2008 IEEE International Conference on
Robotics and Automation
Pasadena, CA, USA, May 19-23, 2008

Observability Index Selection for Robot Calibration

Yu Sun and John M. Hollerbach

Abstract— **This paper relates 5 observability indexes for robot calibration to the "alphabet optimalities" from the experimental design literature. These 5 observability indexes are shown to be the upper and lower bounds of one another. All observability indexes are proved to be equivalent when the design is optimal after a perfect column scaling. It is shown that when the goal is to minimize the variance of the parameters, D-optimality is the best criterion. When the goal is to minimize the uncertainty of the end-effector position, E-optimality is the best criterion. It is proved that G-optimality is equivalent to E-optimality for exact design.**

I. INTRODUCTION

Positioning the end-effector of a robot accurately requires precise calibration of the robot's kinematic parameters. Since kinematic models are ordinarily mildly nonlinear, linearization is an effective means for iteratively solving the nonlinear optimization and for qualitative analysis [2]. With a regression matrix $\mathbf{X}$, which is the Jacobian of the forward kinematics equation in terms of the kinematic parameters, most of the kinematic parameter estimation methods are based on least-squares estimation through redundant measurements on the robot position [11].

Redundant measurements serve as constraint equations to minimize the effect of noise associated with sensors. It is well known that different poses and different combinations of poses contribute to the calibration differently. Choosing a set of poses that can optimally increase the calibration goal has been an active area of research.

In order to measure the goodness of a pose set, several observability indexes have been proposed. Menq and Borm [10], [3] proposed a measure, here termed O_1, that is based on the product of the non-zero singular values of the Jacobian matrix $\mathbf{X}$. O_1 represents the volume of the data scatter, and maximizing it means that the errors of the parameters can be best observed.

Driels and Pathre [5] proposed to use the inverse of the condition number as an observability index, termed O_2 here. The condition number is the smallest non-zero singular value divided by the largest singular value of $\mathbf{X}$. O_2 reaches its maximal value at 1 when all singular values are equal; then the regression matrix $\mathbf{X}$ is well-conditioned. The error of the kinematic parameters can be observed equally well from the measurements.

Whereas O_1 emphasizes the volume of a hyper-ellipsoid whose directions are represented by the singular values, the condition number O_2 emphasizes that the singular values should all be close to each other. One criticism of O_1 has been that considering the volume of the regression matrix may result in favoring one direction over another, and this has been an argument for using O_2.

Nahvi and Hollerbach [13] proposed the minimum singular value of $\mathbf{X}$ as an observability index, here termed O_3. The minimum singular value uses the worst observability of the parameter error as the criterion. Nahvi and Hollerbach [12] also proposed as an observability index the square of smallest non-zero singular value divided by the largest singular value of $\mathbf{X}$, called the noise amplification index and referred to here as O_4. Nahvi and Hollerbach [12] compared different observability indexes, and found that O_1 was worse than the others. Indexes O_2 to O_4 were not significantly different.

In statistics, optimal experiment design theory has given rise to several data measures, of which the most significant are:

- A-optimality minimizes the trace of $(\mathbf{X}'\mathbf{X})^{-1}$ to obtain regression designs, and was first proposed by [7], [4]. There is no counterpart in the robot calibration literature.
- D-optimality maximizes the determinant of $\mathbf{X}'\mathbf{X}$, and was proposed by Wald [17] and later given the D label by Kiefer and Wolfowitz [8]. It is similar to O_1.
- E-optimality maximizes the minimum singular value of $\mathbf{X}'\mathbf{X}$ and was proposed by [6]. It is similar to O_3.
- G-optimality minimizes the maximum prediction variance, and does not have a simple expression in terms of singular values. It was first proposed by Smith [15], and given the G label by Kiefer and Wolfowitz [8].

Since these early works, other less popular alphabet optimalities have been proposed for different purposes. Optimal experimental design theory provides another point of view and a theoretical foundation for robot calibration observability indexes.

There are few papers on a basic comparisons of observability indexes, and their differences and properties are not fully understood. There is no convincing guidance on which observability indexes should be used under what conditions. In this paper, after studying the physical meanings and properties of each observability index, several theoretical conclusions are reached:

1) O_1 (D-optimality) is the best observability index to minimize the variance of the parameters.
2) Since a perfectly column scaled covariance matrix is the correlation coefficient matrix, the observability index can be computed from the correlation coefficient matrix without worrying about scaling.
3) After a perfect column scaling, all existing observ-

Y. Sun and J. M. Hollerbach are with School of Computing of University of Utah. {ysun,jmh}@cs.utah.edu

ability indexes are equivalent for optimal experimental design.
4) G-optimality is equivalent to E-optimality for exact design.
5) O_3 (E-optimality) is the best criteria to minimize the uncertainty of the end-effector position of a robot.

II. Observability Indexes

The parameters of the robot model can be identified from measurement of the end-effector pose and the joint encoder reading. Robot accuracy can be substantially improved by judicious choice of robot poses used for redundant measurements. A robot kinematic model is expressed as

$$\mathbf{f} = \mathbf{f}(\boldsymbol{w}, \boldsymbol{\phi}) + \mathbf{e_0} \tag{1}$$

where $\mathbf{f} = [x, y, z, \theta_x, \theta_y, \theta_z]'$ is the pose of the end effector, $\boldsymbol{w} = [w_1, \cdots, w_m]'$ are the robot model parameters, $\boldsymbol{\phi} = [\phi_1, \cdots, \phi_m]'$ are the readings of the encoders, and $\mathbf{e_0}$ is the measurement error of $\mathbf{f}$.

The linearized calibration equations around the nominal values of the parameters are expressed at iteration i:

$$\Delta\mathbf{f}^i = \mathbf{X}^i \Delta\boldsymbol{w}^i, \tag{2}$$

where $\Delta\mathbf{f}^i = \hat{\mathbf{f}} - \mathbf{f}^i$ is the difference between the measured pose $\mathbf{f}_i$ of the end effector and the pose $\hat{\mathbf{f}}$ calculated by forward kinematics with current parameters, $\Delta\boldsymbol{w}^i$ is the correction to be applied to the current values of the parameters, and $\mathbf{X}^i$ is the Jacobian matrix computed from the current parameters and encoder readings. After each step, the parameters are updated with the estimated differences $\Delta\boldsymbol{w}^i$, as is the Jacobian matrix. The iteration stops when a sufficient accuracy is reached.

At each step, there is a regression problem. Equation 2 can be expressed as

$$\mathbf{y} = \mathbf{X}\boldsymbol{w} + \boldsymbol{\epsilon}, \tag{3}$$

where $\mathbf{y} = \Delta\mathbf{f}$, and $\boldsymbol{w} = \Delta\boldsymbol{w}$. ϵ is the combination of measurement error and linearization error.

The end effector usually has 6 components: 3 positions and 3 orientations. The variances of the measurement error for positions and orientations are usually different, and so weighted least squares is typical used. For optimal experimental design, many techniques have been developed that use weighting with the standard deviations of the endpoint measurements. To simplify the discussion, here we assume the end-effector measurement errors have the same variance.

Using the singular value decomposition, equation 3 becomes

$$\mathbf{y} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}'\boldsymbol{w} + \boldsymbol{\epsilon}, \tag{4}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices, and for n measurements and m parameters,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_m \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

where the σ_i are the singular values of $\mathbf{X}$ with $\sigma_1 \geq \sigma_2 \cdots \geq \sigma_m \geq 0$. Let σ_{MIN} be the smallest non-zero singular value. The maximum singular value $\sigma_{MAX} = \sigma_1$.

Borm and Menq [3] showed that

$$\sigma_m \leq \frac{|\mathbf{y}|}{|\boldsymbol{w}|} \leq \sigma_1. \tag{5}$$

$\mathbf{y}$ forms an ellipsoid when $|\boldsymbol{w}|$ is constant. The ellipsoid provides a sensitivity measure of the position error resulting from parameter error.

In terms of this terminology,

- Observability index O_1 [10], [3] is the root of the product of the singular values:

$$O_1 = \frac{(\sigma_1\sigma_2\cdots\sigma_m)^{1/m}}{\sqrt{m}} \tag{6}$$

- Observability index O_2 [5] is the inverse of the condition number:

$$O_2 = \frac{\sigma_{MIN}}{\sigma_{MAX}} \tag{7}$$

- Observability index O_3 [12] is the minimum singular value:

$$O_3 = \sigma_{MIN} \tag{8}$$

- Observability index O_4 [12] is

$$O_4 = \sigma_{MIN}^2/\sigma_{MAX} \tag{9}$$

III. Physical Meanings of Observability Indexes

Depending on applications, there are two different goals of kinematic calibration. One goal is to obtain the structure model of the robot. It is to minimize the variance of the estimated unknown parameters in the robot model. The other is to build a prediction model to accurately place the robot end-effector to a certain position. It is to minimize the maximum possible uncertainty of the end-effector pose.

A. Parameter Estimation Optimization

Obtaining an unbiased estimate of the parameters with smallest variance is an optimization problem [14], [1]. The goal is to minimize the variance Var[$\hat{\boldsymbol{w}}$]. With least squares, $\boldsymbol{w}$ can be estimated as

$$\hat{\boldsymbol{w}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{10}$$

where $(\mathbf{X}'\mathbf{X})$ is assumed to be invertible.

If we assume the ϵ_i's are independent and identically distributed, i.e., Var[$\boldsymbol{\epsilon}$] $= \sigma^2\mathbf{I}_{6\times n}$, then

$$\text{Var}[\hat{\boldsymbol{w}}] = \sigma(\mathbf{X}'\mathbf{X})^{-1}. \tag{11}$$

Define

$$\mathbf{S} = (\mathbf{X}'\mathbf{X})^{-1}, \text{ and } \mathbf{M} = \mathbf{X}'\mathbf{X}. \tag{12}$$

$\mathbf{M}$ is called the information matrix, moment matrix, or covariance matrix in different literatures. The minimization of a matrix has many expressions corresponding to different physical meanings.

One expression is the determinant, which represents the volume of the confidence hyper-ellipsoid for the parameters. The minimization is expressed as

$$\min \det(\mathbf{S}) = \min \prod_{i=1}^{m} \frac{1}{\sigma_i^2} \tag{13}$$

where the squared singular values σ_i^2 are the eigenvalues of $\mathbf{M}$. This is called D-optimality in the experimental design literature and is similar to observability index O_1.

Another expression is minimizing the norm of the covariance matrix, i.e., $\min(||\mathbf{S}||)$. Let $\mathbf{v}_i$ be an eigenvector of $\mathbf{M}$ and σ_i^2 its corresponding eigenvalue. Since

$$\mathbf{M}\mathbf{v}_i = \sigma_i^2 \mathbf{v}_i,$$

the norm is

$$||\mathbf{M}\mathbf{v}_i|| = \sigma_i^2 ||\mathbf{v}_i||.$$

Due to the property of a norm,

$$||\mathbf{M}\mathbf{v}_i|| \leq ||\mathbf{M}||\ ||\mathbf{v}_i||.$$

Therefore

$$\sigma_i^2 ||\mathbf{v}_i|| \leq ||\mathbf{M}||\ ||\mathbf{v}_i||.$$

This shows that $||\mathbf{M}|| \geq \sigma_i^2$ for any σ_i, and therefore

$$||\mathbf{S}|| \leq \frac{1}{\sigma_i^2} \leq \frac{1}{\sigma_m^2}. \tag{14}$$

This shows that for any matrix norm, $1/\sigma_m^2$ is the upper bound of $||\mathbf{S}||$. To minimize $||\mathbf{S}||$ is the same as to minimize $1/\sigma_m$, which is E-optimality and also the observability index O_3.

Minimizing the trace of the covariance matrix is A-optimality, which has no existing corresponding observability index. It is

$$\min\{\sum_{i=1}^{m} \frac{1}{\sigma_i}\}. \tag{15}$$

which we define as O_5.

B. End-effector Estimation Optimization

After the parameters are estimated, the response variable can be estimated with $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{w}}$. For a robot, with the estimated parameters, the forward kinematic model produces the estimated position of the end-effector. Since there are errors in the estimated parameters, the estimated position of the end-effector has error too. The variance of the error affects the accuracy of the end-effector.

Robot calibration, in some circumstances, is carried out to achieve a much smaller end effector position error than the variance of the measurement of the end-effector for any poses. So the optimal goal in parameter estimation becomes to minimize the possible error of the end-effectors for any poses. From the definition of G-optimality [15], it is the right criterion for such circumstances.

The optimal design that minimizes the variance of parameters (D-optimality) is not necessarily the optimal design that minimizes the uncertainty of the end-effector (G-optimality). The relation between G-optimality and D-optimality has been studied in [9].

The position of an end-effector $\mathbf{f}$ is calculated with imprecise parameters $\hat{\boldsymbol{w}}$.

$$\mathbf{f} = \mathbf{f}(\boldsymbol{w}, \boldsymbol{\phi}) \tag{16}$$

where $\boldsymbol{\phi}$ are the joint angles. With linearization,

$$\mathbf{f} = \Delta\mathbf{f}(\hat{\boldsymbol{w}}, \boldsymbol{w})\hat{\boldsymbol{\phi}} + \Delta\mathbf{f}. \tag{17}$$

If we assume the model is correct and the input readings are accurate, the difference between a real position of an end-effector and a computed position of the end-effector comes from two places. One is linearization error e_1 and the other is the error e_o caused by imprecisely estimated parameters. If a linearization converges after several steps of iteration, the linearization error e_1 can be much lower than the error e_0 of the end-effector. The range of the error depends on the accuracy of the estimated parameter $\hat{\boldsymbol{w}}$ and on the pose. The offset error cannot be eliminated in the linearization iteration since the measurements of the end-effector are noisy and have much bigger variance.

The linearized model is

$$\Delta\mathbf{f} = \mathbf{X}\Delta\boldsymbol{w}, \tag{18}$$

where $\Delta\mathbf{f} = \hat{\mathbf{f}} - \mathbf{f}$ and $\mathbf{f}$ is the set of pose measurements. To minimize the maximal possible variance of $\hat{\mathbf{f}}$ is the same as to minimize $\max\{\text{Var}(\hat{\Delta\mathbf{f}})\}$ for any row $\mathbf{x}$ of Jacobian $\mathbf{X}$, where $\|\mathbf{x}\| \leq 1$. The formula is

$$\begin{aligned} &\min\{\max\{\text{Var}(\hat{\Delta\mathbf{f}}|\mathbf{x})|\mathbf{x}\}|\mathbf{X}\} \\ &= \min\{\max\{\mathbf{x}'\text{Var}[\hat{\Delta\boldsymbol{w}}]\mathbf{x}|\mathbf{x}\}|\mathbf{X}\} \\ &= \min\{\max\{\sigma^2\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}|\mathbf{x}\}|\mathbf{X}\} \end{aligned} \tag{19}$$

subject to $\|\mathbf{x}\| \leq 1$.

Since in practice, all experimental designs are exact, we suppose that it is an n-point design with one observation at $\mathbf{x}$. For an exact design, the moment matrix is

$$\mathbf{M} = \frac{1}{n}\mathbf{X}'\mathbf{X}. \tag{20}$$

The G-optimality is

$$\min\{\max\{n\mathbf{x}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}'|\mathbf{x}\}|\mathbf{X}\}. \tag{21}$$

Theorem 1: If the input variable is bounded as $\|\mathbf{x}\| \leq 1$, the exact design G-optimality is the equivalent to E-optimality

$$\min\{\max\{n\mathbf{x}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}'|\mathbf{x}\}|\mathbf{X}\} \propto \max\{\sigma^2_{MIN}\}. \tag{22}$$

Proof: First, we look at the maximization problem. After including the constraint as an additional linear term with Lagrange multiplier η, we have a regularized linear problem,

$$F(\mathbf{X}, \mathbf{x}) = \max\{n\ (\mathbf{x}'\mathbf{X}'\mathbf{X})^{-1}\mathbf{x} - \eta(\mathbf{x}'\mathbf{x} - 1)\} \tag{23}$$

To maximize for any $\mathbf{x}$, derivatives are taken:

$$\frac{\partial F(\mathbf{X}, \mathbf{x})}{\partial \mathbf{x}} = n(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x} - \eta\mathbf{x} = 0 \tag{24}$$

$$\frac{\partial F(\mathbf{x}, \mathbf{X})}{\partial \eta} = \mathbf{x}'\mathbf{x} - 1 = 0 \tag{25}$$

By the definition of eigenvalues and eigenvectors, the extrema of Equation 24 are eigenvectors of $(\mathbf{X}'\mathbf{X})^{-1}$ when $\mathbf{x}'\mathbf{x} = 1$. Some of the extrema are local maxima. Let $\mathbf{v}$ be an eigenvector of $(\mathbf{X}'\mathbf{X})^{-1}$. Then

$$\begin{aligned} F(\mathbf{X}, \mathbf{x}) &= \max\{n\mathbf{v}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{v}\} \\ &= n\ \max\{\mathbf{v}'\rho\mathbf{v}\} \\ &= n\ \max\{\rho\} \\ &= n\ \rho_{MAX} \end{aligned} \tag{26}$$

where ρ is the corresponding eigenvalue of $(\mathbf{X}'\mathbf{X})^{-1}$.

$F(\mathbf{X}, \mathbf{x})$ reaches a maximum for any design matrix when the input $\mathbf{x}$ is the eigenvector corresponding the largest eigenvalue of the information matrix.

$$\begin{aligned} \min\{F(\mathbf{X}, \mathbf{x})\} &= \min\{F(\mathbf{X}, \mathbf{v})\} \\ &= n\ \min\{\rho_{MAX}\} \\ &= \frac{n}{\max\{\sigma^2_{MIN}\}}, \end{aligned} \tag{27}$$

since $\rho_{MAX} = 1/\sigma^2_{MIN}$.

■

For the same number of measurements, G-optimality is equivalent to E-optimality (O_3). This can be verified with the well know conclusion of G-optimality, which is a sufficient condition for an experimental design $\mathbf{X}$ to satisfy the G-optimality criterion,

$$\min\{F(\mathbf{X}, \mathbf{x})\} = m. \tag{28}$$

where m is the number of parameters.

Theorem 2:

$$\frac{n}{\max\{\sigma^2_{MIN}\}} = m. \tag{29}$$

Proof: First we observe that $Tr(\mathbf{X}'\mathbf{X}) = Tr(\mathbf{X}\mathbf{X}')$. Let $\mathbf{x}'_i$ be the ith row of $\mathbf{X}$, or the ith measurement of variable in the experiment design. For an n-point experimental design,

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}'_1 \\ \vdots \\ \mathbf{x}'_i \\ \vdots \\ \mathbf{x}'_n, \end{bmatrix} \tag{30}$$

and

$$\mathbf{X}\mathbf{X}' = \begin{bmatrix} \mathbf{x}'_1 \\ \vdots \\ \mathbf{x}'_i \\ \vdots \\ \mathbf{x}'_n \end{bmatrix} [\mathbf{x}_1 \dots \mathbf{x}_i \dots \mathbf{x}_n] \tag{31}$$

The diagonal of the matrix $\mathbf{X}\mathbf{X}'$ is

$$diag(\mathbf{X}\mathbf{X}') = [\mathbf{x}'_1\ \mathbf{x}_1 \dots \mathbf{x}'_i\ \mathbf{x}_i \dots \mathbf{x}'_n\ \mathbf{x}_n] \tag{32}$$

Since the input variables are bounded, we have $\mathbf{x}'_i\mathbf{x}_i \leq 1$. Thus $Tr(\mathbf{X}'\mathbf{X}) \leq n$ and

$$\sum_{i=1}^{m} \sigma_i^2 = tr(\mathbf{X}'\mathbf{X}) \leq n. \tag{33}$$

For any moment matrix, the maximum of the minimum eigenvalue is $\sigma^2_{MIN} \leq \frac{n}{m}$. When $\|\mathbf{x}_i\| = 1$ for all i's and all the eigenvalues are equal, then $\sigma^2_{MIN} = \frac{n}{m}$. So $\frac{n}{\max\{\sigma^2_{MIN}\}} = m$ and $\frac{n}{\sigma^2_{MIN}} \geq m$. ■

Since G-optimality uses the variance of the end-effector as the criterion, it is not affected by scaling of the design matrix or the parameters. Selection of units or scales of the variables does not influence the end-effectors. For approximate experimental design, as proved in [9], G-optimality is equal to D-optimality (O_1). Robot calibration seeks designs with a small number of trials, and D-optimality is usually not close to being G-optimal.

IV. Properties of observability indexes

For parameter estimation optimization, there are a number of observability indexes to choose from. Each of them has a different physical meaning and properties. To best select an observability index for an application, several of their properties are studied.

A. Inequality

The five observability indexes are

$$O_1 = (\sigma_1\ \sigma_2\ \dots\ \sigma_m)^{\frac{1}{m}} \tag{34}$$

$$O_2 = \frac{\sigma_m}{\sigma_1} \tag{35}$$

$$O_3 = \sigma_m \tag{36}$$

$$O_4 = \frac{\sigma_m^2}{\sigma_1} \tag{37}$$

$$O_5 = \frac{1}{\frac{1}{\sigma_1} + \frac{1}{\sigma_2} + \dots + \frac{1}{\sigma_m}} \tag{38}$$

All the observability indexes described above use singular values of $\mathbf{X}$. Since $\mathbf{X}'\mathbf{X}$ is a covariance matrix and all variables are linearly independent, it is an $m \times m$ positive-definite symmetric matrix. All singular values $\sigma_1, \sigma_2, \ldots, \sigma_m$ are positive. According to arithmetic-geometric-harmonic means inequality, we have

$$\begin{array}{rcll} \sigma_{MAX} & \geq & \frac{\sigma_1+\sigma_2+\ldots+\sigma_m}{m} & \\ & \geq & (\sigma_1\ \sigma_2\ \ldots\ \sigma_m)^{\frac{1}{m}} & \text{(D-optimum/}O_1\text{)} \\ & \geq & \frac{1}{\frac{1}{\sigma_1}+\frac{1}{\sigma_2}+\ldots+\frac{1}{\sigma_m}} & \text{(A-optimum/}O_5\text{)} \\ & \geq & \sigma_{MIN} & \text{(E-optimum/}O_3\text{),} \end{array} \tag{39}$$

The relation of the observability indexes is

$$\text{mean}(\sigma) \geq O_1 \geq O_5 \geq O_3 \tag{40}$$

When $\sigma_1 = \sigma_2 = \ldots = \sigma_m$, the observability indexes O_1, O_3, and O_5 are equal. When $\sigma_{MAX} \geq 1$,

$$\sigma_{MIN}\ (O_3) \geq \frac{\sigma_{MIN}}{\sigma_{MAX}}\ (O_2). \tag{41}$$

If we can further assume $\sigma_{MIN} \leq 1$,

$$\frac{\sigma_{MIN}}{\sigma_{MAX}}\ \ (O_2)\ \geq \frac{\sigma_{MIN}^2}{\sigma_{MAX}}\ \ (O_4). \tag{42}$$

It suggests that all observability indexes are related. The indexes O_1, O_5, O_3, O_2, and O_4 are the upper bounds of their following indexes by number and lower bounds of their previous indexes.

B. Scaling

For parameter estimation, the observability index selection also highly depends on the properties of robot models. In a robot model, parameters for different variables can be vastly different. They may have different units and different scales. Kinematic parameters can be on the order of 10^{-3}, elasticity can be on the order of 10^{-6}, and actuator model parameters can be on the order of 10^{-9}.

Column scaling is usually applied to the variables. The scaling can be model based or experimental based. The approaches usually use the ranges of the variables to scale them to be bounded in $[-0.5\ 0.5]$ (for simplicity, we omit the centering). Equation 3 becomes

$$\mathbf{y} = \mathbf{X}\mathbf{P}\mathbf{P}'\boldsymbol{w} \tag{43}$$

where $\mathbf{P}$ is a scaling matrix.

However, different scaling approaches may produce different results if the observability index used is sensitive to scaling. An observability index that is invariant to scaling is of interest for robot calibration.

O_1 (D-optimality) is an observability index that is invariant under any non-singular linear transformation [16]. For any non-singular linear transformation $\mathbf{P}_{[m\times m]}$, the criterion of O_1 is the determinant of $\mathbf{P}'\mathbf{X}'\mathbf{X}\mathbf{P}$. Since

$$\det(\mathbf{P}'\mathbf{X}'\mathbf{X}\mathbf{P}) = \det(\mathbf{P}'\mathbf{P})\det(\mathbf{X}'\mathbf{X}), \tag{44}$$

and $\det(\mathbf{P}'\mathbf{P})$ is a scaler that does not change with experimental design. For any two experimental designs, the scaler does not affect the comparison result. Therefore any non-singular linear transformation, including scaling, does not affect the criterion.

The other observability indexes do not have such an attractive property. To optimize the estimation of the parameters for an unscaled robot model or a robot model without a convincing scaling approach, O_1 (D-optimality) is the best criterion.

C. Column Scaling and Correlation Matrix

One typical approach to prevent the artifact of unit selection in robot calibration is to scale the Jacobian matrix column by column. Suppose that the columns of $\mathbf{X}$ have already been centered. To scale them to unit length, let $D_j^2 = \sum_i X_{ij}^2$ and consider the new variables $X_{ij}^* = X_{ij}/D_j$.

Let

$$\mathbf{D} = \begin{bmatrix} D_1 & & 0 \\ & \ddots & \\ 0 & & D_m \end{bmatrix},\ \mathbf{D}^{-1} = \begin{bmatrix} \frac{1}{D_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{D_m} \end{bmatrix} \tag{45}$$

The regression model 4 becomes

$$\mathbf{y} = \mathbf{X}^*\boldsymbol{\psi}, \tag{46}$$

where $\mathbf{X}^* = \mathbf{X}\mathbf{D}^{-1}$ and $\boldsymbol{\psi} = \mathbf{D}\boldsymbol{w}$. Instead of estimating $\boldsymbol{w}$, we want to estimate the standardized parameter $\boldsymbol{\psi}$.

$$\boldsymbol{\psi} = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime} = \mathbf{R}_X^{-1}\mathbf{X}^{*\prime}\mathbf{y}, \tag{47}$$

where $\mathbf{R}_X = \mathbf{X}^{*\prime}\mathbf{X}^*$. Since $X_{ij}^* = X_{ij}/D_j$, the element of $\mathbf{R}_X$ at (i,j) is

$$r_{j,k} = \sum_i \frac{(X_{i,j})(X_{i,k})}{D_j D_k}. \tag{48}$$

We can see that the information matrix for a perfectly scaled design matrix is its correlation coefficient matrix

$$\mathbf{R}_J = \begin{bmatrix} 1 & r_{1,2} & \ldots & r_{1,m} \\ r_{2,1} & 1 & \ldots & r_{2,m} \\ \vdots & \vdots & & \vdots \\ r_{m,1} & r_{m,2} & \ldots & 1 \end{bmatrix}. \tag{49}$$

A nice property of a correlation coefficient matrix is that the diagonal elements are all 1's. With such a property, the scaled design matrix $\mathbf{R}_X$ introduces a constraint for all the observability indexes based on eigenvalues. It is

$$\text{Tr}(\mathbf{R}_J) = \sum_{i=1}^{m} \sigma_i^2 = m. \tag{50}$$

For O_1 (D-optimality), with $\det(\mathbf{R}_J) = \prod_{i=1}^m \sigma_i^2$,

$$\begin{array}{rcl} \log(O_1) & = & \log(det(\mathbf{R}_j) \\ & = & \sum_{i=1}^m (\log(\sigma_i^2)) \\ & = & m\text{mean}(\log(\sigma_i^2)) \\ & \leq & m\ \log(\text{mean}(\sigma_i^2)) \end{array} \tag{51}$$

which is Jensen's inequality. Since

$$\text{mean}\{\sigma_i^2 | i = [1, m]\} = \frac{1}{m}\sum_{i=1}^{m}\sigma_i^2 = 1,$$

$$m \ \log(\text{mean}(\sigma_i^2)) = 0.$$

Therefore,

$$\log(\det(\mathbf{R}_J)) \leq 0 \tag{52}$$

and

$$\sigma_{MIN}^2 \leq O_1 \leq 1 \tag{53}$$

It indicates that O_1 is bounded with $[\sigma_{MIN}, 1]$ and its maximum is 1. To find the eigenvalues to maximize the O_1, the optimal problem is expressed as

$$\max\{\prod_{i=1}^{m}\sigma_i\}, \text{subject to } \sum_{i=1}^{m}\sigma_i^2 = m, \text{ and } \forall\sigma_i > 0.$$

According to inequality of arithmetic and geometric means, the solution of the optimization problem is

$$\sigma_1 = \sigma_2 = \ldots = \sigma_m = 1. \tag{54}$$

When the hyper-ellipsoid becomes a hyper-sphere, the robot calibration reaches its maximal observability and the confidence region of the estimated parameters reaches the minimum. The O_1 (D-optimality) criterion reaches its optimum when the design matrix favors all directions. With its invariance property to scaling, we can conclude that O_1 (D-optimality) is isotropic even for unscaled data.

For E-optimality or O_3, the solution of the optimization problem of

$$\max\{\sigma_{MIN}\}, \text{ subject to } \sum_{i=1}^{m}\sigma_i^2 = m, \tag{55}$$

is also

$$\sigma_1 = \sigma_2 = \ldots = \sigma_m = 1. \tag{56}$$

For O_2, O_4 and O_5 with the constraint, $\sum_{i=1}^{m}\sigma_i^2 = m$, the optimal solutions are all the same as Equation 56. For a perfectly scaled design matrix, no matter what criterion is used, the optimal experimental designs are the same.

V. Conclusion

All 5 observability indexes are related. They are the upper and lower bounds of one another. For minimizing the combined uncertainty of the parameters of robot models, due to different interpretations of minimizing a matrix, there are a number of observability indexes with valid physical meanings. For circumstances when convincing scaling factors are hard to obtained, O_1 (D-optimality) is the best observability index using the covariance matrix for its scaling invariant property. The coefficient matrix in nature is scaling invariant. It is proved to be equal to the perfectly column-scaled covariance matrix. With such an information matrix, all 5 observability indexes produce the same optimal design.

To minimize the possible uncertainty of the end-effector pose of the robot, O_3 (E-optimality) is proved to be the best observability index. Since O_3 is also the best optimal criterion to minimize the variance of the parameters when the robot model is perfectly scaled, O_3 in general is the best observability for robot calibration with correlation coefficient matrix.

Acknowledgments

This research is supported by NIH Grant 1R21EB004600-01A2.

References

[1] A. C. Atkinson. The usefulness of optimum experimental designs. *Journal of Royal Statistical Society B*, 58(1):59–76, 1996.

[2] R. Bernhardt and S. Albright, editors. *Theory of kinematic modelling and numerical procedures for robot calibration*. Morgan Kaufmann, San Francisco, 1990.

[3] J. H. Borm and C. H. Menq. Determination of optimal measurement configurations for robot calibration based on observibility measure. *International Journal of Robotics Research*, 10(1):51–63, 1991.

[4] H. Chernoff. Locally optimum designs for estimating parameters. *Annals of Mathematical Statistics*, 24:586–602, 1953.

[5] M. R. Driels and U. S. Pathre. Significance of observation strategy on the design of robot calibration experiments. *Journal of Robotic Systems*, 7(2):197–223, 1990.

[6] S. Ehrenfeld. On the efficiencies of experimental designs. *Annals of Mathematical Statistics*, 26:247–255, 1955.

[7] G. Elfving. Optimum allocation in linear regression theory. *Annals of Mathematical Statistics*, 23:255–262, 1952.

[8] J. Fiefer and J. Wolfowitz. Optimum designs in regression problems. *Annals of Mathematical Statistics*, 30:271–294, 1959.

[9] J. Kiefer. General equivalence theory for optimum designs (approximate theory). *Annals of Statistics*, 2:849–879, 1974.

[10] C. H. Menq, J. H. Borm, and J. Z. Lai. Identification and observability measure of a basis set of error parameters in robot calibration. *ASME Journal Mechamissions, and Automation in Design*, 111(4):513–518, 1989.

[11] B. W. Mooring and M. R. D. Z. S. Roth, editors. *Fundamentals of Manipulator Calibration*. John Wilery & sons, New York, 1991.

[12] A. Nahvi and J. Hollerbach. The noise amplification index for optimal pose selection in robot calibration. In *Proc. IEEE International Conference on Robotics and Automation*, pages 647–654, 1996.

[13] A. Nahvi, J. Hollerbach, and V. Hayward. Calibration of a parallel robot using multiple kinematic closed loops. In *Proc. IEEE International Conference Robotics and Automation*, pages 407–412, 1994.

[14] F. Pukelsheim. *Optimal design of experiments*. Wiley, New York, 1993.

[15] K. Smith. On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika*, 12(1&2):1–85, 1918.

[16] R. C. St John and N. R. Draper. D-optimality for regression designs: a review. *Technometrics*, 17(1):1523, 1975.

[17] A. Wald. On the efficient design of statistical investigations. *Annals of Mathematical Statistics*, 14(2):134–140, 1943.