

Consistent Parameter Clustering: Definition and Analysis

Ulrich Hillenbrand
Institute of Robotics and Mechatronics
German Aerospace Center (DLR)
82234 Wessling, Germany
Ulrich.Hillenbrand@dlr.de

Abstract

Parameter clustering is a popular robust estimation technique based on location statistics in a parameter space where parameter samples are obtained from data samples. A problem with clustering methods is that they produce estimates not invariant to transformations of the parameter space. This article presents three contributions to the theoretical study of parameter clustering. First, it introduces a probabilistic formalization of parameter clustering. Second, it uses the formalism to define consistency in terms of a symmetry requirement and to derive criteria for a consistent choice of parameterization. And third, it applies the criteria to the practically relevant cases of motion and pose estimation of three-dimensional shapes. Bias and error statistics on random data sets demonstrate a significant advantage of using a consistent parameterization for rotation clustering. Moreover, clustering parameters of analytic shapes is discussed and a real application example of circle estimation given.

Keywords: robust estimation, clustering, Hough transform, statistical consistency.

Published as *Pattern Recognition Letters* **28** (2007), 1112–1122.

1 Introduction

Parameter estimation is generally based upon some parametric model whose best match to the data is sought. Parameter estimators may be categorized according to whether they optimize a robust or non-robust match criterion,

whether they utilize statistics in the data space or in a parameter space, and whether or not they assume a parametric probability density for the data or parameters. Parameter clustering (PC) is a technique distinguished by computing robust location and perhaps dispersion statistics in a parameter space. Location and dispersion estimates do not require a density model and, hence, PC is usually realized without assuming specific densities as a so-called non-parametric technique.

The notion of clustering is used by some authors without relation to location statistics on parameters or features but with an emphasis on partitioning a data set into distinct groups. In the present context of parameter estimation, however, partitioning may be a side effect but not the goal.

The general strategy of PC has been exploited for a long time in numerous variations [2, 15, 14, 7, 9], perhaps its most popular incarnation being the many variants of the Hough transform. Common to all these approaches is that data samples are drawn from which parameter samples are computed, often called ‘votes’, that satisfy constraints posed by each data sample. The intuition is that significant data populations matching an instance of the model constraint will produce many parameter samples that coincide approximately, hence localize in a cluster.

For methods that are based directly on the data statistics [13], such as least-squares or M-estimators, the choice of parameterization of the model constraint does not affect the parameter estimate, as long as the mapping between parameters and constraint is invertible and sufficiently smooth within the relevant domain. The situation for parameter-space methods, however, is quite different in general. For example, the maximum-a-posterior estimate of a parameter does not transform as the parameters do when changing the parameterization of the model constraint. Hence, the estimates obtained in different parameterizations are not equivalent. Likewise, the result of PC is not invariant to parameter transformation.

This fact brings up the question of a proper choice of parameterization for clustering, which has not received adequate treatment in the literature. It is the purpose of this article to analyze the problem from the viewpoint of statistical consistency. The latter requires of an estimate that it matches some property of the underlying data population in the limit of having infinite data samples. If one assumes a specific parametric form of the data population, one requires that the estimated parameter should take the value underlying the given population. For PC as a non-parametric technique, on the other hand, consistency shall here be defined to require that symmetry in the data be reflected in the estimates. This will yield general criteria for choosing a consistent parameterization. The criteria will be applied to the

practically relevant cases of motion and pose estimation of three-dimensional (3D) shapes, and of straight line and circle estimation.

In the next section, I point out the difference of the present analysis to some major previous studies on consistency or parameterization in PC. Section 3 introduces the general estimation problem considered and the mathematical framework for its analysis; the PC algorithm and the concept of consistency are defined; criteria for consistency are derived. In section 4, the cases of motion and pose clustering are treated; bias and error statistics on random data sets demonstrate the advantage of using a consistent parameterization. Clustering parameters of analytic shapes is discussed in section 5; a real application example of circle estimation is given. In section 6, I note the main questions that are not addressed in the analysis. Section 7 summarizes and concludes this study.

2 Relation to previous work

By far most studies of systematic errors in PC relate to its most popular variant, the Hough transform. Among all these studies, however, only few address the issue of consistency or parameterization.

The work in [12] discusses the inconsistency of a number of robust estimators for special mixture populations of data. Although the author includes the Hough transform in his analysis, he treats it as equivalent to random sample consensus (RANSAC) and, hence, does not capture effects of the parameter space. The inconsistency analyzed in the present work is related to the parameter space and not to the data population.

Assuming a specific distribution of background data additional to the structure of interest, one may find a parameterization or parameter-space quantization such that the background contribution is uniform in parameter space. This idea is pursued in [3, 1, 8, 6] for the detection of simple planar structures. In fact, the resulting algorithm can be viewed as a statistical significance test against the null hypothesis of a pure background-data distribution. In most applications, however, the actual distribution of background data in each individual data set will itself differ significantly from any a-priori assumption, and structure detection may thus not be enhanced; cf. [6]. In the present work, the goal of parameterization is not optimal signal detection, but consistency of the parameter estimate. Adequate parameterizations do not depend on the background-data distribution, but on intrinsic properties of the parametric model constraint.

For the Hough transform, there is a relationship of the parameter-space

quantization to an equivalent template in data space, which is analyzed in [10]. The authors conclude that an adequate choice of parameterization and quantization should yield a template shape that matches the sought data populations. Their analysis does not extend to continuous parameter spaces. As above, the proposed criterion for a good parameterization is dependent upon the data population and not directly related to the consistency of a parameter estimate. In the present study, continuous parameter spaces are considered, while their uniform quantizations are trivially covered by the analysis as well. The derived criteria for a consistent parameterization are independent of the underlying data population.

The work most closely related to the present one is found in [5]. The authors discuss the adequacy of various line parameterizations for yielding consistent estimates of line orientation in the plane by the Hough transform. However, they do not provide the mathematical framework required to generalize the analysis to other estimation problems. Here, we start out from the mathematical framework, provide general criteria for consistent parameterizations, and then derive some practically relevant cases.

3 Parameter clustering

In this section, the general estimation problem and the PC approach considered are formalized. Based on the formalization, a consistency requirement is defined and necessary and sufficient conditions for consistent parameterizations are derived.

3.1 The estimation problem

Suppose we want to estimate a transformation T from a model- or data-point set $X \subset \mathbb{R}^m$ to a data-point set $Y \subset \mathbb{R}^n$. The transformation of a point $x \in \mathbb{R}^m$ is assumed to have the general parametric form

$$T(x, \alpha) = F(G_\alpha(x)) = F \circ G_\alpha(x) , \quad (1)$$

where $\{G_\alpha \mid \alpha \in \mathcal{P}\}$ is a d -dimensional Lie group charted in a parameter space $\mathcal{P} \subset \mathbb{R}^d$ and F is a continuously differentiable function on \mathbb{R}^m . For a set of corresponding point pairs $(x, y) \in C \subset X \times Y$ and a unique parameter value $\alpha \in \mathcal{P}$, we thus have the relation

$$y = T(x, \alpha) + \epsilon_{x,y} , \quad (2)$$

where $\epsilon_{x,y} \in \mathbb{R}^n$ are measurement errors. The estimation goal is to uncover the transformation $T = F \circ G_\alpha$ between the point sets X and Y .

For instance, in motion or pose estimation $\{G_\alpha \mid \alpha \in \mathcal{P}\}$ is the 6D Euclidian group acting on points $x \in \mathbb{R}^3$; or in shape estimation it may be the 3D group of deformations that maps the unit sphere onto the set of ellipsoids. The function F is the identity, if the points Y are range data, or perspective projection with lens distortion, if Y is a set of 2D image points. However, the present analysis applies equally to cases of non-visual data and estimation problems, as long as the transformation is of the general form (1).

If the point correspondences C between X and Y are known and the measurement error $\epsilon_{x,y}$ is isotropic Gaussian, the maximum-likelihood estimate is obtained through solving the least-squares problem,

$$\hat{\alpha} = \arg \min_{\alpha \in \mathcal{P}} \sum_{(x,y) \in C} \|y - T(x, \alpha)\|^2, \quad (3)$$

where $\|\cdot\|$ denotes the Euclidean norm. More general Gaussian error assumptions require usage of the Mahalanobis distance in place of the Euclidean norm.

However, point correspondences are often unknown. By far most of the point pairs $X \times Y$ will not follow the transformation (2) for any parameter value $\alpha \in \mathcal{P}$. This is true even in the ideal case that the whole data set X is mapped onto Y by a single transformation $T = F \circ G_\alpha$, and often only a subset of X will match a subset of Y . Estimating the transformation from all point pairs $X \times Y$, one is thus faced with a typically very high proportion of outliers

$$\frac{\text{card}(X) \text{card}(Y)}{\text{card}(C)} - 1 > \max(\text{card}(X), \text{card}(Y)) - 1 \gg 1, \quad (4)$$

which is why a robust estimator is needed. Moreover, accurate measurement-error models are rarely available, making a non-parametric technique desirable. This is where PC comes in and why it is so popular. For its analysis, we need a formal definition of the PC estimate.

A definition of a parameter estimate without reference to corresponding points or an error model may be given in terms of the associated parameter-probability density, which is derived as follows. Let a unique transformation $T = F \circ G_\alpha$, and hence a unique parameter value $\alpha \in \mathcal{P}$, be determined by posing the Nn constraints

$$y_i - T(x_i, \alpha) = 0, \quad i = 1, 2, \dots, N, \quad (5)$$

for any non-degenerate set of N point pairs $(x_i, y_i) \in \mathbb{R}^{m+n}$. In order to satisfy all the constraints, it may be necessary to enlarge the group $\{G_\alpha \mid \alpha \in$

\mathcal{P} by inclusion of nuisance parameters. For instance, when estimating rigid motions, such extra parameters may describe virtual deformations needed to match three point pairs. In the final estimate, however, these nuisance parameters are constrained to the values defining the original group.

In the following, we shall write bold Greek symbols $\boldsymbol{\alpha} = (\alpha, \bar{\alpha}) \in \mathcal{P} \times \mathcal{N}$ to denote the parameters $\alpha \in \mathcal{P}$ together with any required nuisance parameters $\bar{\alpha} \in \mathcal{N}$. The original group $\{G_\alpha \mid \alpha \in \mathcal{P}\}$ is a subgroup of $\{G_\alpha \mid \alpha \in \mathcal{P} \times \mathcal{N}\}$. Without loss of generality, we may assume that $G_\alpha = G_{(\alpha,0)}$ for all $\alpha \in \mathcal{P}$. We will sometimes write $\alpha \in \mathcal{P}$ as a shorthand for $(\alpha, 0) \in \mathcal{P} \times \{0\}$. For the data, we will use the bold notations $\mathbf{x} = (x_1, x_2, \dots, x_N) \in \mathbb{R}^{Nm}$, $\mathbf{y} = (y_1, y_2, \dots, y_N) \in \mathbb{R}^{Nn}$, and

$$\mathbf{y} = (T(x_1, \boldsymbol{\alpha}), T(x_2, \boldsymbol{\alpha}), \dots, T(x_N, \boldsymbol{\alpha})) \equiv T(\mathbf{x}, \boldsymbol{\alpha}) \equiv F(G_\alpha(\mathbf{x})) . \quad (6)$$

Let $p(\mathbf{x}, \mathbf{y})$ be the probability density on $\mathbb{R}^{N(m+n)}$ for N data-point pairs from $X \times Y$. Because, given N data-point pairs (\mathbf{x}, \mathbf{y}) , we have $\mathbf{y} = T(\mathbf{x}, \boldsymbol{\alpha})$ for a unique parameter value $\boldsymbol{\alpha} \in \mathcal{P} \times \mathcal{N}$, we can transform the data-pair density $p(\mathbf{x}, \mathbf{y})$ to the data-parameter density

$$\rho(\mathbf{x}, \boldsymbol{\alpha}) = p(\mathbf{x}, T(\mathbf{x}, \boldsymbol{\alpha})) |\det \partial_2 T(\mathbf{x}, \boldsymbol{\alpha})| \quad (7)$$

on $\mathbb{R}^{Nm} \times \mathcal{P} \times \mathcal{N}$. Here $\partial_2 T(\mathbf{x}, \boldsymbol{\alpha})$ is the $Nn \times Nn$ derivative of $T(\mathbf{x}, \boldsymbol{\alpha})$ w.r.t. $\boldsymbol{\alpha}$, \det denotes the determinant, and $|\cdot|$ the modulus. By integrating (7) over the data space \mathbb{R}^{Nm} , we obtain the probability density on $\mathcal{P} \times \mathcal{N}$ of measuring N point pairs related through the transformation $T = F \circ G_\alpha$,

$$\rho(\boldsymbol{\alpha}) = \int_{\mathbb{R}^{Nm}} d\mathbf{x} \rho(\mathbf{x}, \boldsymbol{\alpha}) = \int_{\mathbb{R}^{Nm}} d\mathbf{x} p(\mathbf{x}, T(\mathbf{x}, \boldsymbol{\alpha})) |\det \partial_2 T(\mathbf{x}, \boldsymbol{\alpha})| . \quad (8)$$

We shall refer to $\rho(\boldsymbol{\alpha})$ as the *parameter density*. This is the key quantity for analyzing PC.

A reasonable parameter estimate can now be defined as the location in parameter space \mathcal{P} of the maximum of the parameter density (8),

$$\hat{\alpha} = \arg \max_{\boldsymbol{\alpha} \in \mathcal{P} \times \{0\}} \rho(\boldsymbol{\alpha}) . \quad (9)$$

Indeed, the estimate (9) is a formalization of the objective of PC.

One type of PC, notably the classic-style Hough transform, does not exploit the N -fold constraints (5), but only a single one. In that case, the parameter manifold defined by the constraint has to be mapped out in \mathcal{P} . The resulting parameter density $\rho(\alpha)$ depends upon the mapping procedure, and the parameter estimate (9) is not necessarily the same as when exploiting N -fold constraints.

3.2 The algorithm

The ideal PC estimate (9) is not feasible in practice since it depends on the parameter density which in turn derives from the unknown data density. Any PC algorithm must approximate the parameter estimate (9) by drawing a finite data sample and computing the corresponding parameter sample from it. Hence, PC may proceed as follows.

1. Do until a significant sample $\{\alpha_1, \alpha_2, \dots\} \subset \mathcal{P}$ of parameter values is collected.
 - (a) Randomly draw N point pairs $\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ from $X \times Y$.
 - (b) Compute the parameter value $\alpha \in \mathcal{P} \times \mathcal{N}$ that satisfies the constraints (5) posed by the drawn point pairs. If all nuisance parameters are within the tolerance range, collect the associated parameter value $\alpha \in \mathcal{P}$.
2. Compute location and perhaps dispersion statistics for significant clusters of the collected parameter sample $\{\alpha_1, \alpha_2, \dots\} \subset \mathcal{P}$.

The estimated centers of parameter clusters may be taken as the parameter estimates for independent model instances in the data, or all the data-point pairs that contribute to a cluster may be used to re-estimate a parameter by some local, non-robust technique, such as (3). Rather than exploiting all clusters found in a single run, one may remove the data of the most prominent cluster and re-run the PC algorithm on the remaining data. This will improve the statistics for the weaker clusters.

In step 1b of the PC algorithm, we may discard a lot of parameter samples because they violate the constraint on a nuisance parameter more than tolerable. Indeed, we only want to sample the parameter density (8) within the parameter subspace \mathcal{P} . Therefore, to keep the total sampling time at a minimum, it is desirable to already constrain the data sampling procedure of step 1a to adequate sets of point pairs.

In practice, an alternative approach to actually computing nuisance parameters from the constraints (5) is to solve only for the sought parameters $\alpha \in \mathcal{P}$ by minimizing an error function analogous to (3). Samples can then be selected based on a low residual error, or properly constrained data sampling ensures that (5) is approximately satisfied.

3.3 Dependence of the estimate on the parameterization

In all of the above formalization of PC, a free choice of the parameterization of the transformation group has remained. However, the outcome of PC, i.e., the estimated transformation $T = F \circ G_{\hat{\alpha}}$ with parameter estimate $\hat{\alpha}$ given by (9), depends upon this choice. While this fact is well known, it is not always taken care of. It will be illustrated here by an example.

Suppose we estimate some positive transformation parameter $\alpha \in [0, \infty)$, say a scale factor. The parameter density for a given data population may be

$$\rho(\alpha) = \alpha e^{-\alpha} , \quad (10)$$

which is a special gamma distribution. According to (9), the ideal PC estimate is then $\hat{\alpha} = 1$.

Alternatively, we could parameterize scale with $\beta = \sqrt{\alpha} \in [0, \infty)$. The parameter density then is

$$\rho'(\beta) = 2\beta^3 e^{-\beta^2} , \quad (11)$$

another gamma distribution, and the ideal PC estimate (9) turns out to be $\hat{\beta} = \sqrt{3/2}$.

Now, while $\beta = \sqrt{\alpha}$, we are faced with $\hat{\beta} \neq \sqrt{\hat{\alpha}}$. Hence, the transformations estimated in the two parameterizations differ,

$$F \circ G_{\hat{\alpha}} \neq F \circ G_{\hat{\beta}^2} . \quad (12)$$

3.4 Consistent parameterizations

Statistical consistency usually requires of an estimator that, for every data population with underlying parameter value $\alpha \in \mathcal{P}$, the estimate would be this value α in the limit of infinite data samples. In general, this is impossible to achieve without knowledge of the specific parametric data density. In PC, we do not assume such detailed model knowledge, which makes PC a non-parametric technique.

While consistency in the usual sense is defined with reference to the specific data population for all $\alpha \in \mathcal{P}$, we now consider the opposite extreme of an α -invariant data population. This will provide us with a weaker form of consistency that does not depend upon the parametric data population.

Hence, suppose we were faced with a population of data points that is symmetric w.r.t. the group $\{G_{\alpha} \mid \alpha \in \mathcal{P}\}$, i.e., the data density is invariant to transformation with G_{α} ,

$$p(\mathbf{x}, \mathbf{y}) = p(G_{\alpha}(\mathbf{x}), \mathbf{y}) |\det \partial G_{\alpha}(\mathbf{x})| \quad \forall \alpha \in \mathcal{P} . \quad (13)$$

Consistency of PC then requires that this symmetry be reflected in the estimates, which should not be biased towards any particular transformation $T = F \circ G_{\hat{\alpha}}$ with $\hat{\alpha}$ given by (9). Clearly, the only way of insuring this is by making the parameter density associated with the symmetric data density uniform,

$$\rho(\boldsymbol{\alpha}) = \text{const.} \quad \forall \boldsymbol{\alpha} \in \mathcal{P} \times \{0\} . \quad (14)$$

In this spirit, we now define consistency for a PC estimator.

Definition 1: *Let $\{G_{\alpha} \mid \alpha \in \mathcal{P}\}$ be a parametric group and let $\{G_{\boldsymbol{\alpha}} \mid \boldsymbol{\alpha} \in \mathcal{P} \times \mathcal{N}\}$ be the super-group including the nuisance parameters required to satisfy the constraints (5). A PC estimator of parameter $\alpha \in \mathcal{P}$ is consistent, if the parameter population arising from a group-symmetric data population is uniform, that is, (14) is satisfied whenever (13) holds.*

Note that condition (13) is independent of the parameterization, while (14) is not. Therefore, requiring the relation between the two is in fact a *selection criterion for parameterizations*.

The relevance of consistency in the sense of Definition 1 is that it avoids a fundamental bias in the parameter population which otherwise derives from the topology of the underlying group. Such a bias would show up in the PC estimate (9), even in the limit of infinite data samples and for all data populations, not just for group-symmetric ones; see section 4.2.

For many groups, there cannot really be an invariant data population. This is impossible, in particular, for the non-compact groups, such as the translation group, as there cannot be an infinitely extended, homogeneous data population. Put mathematically, the associated data and parameter densities could not be normalized. However, we are not concerned here with the question of whether such a data population exists. We only want to know whether PC would deal with it the right way, if it existed. This leads us to a criterion for parameterizations consistent with the group $\{G_{\alpha} \mid \alpha \in \mathcal{P}\}$ that applies regardless of whether or not group-symmetric data exist.

Assertion 1: *Let the group and super-group be given as in Definition 1. Furthermore, let the function*

$$g : (\mathcal{P} \times \mathcal{N})^2 \rightarrow \mathcal{P} \times \mathcal{N} \quad (15)$$

be defined through the composition law in the super-group,

$$G_{g(\boldsymbol{\alpha}, \boldsymbol{\beta})} = G_{\boldsymbol{\alpha}} \circ G_{\boldsymbol{\beta}} . \quad (16)$$

For a group-invariant data density, cf. eq. (13), the associated parameter density is uniform, cf. eq. (14), if and only if the condition

$$|\det \partial_1 g(\boldsymbol{\alpha}, \boldsymbol{\beta})| = 1 \quad \forall \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathcal{P} \times \{0\} \quad (17)$$

holds for the chosen parameterization.

Proof: Transforming the points X according to $x \mapsto G_\beta(x)$, we obtain the data density

$$p'(\mathbf{x}, \mathbf{y}) = p(G_\beta^{-1}(\mathbf{x}), \mathbf{y}) \left| \det \partial G_\beta^{-1}(\mathbf{x}) \right| \quad (18)$$

and the associated parameter density

$$\begin{aligned} \rho'(\boldsymbol{\alpha}) &= \int_{\mathbb{R}^{Nm}} d\mathbf{x} p'(\mathbf{x}, T(\mathbf{x}, \boldsymbol{\alpha})) \left| \det \partial_2 T(\mathbf{x}, \boldsymbol{\alpha}) \right| \\ &= \int_{\mathbb{R}^{Nm}} d\mathbf{x} p(G_\beta^{-1}(\mathbf{x}), T(\mathbf{x}, \boldsymbol{\alpha})) \left| \det \partial G_\beta^{-1}(\mathbf{x}) \right| \left| \det \partial_2 T(\mathbf{x}, \boldsymbol{\alpha}) \right| ; \end{aligned} \quad (19)$$

cf. eq. (8). Transforming now back, $x \mapsto G_\beta^{-1}(x)$, we get

$$\rho'(\boldsymbol{\alpha}) = \int_{\mathbb{R}^{Nm}} d\mathbf{x} p(\mathbf{x}, T(G_\beta(\mathbf{x}), \boldsymbol{\alpha})) \left| \partial_2 T(G_\beta(\mathbf{x}), \boldsymbol{\alpha}) \right| . \quad (20)$$

We can exploit the super-group composition to write the relation

$$T(G_\beta(\mathbf{x}), \boldsymbol{\alpha}) = F(G_\alpha \circ G_\beta(\mathbf{x})) = T(\mathbf{x}, g(\boldsymbol{\alpha}, \beta)) , \quad (21)$$

and, taking its derivative w.r.t. $\boldsymbol{\alpha}$,

$$\partial_2 T(G_\beta(\mathbf{x}), \boldsymbol{\alpha}) = \partial_2 T(\mathbf{x}, g(\boldsymbol{\alpha}, \beta)) \partial_1 g(\boldsymbol{\alpha}, \beta) . \quad (22)$$

Let us substitute eqs. (21) and (22) into eq. (20) to find

$$\begin{aligned} \rho'(\boldsymbol{\alpha}) &= \int_{\mathbb{R}^{Nm}} d\mathbf{x} p(\mathbf{x}, T(\mathbf{x}, g(\boldsymbol{\alpha}, \beta))) \\ &\quad \times \left| \det \partial_2 T(\mathbf{x}, g(\boldsymbol{\alpha}, \beta)) \right| \left| \det \partial_1 g(\boldsymbol{\alpha}, \beta) \right| \\ &= \rho(g(\boldsymbol{\alpha}, \beta)) \left| \det \partial_1 g(\boldsymbol{\alpha}, \beta) \right| . \end{aligned} \quad (23)$$

Equation (23) states that the parameter density of the group-transformed data is equal to the super-group-transformed parameter density.

Now, let us consider the case relevant to parameter consistency. For a group-invariant data population, cf. eq. (13), we get $\rho' = \rho$, and hence from (23)

$$\rho(\boldsymbol{\alpha}) = \rho(g(\boldsymbol{\alpha}, \beta)) \left| \det \partial_1 g(\boldsymbol{\alpha}, \beta) \right| . \quad (24)$$

If the parameter density is uniform, cf. eq. (14), it follows immediately from (24) that (17) must hold. Conversely, if (17) holds, then (24) implies uniformity on $\mathcal{P} \times \{0\}$, since for every $\alpha, \gamma \in \mathcal{P}$ there is a $\beta \in \mathcal{P}$ such that $\gamma = g(\alpha, \beta)$. **q.e.d.**

In the light of Assertion 1, we can now define the consistency of a parameterization as follows.

Definition 2: *Let the group and super-group be given as in Definition 1. The parameterization with $\alpha \in \mathcal{P}$ is consistent, if condition (17) is satisfied.*

The connection with Definition 1 is, of course, that a PC estimator is consistent when clustering consistent parameters, if a group-symmetric data population exists. The latter condition is a technical requirement that is made obsolete by only discussing parameter consistency in the sense of Definition 2.

It is interesting to note that condition (24) implies that

$$\begin{aligned} \int_{\mathcal{A} \times \{0\}} d\alpha \rho(\alpha) &= \int_{\mathcal{A} \times \{0\}} d\alpha \rho(g(\alpha, \beta)) |\det \partial_1 g(\alpha, \beta)| \\ &= \int_{g(\mathcal{A}, \beta) \times \{0\}} d\gamma \rho(\gamma) , \end{aligned} \quad (25)$$

for any measurable parameter set $\mathcal{A} \subset \mathcal{P}$ and its transformation by G_β ,

$$g(\mathcal{A}, \beta) = \{\gamma \in \mathcal{P} \mid \gamma = g(\alpha, \beta) , \alpha \in \mathcal{A}\} . \quad (26)$$

Condition (25) is known as the right invariance of the measure on the group $\{G_\alpha \mid \alpha \in \mathcal{P}\}$, and it means by definition that $\rho(\alpha) d\alpha$ on $\mathcal{P} \times \{0\}$ is the right-invariant Haar measure of this group. Intuitively, an invariant Haar measure assigns equal weight to all group elements G_α . Left- and right-invariant Haar measures exist and are each unique (but not necessarily the same) up to a scale factor for all locally compact topological groups; for details see, e.g., [11], or any textbook on topological groups. In practice, the left- and right-invariant Haar measures will uniquely exist for all parameter estimation problems.

We can now state that the parameter density for group-invariant data density is equal to the right-invariant Haar measure of the group. This result is very intuitive, as every group element should have equal weight in the density. The functional form of the Haar measure, in turn, depends upon the parameterization. This insight establishes another useful formulation of the condition for parameter consistency.

Assertion 2: *Let the group be given as in Definition 1. The parameterization with $\alpha \in \mathcal{P}$ is consistent, if and only if the right-invariant Haar measure of the group is uniform in this parameterization.*

Note that only for a consistent parameterization, the invariant Haar measure can be transferred to the final PC estimates (9), such that equal

weights are obtained for the group estimates given group-invariant data density.

Assertion 2 has the special appeal that it makes no reference to any nuisance parameters or to the function F that is part of the sought transformation $T = F \circ G_\alpha$. It thus makes explicit the fact that consistency of a parameterization does not depend upon a particular choice of nuisance parameters or upon F , as long as a unique transformation T can be determined from a finite number N of data pairs from $X \times Y$; cf. eq. (5).

4 Motion and pose clustering

In this section, we specialize $\{G_\alpha | \alpha \in \mathcal{P}\}$ to the Euclidean group of motions in 3D, that is, for a point $x \in \mathbb{R}^3$,

$$G_\alpha(x) = R(\alpha)x + t(\alpha) , \quad (27)$$

with a rotation $R(\alpha) \in SO(3)$ and a translation $t(\alpha) \in \mathbb{R}^3$, parameterized by $\alpha \in \mathcal{P} \subset \mathbb{R}^6$. We want to estimate transformations $T = F \circ G_\alpha$ from points $X \subset \mathbb{R}^3$ to points Y , where F is either the identity, hence $Y \subset \mathbb{R}^3$ are range data, or a perspective projection, including any distortions, to image points $Y \subset \mathbb{R}^2$.

4.1 A consistent parameterization

A rigid motion between 3D point sets can be uniquely determined from three point pairs, hence $N = 3$ for the constraints (5). For uniquely mapping 3D model points to 2D image points, $N = 4$ point pairs are required. In both cases, we need nuisance parameters for deformations in order to satisfy (5).

Let us consider the consistency condition (17) for parameterizations. We have to analyze the parameters $\gamma = g(\alpha, \beta) \in \mathcal{P} \times \mathcal{N}$ of a composition $G_\alpha \circ G_\beta$ of a rigid motion, described by $\beta \in \mathcal{P}$, and a motion with deformations, described by $\alpha \in \mathcal{P} \times \mathcal{N}$. Under this composition, a point $x \in \mathbb{R}^3$ is transformed according to

$$G_\gamma(x) = G_\alpha \circ G_\beta(x) = M(\alpha)R(\beta)x + M(\alpha)t(\beta) + t(\alpha) , \quad (28)$$

where $M(\alpha) \in GL(3)$ represents a rotation with deformation. We can make use of the polar factorization of $M(\alpha)$ to write

$$M(\alpha) = D(\bar{\alpha})R(\alpha) , \quad (29)$$

with unique rotation $R(\alpha) \in SO(3)$, $\alpha \in \mathcal{P}$, and symmetric matrix $D(\bar{\alpha})$, $\bar{\alpha} \in \mathcal{N}$, representing the deformation.

Let us introduce the notation $\alpha = (\alpha_r, \alpha_t)$, where $\alpha_r \in \mathbb{R}^3$ and $\alpha_t \in \mathbb{R}^3$ are rotational and translational parameters, respectively. Substituting (29) into (28), we then obtain

$$\begin{aligned} G_\gamma(x) &= D(\bar{\alpha}) R(\alpha_r) R(\beta_r) x + D(\bar{\alpha}) R(\alpha_r) t(\beta_t) + t(\alpha_t) \\ &= D(\bar{\alpha}) R(\gamma_r) x + D(\bar{\alpha}) R(\alpha_r) t(\beta_t) + t(\alpha_t) , \end{aligned} \quad (30)$$

where the rotation parameter γ_r describes the composition of rotations,

$$R(\gamma_r) = R(\alpha_r) R(\beta_r) . \quad (31)$$

We can read the structure of the sought parameter dependencies from (30). Assuming that we have k nuisance parameters describing deformations, the $(6+k) \times (6+k)$ matrix of parameter derivatives has the form

$$\partial_1 g(\alpha, \beta) = \begin{pmatrix} \partial\gamma_r/\partial\alpha_r & \mathbf{0}_{3,3} & \mathbf{0}_{3,k} \\ \partial\gamma_t/\partial\alpha_r & \partial\gamma_t/\partial\alpha_t & \partial\gamma_t/\partial\bar{\alpha} \\ \mathbf{0}_{k,3} & \mathbf{0}_{k,3} & \mathbf{1}_k \end{pmatrix} , \quad (32)$$

where $\mathbf{0}_{r,s} \in \mathbb{R}^{r \times s}$ is a matrix with all coefficients zero, $\mathbf{1}_r \in \mathbb{R}^{r \times r}$ is the identity matrix, and the other terms represent, in a somewhat sloppy but intuitive notation, the α - and β -dependent partial derivatives with their respective dimensions. The determinant of (32) turns out to be simply

$$\det \partial_1 g(\alpha, \beta) = (\det \partial\gamma_r/\partial\alpha_r) (\det \partial\gamma_t/\partial\alpha_t) . \quad (33)$$

Choosing the translational parameters to be the translation vector itself, i.e., $\alpha_t \equiv t(\alpha_t)$, we can see from (30) that $\partial\gamma_t/\partial\alpha_t = \mathbf{1}_3$, such that (33) simplifies further to

$$\det \partial_1 g(\alpha, \beta) = \det \partial\gamma_r/\partial\alpha_r . \quad (34)$$

Thus, with the natural choice of translation parameters, we obtain consistent parameters for the Euclidean group, if we choose consistent parameters for the rotational subgroup.

We shall find consistent rotation parameters by the criterion given in Assertion 2. Let us consider the canonical parameters $\alpha_r = (\alpha_1, \alpha_2, \alpha_3) \in \mathcal{P}_r \subset \mathbb{R}^3$ of $SO(3)$. The canonical parameters are related to rotations through the exponential function of matrices,

$$R(\alpha_r) = \exp(\alpha_1 \Lambda_1 + \alpha_2 \Lambda_2 + \alpha_3 \Lambda_3) , \quad (35)$$

where $\Lambda_1, \Lambda_2, \Lambda_3$ are infinitesimal rotations about three orthogonal axes. In fact, $\|\alpha_r\| \in [0, \pi]$ is the angle and $\alpha_r/\|\alpha_r\|$ the oriented axis of the rotation $R(\alpha_r)$. This parameterization is one-to-one within the sphere $\mathcal{P}_r = \{\alpha_r \in \mathbb{R}^3 \mid \|\alpha_r\| < \pi\}$. Rotations with parameters $\|\alpha_r\| = \pi$ are identical to those with parameters $-\alpha_r$.

It is easy to derive consistent parameters, starting out from the canonical parameters. The rotation group is a so-called unimodular group, which means that left- and right-invariant Haar measures are the same. For symmetry reasons, it is clear that the invariant Haar measure of rotations can be written in polar form as

$$\mu_{\text{Haar}}(\alpha_1, \alpha_2, \alpha_3) d\alpha_1 d\alpha_2 d\alpha_3 = \mu(a) a^2 \sin \theta da d\phi d\theta , \quad (36)$$

where μ is an appropriate density function on \mathbb{R}_+ , $a = \|\alpha_r\|$ is the rotation angle, and ϕ and θ are azimuthal and polar angles, respectively, in the parameter space \mathcal{P}_r of canonical parameters. In consistent parameters $\alpha_r^c = (\alpha_1^c, \alpha_2^c, \alpha_3^c) \in \mathcal{P}_r^c \subset \mathbb{R}^3$, the Haar measure is uniform, that is,

$$d\alpha_1^c d\alpha_2^c d\alpha_3^c = (a^c)^2 \sin \theta da^c d\phi d\theta , \quad (37)$$

with $a^c = \|\alpha_r^c\|$. Equating (36) with (37) and solving for a^c , we arrive at

$$a^c \propto \left[\int_0^a da' \mu(a') a'^2 + \text{const.} \right]^{1/3} . \quad (38)$$

The integral is just the cumulative distribution function of the rotation angle a associated with the rotational Haar measure. Now, noting that the Haar measure on the 3D unit sphere of quaternions

$$\left(\cos a/2, \frac{\sin a/2}{a} \alpha_1, \frac{\sin a/2}{a} \alpha_2, \frac{\sin a/2}{a} \alpha_3 \right) \in S^3 \quad (39)$$

is the Lebesgue measure, we can straightforwardly calculate the cumulative distribution function as an integral over quaternions. On the quaternion sphere, we have to integrate over a polar cap around $(1, 0, 0, 0)$ down to a polar angle of $\chi = a/2$ to get the cumulative distribution function

$$P(a) = \int_0^a da' \mu(a') a'^2 \propto \int_0^{a/2} d\chi (\sin \chi)^2 = \frac{a - \sin a}{4} . \quad (40)$$

Finally, if we choose to map $\alpha_r = 0$ onto $\alpha_r^c = 0$ and $\|\alpha_r\| = \pi$ onto $\|\alpha_r^c\| = 1$, we obtain from (37), (38), and (40) the consistent parameters

$$\alpha_r^c = \left(\frac{\|\alpha_r\| - \sin \|\alpha_r\|}{\pi} \right)^{1/3} \frac{\alpha_r}{\|\alpha_r\|} . \quad (41)$$

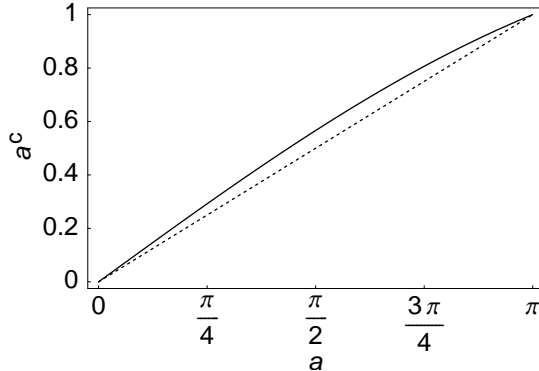


Figure 1: Consistently mapped angle $a^c = \|\alpha_r^c\|$ vs. rotation angle $a = \|\alpha_r\|$ as given by eq. (41). The straight dashed line indicates a pure rescaling of the angle and is drawn for comparison. Only the nonlinearity is effective for PC.

As in canonical parameters, $\alpha_r^c/\|\alpha_r^c\| = \alpha_r/\|\alpha_r\|$ is the oriented axis of the rotation. The consistent parameterization is one-to-one within the unit sphere $\mathcal{P}_r^c = \{\alpha_r^c \in \mathbb{R}^3 \mid \|\alpha_r^c\| < 1\}$. Rotations with parameters $\|\alpha_r^c\| = 1$ are identical to those with parameters $-\alpha_r^c$.

The transformation (41) from canonical parameters to consistent parameters is a nonlinear mapping of the rotation angle, leaving the rotation axis unchanged. Figure 1 shows a plot of the angular map. Note that a pure rescaling of the angle has no effect on the PC estimate of a transformation $T = F \circ G_{\hat{\alpha}}$ with $\hat{\alpha}$ given by (9).¹ All that matters here is the nonlinearity in the mapping. Although the nonlinear difference between canonical and consistent parameters may seem small, it will be demonstrated in section 4.2 that the effect of using consistent parameters may be quite significant.

There are other families of consistent rotation parameters which may be derived in an analogous fashion. For instance, from Euler angles (ψ, θ, ϕ) , one simply obtains $(\psi, \cos \theta, \phi)$. However, the ambiguity in ψ and ϕ at $\cos \theta = 1$ may spoil the location statistics for clustering. Therefore, the consistent parameters (41) are considered in the following.

¹Rescaling is just changing the units of measurement.

4.2 Relation of parameter consistency to estimation accuracy

In section 3.4, we have defined consistency as a symmetry requirement: symmetry in the data population w.r.t. the parametric group should be reflected in the parameter population, such that the PC estimate is uniformly distributed. There is hence no bias in a sample of consistent parameters incurred from the group topology. Conversely, a sample of inconsistent parameters will have such a fundamental bias which shows up as a bias in the PC estimate (9), even in the limit of infinite data samples.

The relation between consistency and accuracy is now illustrated for the case of rotation estimation. Bias and error statistics are presented on random point sets that are subject to random rotations with various levels of noise. Three variants of PC are tested to emphasize the universal nature of the effect of parameter consistency.

4.2.1 The data

As test data, 500 points were randomly drawn from the uniform distribution in the 3D unit sphere. The points were rotated through a random rotation, drawn from the uniform distribution on the quaternion unit sphere. Finally, independent isotropic Gaussian noise was added to each of the original and rotated points, yielding the data-point sets X and Y , respectively. The noise standard deviation was varied from $\sigma = 0.01$ to $\sigma = 0.02$ in steps of 0.001. For each value of the noise standard deviation σ , 1000 data sets X and Y were generated.

4.2.2 The algorithms

The rotation from the points X to the points Y was estimated by PC in canonical parameters (35) and in consistent parameters (41). For each parameterization, three variants of the general PC algorithm, as defined in section 3.2, were applied.

One variant of PC is in the style of an enhanced Hough transform. Parameter samples are counted in a discrete parameter space (100 bins in each dimension) and the bin that has received the maximum count is picked. All the data-point pairs that contribute to this bin are then used for a final least-squares estimate of rotation. We refer to this PC variant as the *least-squares PC*.

The second variant of PC returns the component-wise median of the parameter samples as the rotation estimate. Because rotation parameters

have a cyclic topology, however, computation of a meaningful median requires that the parameter space be reasonably centered. Thus, as in least-squares PC, we start by picking the parameter bin that has received the highest count. We then transform the rotation parameters such that this bin is at the origin of the parameter space. The median is computed in the transformed parameter space. We refer to this PC variant as the *median PC*.

The third variant of PC computes a mean-shift estimate of the location of the maximum parameter density [4]. The window size for the local averaging is adapted to the local parameter density so as to keep the number of averaged samples approximately constant. We refer to this PC variant as the *mean-shift PC*.

For all PC variants, clustering was based on 10^6 pairs of point triples, randomly sampled from X and Y , respectively. Approximate congruence of the point triples was enforced by drawing the Y -triple from a hash table with the three X -point distances as a key. The bin width in the hash table was 0.02, i.e., equal to the maximum noise standard deviation σ , in each distance dimension.

4.2.3 Angle bias

From the 1000 data sets with underlying rotations R_i , we obtained 1000 rotation estimates \hat{R}_i , $i = 1, 2, \dots, 1000$, for each noise level and PC variant. The rotation angles $\text{ang}(R_i)$ and $\text{ang}(\hat{R}_i)$ are a measure of the size of rotations, and their sample bias is

$$B^{\text{ang}} = \frac{1}{1000} \sum_{i=1}^{1000} \left[\text{ang}(\hat{R}_i) - \text{ang}(R_i) \right]. \quad (42)$$

Figure 2 shows that with increasing noise level, a significant angle bias develops for all three PC variants with canonical parameters. In fact, the size of rotation is systematically under-estimated. On the other hand, the same plots show that no such bias exists with consistent parameters.

4.2.4 Rotation error

The quantity of primary interest for evaluating an estimator is, of course, the estimation error. The angle of the rotations $R_i^{-1} \hat{R}_i$ that join the estimated and true rotations, $i = 1, 2, \dots, 1000$, is a natural error measure for rotations.

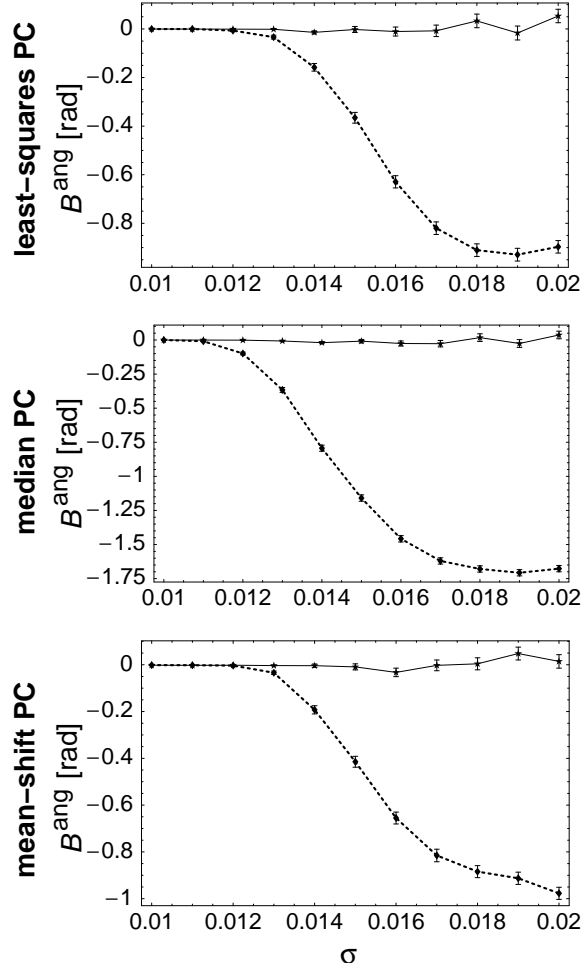


Figure 2: Angle bias B^{ang} for least-squares, median, and mean-shift PC in canonical (dashed lines) and consistent parameters (solid lines) as a function of the noise standard deviation σ . The error bars are standard errors of the mean.

The sample-mean error thus is

$$E^{\text{rot}} = \frac{1}{1000} \sum_{i=1}^{1000} \text{ang}(R_i^{-1} \hat{R}_i). \quad (43)$$

As can be seen in Fig. 3, the accuracy of all estimates degrades with increasing data noise, with the rotation error increasing on sigmoidal curves. Indeed, all tested PC algorithms are driven from a regime of accurate estimation to total break down within the range of noise applied. The accuracy of PC with consistent parameters, however, degrades more gracefully than with canonical parameters.

To more easily appreciate the advantage of consistent over canonical parameters, Fig. 4 shows the sample-mean ratio of canonical and consistent rotation errors,

$$Q^{\text{rot}} = \frac{1}{1000} \sum_{i=1}^{1000} \frac{\text{ang}(R_i^{-1} \hat{R}_i)}{\text{ang}(R_i^{-1} \hat{R}_i^{\text{c}})}, \quad (44)$$

where \hat{R}_i is now the estimate with canonical and \hat{R}_i^{c} with consistent parameters, as a function of the sample-mean rotation error (43) with consistent parameters. It can be seen that consistent parameters outperform canonical parameters throughout, from the regime of accurate estimation to total break down. In fact, the canonical error grows more than an order of magnitude larger than the consistent error for all PC variants. At a reasonably accurate consistent estimate with error ≤ 0.1 rad, the canonical estimate has an error of up to 0.8 rad for least-squares PC, up to 2.5 rad for median PC, and up to 2.0 rad for mean-shift PC. Only at the very lowest error close to zero is the advantage of using consistent parameters negligible. Note that the dramatic improvement in estimation accuracy results entirely from a moderate change in parameterization; cf. Fig. 1.

5 Clustering parameters of analytic shapes

There are estimation problems that may seem not to fit under the general form defined in eq. (2). In particular, one often seeks in a data set the parameters of a family of shapes that is given in an analytic form, such as lines, circles, or quadrics. In such cases, there is no other data set to be matched to the given one.

A parametric family of shapes, however, may usually be generated by applying a group of transformations to a single ‘root’ shape from that family. Estimating the shape parameters is hence equivalent to estimating this

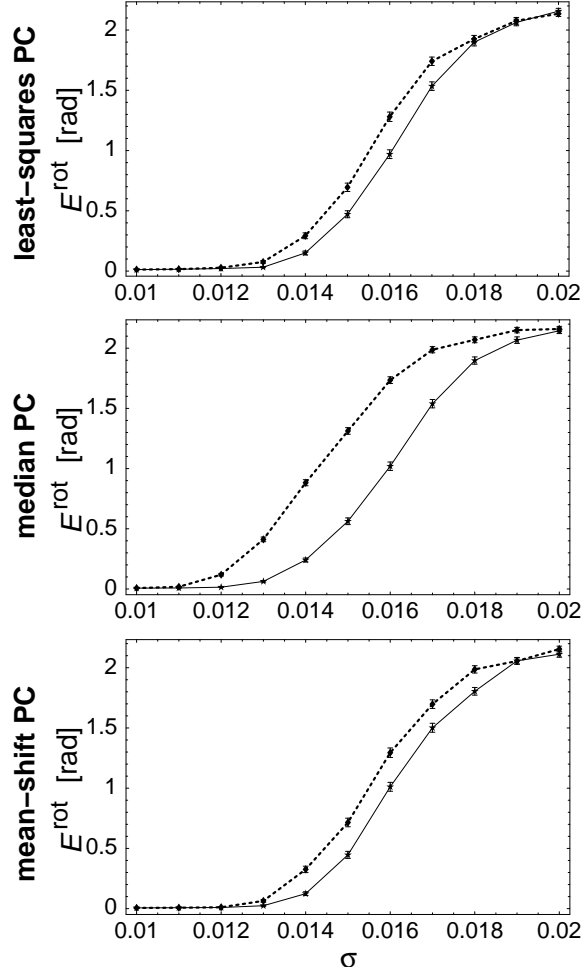


Figure 3: Mean rotation error E^{rot} for least-squares, median, and mean-shift PC in canonical (dashed lines) and consistent parameters (solid lines) as a function of the noise standard deviation σ . The error bars are standard errors of the mean.

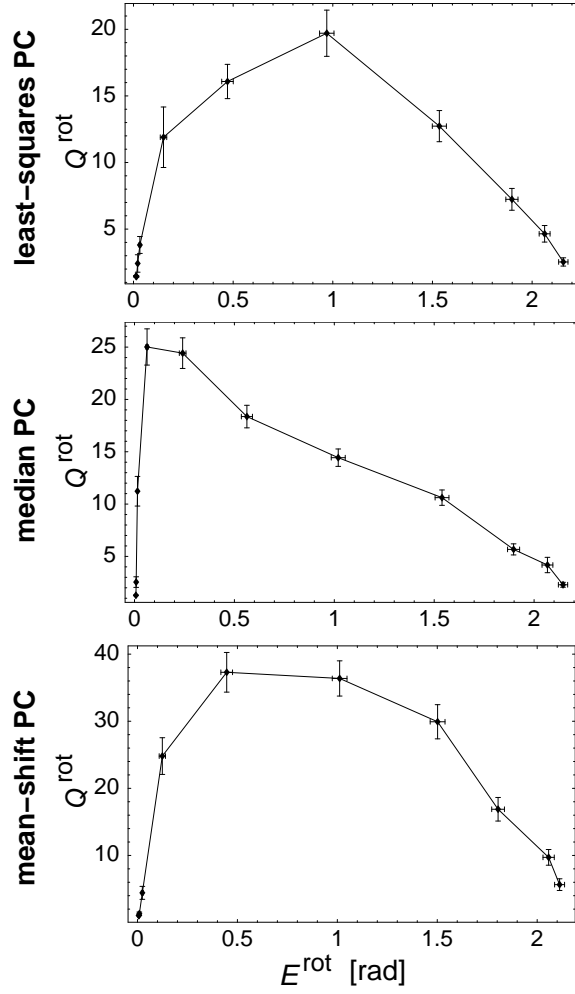


Figure 4: Mean ratio Q^{rot} of canonical and consistent rotation errors for least-squares, median, and mean-shift PC as a function of the mean consistent rotation error E^{rot} . The error bars are standard errors of the mean.

transformation. Moreover, the analytic form of the root shape defines a (usually continuous) point set. In this way, matching of analytic shapes to data is just a special case of the general framework developed here.

A prominent example of analytic shape matching is straight line detection in the plane. Any straight line may be generated from, say, the line $\{(x, y) \in \mathbb{R}^2 \mid y = 0\}$ through rotations and translations along the rotated normal. Choosing the rotation angle $\phi \in [0, \pi)$ and translation distance $t \in \mathbb{R}$ as the parameters, a combination of two such motions (ϕ_1, t_1) and (ϕ_2, t_2) yields the parameters $(\phi_3, t_3) = (\phi_1 + \phi_2, t_1 + t_2)$, which is just a translation in parameter space. Evidently, this parameterization fulfills condition (17) and is thus consistent. It turns out that insuring consistency leads to the usual recommendations for line parameterization [5].

As another example, consider the case of circle estimation in the plane. A general circle may be generated from the unit circle through translation and scaling. The naive parameterization by center $(c_x, c_y) \in \mathbb{R}^2$ and radius $r \in \mathbb{R}_+$ does not fulfill condition (17), but $(c_x, c_y, \ln r)$ does. The potential effect of adopting the consistent parameters instead of the naive ones is now illustrated by an example.

For automated satellite capturing, it is useful to process images of the satellite nozzle; see Fig. 5. The goal is to estimate the parameters of the nozzle edge from points determined by some edge operator. A PC algorithm operating in the naive parameter space is based upon the parameter density $\rho(c_x, c_y, r)$, one operating in the consistent parameter space is based upon $\rho_c(c_x, c_y, \ln r)$. According to the general rules for transforming densities, they are related through

$$\rho_c(c_x, c_y, \ln r) = r \rho(c_x, c_y, r) . \quad (45)$$

Let (c_x^*, c_y^*, r^*) denote the correct values for the nozzle center and radius. In order to visualize the effect of the different radial parameterizations, Fig. 6 shows the conditional densities

$$\rho(r | c_x^*, c_y^*) = \frac{\rho(c_x^*, c_y^*, r)}{\int_{\mathbb{R}_+} dr' \rho(c_x^*, c_y^*, r')} , \quad (46)$$

$$\rho_c(\ln r | c_x^*, c_y^*) = \frac{\rho_c(c_x^*, c_y^*, \ln r)}{\int_{\mathbb{R}} d(\ln r') \rho_c(c_x^*, c_y^*, \ln r')} , \quad (47)$$

and the marginal densities

$$\rho(r) = \int_{\mathbb{R}^2} dc_x dc_y \rho(c_x, c_y, r) , \quad (48)$$

$$\rho_c(\ln r) = \int_{\mathbb{R}^2} dc_x dc_y \rho_c(c_x, c_y, \ln r) , \quad (49)$$

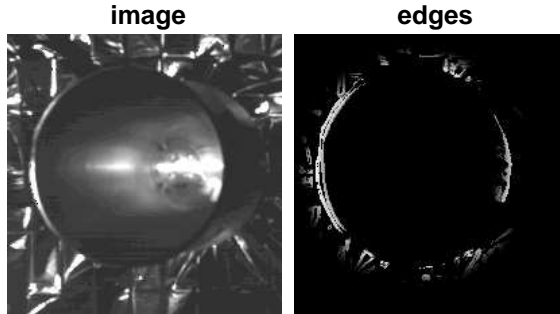


Figure 5: Image of a satellite nozzle and its edge density. The latter was computed from local intensity variations, exploiting prior knowledge on nozzle position as predicted from previous image frames.

or, more precisely, the respective empirical frequencies of parameter samples. Evidently, while samples with the correct radius r^* or log-radius $\ln r^*$ dominate at the correct center position (c_x^*, c_y^*) , there is a dominant mode of smaller circles among all samples. These outlier samples derive largely from reflections on the right inner side of the nozzle that contribute a significant share of edge points; cf. Fig. 5. The problem for a PC algorithm is to find the correct mode around (c_x^*, c_y^*, r^*) or $(c_x^*, c_y^*, \ln r^*)$, respectively, without getting biased by the many outlier samples. In this situation, it is significant that the bias towards smaller circles is just slightly weaker among the log-radial samples than among the radial samples; cf. the lower plot in Fig. 6. Indeed, a mean-shift procedure estimates a circle radius of 71.7 pixels in the (c_x, c_y, r) -parameter space and of 72.7 pixels in the $(c_x, c_y, \ln r)$ -parameter space, the correct value being $r^* = 74.1$ pixels. The radial estimation error made with naive parameterization is hence about 70% larger than with consistent parameterization. For the center position, the estimate obtained with naive parameterization is a non-significant 0.04% worse than that obtained with consistent parameterization.

Note that, although the outliers form the dominant mode in the marginal parameter densities (48) and (49), they do not dominate the joint parameter densities $\rho(c_x, c_y, r)$ and $\rho_c(c_x, c_y, \ln r)$ that respectively drive the naive and consistent PC algorithms. In fact, if they did, it would be impossible to get close to the correct parameters with any parameterization.

Other images may, of course, give rise to a more favorable distribution of parameter samples than in the example given here. In such cases, using the consistent instead of the naive parameterization may not have a significant

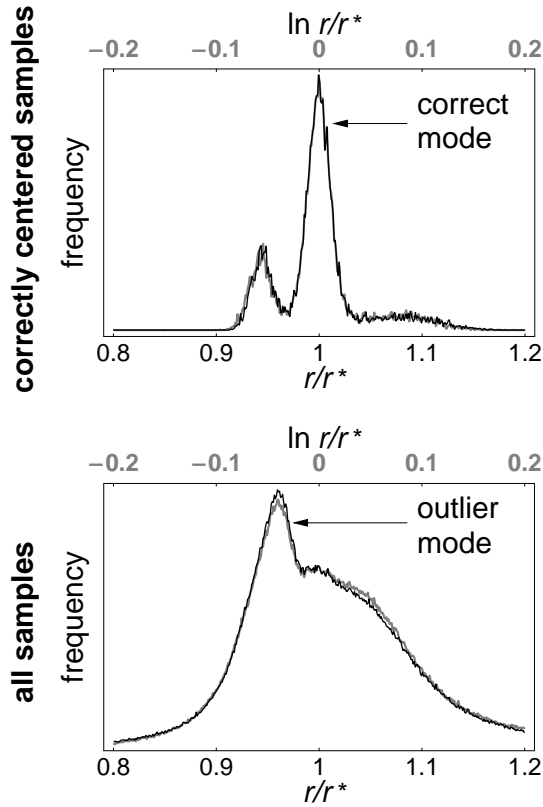


Figure 6: Relative frequencies of radial parameter samples (black thin curves) and log-radial parameter samples (gray thick curves) computed from the edge points shown in Fig. 5. The upper plot shows the frequencies for the samples with correct (± 1 pixel) circle center, corresponding to the conditional densities (46) and (47); the lower plot shows the frequencies for all samples, corresponding to the marginal densities (48) and (49). The radial values are plotted relative to the correct nozzle radius r^* .

effect on the estimation result.

6 Open questions

Two main questions remain open, a theoretical and an empirical one. Theoretically, it is interesting to know whether a set of consistent parameters exists for all estimation problems. From the argument in section 3.4, it is clear that such parameters exist at least for all compact Lie groups. Such groups have an invariant Haar measure on a bounded parameter space. Consistent parameters are those in which this measure is uniform; cf. Assertion 2. Moreover, in section 4.1, the Euclidean group has been treated as an example of a non-compact group with consistent parameters.

In this theoretical study, we could not address the empirical question of how large the benefit of applying consistent parameters to various real cases of PC actually is. There is no representative example application. Rather, the size of the effect will vary with the nature of the estimation problem and of the data. Indeed, a large range of canonical-to-consistent error ratios has been seen in section 4.2.4 by varying the noise level of random data. This study demonstrates, however, that the improvement gained by using consistent parameters can be quite significant. Checking the effect of parameter consistency seems therefore worthwhile in every application of PC.

7 Conclusion

In this article, a probabilistic framework has been introduced for analyzing PC algorithms. A notion of consistency has been defined that requires the parameter population and, hence, the parameter estimates to reflect the symmetry of the data population. A consistent parameter space is thus characterized by the absence of bias in the parameter sample. Criteria for consistency have been derived and it has been shown that consistent parameterizations depend only upon intrinsic properties of the underlying parametric group, notably its invariant Haar measure. As an example, the practically relevant cases of motion and pose estimation of 3D shapes have been treated and a consistent parameterization derived. Moreover, estimation of analytic shapes has been discussed.

The relation of parameter consistency to estimation bias and error has been illustrated for rotation estimation with statistics on random data sets. In particular, a dramatic improvement in accuracy of more than an order of

magnitude has been demonstrated when using consistent instead of canonical parameters. Furthermore, a significant improvement has been seen in a real application example of consistent circle estimation.

Although optimality in terms of expected estimation error cannot be generally proven, the theoretical and experimental results presented suggest that a consistent parameterization should be tried when estimating parameters by PC.

Two important research issues remain. One is to answer the theoretical question of whether a consistent parameterization exists for all estimation problems. The other is to apply consistent parameterizations to various kinds of estimation problems and empirically study the improvement in accuracy gained in each case.

References

- [1] V. S. Alagar and L. H. Thiel. Algorithms for detecting M-dimensional objects in N-dimensional spaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 3:245–256, 1981.
- [2] D. H. Ballard. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, 13:111–122, 1981.
- [3] M. Cohen and G. T. Toussaint. On the detection of structures in noisy pictures. *Pattern Recogn.*, 9:95–98, 1977.
- [4] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24:603–619, 2002.
- [5] Z. Y. Hu and S. D. Ma. The three conditions of a good line parameterization. *Pattern Recogn. Let.*, 16:385–388, 1995.
- [6] Z. Y. Hu and S. D. Ma. Uniform line parameterization. *Pattern Recogn. Let.*, 17:503–507, 1996.
- [7] J. Illingworth and J. Kittler. A survey of the Hough transform. *CVGIP*, 44:87–116, 1988.
- [8] H. Maitre. Contribution to the prediction of performances of the Hough transform. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8:669–674, 1986.
- [9] S. Moss, R. C. Wilson, and E. R. Hancock. A mixture model for pose clustering. *Patt. Recogn. Let.*, 20:1093–1101, 1999.
- [10] J. Princen, J. Illingworth, and J. Kittler. A formal definition of the Hough transform: properties and relationships. *J. Math. Imaging Vision*, 1:153–168, 1992.

- [11] L. A. Santaló. *Integral Geometry and Geometric Probability*. Addison-Wesley, Reading, MA, 1976.
- [12] C. V. Stewart. Bias in robust estimation caused by discontinuities and multiple structures. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19:818–833, 1997.
- [13] C. V. Stewart. Robust parameter estimation in computer vision. *SIAM Review*, 41:513–537, 1999.
- [14] G. Stockmann. Object recognition and localization via pose clustering. *CVGIP*, 40:361–387, 1987.
- [15] G. Stockmann, S. Kopstein, and S. Benett. Matching images to models for registration and object detection via clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 4:229–241, 1982.