions for the parameters that minimize the constrained Lagrangian function require the solution of the first eigenvalue problem [equation (13)]. The first order necessary conditions for the parameters that minimize the unconstrained Lagrangian function require the solution of a set of $N+1$ homogeneous equations. The approximate solution of the set of homogeneous equations requires the solution of the second eigenvalue problem [equation (18)].

We conclude this section by showing that our current method satisfies the necessary conditions for a local minimum. The second partial derivative matrix of the Lagrangian function is called the Hessian matrix. Sufficient conditions for a local minimum are that the Hessian matrix is positive definite:

$$Q = \sum_{i=0}^{N}\sum_{j=0}^{N} v_i v_j H_{ij} > 0 \tag{19}$$

where $\mathbf{v}$ is an arbitrary vector (with a positive magnitude) and the Hessian matrix is defined by:

$$H_{ij} = \frac{\partial^2 L}{\partial s_i\, \partial s_j} \tag{20}$$

where $s_0 = d$ and $s_i = a_i$ for $i = 1$ to N.

For our Lagrangian function [equation (7)], the Hessian matrix is given by:

$$H_{00} = 2 \tag{21}$$

$$H_{0i} = 2B_i \tag{22}$$

$$H_{ij} = 2(A_{ij} - \lambda\delta_{ij}) \tag{23}$$

where δ_{ij} is the Kronecker delta (the elements of the identity matrix).

The quadratic form for the Hessian matrix (Q) is given by

$$Q = 2\left\{(v_0)^2 + 2v_0\sum_{i=1}^{N} v_i B_i + \sum_{i=1}^{N}\sum_{j=1}^{N} v_i v_j A_{ij} - \lambda\sum_{i=1}^{N}(v_i)^2\right\} \tag{24}$$

We can modify the third term in equation (24) by using equation (14), the definition of the $\mathbf{G}$ matrix:

$$\mathbf{v^T A v} = \mathbf{v^T}(\mathbf{G} + \mathbf{BB^T})\mathbf{v} = \mathbf{v^T G v} + (\mathbf{B^T v})^2 \tag{25}$$

To transform the quadratic form for the $\mathbf{G}$ matrix, we will make a coordinate transformation. Let the $\psi_\mathbf{n}$ be the eigenvectors for the $\mathbf{G}$ matrix:

$$\mathbf{G}\psi_\mathbf{n} = \lambda_n\psi_\mathbf{n} \tag{26}$$

Since $\mathbf{G}$ is a real symmetrical matrix, we can construct a set of eigenvectors that are an orthonormal basis for the space of N vectors. Let the columns of the matrix $\mathbf{T}$ be the eigenvectors for the $\mathbf{G}$ matrix:

$$\mathbf{T} = [\psi_\mathbf{1} \quad \psi_\mathbf{2} \quad \cdots \quad \psi_\mathbf{N}] \tag{27}$$

$\mathbf{T}$ is an orthogonal matrix. We will make a coordinate transformation from the $\mathbf{v}$ vector to a new vector ($\mathbf{r}$):

$$\mathbf{v} = \mathbf{Tr} \tag{28}$$

Using our new coordinates, the quadratic form for the $\mathbf{G}$ matrix becomes:

$$\mathbf{v^T G v} = \mathbf{r^T T^T G T r} \tag{29}$$

Since the columns of the matrix $\mathbf{T}$ are the eigenvectors for the $\mathbf{G}$ matrix:

$$\mathbf{GT} = \mathbf{TE} \tag{30}$$

where:

$$\mathbf{E} = \text{diag}\,[\lambda_1, \lambda_2, \lambda_N] \tag{31}$$

Since $\mathbf{T}$ is an orthogonal matrix, the right side of equation (29) becomes:

$$\mathbf{v^T G v} = \mathbf{r^T E r} = \sum_{i=1}^{N}\lambda_i(r_i)^2 \tag{32}$$

An orthogonal transformation preserves length:

$$\sum_{i=1}^{N}(v_i)^2 = \mathbf{v^T v} = \mathbf{r^T T^T T r} = \mathbf{r^T r} = \sum_{i=1}^{N}(r_i)^2 \tag{33}$$

Our goal is to show that the Hessian matrix is positive definite. We have modified several terms on the right side of equation (24). Using equations (25), (32) and (33), equation (24) becomes:

$$Q = 2\left\{(v_0 + \mathbf{B^T v})^2 + \sum_{i=1}^{N}(\lambda_i - \lambda)(r_i)^2\right\} \tag{34}$$

The parameter λ in equation (34) is one of the eigenvalues of G [see equation (13)]. If λ is the smallest of the eigenvalues, the second term in equation (34) will be nonnegative. The Hessian matrix is positive definite except for the following special case: assume $v_0 = 0$ and λ_1 is the smallest of the eigenvalues. Choose the $\mathbf{r}$ vector with $r_1 = 1$ and the other $r_i = 0$. Then the second term in

https://doi.org/10.1017/S0263574797000556 Published online by Cambridge University Press

equation (34) will be zero. The **v** vector is determined by equation (28). The first term will be positive unless **B** and **v** are orthogonal ($\mathbf{B}^{\mathbf{T}}\mathbf{v} = 0$).

For our previous method, the solution of equation (18) that corresponds to the eigenvalue with the smallest magnitude is our best approximate solution to equation (15). For our current method, we chose the smallest of the eigenvalues. Thus, λ can be negative and may not have the smallest magnitude.

3. NUMERICAL EXPERIMENTS

We have developed a method for obtaining the least squares fit of a hyperplane to uncertain data. In this section, we will apply our method to synthetic data sets in three dimensions. We will compare the results for our eigenvalue method with three other methods: our previous method, Pratt's method, and a partial information approach [neglecting one equation at a time in equation (15)]. The partial information approach produces four different solutions for the unknown parameters, rather than a single solution. For each method, we can calculated an error measure [the error is the square root of the first term on the right side of equation (6)]. We will find that our method produces the smallest errors.

We consider a plane in three dimensional space with parameters: $a = (0.57735,\ 0.57735,\ 0.57735)$ and $d = -1$. To create a synthetic data set, we calculate 81 data points with random errors that are proportional to a parameter (σ). Given the data set and the weights ($w_j = 1/81$), we can calculate the B vector and the A matrix and solve the eigenvalue problem. The error measure is displayed in Table I for five values of σ for the four methods (New, Old, Pratt, and Partial information). In the last section, we proved that our new method should have the minimum error. In Table I, our new method always has the lowest error, followed by: the old method, the Pratt method, and the partial information method. As the data errors increase (σ becomes larger), the errors for our new method become substantially less than the errors for the other methods (when there are no errors in the data, all four of the methods have zero errors).

The parameters a and d calculated by our new method are displayed in Table II for five values of σ. As σ increases, the magnitude of the parameter d increases.

Table I. The error measure for synthetic data set for four methods

Method	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.4$	$\sigma = 0.8$
New	0.034953	0.070039	0.140247	0.279550
Old	0.034966	0.070148	0.141092	0.285753
Pratt	0.034968	0.070182	0.141465	0.287659
P1	0.035002	0.070398	0.142629	0.293255
P2	0.035030	0.070597	0.143843	0.297911
P3	0.035031	0.070608	0.143947	0.298470
P4	0.034983	0.070259	0.141706	0.287899

Table II. The least squares estimates of parameters a and d for five synthetic data sets

σ	a_1	a_2	a_3	d
0.0	0.57735	0.57735	0.57735	−1.00000
0.1	0.57704	0.57576	0.57924	−1.05197
0.2	0.57733	0.57513	0.57958	−1.10130
0.4	0.57906	0.57591	0.57708	−1.19361
0.8	0.58441	0.58196	0.56550	−1.36040

There is not a trend in the direction parameters (a_i); sometimes they are greater than the true values and sometimes they are less.

4. CONCLUSIONS

Sensor based robotic systems are an important emerging technology. When robots are working in unknown or partially known environments, they need range sensors that will measure the Cartesian coordinates of surfaces of objects in their environment. Like any sensor, range sensors must be calibrated. The range sensors can be calibrated by comparing a measured surface shape to a known surface shape. The most simple surface is a plane and many physical objects have planar surfaces. Thus, an important problem in the calibration of range sensors is to find the best (least squares) fit of a plane to a set of 3D points.

We have developed a method for obtaining the least squares fit of a hyperplane to uncertain data. The first order necessary conditions require the solution of an eigenvalue problem. We have shown that the solution satisfies the second order conditions (the Hessian matrix is positive definite). Thus, our solution satisfies the sufficient conditions for a local minimum. We have performed numerical experiments that demonstrate that our solution is superior to alternative methods.

Acknowledgements

After reading reference 2, E.M. Oblow encouraged the author to find the exact (rather than approximate) solution to this problem. This research was performed under sponsorship of the Engineering Research Program, Office of Basic Energy Sciences, of the U.S. Department of Energy, under contract DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc.

References

1. C.L. Lawson and R.J. Hanson, *Solving Least Squares Problems,* (Prentice-Hall, Englewood Cliffs, New Jersey, 1974).
2. D.B. Reister and M.D. Morris, "A Method for Obtaining a Least Squares Fit of a Hyperplane to Uncertain Data" *ORNL/TM-12686* (Oak Ridge National Laboratory, 1994).
3. V. Pratt, "Direct Least-Squares Fitting of Algebraic Surfaces" *Computer Graphics,* **21,** 145–152 (1987).
4. Y. Nievergelt, "Total Least Squares: State-of-the-Art Regression in Numerical Analysis" *SIAM Review,* **36,** 258–264 (1994).

https://doi.org/10.1017/S0263574797000556 Published online by Cambridge University Press