On Task-Oriented Criteria for Configurations Selection in Robot Calibration

Henry Carrillo, Oliver Birbach, Holger Täubig, Berthold Bäuml, Udo Frese and José A. Castellanos

Abstract— **This paper studies different criteria for selecting configurations for the task of calibrating a robotic system. Given an automatic and self-contained procedure which allows the robot to calibrate itself without the need of external tools, we are interested in how to select the set of configurations that maximize calibration accuracy while minimizing calibration time. We experiment with the active calibration of a multi-sensorial humanoid's upper body and report that determinant-based criteria should be preferred when a greedy selection is used. In addition to criteria comparison, we further propose a new criterion for configuration selection. Its novelty stems from a direct treatment of the robot's end-effector tool variance. This is contrary to previous approaches which target the variance indirectly via calibration parameters. Our proposed objective function is derived as a compact formulation from the mean error of the robot's end-effector tool from which its variance can be computed using traditional criteria known from the theory of optimal experimental design (e.g. A-optimality).**

I. INTRODUCTION

Accurately determining the values for the various parameters of the model that describes a complex system is a precondition for successful operation of numerous robotic and related tasks, such as in control, artificial intelligence and computer vision. This procedure is known as *calibration* and involves numerous repetitive tasks, e.g. designing experiments to obtain the measurements and measuring the in-out relations of the system. Moreover, the calibration procedures often require human interaction and external tools.

Keeping humans as well as external tools out of the task of robot calibration is a major concern when operating a robotic systems as it allows reduction of the maintenance time and its associated monetary costs.

Recent advances to alleviate the downtime of the robot due to calibration have focused on designing the problem as a joint calibration by using only measurements gathered from on-board sensors. An example of this trend is Birbach et al. [1], who proposed to calibrate a humanoid upper-body equipped with heterogeneous sensors, such as stereo cameras and an IMU, in an automatic and self-contained fashion by observing a marker on the robot's wrist while performing

H. Carrillo and J. A. Castellanos are with *Departamento de Informática e Ingeniería de Sistemas, Instituto de Investigación en Ingeniería de Aragón, Universidad de Zaragoza*, C/ María de Luna 1, 50018, Zaragoza, Spain. {hcarri, jacaste}@unizar.es

O. Birbach, H. Täubig and U. Frese are with German Center for Artificial Intelligence (DFKI), Cyber Physical Systems, 28359 Bremen, Germany {oliver.birbach, holger.taeubig, udo.frese}@dfki.de

B. Bäuml is with DLR Institute of Robotics and Mechatronics, Münchnerstr. 20, 82234 Wessling, Germany berthold.baeuml@dlr.de

Fig. 1. Which configuration should we use for calibration? The figure depicts three plausible robot configurations for an automatic and self-contained calibration task. The quest for minimum time but maximum accuracy in a calibration task raises the question about the effect of robot configurations on this inherent estimation problem. The above images are from DLR's mobile humanoid Justin that was used for the experiments.

heuristic configurations (i.e. manually predetermined fixed values of joint angles).

Although this proposal presents a step forward in automating the tedious task of calibrating robotic systems, the common problem of which measurements to select to carry out the calibration task is still overlooked, see also Fig. 1.

The set of measurements chosen within the procedure is crucial for the outcome of the calibration, especially when the calibration task is cast as a least-squares problem [2]–[4]. Not only does a "good" set of measurements prevents an ill-conditioned calibration problem, but taking also the impact of measurements into account allows to reduce their number while maintaining the same calibration accuracy, therefore reducing the overall calibration time.

In this paper we tackle the problem of *how to select the set of measurements for calibrating a robot* under the constraints of minimum calibration time (i.e. minimum number of measurements) and maximum calibration accuracy.

We make use of an automatic and self-contained calibration procedure [1] that allows us to specify the measurements by the robot configuration. Therefore, the criterion to optimize the calibration relates to the robot configurations. This stems from having all sensors on-board, hence the robot's configuration governs which measurements are taken.

Previous approaches [5]–[7] select the configurations using indexes or criteria that target the variance of the estimated parameters and expect to improve indirectly the error in the robot's end-effector tool, a.k.a Tool Center Point (TCP). In contrast, in this paper we propose *a selection criterion that chooses the set of configurations by directly optimizing the error at the TCP*. This also allows us to monitor the probable error at the TCP and to decide when the number of configurations is sufficient for our calibration task.

TABLE I

SUMMARY OF SELECTION CRITERIA.

Notice that given a Jacobian matrix $\mathbf{X}$ and a covariance matrix $\Sigma = (\mathbf{X}^\intercal \mathbf{X})^{-1}$, the singular values (i.e. $\sigma_1, \ldots, \sigma_m$) of $\mathbf{X}$ and the eigenvalues (i.e. $\lambda_1, \ldots, \lambda_m$) of Σ are related by $\lambda_i = \sigma_i^{-2}$.

Previous criteria	Corresponding TOED criteria
$O_A = \frac{1}{\frac{1}{\sigma_1}+\frac{1}{\sigma_2}+\cdots+\frac{1}{\sigma_m}}$	*A-opt* = $m^{-1}(\lambda_1 + \lambda_2 + \cdots + \lambda_m)$
$O_D = \frac{(\sigma_1 \sigma_2 \cdots \sigma_m)^{1/m}}{\sqrt{l}}$	*D-opt* = $(\lambda_1 \lambda_2 \cdots \lambda_m)^{1/m}$
$O_E = \sigma_{min}$	*E-opt* = λ_{max}
$O_C = \frac{\sigma_{min}}{\sigma_{max}}$	-

In addition to the proposed selection criterion, we also study the effect of using the selection criteria in conjunction with a greedy optimization to select the set of configurations. Our experiments show that, (*i*) using a greedy optimization leads most of the time to sub-optimal results in the selection of configurations, and (*ii)* that determinant based selection criteria should be preferred regarding global optimality.

The reminder of the paper is structured as follows. Section II discusses related work, emphasizing previous indexes or criteria to select the measurements for a calibration task. Section III presents the proposed selection criterion. Section IV outlines the active calibration procedure used for the experiments and Sec. V shows experimental results of the procedure in simulation and real environments. Finally, Sec. VI discusses the results and concludes the paper.

II. Related work

It is well known that in a calibration task the measurements taken will influence the accuracy of the estimated kinematic or inertial parameters of a robot, see [2] or [4]. Aiming at determining the optimal set of measurements from which the calibration parameters should be estimated, several criteria (a.k.a metrics or indexes) have been proposed, mainly to quantify the reduction in the variance of the estimated parameters.

Historically, the computation of indexes has been based on singular values of the Jacobian matrix ($\mathbf{X}$) after casting the calibration task as a least-squares problem. The general idea is to predict through the singular values (i.e. $\sigma_1, \sigma_2, \ldots, \sigma_m$) of $\mathbf{X}$, which configurations will permit better observability, hence reducing the variance of the estimated parameters.

In the above context, a first proposal called O_C was introduced by Gautier and Khalil [5]. They proposed that the criterion for selecting the measurements was based on the condition number of $\mathbf{X}$. The rationale behind that is to have a well-conditioned problem, where each measurement gives an uniform amount of information in this case. The formulation of O_C as well as the other criteria are summarized in Tab. I.

A second approach O_D is due to Borm and Menq [2], and their proposal was based on the minimum ellipsoid's error of the estimated parameters. Nahvi et al. [6] introduced a third index O_E by using the minimum singular value.

Sun and Hollerbach [8] presented a detailed study about the above indexes and discussed their relation with the criteria proposed in the theory of optimum experimental design (TOED) [9]. In the TOED the idea is to measure the uncertainty encompassed in a covariance matrix Σ resulting from a particular design ξ. Particularly, for our calibration task each of the used sets of measurements defines the design and the covariance matrix is obtained by $\Sigma = (\mathbf{X}^\intercal \mathbf{X})^{-1}$ assuming identical measurement variances. In contrast to the observability indexes, the objective of the criteria stemming from the TOED is to minimize the eigenvalues (i.e. $\lambda_1, \lambda_2, \ldots, \lambda_m$ with $\lambda_i = \sigma_i^{-2}$) of $(\mathbf{X}^\intercal \mathbf{X})^{-1}$ which results in minimizing the variance of the estimated parameters. Among the most used criteria are A-optimality [10] (based on the trace), D-optimality [10], [11] (based on the determinant) and E-optimality [12] (based on the maximum eigenvalue).

In [8], the authors also proposed the counterpart of A-optimality as an observability index, called O_A. Table I shows row-wise the relation between the observability indexes and the TOED criteria as reported in [8]. One drawback of O_A, as well O_E, is that their values are a mixture (sum-wise) of the eigenvalues which units depend on the parameters. Therefore, they often have a difficult physical interpretation.

Several previous approaches have used the above observability indexes with Jacobian matrices to determine the optimal set of measurements in order to estimate the parameters in a calibration task accurately. These include dividing randomly sampled configurations from the workspace in sets and selecting the set with best observability index [2], or a greedy search with an exchange scheme over a sampled set of configurations until a maximum number of configurations or desired accuracy is reached [6], [7], [13], [14]. The latter approach can also be cast as a variant of the algorithm for the design of optimal experiments known as DETMAX [15]. Measurement selection using an evolutionary scheme was introduced in [16], [17].

In general, the above algorithms that actively select the set of measurements for a calibration task are called active robot calibration algorithms [7].

To the best knowledge of the authors, all prior approaches for selecting the set of optimal measurements for robot calibration optimize *only* the variance of the calibration parameters and *not* directly the TCP variance, although it is the main concern in a real task: A mobile manipulator commanded to open a door using the knob needs to be calibrated to precisely reach the door knob but not to have low variance model's parameters.

With the above in mind, we propose a selection criterion for choosing the set of measurements that reduce the variance of the TCP and develop a compact formulation for the mean error of the TCP in the next section.

III. Proposed Selection Criterion

Let us suppose that the TCP position is computed by a forward kinematic function $f(q, \alpha)$, that depends on the configuration of the robot q and a set of parameters α. Also let us assume that the configuration q is given (i.e. it is a measured variable) and that the set of parameters

TABLE II

SUMMARY OF TASK ORIENTED SELECTION CRITERIA.

Notice that l is the number of Jacobian matrices $\mathbf{J}$ and $\mathbf{\Sigma} = \text{cov}(\hat{\alpha})$ is a covariance matrix.

Task oriented criteria
Task *A-opt* = $\frac{1}{l}\sum_i \text{tr}(\mathbf{J}^{(i)}\mathbf{\Sigma}\mathbf{J}^{(i)\intercal}) = \frac{1}{l}\text{tr}\left(\mathbf{\Sigma}\sum_i \mathbf{J}^{(i)\intercal}\mathbf{J}^{(i)}\right)$
Task *D-opt* = $\frac{1}{l}\sum_i \det(\mathbf{J}^{(i)}\mathbf{\Sigma}\mathbf{J}^{(i)\intercal})$
Task *E-opt* = $\frac{1}{l}\sum_i \lambda_{max}(\mathbf{J}^{(i)}\mathbf{\Sigma}\mathbf{J}^{(i)\intercal})$

α is estimated in the calibration task via least squares. Consequently, a natural criterion to assess the mean TCP error, given a set of parameters from a calibration task, is the expected value of the Euclidean (ℓ_2) error of $f(q,\alpha)$ with different parameters over the workspace of q:

$$\mathbb{E}_q[\|f(q,\hat{\alpha}) - f(q,\alpha)\|_2^2] \tag{1}$$

where $f(q,\alpha)$ refers to the true TCP position after measuring a configuration q and $f(q,\hat{\alpha})$ is the TCP position obtained with the estimated set of parameters $\hat{\alpha}$ using the same configuration.

For simplicity let's denote $f(q,\hat{\alpha})$ by r and its counterpart by $\bar{r}$. Then by properties of the trace and by the definition of the covariance it follows that:

$$\mathbb{E}_q[\|r-\bar{r}\|_2^2] = \mathbb{E}_q[\text{tr}((r-\bar{r})(r-\bar{r})^\intercal)] \tag{2}$$

$$= \text{tr}(\mathbb{E}_q[(r-\bar{r})(r-\bar{r})^\intercal]) \approx \frac{1}{l}\sum_{i=1}^{l}\text{tr}(\text{cov}(f(q^{(i)},\hat{\alpha}))) \tag{3}$$

$\mathbb{E}_q$ means the average over configurations q, which is approximated by a sum over a fixed number (e.g. l) of $q^{(i)}$ configurations samples.

One drawback with the above formulation of the criterion is that it requires to compute each covariance matrix for each TCP selected for evaluation. By carrying out a full simulation of the model such type of simulation could be computationally expensive, e.g. if we desire to test with hundreds or thousands configurations $q^{(i)}$.

One way to overcome the above is to "decouple" the covariance matrix. With the above in mind we use the error propagation law

$$\frac{1}{l}\sum_{i=1}^{l}\text{tr}(\text{cov}(f(q^{(i)},\hat{\alpha}))) \approx \frac{1}{l}\sum_{i=1}^{l}\text{tr}(\mathbf{J}^{(i)}\text{cov}(\hat{\alpha})\mathbf{J}^{(i)\intercal}) \tag{4}$$

where,

$$\mathbf{J}^{(i)} = \frac{\partial f(q^{(i)},\hat{\alpha})}{\partial \hat{\alpha}} \tag{5}$$

is the Jacobian of the model response with respect to the parameters. It should be noticed that given that $\mathbf{J}^{(i)}$ is a known constant, (4) is linear in $\text{cov}(\hat{\alpha})$. Hence, it can be formulated as a compact weighted sum of coefficients. Using the definition of vector product, indexes constrained for each particular matrix, and the trace's product property, it follows from (4):

$$\frac{1}{l}\sum_{i=1}^{l}\text{tr}(\mathbf{J}^{(i)}\text{cov}(\hat{\alpha})\mathbf{J}^{(i)\intercal}) = \frac{1}{l}\sum_{i,j,k,l} J_{jk}^{(i)}\text{cov}(\hat{\alpha})_{kl}J_{jl}^{(i)} = \tag{6}$$

$$\frac{1}{l}\sum_{k,l}\text{cov}(\hat{\alpha})_{kl}\left(\sum_{i,j}J_{jk}^{(i)}J_{jl}^{(i)}\right) = \frac{1}{l}\sum_{k,l}\text{cov}(\hat{\alpha})_{kl}\left(\sum_{i=1}^{l}\mathbf{J}^{(i)\intercal}\mathbf{J}^{(i)}\right)_{kl} \tag{7}$$

$$= \frac{1}{l}\text{tr}\left(\text{cov}(\hat{\alpha})\sum_{i=1}^{l}\mathbf{J}^{(i)\intercal}\mathbf{J}^{(i)}\right) \tag{8}$$

therefore,

$$\frac{1}{l}\sum_{i=1}^{l}\text{tr}\left(\text{cov}\left(f(q^{(i)},\hat{\alpha})\right)\right) \approx \frac{1}{l}\text{tr}\left(\text{cov}(\hat{\alpha})\sum_{i=1}^{l}\mathbf{J}^{(i)\intercal}\mathbf{J}^{(i)}\right) \tag{9}$$

In (9) a compact formulation of a first degree approximation of the proposed criterion is presented. Also in (9) the term $\sum_i \mathbf{J}^{(i)\intercal}\mathbf{J}^{(i)}$ can be precomputed. Hence, despite its approximate nature, a key advantage of using (9) over (4) is achieving a reduction in complexity, i.e. for 100 test candidate configurations $q_{(i)}$, (4) has to be computed 100 times for each set $\hat{\alpha}$ but (9) needs to be computed only once.

Another interpretation of the proposed criterion (9) can be given using the TOED [9]. There, the mean trace (i.e. A-optimality) gives the mean uncertainty encompassed in the covariance matrix, that, for our case, represents *the mean squared TCP error in meters*. One may replace the trace operator in (4) for, in this case, another uncertainty measuring operator, such as the determinant (i.e. D-optimality) or the maximum eigenvalue (i.e. E-optimality) as stated in Tab. II. We explore this possibility in Sec. V.

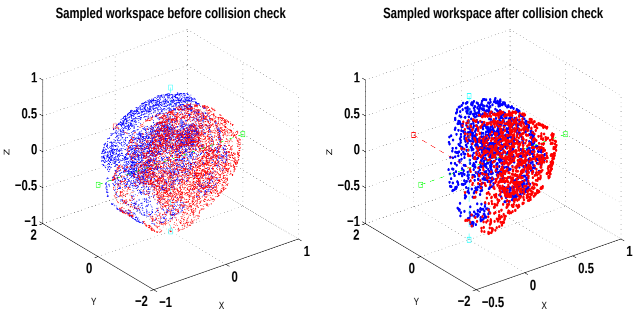
IV. EXPERIMENTAL CALIBRATION PROCEDURE

In order to test the proposed selection criterion and study its relation with previous proposals, we use a calibration procedure which actively chooses the configurations used in the calibration task. First, we produce a set of candidates configurations and check it for collisions. We then apply a greedy optimization using a selection criterion and select the configurations for calibration. Finally, the selected configurations are used to calibrate the robot.

A. Sampling and Checking Configurations

The first step is to produce feasible robot configurations for the calibration task. These configurations must be constrained by the mechanical limits of the robot and be safe, i.e. collision free. Also they should cover most of the robot's workspace to account for all possible configurations.

The generation of robot's configurations is easily achievable using the inverse kinematics of the robot, but computing the inverse kinematics of a complex system (e.g. the humanoid robot we used in the experiments) is a difficult problem. Hence, to keep the calibration problem computationally



a) b)

Fig. 2. The pose cloud obtained by the first stage of the active calibration algorithm used for experimentation. Blue markers belong to left arm and red markers to right arm. **a)** The pose cloud before checking for collisions. **b)** The remaining pose cloud after checking for collisions.

tractable, we opted for sampling the joint angle parameter space and generate a cloud of possibles configurations. From this cloud, using a nearest neighborhood algorithm, we obtain the joint angles that produce the configurations over the maximum cube circumscribed in the workspace.

We checked every configuration by applying the algorithm by Täubig et al. [18] to detect self collisions. The algorithm computes so-called swept volumes of the robot's bodies and checks them for pairwise collisions. Designed to detect collisions while moving rapidly, it operates on swept volumes defined by joint angle intervals. For checking our static configurations, we make use of these intervals to accommodate for unknown joint angle deviations. This allows us to generate safe trajectories, even though the robot has mild joint angle offsets (of up to $0.5°$).

As an example of this stage Fig. 2 shows the resulting cloud before and after checking for collision of the humanoid robot used for experiments in Sec. V.

Given that the task oriented selection criteria need feasible robot configurations for computation, the sampled configurations are divided in two groups: one of candidate configurations (A_c) and one of testing configurations (A_t). The first group contains configurations that could be selected by the greedy optimization, hence becoming configurations of the final calibration task. The second set of configurations are used by the task oriented criteria to form Jacobians.

B. Greedy Optimization

Limitations of the number of configurations used for calibration affect almost any real calibration task, which mainly arises due to time or monetary reasons. If we are given a limit of N configurations out of a set of M candidates, an exhaustive search has $\frac{M!}{N!(M-N)!}$ possibilities of choosing the N configurations. For $N = 8$ and $M = 114$ this would count for 5.5034×10^{11} needed trials.

Given the computational complexity of the exhaustive search we need to rely on sub-optimal approaches, such as greedy optimization, to select the configurations. In greedy optimization we select the configuration that gives the best output of the selection criterion each step until the maximum number of configurations has been reached. With this approach $N * M$ trials are need. For $N = 8$ and $M = 114$ this count for 912 needed trials. A practical aspect to remark is that in order to avoid rank deficient problems with infinite Σ, a plausible prior is used at the first step of the greedy optimization.

More formally, our greedy optimization is as follows:

Algorithm 1 Configuration selection by greedy optimization

Require:
- A candidate set of configurations $A_c : \{a_{c1}, \ldots, a_{cm}\}$.

Ensure:
- The optimized set of configurations $S : \{s_1, \ldots, s_n\}$.

```
Initialize S : {} .
for i = 1 to N do
    s_i := argmax_{a ∈ A_c} SelCriterion ({a} ∪ S)
end for
```

C. Calibration

Finally, after selecting the configurations, we perform the actual calibration. This part is based on the earlier mentioned calibration approach [1], where the idea is to perform a joint calibration of all the sensors by using all the data and solving the estimation problem via least squares.

The calibration has three steps. Firstly, the relation between the measured variables and the calibration parameters is stated in a set of equations. Secondly, a plausible initial estimate is given to all the parameters from a previous calibration result. Finally, the calibration parameters are estimated using least squares solving techniques, e.g. Levenberg-Marquardt. Further details on the implementation can be obtained in [1] and [19].

V. Experimental Results

A. Prerequisites

For our experiments we consider the calibration of the upper torso of a multi-sensor humanoid robot [20]. Specifically, we jointly calibrate the robot's stereo cameras and the kinematic chain. In detail, we estimate both cameras' intrinsic parameters (focal length $f_{L/R}$, principal point $C_{L/R}$ and radial distortion ($\kappa_{L/R}$)), poses of the left camera (T_H^L), the left camera relative to the right camera (stereo, T_R^L) and the correction parameters (angle offset $\theta_{\text{off},i}$, elasticity K_i^{-1}) for the joints i in the kinematic chain.

The optimization of the TCP-error in world frame is not affected by the camera calibration, therefore we use $f(q, \alpha)$ as the TCP in camera frame. This way the TCP-error includes the camera-pose error and the configuration selection also takes care of a proper camera calibration.

The formulation of the measurement functions used for the calibration follows earlier work of the authors (see [20] and [1]).

B. Calibration Result

We executed the calibration algorithm described in section IV using the proposed Task A-opt as the selection criterion. Specifically, we set up the active calibration procedure with a set of 30 candidate configurations (A_c), a

TABLE III

CALIBRATION RESULTS (EXCERPT)

	Intrinsic left			Intrinsic right		
	f_L (px)	C_L (px)	κ_L	f_R (px)	C_R (px)	κ_R
μ	1876.5	768.3, 624.2	0.09	1877.6	818.3, 620.9	0.09
σ	1.72	1.83, 1.62	$1.8 \cdot 10^{-3}$	1.93	2.05, 1.95	$2.8 \cdot 10^{-3}$

	Transformation T_R^L			Transformation T_H^L		
	Translation (m)			Translation (m)		
	x	y	z	x	y	z
μ	-0.203	-0.001	-0.002	0.071	0.104	0.129
σ	0.0001	0.0002	0.0006	0.0006	0.0006	0.0006

	Residual rms					
	Left camera (px)			Right camera (px)		
	x	y		x	y	
	0.75	0.93		0.61	1.22	

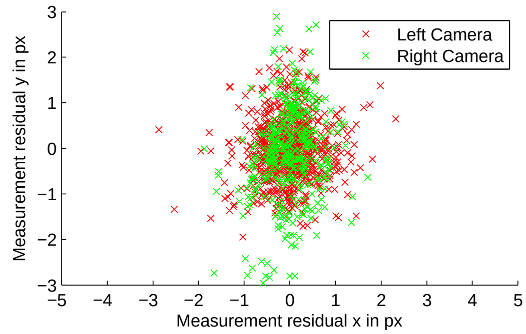


Fig. 3. Measurement residuals of left and right camera measurements.

set of 3115 testing configurations (A_t) and we selected 8 configurations for calibration (S). We uniformly sampled from the joint angle parameter space every 30 degree to obtain the configurations.

Table III presents an excerpt of the calibration results. The estimates (μ) and their σ-bounds are shown and indicate that the sensors are estimated near to their physical position and within a plausible bound. The rest of the parameters not mentioned in the table (e.g. joint angles offset) have also plausible estimates.

Figure 3 depicts the distribution of the cameras residuals and shows a fair sampled Gaussian distribution. The above in conjunction with the low camera residuals suggest that the used models fit well.

Finally, an animation of the selected configurations for calibration can be seen in the accompanying multimedia material.

C. Results on the Effect of Using the Selection Criteria with a Greedy Optimization

It is known that in general a greedy optimization cannot guarantee global optimal results. To the best of the authors' knowledge, there are no mathematical proofs that guarantee global optimal results using a greedy selection for the criteria described in Sec. II.

In order to assess the effect of using the selection criteria with a greedy optimization, we tabulated a cross-evaluation of each set of robot configuration selected with all the selection criteria. For the set of configurations obtained with a particular criterion (e.g. E-opt), we evaluated this set of configurations with the rest of criteria (e.g. Task A-opt, Task D-opt, Task E-opt and so on).

We executed the calibration algorithm described in Section IV for six selection criteria named: A-opt, D-opt, E-opt, Task A-opt, Task D-opt and Task E-opt. The first three are explained in Sec. II and its formulation is in Tab. I. The last three are explained in Sec. III and its formulation is in Tab. II. Regarding the active calibration procedure, we used a set of 114 candidate configurations (A_c), a set of 3115 testing configurations (A_t) and selected 8 configurations for calibration (S).

Table IV shows all the cross-evaluation results. Names in columns represent the selection criteria used with a greedy optimization to obtain the set of configurations. Names in rows represent the selection criteria used for evaluation. The numbers in the table can only be compared row-wise and they represent the final value of the criterion after evaluating a set of robot configurations.

Also in Tab. IV each cell has the error relative to the minimum value row-wise in parenthesis. This permits to assess the degree of sub-optimality due to the use greedy optimization for each selection criteria.

The last column of the Tab. IV shows the cross-evaluation for the set of configurations chosen heuristically. It is worth to point out that its errors, relative to the minimum row-wise, are by far the biggest. Despite using 14 instead of 8 configurations, these results suggest that it is better to optimize the selection of configurations using a mathematically developed criterion rather than optimize it heuristically or randomly, given the complex structure of the problem.

To complete the information in Tab. IV, Fig. 4 depicts the evolution of the cross-evaluation values over the set of configurations $|S|$. It can be noticed that the improvement of the criteria is roughly exponential. This indicates that with $|S| = 8$ we are still covering "new dimensions" in the parameter space and not yet just "averaging out" the error with the least squares, which would have been a value proportional to the inverse of the size of the data used in the least squares and a less rapid decay.

VI. DISCUSSION

In the following we discuss the experimental results regarding the effect of using different selection criteria and also discuss the proposed task oriented criterion.

A. Greedy Optimization and Active Robot Calibration

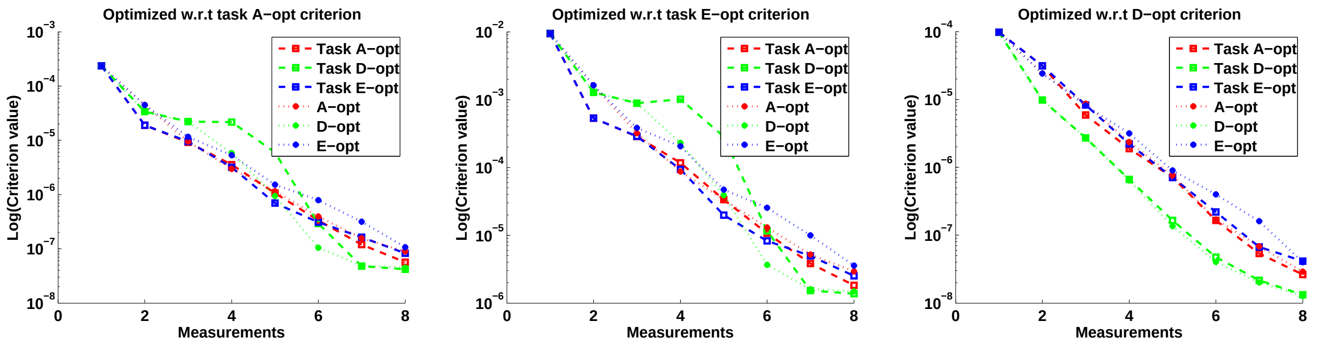
Prior work in active robot calibration (c.f. Sec. II) are mostly based on greedy like optimization algorithms, such as DETMAX, that converge quickly but cannot guarantee global optimal results.

We studied the influence of the greedy optimization in the selection criteria by cross evaluating the set of configurations selected with one criterion (e.g. E-opt) to the others (e.g. A-opt, D-opt,...). The results are tabulated in Tab. IV which show several examples were greedy selection leads to sub-optimal results. For instance, the E-opt value of the set

TABLE IV

RESULTS OF CROSS EVALUATING THE SELECTION CRITERIA WITH $|A_c| = 114$, $|A_t| = 3115$ AND $|S| = 8$. THE MINIMUM VALUE ROW-WISE IS HIGHLIGHTED IN GRAY AND THE ERROR RELATIVE TO THE MINIMUM VALUE ROW-WISE IS IN PARENTHESIS IN EACH CELL.

	Configurations						
Criteria	Task A-opt	Task D-opt	Task E-opt	A-opt	D-opt	E-opt	Heuristic @ $\|S\| = 14$
Task A-opt $\times 10^{-8}$	5.65 (1.33x)	4.24	8.09 (1.90x)	8.94 (2.10x)	4.47 (1.05x)	10.68 (2.51x)	67.16 (15.83x)
Task D-opt $\times 10^{-20}$	2.39 (1.32x)	1.81	3.28 (1.81x)	3.52 (1.94x)	1.88 (1.03x)	4.09 (2.25x)	25.07 (13.85x)
Task E-opt $\times 10^{-6}$	1.83 (1.32x)	1.38	2.51 (1.81x)	2.94 (2.13x)	1.48 (1.07x)	3.57 (2.58x)	5.16 (3.73x)
A-opt $\times 10^{-7}$	3.55 (2.55x)	1.62 (1.16x)	17.43 (12.53x)	2.71 (1.94x)	1.39	4.75 (3.41x)	359.48 (258.6x)
D-opt $\times 10^{-8}$	2.64 (2.09x)	1.97 (1.56x)	6.04 (4.79x)	2.89 (2.29x)	1.26	3.99 (3.16x)	56.12 (44.53x)
E-opt $\times 10^{-7}$	25.59 (3.07x)	8.31	215.14 (25.88x)	10.88 (1.30x)	9.33 (1.12x)	35.69 (4.29x)	4970.2 (598.0x)



a) b) c)

Fig. 4. Evolution of the cross-evaluation presented in Tab. IV over the set of configurations $|S|$. The X axis indicates the size of $|S|$ and the Y axis indicates the log value of each criteria. **a)** Set of configurations evaluated with Task A-opt. **b)** Set of configurations evaluated with Task E-opt. **c)** Set of configurations evaluated with D-opt.

chosen greedily with E-opt criterion (row 6, column 6) is 35.69×10^{-7} but the E-opt value of the set chosen greedily with Task D-opt criterion (row 6, column 2) is 8.31×10^{-7}, hence the E-opt set selection is sub-optimal against Task D-opt because the greedy optimization algorithm was trying to minimize the selection criterion E-opt.

To highlight the differences between greedy and exhaustive based active calibration, we performed a small scale example with a set of 18 candidate configurations (A_c) and selected 4 configurations for calibration. The results are tabulated in Tab. V and show that greedy optimization yields sub-optimal results while exhaustive optimization produces global optimal results. Again, the greedy optimization mislead the results of both criteria by selecting sub-optimal sets. It is worth pointing out that both criteria selected the same configurations in the exhaustive search.

From the cross-evaluation, we also observe that determinant based criteria (D-opt and Task D-opt) consistently outperform other criteria when the greedy optimization is used. These results agree with the TOED [9] where the D-opt is regarded as the criterion which measures more accurately the uncertainty in the covariance matrix. They also agree with previous works in active calibration such as [8], which reports from a theoretical perspective that determinant based criteria are the best choice for precise calibration. Also, in a loosely related field of research, such as SLAM [11], [21], the same behavior have been previously reported.

The above results suggest that for an active calibration algorithm using greedy selection, the best choice in order to have a near global optimal result is to use a determinant based criterion.

TABLE V

RESULTS OF CROSS EVALUATING THE SELECTION CRITERIA WITH $|A_c| = 18$ AND $|S| = 4$.

	Configurations			
	Greedy		Exhaustive	
Criteria	A-opt	E-opt	A-opt	E-opt
A-opt $\times 10^{-5}$	2.68 (1.44x)	4.29 (2.30x)	1.86 (1x)	1.86 (1x)
E-opt $\times 10^{-4}$	2.42 (2.14x)	3.27 (2.89x)	1.13 (1x)	1.13 (1x)

To conclude this part, it is worth to point out that the optimization algorithm used for selecting the robot configurations for a calibration task is as important as the criteria or metrics used to determine the adequacy of a robot configuration, and further research efforts should going toward global optimum procedures with low computational complexity.

B. Task Oriented Selection Criteria

A key question in active calibration of robots is how many measurements to use. This number is often limited by time or monetary reasons but beyond these, it would be ideal to select the number of measurements by the adequacy of the measurements themselves, i.e. how good is the set of measurements with regards to the calibrated parameters?

The studied selection criteria come in handy to determine the number of configurations but some of them have difficult physical interpretation that could lead to a heuristic use. For example, A-opt uses the covariance matrix Σ of the estimated parameters. Therefore, their values are a mixture (sum-wise) of the diagonal of Σ, which units depend on the parameters. In particular for our experimental calibration, A-opt's values have no intuitive physical representation as they are obtained from summing different units, such as angular units ($rads^{-1}$), translational units (m) and image units ($pixel$).

In contrast, task oriented criteria that use a covariance matrix related with the TCP have a more intuitive physical meaning. For example, the proposed task A-opt also computes its values as a mixture (sum-wise) of the diagonal. If the TCP is parametrized, as in our experiments, in a Cartesian form, the task A-opt results will be in translational units (m) allowing to know the accuracy of the TCP each time the selection criterion is used. Knowing the probable TCP accuracy allows to determine the number of configurations rationally.

Consequently, using the proposed task A-opt with a plausible selection technique will allow us to monitor the adequacy of the set of configurations and choose the number of configurations used in the calibration according to the task.

VII. Conclusion

We proposed a task oriented selection criterion that permits the assessment of the mean TCP error variance directly, which is contrary to previous approaches that target it indirectly via the calibration parameters. By that, it provides a task oriented indicator of the adequacy of the set of configuration used for calibration. This indicator may be used to limit the number of configurations, hence reducing the calibration time.

In addition to the criterion proposal, we studied the effect of using a selection criteria with a greedy like optimization algorithm as the selection procedure in an active calibration algorithm. Greedy optimization algorithms were used in many previous works because of their low computational complexity compared to an exhaustive search, but they do not provide a guarantee about global optimal results.

From our experiments, we found that (*i*) a greedy selection of configurations leads most of the time to sub-optimal set selection, thus misleading the effect of using a particular criterion, (*ii*) that determinant based criteria should be preferred regarding optimality when a greedy selection is used, and (*iii*) that automatic selection of configurations impressively outperforms previous configurations that have been chosen manually.

A. Future Work

A direct extension of this work is to integrate in a online fashion the calibration procedure in the selection of the configurations. Another direction of future research is the use of different optimization methods, such as dynamic programming, in the selection of configurations.

VIII. Acknowledgments

This work was supported by the MICINN-FEDER projects DPI2009-13710 and DPI2012-36070, research grant BES-2010-033116, DGA-FSE (grupo T04) and under DFG grant FR2620/1-1. We thank DLR's *Agile Justin* team for their inevitable technical support.

References

[1] O. Birbach, B. Bäuml, and U. Frese, "Automatic and self-contained calibration of a multi-sensorial humanoids upper body," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2012, pp. 3103–3108.

[2] J.-H. Borm and C.-H. Meng, "Determination of optimal measurement configurations for robot calibration based on observability measure," *The International Journal of Robotics Research*, vol. 10, no. 1, pp. 51–63, 1991.

[3] J. M. Hollerbach and C. W. Wampler, "The calibration index and taxonomy for robot kinematic calibration methods," *The International Journal of Robotics Research*, vol. 15, no. 6, pp. 573–591, 1996.

[4] J. Hollerbach, W. Khalil, and M. Gautier, "Model identification," in *The Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Springer, 2008.

[5] M. Gautier and W. Khalil, "Direct calculation of minimum set of inertial parameters of serial robots," *IEEE Transactions on Robotics and Automation*, vol. 6, no. 3, pp. 368–373, 1990.

[6] A. Nahvi, J. M. Hollerbach, and V. Hayward, "Calibration of a parallel robot using multiple kinematic closed loops," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 1994, pp. 407–412, Vol.1.

[7] Y. Sun and J. Hollerbach, "Active robot calibration algorithm," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2008, pp. 1276–1281.

[8] Y. Sun and J. Hollerbach, "Observability index selection for robot calibration," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2008, pp. 831–836.

[9] F. Pukelsheim, *Optimal Design of Experiments*, ser. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 2006.

[10] J. Kiefer, "General Equivalence Theory for Optimum Designs (Approximate Theory)," *The Annals of Statistics*, vol. 2, no. 5, pp. 849–879, 1974.

[11] H. Carrillo, I. Reid, and J. A. Castellanos, "On the comparison of uncertainty criteria for active slam," in *IEEE International Conference on Robotics and Automation*, 2012, pp. 2080–2087.

[12] S. Ehrenfeld, "On the efficiency of experimental designs," *The Annals of Mathematical Statistics*, vol. 26, no. 2, pp. 247–255, 1955.

[13] A. Nahvi and J. M. Hollerbach, "The noise amplification index for optimal pose selection in robot calibration," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 1996, pp. 647–654 Vol.1.

[14] D. Daney, "Optimal measurement configurations for Gough platform calibration," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2002, pp. 147–152 Vol.1.

[15] T. J. Mitchell, "An algorithm for the construction of "D-Optimal" experimental designs," *Technometrics*, vol. 16, no. 2, pp. 203–210, 1974.

[16] G. Calafiore, M. Indri, and B. Bona, "Robot dynamic calibration: Optimal excitation trajectories and experimental parameter estimation," *Journal of Robotic Systems*, vol. 18, no. 2, pp. 55–68, 2001.

[17] D. Daney, Y. Papegay, and B. Madeline, "Choosing measurement poses for robot calibration with the local convergence method and tabu search," *The International Journal of Robotics Research*, vol. 24, no. 6, pp. 501–518, 2005.

[18] H. Täubig, B. Bäuml, and U. Frese, "Real-time swept volume and distance computation for self collision detection," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011, pp. 1585–1592.

[19] R. Wagner, O. Birbach, and U. Frese, "Rapid development of manifold-based graph optimization systems for multi-sensor calibration and SLAM," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011, pp. 3305–3312.

[20] B. Bäuml, O. Birbach, T. Wimböck, U. Frese, A. Dietrich, and G. Hirzinger, "Catching flying balls with a mobile humanoid: System overview and design considerations," in *Proceedings of the IEEE-RAS International Conference on Humanoid Robotics*, 2011, pp. 513–520.

[21] H. Carrillo, Y. Latif, J. Neira, and J. A. Castellanos, "Fast minimum uncertainty search on a graph map representation," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 2504–2511.