On the optimality of DIA-estimators theory and applications

Teunissen, P. J.G.

DOI
10.1007/s00190-024-01859-w

Publication date
2024

Document Version
Final published version

Published in
Journal of Geodesy

Citation (APA)
Teunissen, P. J. G. (2024). On the optimality of DIA-estimators: theory and applications. *Journal of Geodesy*, *98*(5), Article 43. https://doi.org/10.1007/s00190-024-01859-w

Journal of Geodesy (2024) 98:43
https://doi.org/10.1007/s00190-024-01859-w

ORIGINAL ARTICLE

On the optimality of DIA-estimators: theory and applications

P. J. G. Teunissen[1,2]

Received: 26 December 2023 / Accepted: 13 April 2024

Abstract

In this contribution, we introduce, in analogy to penalized ambiguity resolution, the concept of penalized misclosure space partitioning, with the goal of directing the performance of the DIA-estimator towards its application-dependent tolerable risk objectives. We assign penalty functions to each of the decision regions in misclosure space and use the distribution of the misclosure vector to determine the optimal partitioning by minimizing the mean penalty. As each minimum mean penalty partitioning depends on the given penalty functions, different choices can be made, in dependence of the application. For the DIA-estimator, we introduce a special set of penalty functions that penalize its unwanted outcomes. It is shown how this set allows one to construct the optimal DIA-estimator, being the estimator that within its class has the largest probability of lying inside a user specified tolerance region. Further elaboration shows how these penalty functions are driven by the influential biases of the different hypotheses and how they can be used operationally. Hereby the option is included of extending the misclosure partitioning with an additional undecided region to accommodate situations when it will be hard to discriminate between some of the hypotheses or when identification is unconvincing. By extending the analogy with integer ambiguity resolution to that of integer-equivariant ambiguity resolution, we also introduce the maximum probability estimator within the similar larger class.

Keywords Detection-Identification-Adaptation (DIA) · Misclosure partitioning · Minimum mean penalty partitioning · DIA-estimator · DIA-penalty function · Influential bias · Maximum probability estimators

1 Introduction

DIA-estimation captures the overall problem of detection, identification and adaptation (DIA) as one of estimation (Teunissen 2018). As its structure is similar to that of mixed integer estimation, one can cast its parameter solution in a similar form. In case of GNSS mixed integer estimation, the integer map $\mathcal{I} : \mathbb{R}^n \mapsto \mathbb{Z}^n$ defines the ambiguity pull-in regions as $\mathcal{P}_{z\in\mathbb{Z}^n} = \{u \in \mathbb{R}^n \mid z = \mathcal{I}(u)\}$, resulting in the integer ambiguity-resolved baseline $\breve{\underline{b}} = \sum_{z\in\mathbb{Z}^n} \hat{\underline{b}}(z)\, p_z(\hat{\underline{a}})$, with $\hat{\underline{a}}$ the ambiguity-float estimator, $\hat{\underline{b}}(z)$ the conditional baseline estimator, and $p_z(.)$ the indicator function of $\mathcal{P}_z$. Similarly for DIA-estimation, the hypothesis map $\mathcal{H} : \mathbb{R}^r \mapsto [0, 1, \ldots, k]$ defines the partitioning of misclosure space as $\mathcal{P}_{i\in[0,\ldots,k]} = \{t \in \mathbb{R}^r \mid i = \mathcal{H}(t)\}$ and results in the DIA-estimator $\bar{\underline{x}} = \sum_{i=0}^{k} \hat{\underline{x}}_i\, p_i(\underline{t})$, with $\underline{t}$ the misclosure vector, $\hat{\underline{x}}_i$ the hypothesis-conditioned BLUE, and $p_i(.)$ the indicator function of $\mathcal{P}_i$.

This analogy is extended in this contribution to penalized ambiguity resolution (Teunissen 2004). By assigning penalty functions to each of the decision regions in misclosure space, $\mathcal{P}_{i\in[0,\ldots,k]} \subset \mathbb{R}^r$, the mean penalty of any chosen misclosure space partitioning can be determined and compared. As a result, we determine and study the optimal DIA-estimator, being the estimator that within its class has the highest probability of lying inside a user defined tolerance region.

Although the presented theory is non-Bayesian throughout in the parameters, we also show how distributional information on the biases can be incorporated if available. The theory is applicable to a wide range of applications, for example, quality control of geodetic networks (DGCC 1982; Zaminpardaz and Teunissen 2019; Yang et al. 2021), geophysical and structural deformation analysis (Lehmann and Lösler 2017; Nowel 2020; Zaminpardaz et al. 2020), different GNSS applications (Perfetti 2006; Yu et al. 2023), and various configurations of integrated navigation systems (Gillis-

✉ P. J. G. Teunissen
p.j.g.teunissen@tudelft.nl

[1] Department of Geoscience and Remote Sensing, Delft University of Technology, Delft, The Netherlands

[2] School of Earth and Planetary Sciences, Curtin University, Perth, Australia

Published online: 21 May 2024

sen and Elema 1996; Teunissen 1989; Salzmann 1991). For all these different applications, user-derived penalties can be set to direct the DIA-estimator perform according to its application-dependent tolerable risk objectives.

This contribution is organized as follows: After a brief review of DIA-estimation in Sect. 2, we introduce in Sect. 3, in analogy with penalized ambiguity resolution, the concept of penalized testing for estimation. As any testing procedure is unambiguously described by its partitioning of misclosure space, we show how the mean penalty of such partitionings can be evaluated, thus giving users the tool to compare different testing procedures using their own assigned penalties. We also determine the partitioning of misclosure space that results in the minimum mean penalty. Its operational use under known and unknown biases is discussed, and it is shown how the penalties need to be chosen to recover the misclosure space partitioning of classical multi-hypotheses datasnooping (Baarda 1968a; Teunissen 2000; Lehmann and Lösler 2016).

In Sect. 4, we focus attention to the *consequences* of testing decisions, rather than only to the correctness of the decisions. This is particularly of importance when the goal is not per se the correct identification of the active hypothesis, but rather being able to direct the performance of the DIA-estimator towards its application-dependent tolerable risk objectives. For that purpose, we introduce a special DIA-penalty function that penalizes unwanted outcomes of the estimator. We show how this penalty function maximizes the probability $\mathsf{P}[\underline{\bar{x}} \in \Omega_x]$, thereby enabling the construction of the optimal DIA-estimator. By extending the analogy with integer estimation to that of integer-equivariant estimation, we also introduce and derive the maximum probability estimator in the similar larger class.

Further elaboration of the DIA-penalty functions is conducted in Sect. 5, thereby showing the prominent role played by the *influential* biases. Hereby we also present different operational simplifications of the penalty functions and associated minimum mean penalty partitionings. This includes the option of having an additional *undecided* region to accommodate situations where one lacks confidence in the decision making. In such cases, one may rather prefer to state that a solution is unavailable, than providing an actual, but possibly unreliable, parameter estimate. The theory is illustrated and supported by several worked out examples. Finally in Sect. 6, a summary and conclusions are given.

The following notation is used: $\mathsf{E}(.)$ and $\mathsf{D}(.)$ stand for the expectation and dispersion operator, respectively, and $\mathcal{N}_p(\mu, Q)$ denotes a p-dimensional, normally distributed random vector, with mean (expectation) μ and variance matrix (dispersion) Q. We denote a random variable or random vector with an underscore. Thus, $\underline{y}$ is random, while x is not. If the same symbol is used with and without underscore, then the latter is a realisation of the former. Thus, $\hat{x}_0$ is an outcome or realisation of the random $\underline{\hat{x}}_0$. The probability of an event $\mathcal{A}$ is denoted as $\mathsf{P}[\mathcal{A}]$, a proportional to b as $a \propto b$, and the logical characters for and/or as $\wedge/\vee$. For the probability of $\mathcal{H}_\alpha$-hypothesis occurrence, we use the shorthand notation $\pi_\alpha = \mathsf{P}[\mathcal{H}_\alpha] = \mathsf{P}[\underline{\mathcal{H}} = \mathcal{H}_\alpha]$. The probability density function (PDF) of a random vector $\underline{t}$ is denoted as $f_{\underline{t}}(t)$. The noncentral Chi-square distribution with p degrees of freedom and noncentrality parameter λ is denoted as $\chi^2(p, \lambda)$ and its δ-percentage critical value as $\chi^2_\delta(p, 0)$. $\mathbb{R}^p$ and $\mathbb{Z}^p$ denote the p-dimensional spaces of real- and integer numbers, respectively. $\mathbb{R}^r_{\geq 0}$ denotes the space of r-vectors having nonnegative entries and e_r is the r-vector of ones. $||x||^2_Q = (x)^T Q^{-1}(x)$ denotes the squared Q-weighted norm of vector x and $\delta_{i\alpha}$ the Kronecker-delta, with $\delta_{i\alpha} = 1$ if $i = \alpha$ and $\delta_{i\alpha} = 0$ if $i \neq \alpha$. The identity matrix is denoted as I and the projector that projects orthogonally, in the metric of Q, on the range space of matrix M as $P_M = M(M^T Q^{-1} M)^{-1} M^T Q^{-1}$, where $P^\perp_M = I - P_M$. The range space of a matrix M is denoted as $\mathcal{R}(M)$.

2 A brief DIA review

In this section, we give a brief review of DIA-estimation and its properties.

2.1 Hypotheses, BLUEs and misclosure vector

We start by formulating our null-hypothesis $\mathcal{H}_0$ and k alternative hypotheses $\mathcal{H}_i$, $i = 1, \ldots, k$. The null-hypothesis, also referred to as working hypothesis, consists of the model that one believes to be valid under normal working conditions. We assume it to read

$$\mathcal{H}_0 : \underline{y} \sim \mathcal{N}_m(Ax, Q_{yy}) \tag{1}$$

with $A \in \mathbb{R}^{m\times n}$ the given design matrix of rank n, $x \in \mathbb{R}^n$ the to-be-estimated unknown parameter vector, and $Q_{yy} \in \mathbb{R}^{m\times m}$ the given positive-definite variance matrix of $\underline{y}$. The redundancy of $\mathcal{H}_0$ is $r = m - \text{rank}(A) = m - n$.

Although every part of the assumed null-hypothesis can be wrong, we assume that if a misspecification in $\mathcal{H}_0$ occurred that it is confined to an underparametrization of the mean of $\underline{y}$. The alternative hypotheses will therefore only differ from $\mathcal{H}_0$ in their mean of $\underline{y}$. The ith alternative hypothesis is assumed given as:

$$\mathcal{H}_i : \underline{y} \sim \mathcal{N}_m(Ax + C_i b_i, Q_{yy}) \tag{2}$$

for some unknown vector $C_i b_i \in \mathbb{R}^m \backslash \{0\}$, with $[A, C_i] \in \mathbb{R}^{m\times(n+q_i)}$ a known matrix of full rank $n + q_i$. Through $C_i b_i$ one may model, for instance, the presence of one or more outliers in the data, satellite failures, antenna-height errors,

cycle-slips in GNSS phase data, neglectance of atmospheric delays, or any other systematic effect that one failed to take into account under $\mathcal{H}_0$. We will use the lowercase c_i, instead of C_i, when $q_i = 1$, i.e. when b_i is a scalar.

For our further considerations, it is useful to first bring (1) in canonical form. This is achieved by means of the *Tienstra*-transformation and its inverse,

$$\mathcal{T} = [A^{+T}, B]^T \text{ and } \mathcal{T}^{-1} = [A, B^{+T}] \tag{3}$$

in which $A^+ = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}$ and $B^+ = (B^T Q_{yy} B)^{-1} B^T Q_{yy}$ are the BLUE-inverses of A and B, respectively, and B is an $m \times r$ basis-matrix of the null space of A^T, i.e. $B^T A = 0$ and $\text{rank}(B) = r$. Application of $\mathcal{T}$ to $\underline{y}$ gives under the null hypothesis (1),

$$\begin{bmatrix} \underline{\hat{x}}_0 \\ \underline{t} \end{bmatrix} = \mathcal{T}\underline{y} \overset{\mathcal{H}_0}{\sim} \mathcal{N}_m \left(\begin{bmatrix} x \\ 0 \end{bmatrix}, \begin{bmatrix} Q_{\hat{x}_0\hat{x}_0} & 0 \\ 0 & Q_{tt} \end{bmatrix} \right) \tag{4}$$

in which $\underline{\hat{x}}_0 = A^+\underline{y} \in \mathbb{R}^n$ is the best linear unbiased estimator (BLUE) of x under $\mathcal{H}_0$ and $\underline{t} = B^T\underline{y} \in \mathbb{R}^r$ is the *misclosure* vector of $\mathcal{H}_0$, having variance matrices $Q_{\hat{x}_0\hat{x}_0} = (A^T Q_{yy}^{-1} A)^{-1}$ and $Q_{tt} = B^T Q_{yy} B$, respectively. As the misclosure vector $\underline{t}$ is zero-mean under the null-hypothesis and stochastically independent of $\underline{\hat{x}}_0$, it contains all the available information useful for testing the validity of $\mathcal{H}_0$.

Under the alternative hypothesis (2), $\mathcal{T}\underline{y}$ becomes distributed as:

$$\begin{bmatrix} \underline{\hat{x}}_0 \\ \underline{t} \end{bmatrix} = \mathcal{T}\underline{y} \overset{\mathcal{H}_i}{\sim} \mathcal{N}_m \left(\begin{bmatrix} I_n & A^+C_i \\ 0 & B^T C_i \end{bmatrix} \begin{bmatrix} x \\ b_i \end{bmatrix}, \begin{bmatrix} Q_{\hat{x}_0\hat{x}_0} & 0 \\ 0 & Q_{tt} \end{bmatrix} \right) \tag{5}$$

Thus, $\underline{\hat{x}}_0$ and $\underline{t}$ are still independent, but now have different means than under $\mathcal{H}_0$. Due to the canonical structure of (5), it now becomes rather straightforward to infer the BLUEs of x and b_i under $\mathcal{H}_i$. As $\underline{\hat{x}}_0$ and $\underline{t}$ are independent and the mean of $\underline{\hat{x}}_0$ under $\mathcal{H}_i$ depends on more parameters than only those of x, the estimator $\underline{\hat{x}}_0$ will not contribute to the determination of the BLUE of b_i. Hence, it is $\underline{t}$ that is solely reserved for the determination of the BLUE of b_i, which then on its turn can be used in the determination of the BLUE of x under $\mathcal{H}_i$. The BLUEs of x and b_i under $\mathcal{H}_i$ are therefore given as

$$\begin{aligned} \underline{\hat{x}}_i &= \underline{\hat{x}}_0 - A^+C_i\underline{\hat{b}}_i \\ \underline{\hat{b}}_i &= (B^T C_i)^+\underline{t} \end{aligned} \tag{6}$$

in which $(B^T C_i)^+ = (C_i^T B Q_{tt}^{-1} B^T C_i)^{-1} C_i^T B Q_{tt}^{-1}$ denotes the BLUE-inverse of $B^T C_i$. The result (6) shows how $\underline{\hat{x}}_0$ is to be *adapted* when switching from the BLUE of $\mathcal{H}_0$ to that of $\mathcal{H}_i$.

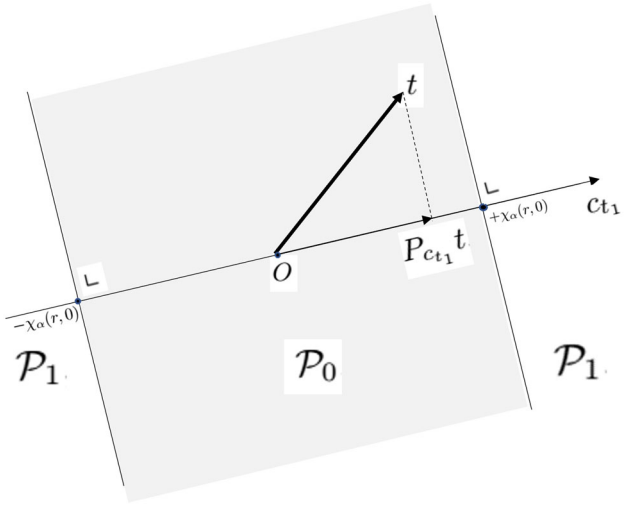

Fig. 1 Misclosure space partitioning $\mathbb{R}^r = \mathcal{P}_0 \cup \mathcal{P}_1$ for the binary testing of $\mathcal{H}_0 : \mathsf{E}(\underline{y}) = Ax$ against $\mathcal{H}_1 : \mathsf{E}(\underline{y}) = Ax + C_1 b_1$

2.2 Testing and misclosure space partitioning

Which of the possible parameter solutions to deliver, $\hat{x}_0$ or one of the $\hat{x}_i$'s, is decided through hypothesis testing, and as mentioned, it is the misclosure vector

$$\underline{t} \overset{\mathcal{H}_i}{\sim} \mathcal{N}_r(C_{t_i} b_i, Q_{tt}), \text{ with } C_{t_i} = B^T C_i \tag{7}$$

that forms the input to hypothesis testing. Would one only have a *single* alternative hypothesis ($k = 1$), one would likely use the uniformly most powerful invariant (UMPI) test statistic (Arnold 1981; Teunissen 2000),

$$\underline{T}_{q_i} = ||P_{C_{t_i}}\underline{t}||^2_{Q_{tt}} \overset{\mathcal{H}_i}{\sim} \chi^2(q_i, \lambda_i = ||C_{t_i} b_i||^2_{Q_{tt}}) \tag{8}$$

where $P_{C_{t_i}} = C_{t_i}(C_{t_i}^T Q_{tt}^{-1} C_{t_i})^{-1} C_{t_i}^T Q_{tt}^{-1}$, to accept $\mathcal{H}_0$ when $T_{q_i} \leq \chi^2_\alpha(q_i, 0)$ and otherwise reject $\mathcal{H}_0$ in favour of $\mathcal{H}_i$. Such binary decision making can be visualized through a corresponding binary partitioning of misclosure space. With the partitioning

$$\begin{aligned} \mathcal{P}_0 &= \{t \in \mathbb{R}^r \mid T_{q_i} = ||P_{C_{t_i}} t||^2_{Q_{tt}} \leq \chi^2_\alpha(q_i, 0)\} \\ \mathcal{P}_i &= \mathbb{R}^r \backslash \mathcal{P}_0 \end{aligned} \tag{9}$$

one would then choose for $\mathcal{H}_0$ if $t \in \mathcal{P}_0$ and for $\mathcal{H}_i$ if $t \in \mathcal{P}_i$, see Fig. 1. If $q_i = 1$, then T_{q_i} can be expressed in Baarda's w-statistic (Baarda 1968a) as $T_{q_i=1} = w_i^2$, with

$$w_i = \frac{c_{t_i}^T Q_{tt}^{-1} t}{\sqrt{c_{t_i}^T Q_{tt}^{-1} c_{t_i}}} \tag{10}$$

We note that the UMPI test statistic (8) can also be expressed in the BLUE of b_i under $\mathcal{H}_i$ as Teunissen (2000)

$$\underline{T}_{q_i} = ||\hat{b}_i(\underline{t})||^2_{Q_{\hat{b}_i\hat{b}_i}} \overset{\mathcal{H}_i}{\sim} \chi^2(q_i, \lambda_i = ||b_i||^2_{Q_{\hat{b}_i\hat{b}_i}}) \tag{11}$$

where $Q_{\hat{b}_i\hat{b}_i} = (C_{t_i}^T Q_{tt}^{-1} C_{t_i})^{-1}$. Here we have written the BLUE of b_i as $\hat{b}_i(\underline{t})$ to explicitly show its dependence on the misclosure vector $\underline{t}$, cf. (6). Expression (11) shows that the binary test between $\mathcal{H}_0$ and $\mathcal{H}_i$ can therefore also be interpreted as a *significance* test: choose $\mathcal{H}_0$ if the bias-estimate is considered insignificant, else choose $\mathcal{H}_i$.

In the multiple alternative hypotheses case ($k > 1$), one cannot generalize the above binary decision making and expect the UMPI property to remain valid. However, although for now it is not yet clear how the actual multiple hypotheses decision making should look like, the idea of partitioning misclosure space for the purpose of such decision making can easily be generalized from the case $k = 1$ to $k > 1$, this in analogy with the *pull-in regions* of integer estimation and integer aperture estimation (Teunissen 2003a). Therefore, if we let the multiple hypotheses testing procedure be captured by the unambiguous mapping $\mathcal{H} : \mathbb{R}^r \mapsto \{0, 1, \ldots, k\}$, the regions

$$\mathcal{P}_{i\in[0,\ldots,k]} = \{t \in \mathbb{R}^r \mid i = \mathcal{H}(t)\}, \tag{12}$$

form a partitioning of the r-dimensional misclosure space, i.e. $\cup_{i=0}^k \mathcal{P}_i = \mathbb{R}^r$ and $\mathcal{P}_i \cap \mathcal{P}_j = \emptyset$ for $i \neq j$. Hence, by specifying (12), one would have automatically and unambiguously defined the *multiple* testing procedure as selecting $\mathcal{H}_i$ if $t \in \mathcal{P}_i$.

Formulation (12) is a very general one and applies in principle to any unambiguous multiple hypotheses testing problem. How the mapping $\mathcal{H}$, or its partitioning $\mathcal{P}_i$, $i = 0, \ldots, k$, is defined determines how the actual testing procedure is executed. The following example shows how Baarda's *datasnooping* (Baarda 1968b), being one of the more familiar outlier testing procedures, fits into the above partitioning framework.

Example 1 (*Detection and 1-dim identification*) Let the design matrices $[A, C_i]$ of the k hypotheses $\mathcal{H}_i$ (cf. 2) be of order $m \times (n + 1)$, with $i = 1, \ldots, k$, denote $C_i = c_i$ and $B^T c_i = c_{t_i}$, and write Baarda's test-statistic (Baarda 1968a; Teunissen 2000) as $|\underline{w}_i| = ||P_{c_{t_i}} \underline{t}||_{Q_{tt}}$. Then,

$$\begin{aligned} \mathcal{P}_0 &= \{t \in \mathbb{R}^r \mid ||t||_{Q_{tt}} \leq \tau \in \mathbb{R}^+\} \\ \mathcal{P}_{i\neq 0} &= \{t \in \mathbb{R}^r \backslash \mathcal{P}_0 \mid i = \arg \max_{j\in\{1,\ldots,k\}} |w_j|\} \end{aligned} \tag{13}$$

form a partitioning of misclosure space, provided not two or more of the vectors c_{t_i} are parallel. The inference induced by this partitioning is thus that the null-hypothesis gets accepted if in the *detection* step the overall model test gets accepted, $||t||_{Q_{tt}} \leq \tau$, while in case of rejection, the largest value of the statistics $|w_j|$, $j = 1, \ldots, k$, say $|w_i|$, is used for *identifying* the ith alternative hypothesis. In the first case, $\hat{x}_0$ is provided as the output estimate of x, while in the second case, $\hat{x}_0$ is *adapted* to provide the output as $\hat{x}_i$, cf. (6). In case $k = m$ and the c_i are canonical unit vectors, the above testing reduces to Baarda's single-outlier data-snooping, i.e. the procedure in which the individual observations are screened for possible outliers (Baarda 1968a; DGCC 1982; Kok 1984).

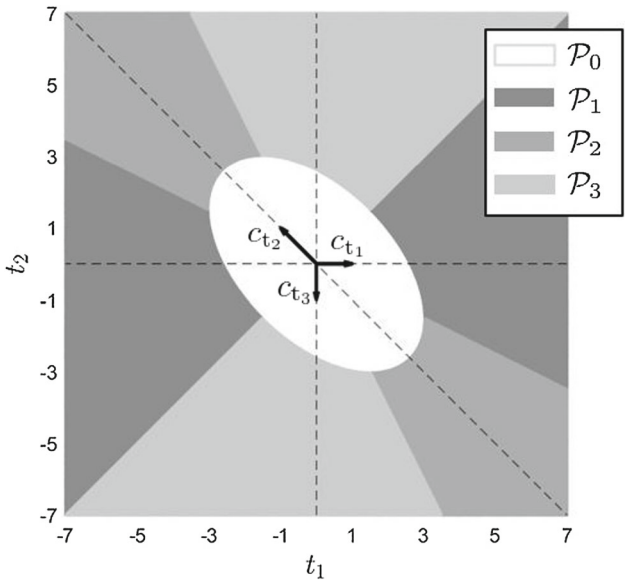

Fig. 2 Misclosure space partitioning in $\mathbb{R}^{r=2}$ (cf. Example 1): elliptical $\mathcal{P}_0$ for detection, with $\mathcal{P}_{i\in[1,2,3]}$ for outlier identification

Figure 2 illustrates the geometry of partitioning (13) for the case $A = [1, 1, 1]^T$, $Q_{yy} = I_3$, $c_1 = [1, 0, 0]^T$, $c_2 = [0, 1, 0]^T$, and $c_3 = [0, 0, 1]^T$, cf. (1) and (2). With

$$B^T = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \tag{14}$$

the inverse variance matrix of the misclosure vector follows as

$$Q_{tt}^{-1} = (B^T Q_{yy} B)^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}^{-1} = \tfrac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \tag{15}$$

This matrix determines the shape of the elliptical detection region $||t||^2_{Q_{tt}} = t^T Q_{tt}^{-1} t < \tau^2$. The *fault lines* along which the $\mathsf{E}(\underline{t}|\mathcal{H}_i) = B^T c_i b_i$ move when b_i varies, $i = 1, 2, 3$, have direction vectors $c_{t_1} = B^T c_1 = [1, 0]^T$, $c_{t_2} = B^T c_2 = [-1, 1]^T$, and $c_{t_3} = B^T c_3 = [0, -1]^T$. □

2.3 The DIA-estimator and its PDF

Once testing has been concluded, one has either accepted the null-hypothesis $\mathcal{H}_0$ and provided $\hat{x}_0$ as the parameter estimate of x, or identified one of the alternative hypotheses, say $\mathcal{H}_i$, $i = 1, \ldots, k$, and provided $\hat{x}_i$ as the parameter estimate of x. The first happens when $t \in \mathcal{P}_0$, while the second when $t \in \mathcal{P}_{i\neq 0}$. This shows that the whole of the detection, identification and adaptation (DIA) procedure to come to a final solution for the unknown parameter vector x, is a combination of *estimation* and *testing*, whereby the uncertainty of both would need to be accommodated for in the quality description of the final result. The actual estimator that DIA produces is therefore not $\underline{\hat{x}}_0$ nor $\underline{\hat{x}}_i$, but

$$\underline{\bar{x}}_{\text{DIA}} = \left\{ \begin{array}{l} \underline{\hat{x}}_0 \text{ if } \underline{t} \in \mathcal{P}_0 \\ \underline{\hat{x}}_i \text{ if } \underline{t} \in \mathcal{P}_{i\neq 0} \end{array} \right\} = \sum_{i=0}^{k} \underline{\hat{x}}_i \, p_i(\underline{t}) \tag{16}$$

in which $p_i(t)$ denotes the indicator function of $\mathcal{P}_i$, (i.e. $p_i(t) = 1$ for $t \in \mathcal{P}_i$ and $p_i(t) = 0$ elsewhere). The DIA-estimator (16) represents a *class* of estimators, with each member in the class unambiguously defined through its misclosure space partitioning. Changing the testing procedure will change the partitioning and consequently also the DIA-estimator.

The structure of (16) resembles that of mixed integer estimation. As mentioned in Teunissen (2018) p.67, this similarity can be extended further to mixed integer-equivariant estimation. This is achieved if one would replace the indicator functions $p_i(t)$ of (16), with misclosure weighting functions $w_i(t) : \mathbb{R}^r \mapsto \mathbb{R}$, satisfying $\omega_i(t) \geq 0$, $i = 0, \ldots, k$, and $\sum_{i=0}^{k} \omega_i(t) = 1$. As a result, we obtain, in addition to the DIA-class, a second class of estimators, namely

$$\underline{\bar{x}}_{\text{WSS}} = \sum_{i=0}^{k} \underline{\hat{x}}_i \omega_i(\underline{t}) \tag{17}$$

which we will call the *weighted solution-separation (WSS)* class. Note, since the indicator functions satisfy the properties $p_i(t) \geq 0$, $i = 0, \ldots, k$, and $\sum_{i=0}^{k} p_i(t) = 1$, that the DIA-class is a subset of the WSS-class, just like the integer-class is a subset of the integer-equivariant class (Teunissen 2003b).

We have named estimators from the class (17) 'weighted solution-separation' estimators, since they can alternatively be represented as

$$\underline{\bar{x}}_{\text{WSS}} = \underline{\hat{x}}_0 + \sum_{i=1}^{k} (\underline{\hat{x}}_i - \underline{\hat{x}}_0) \omega_i(\underline{t}) \tag{18}$$

thus showing how $\underline{\bar{x}}_{\text{WSS}}$ is obtained through a weighted solution-separation sum adjustment of the $\mathcal{H}_0$-solution $\underline{\hat{x}}_0$. Note, as $\underline{t}$ is independent of $\underline{\hat{x}}_0$ and the solution separations $\underline{\hat{x}}_i - \underline{\hat{x}}_0$ are functions of the misclosure vector $\underline{t}$ only (cf. 6), that the weighted solution-separation sum of (18) is also independent of $\underline{\hat{x}}_0$. With formulation (18) one should be aware, however, that the k weights $\omega_i(t)$ sum up to $1 - \omega_0(t)$ and not to 1.

To be able to determine and judge the parameter estimation quality of (16) and (17), we need their probability density function. As (16) can be considered a special case of (17), the use of the subscripts 'DIA' or 'WSS' will only be used in the following if the need arises. We have the following result (Teunissen 2018).

Theorem 1 (PDF of $\underline{\bar{x}}$) *The probability density function of* (17) *is given as*

$$f_{\underline{\bar{x}}}(x) = \int_{\mathbb{R}^r} f_{\underline{\hat{x}}_0}(x + \ell(\tau)) f_{\underline{t}}(\tau) d\tau \tag{19}$$

where $\ell(t) = \sum_{i=1}^{k} L_i t \omega_i(t)$ *and* $L_i = A^+ C_i [B^T C_i]^+$. ■

Proof We first express $\underline{\bar{x}}$ in the two independent vectors $\underline{\hat{x}}_0$ and $\underline{t}$. With $\sum_{i=0}^{k} \omega_i(t) = 1$, substitution of $\underline{\hat{x}}_i = \underline{\hat{x}}_0 - L_i \underline{t}$ (cf. 6) into (17) gives $\underline{\bar{x}} = \underline{\hat{x}}_0 - \ell(\underline{t})$. Application of the PDF transformation rule to the pair $\underline{\bar{x}} = \underline{\hat{x}}_0 - \ell(\underline{t})$, $\underline{t}$, recognizing the Jacobian to be 1, gives then for their joint PDF $f_{\underline{\bar{x}},\underline{t}}(x, t) = f_{\underline{\hat{x}}_0,\underline{t}}(x + \ell(t), t)$. The marginal (19) follows then from integrating t out and recognizing that $\underline{\hat{x}}_0$ and $\underline{t}$ are independent. □

The above result shows how the impact of the hypotheses is felt through the shifting over $\ell(\tau)$ of the PDF of $\underline{\hat{x}}_0$, and thus, how it can be manipulated, either through the choice of $p_i(t)$, i.e. the choice of misclosure space partitioning, or through the choice of the misclosure weighting functions $\omega_i(t)$. Note that (19) can be expressed in terms of an expectation as

$$f_{\underline{\bar{x}}}(x) = \mathsf{E}\left(f_{\underline{\hat{x}}_0}(x + \ell(\underline{t})) \right) \tag{20}$$

thus showing that the PDF equals the average of random shifts $\ell(\underline{t})$ of the PDF of $\underline{\hat{x}}_0$. This expression is useful when one wants to Monte-Carlo simulate $f_{\underline{\bar{x}}}(x)$ or integral-values of it Robert and Casella (2004). For example, to compute $\mathsf{P}[\underline{\bar{x}} \in \Omega \subset \mathbb{R}^n] = \int_{\mathbb{R}^n} f_{\underline{\bar{x}}}(x) i_\Omega(x) dx$, with $i_\Omega(x)$ being the indicator function of Ω, we first express the probability in terms of an expectation, $\mathsf{P}[\underline{\bar{x}} \in \Omega \subset \mathbb{R}^n] = V_\Omega \int_{\mathbb{R}^n} f_{\underline{\bar{x}}}(x) u_{\underline{x}}(x) dx = V_\Omega \mathsf{E}(f_{\underline{\bar{x}}}(\underline{x}))$, with volume $V_\Omega = \int_\Omega dx$ and PDF $u_{\underline{x}}(x) = \frac{i_\Omega(x)}{V_\Omega}$ being the *uniform* PDF over Ω. Then one may use the Monte-Carlo approximation $\mathsf{P}[\underline{\bar{x}} \in \Omega \subset \mathbb{R}^n] \approx \frac{V_\Omega}{k_x} \sum_{j=1}^{k_x} f_{\underline{\bar{x}}}(x_j)$, in which x_j, $j = 1, \ldots, k_x$, are the k_x samples drawn from the uniform PDF over $\Omega \subset \mathbb{R}^n$. This, together with a similar Monte-Carlo

approximation of (20), gives then

$$\mathsf{P}[\underline{\bar{x}} \in \mathcal{Q} \subset \mathbb{R}^n] \approx \frac{V_{\mathcal{Q}}}{k_t k_x} \sum_{i=1}^{k_t} \sum_{j=1}^{k_x} f_{\underline{\hat{x}}_0}(x_j + \ell(t_i)) \tag{21}$$

in which $t_i, i = 1, \ldots, k_t$, are the k_t samples drawn from the PDF $f_{\underline{t}}(t)$. Standard Monte-Carlo simulation can be further improved with importance sampling and other variance-reduction techniques, see, e.g. Kroese et al. (2011).

An important difference between (16) and (17) is the use of *binary* weights $p_i(t)$ in the DIA-estimator. It is through these binary weights that the DIA-estimator is unambiguously linked to hypothesis testing. In fact, the testing procedure is the defining *trait* of the DIA-estimator. No such link exists, however, when the WSS-estimator is based on *smooth* misclosure weight functions $\omega_i(t)$. In that case, a weighted average of *all* $k+1$ parameter solutions $\hat{x}_i$ is taken, instead of the single 'winner-takes-all' solution of (16). Although an explicit testing procedure is absent in case of the WSS-estimator with smooth weights, the estimator does reveal its hypothesis-preference through its choice of weighting functions. As the weight $\omega_i(t)$ can be seen to be a measure of preference that is given to solution $\hat{x}_i$ for a given t, it may be interpreted as the conditional probability $\mathsf{P}[\underline{i} = i|t]$. For the binary weight $\omega_i(t) = p_i(t)$, it would then be the $1-0$ probability of selecting the BLUE $\underline{\hat{x}}_i$ given the outcome of the misclosure vector being t. For the conditional and unconditional expectations of the random weight $\omega_i(\underline{t})$ we then have $\mathsf{E}(\omega_i(\underline{t})|\mathcal{H}_j) = \mathsf{P}[\underline{i} = i|\mathcal{H}_j]$ and $\mathsf{E}(\omega_i(\underline{t})) = \mathsf{P}[\underline{i} = i]$, thus showing how the expectation of the weights can be read as probabilities assigned to the hypotheses. In case of the DIA-estimator, having the binary weight $\omega_i(t) = p_i(t)$, the expectations specialize to $\mathsf{E}(p_i(\underline{t})|\mathcal{H}_j) = \mathsf{P}[\underline{t} \in \mathcal{P}_i|\mathcal{H}_j]$ and $\mathsf{E}(p_i(\underline{t})) = \mathsf{P}[\underline{t} \in \mathcal{P}_i]$, which are the probabilities with which the hypotheses are identified by the testing procedure.

In our description of DIA-estimation, we so far assumed that always one of the estimates $\hat{x}_i$, $i = 0, \ldots, k$, was provided as output, even, for instance, if it would be hard to discriminate between some of the hypotheses or when identification is unconvincing. However, when one lacks confidence in the decision making, one may rather prefer to state that a solution is unavailable, than providing an actual, but possible unreliable, parameter estimate. To accommodate such situations, one can generalize the procedure and introduce an additional *undecided* region $\mathcal{P}_{k+1} \subset \mathbb{R}^r$ in the misclosure space partitioning. This is similar in spirit to the undecided regions of the theory of *integer aperture estimation* (Teunissen 2003a). With the undecided region $\mathcal{P}_{k+1}$ in place, the DIA-estimator generalizes to

$$\underline{\bar{x}} = \begin{cases} \underline{\hat{x}}_i & \text{if } \underline{t} \in \mathcal{P}_i,\ i = 0, \ldots, k \\ \text{unavailable} & \text{if } \underline{t} \in \mathcal{P}_{k+1} \end{cases} \tag{22}$$

As parameter estimates are now only provided when $t \in \mathbb{R}^r \backslash \mathcal{P}_{k+1}$, the evaluation of the DIA-estimator would now need to be based on its conditional PDF $f_{\underline{\bar{x}}|t \notin \mathcal{P}_{k+1}}(x)$, the expression of which can be found in Teunissen (2018).

In practice one is quite often not interested in the complete parameter vector $x \in \mathbb{R}^n$, but rather only in certain functions of it, say $\theta = F^T x \in \mathbb{R}^p$. As its DIA-estimator is then computed as $\underline{\bar{\theta}} = F^T \underline{\bar{x}}$, we need its distribution to evaluate its performance. In analogy with Theorem 1, the PDF of $\underline{\bar{\theta}}$ is given as $f_{\underline{\bar{\theta}}}(\theta) = \int_{\mathbb{R}^r} f_{\underline{\hat{\theta}}_0}(\theta + F^T \ell(\tau)) f_{\underline{t}}(\tau) d\tau$. Although we will be working with $\underline{\bar{x}}$, instead of $\underline{\bar{\theta}}$, in the remaining of this contribution, it should be understood that the results provided can similarly be given for $\underline{\bar{\theta}} = F^T \underline{\bar{x}}$ as well.

We also note, although all our results are formulated in terms of the misclosure vector $\underline{t} \in \mathbb{R}^r$, that they can be formulated in terms of the least-squares residual vector $\underline{\hat{e}}_0 = \underline{y} - A\underline{\hat{x}}_0 \in \mathbb{R}^m$ as well. This follows, since $\underline{t} = B^T \underline{\hat{e}}_0$.

3 Penalized testing

3.1 Minimum mean penalty testing

The DIA-estimator (16) represents a class of estimators, with each member in the class unambiguously defined through its misclosure partitioning. Any change in the partitioning will change the outcome of testing and thus also the quality of the testing procedure and its decision making. As, like in (22), the number of subsets of the partitioning need not be equal to the number of hypotheses, we put in the following no restriction on the number of subsets and thus let misclosure space $\mathbb{R}^r$ be partitioned in $l+1$ subsets $\mathcal{P}_i, i = 0, \ldots, l$, thereby assuming that each subset is unambiguously linked to a decision, i.e. decision i is made when $t \in \mathcal{P}_i$. To be able to compare the quality of different partitionings, we introduce a weighting scheme that weighs the envisioned risk of a decision i. This is done by assigning to decision i, a nonnegative *risk penalizing* function $\mathrm{r}_{i\alpha}(t)$, with $t \in \mathcal{P}_i$, for each of the $k+1$ hypotheses $\mathcal{H}_\alpha$, $\alpha = 0, \ldots, k$. Note that we allow the penalty of the invoked risk depend on where t is located within $\mathcal{P}_i$. Using the indicator function $p_i(t)$ of $\mathcal{P}_i$, we can write the hypothesis $\mathcal{H}_\alpha$-penalty function, for all $t \in \mathbb{R}^r$, as

$$\mathrm{r}_\alpha(t) = \sum_{i=0}^{l} \mathrm{r}_{i\alpha}(t) p_i(t) \tag{23}$$

As the misclosure vector $\underline{t}$ is random, the function values $\mathrm{r}_\alpha(t)$ may now be considered outcomes of a random risk penalty variable $\underline{\mathrm{r}}$ conditioned on $\mathcal{H}_\alpha$. We therefore have the conditional means

$$\mathsf{E}(\underline{\mathrm{r}}|\mathcal{H}_\alpha) = \int_{\mathbb{R}^r} \mathrm{r}_\alpha(t) f_{\underline{t}}(t|\mathcal{H}_\alpha) dt$$

$$\mathsf{E}(\underline{r}|t) = \sum_{\alpha=0}^{k} r_\alpha(t)\mathsf{P}[\mathcal{H}_\alpha|t] \tag{24}$$

and the unconditional mean

$$\mathsf{E}(\underline{r}) = \sum_{\alpha=0}^{k} \mathsf{E}(\underline{r}|\mathcal{H}_\alpha)\mathsf{P}[\mathcal{H}_\alpha] = \int_{\mathbb{R}^r} \mathsf{E}(\underline{r}|t)\, f_{\underline{t}}(t)dt \tag{25}$$

where

$$\mathsf{P}[\mathcal{H}_\alpha|t] = \frac{f_{\underline{t}}(t|\mathcal{H}_\alpha)\mathsf{P}[\mathcal{H}_\alpha]}{\sum_{\beta=0}^{k} f_{\underline{t}}(t|\mathcal{H}_\beta)\mathsf{P}[\mathcal{H}_\beta]} \tag{26}$$

Would we change the partitioning of $\mathbb{R}^r$, i.e. change the values of t for which decision i is made, then the mean penalty $\mathsf{E}(\underline{r})$ would change as well. Hence, we can now think of a best possible partitioning, namely one that would minimize the mean risk penalty.

To minimize $\mathsf{E}(\underline{r})$ in dependence on the $l+1$ subsets $\mathcal{P}_i$, $i = 0, \ldots, l$, we make use of the following lemma:

Lemma 1 (Optimal constrained partitioning) *Let the $l+1$ subsets $\mathcal{P}_i \subset \mathbb{R}^r$, $i = 0, \ldots, l$, form a partitioning of $\mathbb{R}^r$, i.e. $\cup_{i=0}^{l}\mathcal{P}_i = \mathbb{R}^r$ and $\mathcal{P}_i \cap \mathcal{P}_j = \emptyset$ for $i \neq j$, and let $f_i(t) : \mathbb{R}^r \mapsto \mathbb{R}$ be $l+1$ given non-negative functions. If $\mathcal{P}_0$ is known, then the $\mathcal{P}_0$-constrained subsets that* minimize *the sum*

$$S = \sum_{i=0}^{l} \int_{\mathcal{P}_i} f_i(t)dt \tag{27}$$

are given as

$$\mathcal{P}_{i\in[1,..,l]} = \{t \in \mathbb{R}^r \backslash \mathcal{P}_0 \mid i = \arg\min_{j\in[1,..,l]} f_j(t)\} \tag{28}$$

■

Proof From writing the sum S as

$$S = \int_{\mathcal{P}_0} f_0(\tau)d\tau + \sum_{i=1}^{l} \int_{\mathcal{P}_i} f_i(\tau)d\tau \tag{29}$$

and recognizing that the l subsets $\mathcal{P}_i, i = 1, \ldots, l$, now form a partitioning of $\mathbb{R}^r \backslash \mathcal{P}_0$, it follows that the second term of (29) is minimized when each of the subsets $\mathcal{P}_i$ covers that part of the domain $\mathbb{R}^r \backslash \mathcal{P}_0$ for which f_i attains the smallest function values. As a result, the l subsets are to be chosen as given by (28). □

Note, by assuming the l subsets $\mathcal{P}_i$ of (28) to form a partitioning of $\mathbb{R}^r \backslash \mathcal{P}_0$, implicit properties are asked of the functions $f_i(t)$. Such partitioning would for instance not be realized if all functions $f_i(t)$ would be equal. Also note that in the unconstrained case, i.e. if $\mathcal{P}_0$ is also unknown, $\mathbb{R}^r \backslash \mathcal{P}_0$ needs to be replaced by $\mathbb{R}^r$ in (28) and $[1, \ldots, l]$ by $[0, \ldots, l]$. We will have use for both the constrained and unconstrained cases in the sections following. In fact, expression (28) is also very useful for the unconstrained case, since it shows, once the unconstrained minimizer $\mathcal{P}_0$ is found, that the function $f_{j=0}(t)$ need not be considered anymore in the search for the remaining unconstrained minimizers $\mathcal{P}_{i\in[1,..,l]}$. This will allow us, as we will see in the sections following, to provide compact and transparent formulations of the various misclosure partitionings. Finally note, to maximize the sum S, the minimization in (28) needs to be replaced by a maximization. With the minimum and maximum, one can bound the sum S as $S_{\min} \leq S \leq S_{\max}$.

We now apply Lemma 1 so as to find the misclosure space testing partitioning that minimizes the mean penalty $\mathsf{E}(\underline{r})$.

Theorem 2a (Minimum mean penalty testing) *The misclosure space partitioning $\mathcal{P}_{i\in[0,..,l]} \subset \mathbb{R}^r$ that minimizes the mean penalty*

$$\mathsf{E}(\underline{r}) = \sum_{i=0}^{l} \int_{\mathcal{P}_i} \sum_{\alpha=0}^{k} r_{i\alpha}(t)\, f_{\underline{t}}(t|\mathcal{H}_\alpha)\mathsf{P}[\mathcal{H}_\alpha]dt \tag{30}$$

is given by:

$$\mathcal{P}_{i\in[0,..,l]} = \{t \in \mathbb{R}^r | i = \arg\min_{j\in[0,..,l]} \sum_{\alpha=0}^{k} r_{j\alpha}(t) F_\alpha(t)\} \tag{31}$$

where $F_\alpha(t) = f_{\underline{t}}(t|\mathcal{H}_\alpha)\mathsf{P}[\mathcal{H}_\alpha]$.

Proof From writing the mean penalty as

$$\begin{aligned}\mathsf{E}(\underline{r}) &\overset{(25)}{=} \sum_{\alpha=0}^{k} \mathsf{E}(\underline{r}|\mathcal{H}_\alpha)\mathsf{P}[\mathcal{H}_\alpha] \\ &\overset{(24)}{=} \sum_{\alpha=0}^{k} \int_{\mathbb{R}^r} r_\alpha(t)\, f_{\underline{t}}(t|\mathcal{H}_\alpha)dt\mathsf{P}[\mathcal{H}_\alpha] \\ &\overset{(23)}{=} \sum_{i=0}^{l} \int_{\mathcal{P}_i} f_i(t)dt\end{aligned} \tag{32}$$

with $f_i(t) = \sum_{\alpha=0}^{k} r_{i\alpha}(t)\, f_{\underline{t}}(t|\mathcal{H}_\alpha)\mathsf{P}[\mathcal{H}_\alpha]$, the result follows when applying Lemma 1. □

Note that the minimizer in (31) is invariant to a scaling of its objective function with an arbitrary nonnegative function of t. At various places in the following use will be made of this property. For instance, by using a common scaling, the values of the penalty functions may all be considered to lie between 0 and 1. Also note, by normalizing the objective function of (31) with the marginal PDF of the misclosure

vector t, $f_{\underline{t}}(t) = \sum_{\alpha=0}^{k} f_{\underline{t}}(t|\mathcal{H}_\alpha)\mathsf{P}[\mathcal{H}_a]$, and recognizing the result as $\mathsf{E}(\underline{r}_j|t) = \sum_{\alpha=0}^{k} r_{j\alpha}(t)\mathsf{P}[\mathcal{H}_\alpha|t]$, that the optimal partitioning (31) can be written as

$$\mathcal{P}_{i\in[0,..,l]} = \{t \in \mathbb{R}^r \mid i = \arg \min_{j\in[0,..,l]} \mathsf{E}(\underline{r}_j|t)\} \tag{33}$$

thus showing that each decision i, i.e. each subset $\mathcal{P}_i$, is having the smallest possible conditional mean penalty.

Expressing the minimum mean penalty partitioning in $\mathsf{E}(\underline{r}_i|t)$ is also insightful in case one of the penalty functions is simply equal to a constant.

Corollary 1 (A constant penalty function) *Let decision $i = l$ has the constant penalty functions $r_{l\alpha}(t) = \rho$ for $\alpha = 0, \ldots, k$. Then, the minimum mean penalty partitioning follows from (31) as*

$$\begin{aligned}\mathcal{P}_l &= \{t \in \mathbb{R}^r \mid \rho < \min_{j\in[0,..,l-1]} \mathsf{E}(\underline{r}_j|t)\}\\ \mathcal{P}_{i\in[0,..,l-1]} &= \{t \in \mathbb{R}^r \backslash \mathcal{P}_l \mid i = \arg \min_{j\in[0,..,l-1]} \mathsf{E}(\underline{r}_j|t)\}\end{aligned} \tag{34}$$

■

As an application, one can think of decision $i = l$ being the decision to *not* identify one of the hypotheses, thereby declaring the parameter solution *unavailable*, cf. (22). The solution would then be declared unavailable when the smallest misclosure-conditioned mean penalty is still considered too large, i.e. larger than ρ.

Finally note that we considered the occurrence of hypotheses as a discrete random variable, with its probabilities of occurrence given by the function $\mathsf{P}[\mathcal{H}_\alpha]$ (a stricter, but longer notation would have been $\mathsf{P}[\underline{\mathcal{H}}_\alpha = \mathcal{H}_\alpha]$). Specifying these probabilities may not be easy and may require extensive experience on the actual frequencies of their occurrence. In the absence of such experience however, guidance may be taken from considerations of symmetry or complexity. For instance, if there is no reason to believe that one alternative hypothesis is more likely to occur than another, then with $\mathsf{P}[\mathcal{H}_0] = \pi_0$, the probabilities of the alternative hypotheses are given as $\mathsf{P}[\mathcal{H}_\alpha] = (1-\pi_0)/k$ for $\alpha = 1, \ldots, k$. Also, with reference to the *principle of parsimony*, one could consider describing the probabilities of occurrence as decreasing functions of the bias-vector dimensions, q_α. For instance, in case of multiple outlier testing, it seems reasonable to attach a lower probability to the simultaneous occurrence of a higher number of outliers. As an example, having $\pi \ll 1-\pi$ as the probability of a single-outlier occurrence, one could model the probability of occurrence of an m-observation, q_α-outlier hypothesis as

$$\mathsf{P}[\mathcal{H}_\alpha] \propto \pi^{q_\alpha}(1-\pi)^{m-q_\alpha} \tag{35}$$

Although the assignment of probabilities $\mathsf{P}[\mathcal{H}_\alpha]$ may in general not be an easy task, some consolation can perhaps be taken from the following two considerations. First note, as the PDF (19) can be computed rigorously for any partitioning, that the hypothesis-conditioned quality description of the corresponding DIA-estimator will not suffer from inaccuracies in specifying $\mathsf{P}[\mathcal{H}_\alpha]$. Second we note, as $\mathsf{P}[\mathcal{H}_\alpha]$ in (31) occurs in a product with $r_{j\alpha}(t)$, that any inaccuracies in the probability assignment may be interpreted as a variation in the risk penalty assignment.

We now give four simple examples to illustrate the workings of (31). We often make use of the simpler short-hand notation $\pi_\alpha = \mathsf{P}[\mathcal{H}_\alpha]$.

Example 2 (*Detection-only*) Let $k = l = 1$, $r_{10} = r_{11} = \rho$, $\mathsf{E}(\underline{y}) \overset{\mathcal{H}_0}{\sim} \mathcal{N}_m(Ax, Q_{yy})$ and $\mathcal{H}_1 \neq \mathcal{H}_0$. Then (31) simplifies to $\mathcal{P}_0 = \{t \in \mathbb{R}^r \mid r_{00}F_0(t) + r_{01}F_1(t) < \rho(F_0(t) + F_1(t))\}$, from which follows

$$\begin{aligned}\mathcal{P}_0 &= \{t \in \mathbb{R}^r \mid \mathsf{E}(\underline{r}_0|t) < \rho\}\\ \mathcal{P}_1 &= \mathbb{R}^r \backslash \mathcal{P}_0\end{aligned} \tag{36}$$

In this case the null-hypothesis would only be accepted if its misclosure-conditioned mean penalty is small enough. This case is referred to as 'detection-only' as no identification of particular alternative hypotheses is asked for Zaminpardaz and Teunissen (2023). Rejection of the null-hypothesis would thus automatically lead to an unavailability of a parameter solution for x. □

Example 3 (*The $k = l = 1$ case, with varying penalties*) Without the assumption of the same penalty $r_{10} = r_{11}$ for decision $i = 1$, (31) simplifies, with $r_{00} < r_{10}$, to

$$\begin{aligned}\mathcal{P}_0 &= \{t \in \mathbb{R}^r \mid f_{\underline{t}}(t|\mathcal{H}_0) > c f_{\underline{t}}(t|\mathcal{H}_1)\}\\ \mathcal{P}_1 &= \mathbb{R}^r \backslash \mathcal{P}_0\end{aligned} \tag{37}$$

with $c = \frac{r_{01}-r_{11}}{r_{10}-r_{00}} \frac{1-\pi_0}{\pi_0}$. Note that $\mathcal{P}_0$ increases in size when π_0 gets larger at the expense of $\pi_1 = 1-\pi_0$ and/or when the relative penalty ratio $\frac{r_{01}-r_{11}}{r_{10}-r_{00}}$ gets smaller. This is also what one would like to happen: for a larger occurrence probability of the null-hypothesis, a larger acceptance region, with in the limit no rejection at all when $\pi_0 \to 1$. Similarly, also with a decreasing relative risk of making the wrong decision $i = 0$ while $\mathcal{H}_1$ is true, one would like the acceptance region to increase in size. □

Example 4 ($k = l = 2$ *and $\mathcal{P}_0$ is given*) In this case we have three hypotheses and three decisions. We assume $(r_{21} - r_{11})\pi_1 = (r_{12} - r_{22})\pi_2$. Then, with $c = \frac{\pi_0}{(r_{21}-r_{11})\pi_1} > 0$, the partitioning for the three hypotheses reads

$$\mathcal{P}_0 = \text{given}$$

$$\mathcal{P}_1 = \{t \in \mathbb{R}^r \backslash P_0 | \frac{f_{\underline{t}}(t|\mathcal{H}_1)}{f_{\underline{t}}(t|\mathcal{H}_0)} + c\ \mathtt{r}_{20} > \frac{f_{\underline{t}}(t|\mathcal{H}_2)}{f_{\underline{t}}(t|\mathcal{H}_0)} + c\ \mathtt{r}_{10}\}$$
$$\mathcal{P}_2 = \mathbb{R}^r \backslash \{\mathcal{P}_0 \cup \mathcal{P}_1\} \tag{38}$$

This shows that if $\mathtt{r}_{20}$ gets larger, i.e. the penalty of choosing $\mathcal{H}_2$ while $\mathcal{H}_0$ is true gets larger, then the region $\mathcal{P}_1$ gets larger at the expense of $\mathcal{P}_2$. □

Example 5 ($k = 1, l = 2$, *with undecided and* $\mathcal{P}_0$ *is given*) In this case, we have two hypotheses and three decisions. The partitioning for the three decisions follows then from (31), with $\mathtt{r}_{21} > \mathtt{r}_{11}$, as

$$\mathcal{P}_0 = \text{given}$$
$$\mathcal{P}_1 = \{t \in \mathbb{R}^r \backslash \mathcal{P}_0 \mid f_{\underline{t}}(t|\mathcal{H}_1) > c f_{\underline{t}}(t|\mathcal{H}_0)\}$$
$$\mathcal{P}_2 = \mathbb{R}^r \backslash \{\mathcal{P}_0 \cup \mathcal{P}_1\} \tag{39}$$

with $c = \frac{\mathtt{r}_{10} - \mathtt{r}_{20}}{\mathtt{r}_{21} - \mathtt{r}_{11}} \frac{\pi_0}{1-\pi_0}$ and $\mathcal{P}_2$ the undecided region □

3.2 Creating an operational misclosure partitioning

Partitioning (31) is only operational if the PDFs $f_{\underline{t}}(t|\mathcal{H}_\alpha)$ would be completely known. In our case, however, we also have to deal with the bias vectors b_α of $\mathcal{H}_\alpha$, cf. (5), and therefore, we only have the PDFs

$$f_{\underline{t}}(t|b_\alpha, \mathcal{H}_\alpha),\ \ b_\alpha \in \mathbb{R}^{q_\alpha} \tag{40}$$

available. We can now discriminate between the following three cases:

$(a) b_\alpha$ known
$(b) \underline{b}_\alpha$ random, with known PDF
$(c) b_\alpha$ unknown (41)

Case (a): When all the bias vectors are known, also the PDFs $f_{\underline{t}}(t|\mathcal{H}_\alpha) := f_{\underline{t}}(t|b_\alpha, \mathcal{H}_\alpha)$ are known and partitioning (31) can be applied directly.

Case (b): When the bias vectors are considered random with known PDF $f_{\underline{b}_\alpha}(b_\alpha|\mathcal{H}_\alpha)$, the marginal PDF $f_{\underline{t}}(t|\mathcal{H}_\alpha)$ can be constructed from the joint PDF $f_{\underline{t},\underline{b}_\alpha}(t, b_\alpha|\mathcal{H}_\alpha) = f_{\underline{t}|b_\alpha}(t|b_\alpha, \mathcal{H}_\alpha) f_{\underline{b}_\alpha}(b_\alpha|\mathcal{H}_\alpha)$ as

$$f_{\underline{t}}(t|\mathcal{H}_\alpha) = \int_{\mathbb{R}^{q_\alpha}} f_{\underline{t}|b_\alpha}(t|\beta, \mathcal{H}_\alpha) f_{\underline{b}_\alpha}(\beta|\mathcal{H}_\alpha) d\beta \tag{42}$$

where $f_{\underline{t}|b_\alpha}(t|\beta, \mathcal{H}_\alpha) := f_{\underline{t}}(t|\beta, \mathcal{H}_\alpha)$. Using (42), partitioning (31) can again be applied directly. For example, if it is believed that the distributional information on the biases can be captured by $f_{\underline{b}_\alpha}(b|\mathcal{H}_\alpha) = \mathcal{N}_{q_\alpha}(0, Q_\alpha)$, then the marginal PDF (42) becomes $f_{\underline{t}}(t|\mathcal{H}_\alpha) = \mathcal{N}_r(0, Q_{tt} + C_{t_\alpha} Q_\alpha C_{t_\alpha}^T)$, thus showing that the prior on the biases results under $\mathcal{H}_\alpha$ in a variance-inflation of $f_{\underline{t}}(t|\mathcal{H}_\alpha)$ in the hypothesized fault-direction $\mathcal{R}(C_{t_\alpha})$. Would, alternatively, the PDF of $\underline{b}_\alpha$ be so peaked that it becomes equal to a Dirac delta-function, $f_{\underline{b}_\alpha}(\beta|\mathcal{H}_\alpha) = \delta(\beta - b_\alpha)$, with b_α known, then substitution into (42) gives

$$f_{\underline{t}}(t|\mathcal{H}_\alpha) = f_{\underline{t}|\underline{b}_\alpha}(t|b_\alpha, \mathcal{H}_\alpha) := f_{\underline{t}}(t|b_\alpha, \mathcal{H}_\alpha) \tag{43}$$

thus recovering (40), but now with b_α known.

Case (c): As the above two cases, bias-known or bias-random, may generally not be applicable, one will have to work with an alternative approach to cope with the unknown bias vectors. We present two such approaches. If $f_{\underline{t}}(t|\mathcal{H}_\alpha)$ in (30) is replaced by $f_{\underline{t}}(t|b_\alpha, \mathcal{H}_\alpha)$, the mean penalty is obtained as function of the unknown biases, $\mathsf{E}(\underline{r}|b_1, \ldots, b_k)$. To cope with the unknown biases we try to capture the characteristics of the function by using its *average* $\bar{\mathsf{E}}(\underline{r})$ or by using an *estimate* $\hat{\mathsf{E}}(\underline{r})$. The first approach is realized if we replace $f_{\underline{t}}(t|b_\alpha, \mathcal{H}_\alpha)$ in $\mathsf{E}(\underline{r}|b_1, \ldots, b_k)$ by its average

$$\bar{f}_{\underline{t}}(t|\mathcal{H}_\alpha) = \frac{1}{|G_\alpha|} \int_{G_\alpha} f_{\underline{t}}(t|\beta, \mathcal{H}_\alpha) g_\alpha(\beta) d\beta \tag{44}$$

in which $|G_\alpha| = \int_{G_\alpha} d\beta$. To determine this average, we still need to choose the function $g_\alpha(\beta)$. As we generally do not know more about the biases $b_\alpha \in \mathbb{R}^{q_\alpha}$ than that they can occur freely around the origin, it seems reasonable to choose the function $g_\alpha(\beta)$ as a flat function, symmetric about the origin, and having sufficient domain to include all the practically sized biases. In the unweighted case, this would be the function $g_\alpha(\beta) = |G_\alpha|$ over the domain G_α.

In the second approach, we use the bias-estimates $\hat{b}_\alpha$ to estimate the mean penalty as $\hat{\mathsf{E}}(\underline{r}) = \mathsf{E}(\underline{r}|\hat{b}_1, \ldots, \hat{b}_k)$. This approach is generally simpler than constructing the average $\bar{\mathsf{E}}(\underline{r})$. Furthermore, as the following Lemma shows, it provides a *strict* upper bound on the mean penalty function.

Lemma 2 (Maximum mean penalty)*: Let the mean penalty be estimated as* $\hat{\mathsf{E}}(\underline{r}) = \mathsf{E}(\underline{r}|\hat{b}_1, \ldots, \hat{b}_k)$*, where* $\hat{b}_\alpha = \arg\max_{\beta \in \mathbb{R}^{q_\alpha}} f_{\underline{t}}(t|\beta, \mathcal{H}_\alpha)$*,* $\alpha = 1, \ldots, k$*. Then*

$$\hat{\mathsf{E}}(\underline{r}) = \max_{b_1 \in \mathbb{R}^{q_1}, \ldots, b_k \in \mathbb{R}^{q_k}} \mathsf{E}(\underline{r}|b_1, \ldots, b_k) \tag{45}$$

■

Proof The proof follows by noting that the bias-dependent functions $f_{\underline{t}}(t|b_\alpha, \mathcal{H}_\alpha)$ occur in a decoupled form in the non-negative linear combinations of $\mathsf{E}(\underline{r}|b_1, \ldots, b_k)$. Hence, its joint bias-maximizer is provided by the bias-maximizers of the individual functions $f_{\underline{t}}(t|b_\alpha, \mathcal{H}_\alpha)$. □

The relevance of this result is that by replacing in the mean penalty function the unknown bias vectors with their

estimates $\hat{b}_\alpha$, we automatically obtain a strict upperbound on the mean penalty, i.e. for *none* of the possible values that the bias vectors may take will the mean penalty be larger than this upperbound. Hence, by working with $\hat{\mathsf{E}}(\underline{r})$ instead of $\bar{\mathsf{E}}(\underline{r})$, one is assured of a *conservative* approach. As this property may be considered attractive in case of safety-critical applications, we will work in the following, when the biases are unknown, with $\hat{\mathsf{E}}(\underline{r})$. From using the above Lemma, its best partitioning is obtained as follows.

Theorem 2b (Minimum mean penalty testing) *The misclosure space partitioning* $\mathcal{P}_{i\in[0,\ldots,l]} \subset \mathbb{R}^r$ *that minimizes the mean penalty upperbound*

$$\hat{\mathsf{E}}(\underline{r}) = \sum_{i=0}^{l} \int_{\mathcal{P}_i} \sum_{\alpha=0}^{k} r_{i\alpha}(t)\, f_{\underline{t}}(t|\hat{b}_\alpha(t), \mathcal{H}_\alpha)\mathsf{P}[\mathcal{H}_\alpha]dt \tag{46}$$

where $\hat{b}_\alpha(t) = \arg\max_{\beta\in\mathbb{R}^{q_\alpha}} f_{\underline{t}}(t|\beta, \mathcal{H}_\alpha)$, *is given by*

$$\mathcal{P}_{i\in[0,\ldots,l]} = \{t \in \mathbb{R}^r | i = \arg\min_{j\in[0,\ldots,l]} \sum_{\alpha=0}^{k} r_{j\alpha}(t)\hat{F}_\alpha(t)\} \tag{47}$$

where $\hat{F}_\alpha(t) = f_{\underline{t}}(t|\hat{b}_\alpha(t), \mathcal{H}_\alpha)\mathsf{P}[\mathcal{H}_\alpha]$. ■

Note that all the results obtained so far in this section do not require the misclosure vector to be normally distributed. Also in the sections following we provide results that generally do not require such assumption. However, in all the examples following, we will assume the misclosure vector to be normally distributed as (7) and therefore that the bias-maximizer of $f_{\underline{t}}(t|b_\alpha, \mathcal{H}_\alpha)$ is given as $\hat{b}_\alpha = (C_{t_\alpha})^+ t$, cf. (6). The results of (31) and (47) then specialize to the following.

Theorem 2c (Minimum mean penalty testing) *Let the PDF of the misclosure vector be given as*

$$f_{\underline{t}}(t|b_\alpha, \mathcal{H}_\alpha) \overset{\mathcal{H}_\alpha}{\propto} \exp\{-\tfrac{1}{2}||t - C_{t_\alpha} b_\alpha||^2_{Q_{tt}}\} \tag{48}$$

Then, the misclosure space partitionings of (31), for known bias b_α, *and (47), for estimated bias* $\hat{b}_\alpha(t)$, *specialize to*

$$\mathcal{P}_{i\in[0,..,l]} = \{t \in \mathbb{R}^r \,|i = \arg\min_{j\in[0,..,l]} \sum_{\alpha=0}^{k} r_{j\alpha}(t) \exp\{+\tfrac{1}{2}T_\alpha(t)\}\} \tag{49}$$

where

$$T_\alpha(t) \begin{cases} \overset{(31)}{=} T_{q_\alpha}(t) - ||\hat{b}_\alpha(t) - b_\alpha||^2_{Q_{\hat{b}_\alpha\hat{b}_\alpha}} + \ln \pi_\alpha^2 \\ \overset{(47)}{=} T_{q_\alpha}(t) + \ln \pi_\alpha^2 \end{cases} \tag{50}$$

with $T_{q_\alpha}(t) = ||P_{C_{t_\alpha}} t||^2_{Q_{tt}} = ||\hat{b}_\alpha(t)||^2_{Q_{\hat{b}_\alpha\hat{b}_\alpha}}$, $T_{q_\alpha=0}(t) = 0$, *and* $\pi_\alpha = \mathsf{P}[\mathcal{H}_\alpha]$. ■

Proof As $||t - C_{t_\alpha} b_\alpha||^2_{Q_{tt}} = ||P^\perp_{C_{t_\alpha}} t||^2_{Q_{tt}} + ||\hat{b}_\alpha(t) - b_\alpha||^2_{Q_{\hat{b}_\alpha\hat{b}_\alpha}}$ and $||P^\perp_{C_{t_\alpha}} t||^2_{Q_{tt}} = ||t||^2_{Q_{tt}} - ||\hat{b}_\alpha(t)||^2_{Q_{\hat{b}_\alpha\hat{b}_\alpha}}$, we have

$$\begin{aligned} ||t - C_{t_\alpha} b_\alpha||^2_{Q_{tt}} = {} & ||t||^2_{Q_{tt}} - ||\hat{b}_\alpha(t)||^2_{Q_{\hat{b}_\alpha\hat{b}_\alpha}} \\ & + ||\hat{b}_\alpha(t) - b_\alpha||^2_{Q_{\hat{b}_\alpha\hat{b}_\alpha}} \end{aligned} \tag{51}$$

and therefore

$$\begin{aligned} F_\alpha(t) &= \pi_\alpha \exp\{-\tfrac{1}{2}||t - C_{t_\alpha} b_\alpha||^2_{Q_{tt}}\} \\ &= \exp\{-\tfrac{1}{2}||t||^2_{Q_{tt}}\} \exp\{+\tfrac{1}{2}T_\alpha(t)\} \end{aligned} \tag{52}$$

which upon substitution into (31) proves the result. □

Note, if $r^{(31)}_{i\alpha}(t) = r^{(47)}_{i\alpha}(t) \exp\{+\tfrac{1}{2}||\hat{b}_\alpha(t) - b_\alpha||^2_{Q_{\hat{b}_\alpha\hat{b}_\alpha}}\}$, with $r^{(31)}_{i\alpha}(t)$ and $r^{(47)}_{i\alpha}(t)$ being the penalty functions of (31) and (47), respectively, that the bias-known case transforms into the bias-estimated case, thus showing that the switch from the bias-known to the bias-estimated case, cf. (50), can also be interpreted as a use of different penalty functions.

We now give two simple examples to show the workings of (49).

Example 6 (*Detection only*): Let $k = l = 1$, with $\mathcal{H}_1$ being the most relaxed alternative hypothesis, $\mathsf{E}(\underline{t}) \in \mathbb{R}^r \backslash \{0\}$. Then $T_{q_1=r} = ||t||^2_{Q_{tt}}$, from which it follows with (49), and $r_{\alpha\alpha}(t) < r_{i\alpha}(t)$, $i \neq \alpha$, that

$$\begin{aligned} \mathcal{P}_0 &= \{t \in \mathbb{R}^r | \;\; ||t||^2_{Q_{tt}} < \tau^2\} \\ \mathcal{P}_1 &= \mathbb{R}^r \backslash \mathcal{P}_0 \end{aligned} \tag{53}$$

with $\tau^2 = \ln\left[\frac{r_{10}(t) - r_{00}(t)}{r_{01}(t) - r_{11}(t)} \frac{\pi_0}{\pi_1}\right]^2$. This shows how the overall-model test statistic $T_{q_1=r} = ||t||^2_{Q_{tt}}$ is used in the acceptance or rejection of $\mathcal{H}_0$. □

Example 7 (*Undecided included*): Let $k = 1$, $l = 2$ and assume that $\mathcal{P}_0$ is a-priori given. Thus we have two hypotheses and three decisions. As alternative hypothesis, we take $\mathcal{H}_1 : \mathsf{E}(\underline{t}) = C_1 b_1 \neq 0$. Then $T_{q_1} = ||P_{C_1} t||^2_{Q_{tt}} = ||\hat{b}(t)||^2_{Q_{\hat{b}\hat{b}}}$, from which it follows with (49), and $r_{11}(t) < r_{21}(t)$, that

$$\begin{aligned} \mathcal{P}_1 &= \{t \in \mathbb{R}^r \backslash \mathcal{P}_0 \mid ||\hat{b}_1(t)||^2_{Q_{\hat{b}\hat{b}}} > \tau^2\} \\ \mathcal{P}_2 &= \mathbb{R}^r \backslash \{\mathcal{P}_0 \cup \mathcal{P}_1\} \end{aligned} \tag{54}$$

with $\tau^2 = \ln\left[\frac{r_{10}(t) - r_{20}(t)}{r_{21}(t) - r_{11}(t)} \frac{\pi_0}{\pi_1}\right]^2$. The misclosure space partitioning is shown in Fig. 3 for $q_1 = 1$ and $r = 2$. Note how the size of the undecided region $\mathcal{P}_2$ is driven by r_{20} and r_{21}. If both get larger then $\mathcal{P}_1$ gets larger and $\mathcal{P}_2$ smaller. □

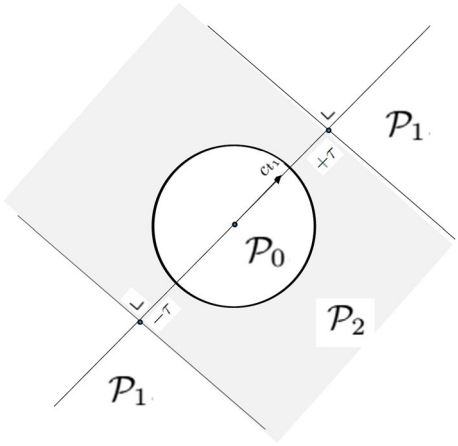

Fig. 3 Misclosure space partitioning in $\mathbb{R}^{r=2}$: $\mathcal{P}_0$ for detection, $\mathcal{P}_1$ for identifying $\mathcal{H}_1$, and $\mathcal{P}_2$ for unavailability decision of Example 7

3.3 Maximizing the probability of correct decisions

The minimum mean penalty partitioning of misclosure space simplifies if an additional simplifying structure is given to the set of penalty functions. This is the case for instance when $l = k$ and all correct decisions are given the same penalty and the penalties for incorrect decisions are symmetrized.

Corollary 2a (Symmetric penalties) *For symmetric and identical correct-decision penalties,* $\mathtt{r}_{i\alpha}(t) = \mathtt{r}_{\alpha i}(t)$ *and* $\mathtt{r}_{ii}(t) = r(t)$, $i, \alpha \in [0, \ldots, l = k]$, *the minimum mean penalty misclosure partitionings of* (31) *and* (49) *simplify, respectively, to*

$$\mathcal{P}_{i\in[0,\ldots,k]} = \{t \in \mathbb{R}^r | F_i(t) > F_j(t) + g_{ij}(t), \forall j \neq i\} \tag{55}$$

and

$$\mathcal{P}_{i\in[0,\ldots,k]} = \{t \in \mathbb{R}^r | T_i(t) > T_j(t) + h_{ij}(t), \forall j \neq i\} \tag{56}$$

with

$$\begin{cases} g_{ij}(t) = \sum\limits_{\alpha=0,\neq i,\neq j}^{k} \mu_{ij\alpha}(t) F_\alpha(t) \\ h_{ij}(t) = \ln[1 + \sum\limits_{\alpha=0,\neq i,\neq j}^{k} \mu_{ij\alpha}(t) \exp\{+\frac{1}{2}(T_\alpha(t) - T_j(t))\}]^2 \end{cases} \tag{57}$$

where $\mu_{ij\alpha}(t) = \frac{\mathtt{r}_{i\alpha}(t) - \mathtt{r}_{j\alpha}(t)}{\mathtt{r}_{ij}(t) - r(t)}$, $\mathtt{r}_{ij}(t) > r(t)$ *for* $i \neq j$. ■

This result shows how the g- and h-functions drive the difference between the individual partitioning subsets. For instance, if $g_{ij}(t) > 0$ and $h_{ij}(t) > 0$, then $\mathcal{P}_i$ can expected to be smaller than $\mathcal{P}_j$ for when their probability of hypothesis occurrence is equal. This happens when the penalties of decision i are larger than those of decision j, $\mathtt{r}_{i\alpha}(t) > \mathtt{r}_{j\alpha}(t)$.

A further simplification is reached when all penalties for incorrect decisions are taken to be equal, since then the g- and h-functions of (55) and (56) vanish, $g_{ij}(t) = h_{ij}(t) \equiv 0$. As an example consider the case that $\mathtt{r}_{\alpha\alpha}(t) = r_\alpha$ and $\mathtt{r}_{i\alpha}(t) = 1$ for $i \neq \alpha$. Then the penalty functions become

$$\mathtt{r}_{i\alpha}(t) = 1 - \delta_{i\alpha}(1 - r_\alpha), \ i, \alpha = 0, \ldots, k \tag{58}$$

with $\delta_{i\alpha} = 1$ for $i = \alpha$ and $\delta_{i\alpha} = 0$ otherwise, from which the mean penalty follows as

$$\begin{aligned} \mathsf{E}(\underline{\mathtt{r}}) &= \sum_{i=0}^{k} \sum_{\alpha=0}^{k} \mathtt{r}_{i\alpha}(t) \mathsf{P}[\underline{t} \in \mathcal{P}_i, \mathcal{H}_\alpha] \\ &= 1 - \sum_{\alpha=0}^{k} \rho_\alpha \mathsf{P}[\underline{t} \in \mathcal{P}_\alpha, \mathcal{H}_\alpha] \end{aligned} \tag{59}$$

with reward $\rho_\alpha = 1 - r_\alpha$. Minimizing the mean penalty is now the same as maximizing a reward-weighted probability sum of correct decisions. This simplification also translates into the solution of the testing partitioning.

Corollary 2b (Maximum correct decision probability) *Let* $l = k$ *and the penalty functions be given as* (58). *Then* (31) *and* (49) *simplify respectively to*

$$\mathcal{P}_{i\in[0,..,k]} = \{t \in \mathbb{R}^r \mid i = \arg \max_{\alpha\in[0,\ldots,k]} \rho_\alpha F_\alpha(t)\} \tag{60}$$

and

$$\mathcal{P}_{i\in[0,..,k]} = \{t \in \mathbb{R}^r | i = \arg \max_{\alpha\in[0,\ldots,k]} (T_\alpha(t) + \ln \rho_\alpha^2)\} \tag{61}$$

with $\rho_\alpha = 1 - r_\alpha$. ■

Note, since $F_\alpha(t) = \pi_\alpha f_{\underline{t}}(t|\mathcal{H}_\alpha)$, that through products $\rho_\alpha \pi_\alpha$, $\alpha = 0, \ldots, k$, credence is given to hypotheses. The larger $\rho_\alpha \pi_\alpha$ gets, the more credence is given to $\mathcal{H}_\alpha$. Although the reward $\rho_\alpha = 1 - r_\alpha$ and the hypothesis occurrence probability π_α both come together as a product, and therefore, as such, can create the same effect on $\mathcal{P}_\alpha$, it is important to realize that they have a different origin, i.e. the reward ρ_α is user-driven, while the probability π_α is model-driven. Furthermore, the π_α's have to sum up to 1, while such is not required for the ρ_α's.

To provide a clearer description of the detection and identification steps in the above testing procedure, we may separate the conditions for $\mathcal{P}_0$ and $\mathcal{P}_{i\in[1,\ldots,k]}$. For (61) this gives,

$$\mathcal{P}_0 = \{t \in \mathbb{R}^r | \max_{\alpha\in[1,\ldots,k]} \left(T_\alpha(t) + \ln \rho_\alpha^2\right) < \ln[\rho_0 \pi_0]^2\}$$

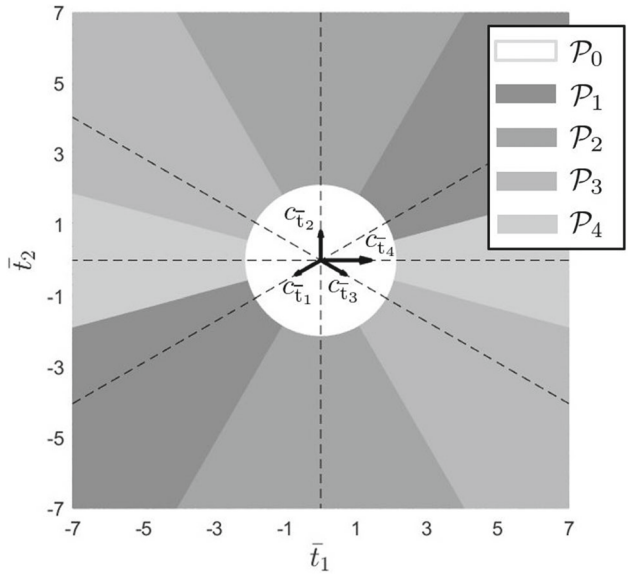

Fig. 4 Misclosure space partitioning in $\mathbb{R}^{r=2}$ for single-outlier data snooping (cf. Example 9)

$$\mathcal{P}_{i\in[1,\ldots,k]} = \{t \in \mathbb{R}^r \backslash \mathcal{P}_0 | i = \arg \max_{\alpha\in[1,\ldots,k]} \left(T_\alpha(t) + \ln \rho_\alpha^2\right)\} \tag{62}$$

This shows, given the misclosure vector t, that the detection step consists of computing the maximum of $T_\alpha(t) + \ln \rho_\alpha^2$ over all k alternative hypotheses and checking whether this is smaller than the constant $\ln[\rho_0\pi_0]^2$. If it is, then the null-hypothesis is accepted. If not, then the maximum determines which of the alternative hypotheses is identified as the reason for the rejection of the null-hypothesis.

In the following examples, we illustrate how the above determined testing partitioning compares or specializes to some of the testing procedures used in practice.

Example 8 (*Bias known vs. bias unknown*) In this example, we illustrate the role the bias plays in the transition from (60) to (61). From the decomposition

$$||t - C_{t_\alpha} b_\alpha||^2_{Q_{tt}} = ||P^\perp_{C_{t_\alpha}} t||^2_{Q_{tt}} + ||\hat{b}_\alpha(t) - b_\alpha||^2_{Q_{\hat{b}_\alpha\hat{b}_\alpha}} \tag{63}$$

it follows, with $F_\alpha(t) = \pi_\alpha f_{\underline{t}}(t|\mathcal{H}_\alpha) \propto \exp\{-\frac{1}{2}||t - C_{t_\alpha} b_\alpha||^2_{Q_{tt}}$, that the objective function $\rho_\alpha F_\alpha(t)$ of (60) is driven by two different measures of inconsistency: the inconsistency of t with the range space of C_{t_α} as measured by $||P^\perp_{C_{t_\alpha}} t||^2_{Q_{tt}}$ and the difference between the estimated and known bias, as measured by $||\hat{b}_\alpha(t) - b_\alpha||^2_{Q_{\hat{b}_\alpha\hat{b}_\alpha}}$. This second discrepancy measure disappears in case of (61), as then the unknown bias is replaced by its estimate, thus giving

$$||t - C_{t_\alpha}\hat{b}_\alpha||^2_{Q_{tt}} = ||P^\perp_{C_{t_\alpha}} t||^2_{Q_{tt}} = ||t||^2_{Q_{tt}} - T_{q_\alpha} \tag{64}$$

and therefore (61). Note, as alternative to the conservative approach (cf. Lemma 2), that one may also consider using the approximation $||t - C_{t_\alpha} b_\alpha||^2_{Q_{tt}} \approx ||P^\perp_{C_{t_\alpha}} t||^2_{Q_{tt}} + q_\alpha$, since $\mathsf{E}(||\hat{\underline{b}}_\alpha - b_\alpha||^2_{Q_{\hat{b}_\alpha\hat{b}_\alpha}} | \mathcal{H}_\alpha) = q_\alpha$. □

Example 9 (*Datasnooping with $\mathcal{P}_0$ known*) In this example, we consider the detection subset $\mathcal{P}_0$ to be given and equal to the acceptance region of the overall model test,

$$\mathcal{P}_0 = \{t \in \mathbb{R}^r \; | ||t||^2_{Q_{tt}} \le \tau^2\} \tag{65}$$

Furthermore, we consider the case that the C-matrices of all alternative hypotheses are one-dimensional, i.e. $q_\alpha = 1$, $C_\alpha = c_\alpha$ for $\alpha = 1, \ldots, k$. This is the case, for instance, when only single blunders in the $k = m$ observations are considered. As there are no differences in the complexities of the k alternative hypotheses and no reason for assuming certain alternative hypotheses to be more likely than others, the choice $\pi_\alpha = \text{constant}, \alpha = 1, \ldots, k$, seems a reasonable one. Additionally, we assume that no penalties are assigned to correct decisions, i.e. $r_\alpha = 0, \alpha = 1, \ldots, k$. Then, $T_\alpha = T_{q_\alpha} + \text{constant}$, which, together with $T_{q_\alpha} = w^2_\alpha$, gives for (61),

$$\mathcal{P}_{i\in[1,..k]} = \{t \in \mathbb{R}^r \backslash \mathcal{P}_0 \mid i = \arg \max_{\alpha\in[1,\ldots,k]} |w_\alpha|\} \tag{66}$$

This is the partitioning corresponding to Baarda's standard datasnooping procedure (Baarda 1968b) in case $k = m$ and the c_α are equal to the canonical unit vectors.

The above partitioning, cf. (65) and (66), is shown in Fig. 4 for the same B-matrix and same Q_{tt}-matrix as used in Example 1. In this case, however, $k = 4$ with $c_1 = [1, 0, 0]^T$, $c_2 = [0, 1, 0]^T$, $c_3 = [0, 0, 1]^T$ and $c_4 = [1, 2, 3]^T$. Furthermore, so as to view misclosure space in the standard metric, the misclosure vector was transformed with

$$R = \begin{bmatrix} -\sqrt{2}/2 & -\sqrt{2}/2 \\ -\sqrt{6}/6 & +\sqrt{6}/6 \end{bmatrix} \tag{67}$$

such that the transformed misclosure vector $\bar{\underline{t}} = R\underline{t}$ has identity variance matrix, $Q_{\bar{t}\bar{t}} = I_2$. The detection-region $\mathcal{P}_0$ shows therefore as a circle instead of an ellipse, cf. Fig. 2. The $c_{\bar{t}}$-vectors are then given as $c_{\bar{t}_i} = RB^T c_i, i = 1, 2, 3, 4$. □

Example 10 (*Datasnooping with $\mathcal{P}_0$ unknown*) In this example, the same settings are used as in the previous example, except that now the detection subset $\mathcal{P}_0$ is assumed unknown. Using $\pi_\alpha = \frac{1}{k}(1 - \pi_0), \alpha = 1, \ldots, k$, and $||P_{c_{t_\alpha}} t||^2_{Q_{tt}} = w^2_\alpha$, $\mathcal{P}_0$ follows from the first expression of (61) as

$$\mathcal{P}_0 = \{t \in \mathbb{R}^r \mid \max_{\alpha\in[1,\ldots,k]} w^2_\alpha \le a\} \tag{68}$$

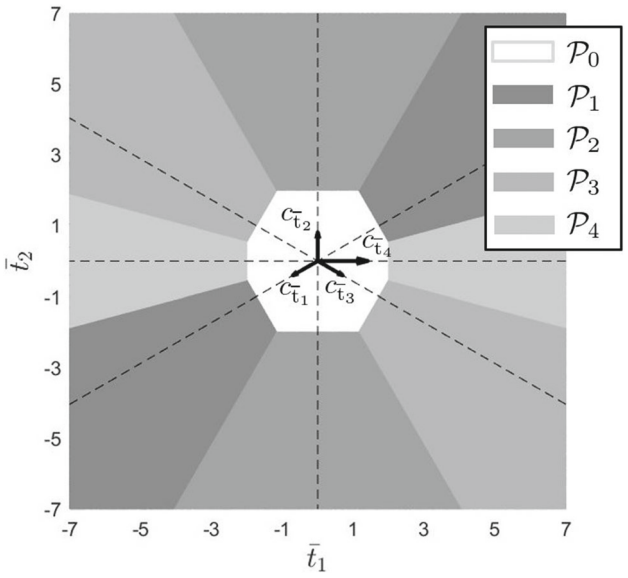

Fig. 5 Misclosure space partitioning in $\mathbb{R}^{r=2}$ for single-outlier data snooping, showing the detection region $\mathcal{P}_0$ as a polygon (cf. Example 10)

with $a = \ln\left(\frac{k\pi_0}{1-\pi_0}\right)^2$. The k corresponding subsets $\mathcal{P}_i$ are given by (66).

Expression (68) shows that $\mathcal{P}_0$ is determined as the intersection of k pairs of $(k-1)$-dimensional hyperplanes, having the direction vectors c_{t_α}, $\alpha = 1, \ldots, k$, as their normals. The distance of the origin to these hyperplanes is governed by the constants $\ln\left(\frac{k\pi_0}{1-\pi_0}\right)^2$, $\alpha = 1, \ldots, k$. These distances, and thereby the volume of $\mathcal{P}_0$, get larger when k and/or π_0 increases. Hence, the acceptance region $\mathcal{P}_0$ increases in size when the probability of $\mathcal{H}_0$-occurrence increases and/or the number of alternative hypotheses increases.

The above partitioning is shown in Fig. 5 for the same model and hypotheses as used in Example 9. Compare this geometry with that of Fig. 4. □

3.4 Including an undecided region $\mathcal{P}_{k+1}$

We now extend Corollary 2b so as to also include an undecided region.

Corollary 2c (Maximum correct decision probability) *Let $l = k+1$, the penalty functions be given as* (58) *and the undecided penalties as $\imath_{k+1,\alpha}(t) = u_\alpha$, $\alpha = 0, \ldots, k$. Then* (31) *and* (49) *simplify, respectively, to*

$$\begin{aligned}\mathcal{P}_{i\in[0,..,k]} &= \{t \in \mathbb{R}^r \backslash \mathcal{P}_{k+1} \mid i = \arg\max_{\alpha\in[0,\ldots,k]} \rho_\alpha F_\alpha(t)\}\\ \mathcal{P}_{k+1} &= \{t \in \mathbb{R}^r \mid \max_{\alpha\in[0,\ldots,k]} \rho_\alpha F_\alpha(t) \le G(t)\}\end{aligned} \tag{69}$$

and

$$\begin{aligned}\mathcal{P}_{i\in[0,..,k]} &= \{t \in \mathbb{R}^r \backslash \mathcal{P}_{k+1} | i = \arg\max_{\alpha\in[0,\ldots,k]} \left(T_\alpha(t) + \ln \rho_\alpha^2\right)\}\\ \mathcal{P}_{k+1} &= \{t \in \mathbb{R}^r \mid \max_{\alpha\in[0,\ldots,k]} \left(T_\alpha(t) + \ln \rho_\alpha^2\right) < H(t)\}\end{aligned} \tag{70}$$

where $\rho_\alpha = 1 - r_\alpha$, $\mu_\alpha = 1 - u_\alpha$, $G(t) = \sum_{\alpha=0}^k \mu_\alpha F_\alpha(t)$ and $H(t) = \ln[\sum_{\alpha=0}^k \mu_\alpha \exp\{+\frac{1}{2}T_\alpha(t)\}]^2$. ■

Note, when the detection region $\mathcal{P}_0$ would be a-priori given, that (69) would change to

$$\begin{cases}\mathcal{P}_{i\in[1,..,k]} = \{t \in \mathbb{R}^r \backslash \{\mathcal{P}_0 \cup \mathcal{P}_{k+1}\} \mid \\ \qquad\qquad i = \arg\max_{\alpha\in[1,\ldots,k]} \rho_\alpha F_\alpha(t)\} \\ \mathcal{P}_{k+1} \quad = \{t \in \mathbb{R}^r \backslash \mathcal{P}_0 \mid \max_{\alpha\in[1,\ldots,k]} \rho_\alpha F_\alpha(t) \le G(t)\}\end{cases} \tag{71}$$

with a similar change to (70). Also note, when comparing Corollary 2c with Corollary 2b, that the defining conditions for $\mathcal{P}_{i\in[0,\ldots,k]}$ look the same, cf. (60) vs (69), but are actually not the same, since they apply, in case of Corollary 2c, to the restricted space $\mathbb{R}^r \backslash \mathcal{P}_{k+1}$, i.e. misclosure space with the undecided region excluded. The characteristics of the undecided region $\mathcal{P}_{k+1} \subset \mathbb{R}^r$ are driven by the undecided penalties u_α one assigns to the hypotheses $\mathcal{H}_\alpha, \alpha = 0, \ldots, k$. One can expect $\mathcal{P}_{k+1}$ to be empty if one assigns the maximum penalty to all. And indeed, if $u_\alpha = 1$ for $\alpha = 0, \ldots, k$, then the inequality in (69) will never be satisfied, implying $\mathcal{P}_{k+1} = \emptyset$. Similarly, if no penalty at all is put on an undecided decision and thus $u_\alpha = 0$ for $\alpha = 0, \ldots, k$, then the inequality of (69) is trivially fulfilled, implying that $\mathcal{P}_{k+1} = \mathbb{R}^r$. Hence, in this case no other decision than an undecided decision will be made.

If all the undecided penalties are equal, $u_\alpha = u$, and all the rewards equal one, $\rho_\alpha = 1$, $\alpha = 0, \ldots, k$, then division by $f_{\underline{t}}(t) = \sum_{\alpha=0}^k \pi_\alpha f_{\underline{t}}(t|\mathcal{H}_\alpha)$ of both sides of (69)'s inequality, gives for the undecided region the inequality

$$\max_{\alpha\in[0,\ldots,k]} \mathsf{P}[\mathcal{H}_\alpha|t] < 1 - u \tag{72}$$

As the probability $\mathsf{P}[\mathcal{H}_\alpha|t]$ for $t \in \mathcal{P}_\alpha$ tends to decrease towards the boundaries of $\mathcal{P}_\alpha$, one can expect the undecided region to be located at the boundaries of these regions and therefore indeed provide an undecided decision if identifiability between two hypotheses becomes problematic.

Example 11 (*Two hypotheses and three decisions*) As an application of Corollary 2c, let $k = 1$, $l = 2$, and assume $\mu_\alpha = \mu$ for $\alpha = 0, 1$. The two hypotheses considered are $\underline{t} \overset{\mathcal{H}_0}{\sim} \mathcal{N}_r(0, Q_{tt})$ and $\underline{t} \overset{\mathcal{H}_1}{\sim} \mathcal{N}_r(C_{t_1}b_1, Q_{tt})$. For the three decisions, we need to determine the partitioning $\mathbb{R}^r =$

$\mathcal{P}_0 \cup \mathcal{P}_1 \cup \mathcal{P}_2$. For $\mathcal{P}_0$ and $\mathcal{P}_1$, we obtain from (70),

$$\begin{cases} \mathcal{P}_0 = \{t \in \mathbb{R}^r \backslash \mathcal{P}_2 \mid T_0(t) + \ln \rho_0^2 > T_1(t) + \ln \rho_1^2\} \\ \mathcal{P}_1 = \mathbb{R}^r \backslash \{\mathcal{P}_0 \cup \mathcal{P}_2\} \end{cases}$$

which can be rewritten as

$$\begin{cases} \mathcal{P}_0 = \{t \in \mathbb{R}^r \backslash \mathcal{P}_2 \mid ||P_{C_{t_1}} t||^2_{Q_{tt}} < \ln \left[\frac{\rho_0 \pi_0}{\rho_1 \pi_1}\right]^2\} \\ \mathcal{P}_1 = \mathbb{R}^r \backslash \{\mathcal{P}_0 \cup \mathcal{P}_2\} \end{cases} \tag{73}$$

We now determine the undecided region $\mathcal{P}_2$. According to (70), $\mathcal{P}_2$ is defined by the two inequalities

$$\begin{cases} T_0(t) + \ln \rho_0^2 < \ln[\mu \exp\{\frac{1}{2} T_0(t)\} + \mu \exp\{\frac{1}{2} T_1(t)\}]^2 \\ T_1(t) + \ln \rho_1^2 < \ln[\mu \exp\{\frac{1}{2} T_0(t)\} + \mu \exp\{\frac{1}{2} T_1(t)\}]^2 \end{cases}$$

which can be rewritten as

$$||P_{C_{t_1}} t||^2_{Q_{tt}} > \text{LB} \quad \text{and} \quad ||P_{C_{t_1}} t||^2_{Q_{tt}} < \text{UB} \tag{74}$$

with the bounds given as

$$\text{LB} = \ln \left[\frac{1-\mu/\rho_0}{\mu/\rho_1} \frac{\rho_0 \pi_0}{\rho_1 \pi_1}\right]^2, \quad \text{UB} = \ln \left[\frac{\mu/\rho_0}{1-\mu/\rho_1} \frac{\rho_0 \pi_0}{\rho_1 \pi_1}\right]^2 \tag{75}$$

It follows from (74) that $\mathcal{P}_2$ is *empty* if LB > UB, in which case $\mathcal{P}_0$ and $\mathcal{P}_1$ follow from (73) as

$$\begin{cases} \mathcal{P}_0 = \{t \in \mathbb{R}^r \mid ||P_{C_{t_1}} t||^2_{Q_{tt}} < \ln \left[\frac{\rho_0 \pi_0}{\rho_1 \pi_1}\right]^2\} \\ \mathcal{P}_1 = \mathbb{R}^r \backslash \mathcal{P}_0 \end{cases} \tag{76}$$

Since LB > UB if $\mu < (\frac{1}{\rho_0} + \frac{1}{\rho_1})^{-1}$, it follows that $\mathcal{P}_{k+1}$ is empty if μ is small enough.

The undecided region $\mathcal{P}_2$ is *nonempty* if LB < UB. As both inequalities of (74) need then to be satisfied, its complement $\mathbb{R}^r \backslash \mathcal{P}_2$ requires that only one of the following two inequalities need to be satisfied,

$$||P_{C_{t_1}} t||^2_{Q_{tt}} < \text{LB} \quad \text{or} \quad ||P_{C_{t_1}} t||^2_{Q_{tt}} > \text{UB} \tag{77}$$

Since $\text{LB} < \ln \left[\frac{\rho_0 \pi_0}{\rho_1 \pi_1}\right]^2$ and $\text{UB} > \ln \left[\frac{\rho_0 \pi_0}{\rho_1 \pi_1}\right]^2$ if LB < UB, it follows from (73) and (77) that in case $\mathcal{P}_2$ is nonempty, the three subsets are given as

$$\begin{cases} \mathcal{P}_0 = \{t \in \mathbb{R}^r \mid ||P_{C_{t_1}} t||^2_{Q_{tt}} < \text{LB}\} \\ \mathcal{P}_1 = \{t \in \mathbb{R}^r \mid ||P_{C_{t_1}} t||^2_{Q_{tt}} > \text{UB}\} \\ \mathcal{P}_2 = \{t \in \mathbb{R}^r \mid \text{LB} < ||P_{C_{t_1}} t||^2_{Q_{tt}} < \text{UB}\} \end{cases} \tag{78}$$

An illustration of this misclosure partitioning is given in Fig. 6. Compare this with the partitioning of Example 7 and Fig. 3. □

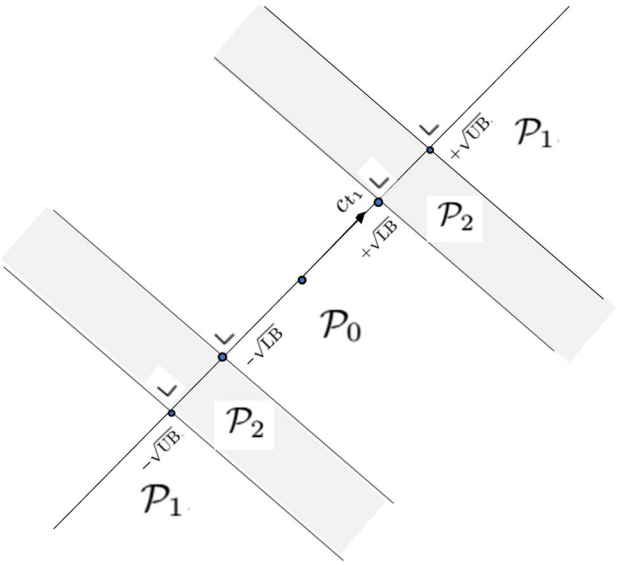

Fig. 6 Misclosure space partitioning of Example 11: $\mathcal{P}_0$ for detection, $\mathcal{P}_1$ for identifying $\mathcal{H}_1$, and $\mathcal{P}_2$ for unavailability decision

Example 12 (*Datasnooping with $\mathcal{P}_0$ given and undecided included*) Consider the null- and alternative hypotheses $\underline{y} \overset{\mathcal{H}_0}{\sim} \mathcal{N}_m(Ax, Q_{yy})$ and $\underline{y} \overset{\mathcal{H}_\alpha}{\sim} \mathcal{N}_m(Ax + c_\alpha b_\alpha, Q_{yy})$, with $\pi_\alpha = \frac{1-\pi_0}{k}$, $\alpha = 1, \ldots, k$, and assume the penalty functions (58) with $r_\alpha = 1 - \rho$, together with the undecided penalties $u_\alpha = 1 - \mu$. As the a-priori chosen detection region we take the acceptance region of the overall model test. Then the misclosure space partitioning follows from Corollary 2c, cf. (71), as

$$\begin{aligned} \mathcal{P}_0 &= \{t \in \mathbb{R}^r \mid ||t||^2_{Q_{tt}} < \tau^2\} \\ \mathcal{P}_{k+1} &= \{t \in \mathbb{R}^r \backslash \mathcal{P}_0 | \max_{\alpha \in [1,\ldots,k]} w_\alpha^2(t) < h(t)\} \\ \mathcal{P}_{i \in [1,\ldots,k]} &= \{t \in \mathbb{R}^r \backslash \{\mathcal{P}_0 \cup \mathcal{P}_{k+1}\} | i = \arg \max_{\alpha \in [1,\ldots,k]} w_\alpha^2(t)\} \end{aligned} \tag{79}$$

where

$$\begin{aligned} h(t) &= \ln[\tfrac{\mu}{\rho} \exp\{\tfrac{1}{2} a\} + \tfrac{\mu}{\rho} \textstyle\sum_{\alpha=1}^k \exp\{\tfrac{1}{2} w_\alpha^2(t)\}]^2 \\ a &= \ln \left[\frac{k \pi_0}{1-\pi_0}\right]^2 \end{aligned} \tag{80}$$

Compare this partitioning with that of Example 9. Would the undecided region $\mathcal{P}_{k+1}$ be empty, then the above partitioning reduces back to that of Example 9, cf. (65) and (66). The undecided region is empty, $\mathcal{P}_{k+1} = \emptyset$, if the undecided reward is zero, $\mu = 1 - u = 0$. The $\mathcal{P}_{k+1}$-defining inequality can then never be satisfied. Also note that $h(t)$ gets larger if the undecided reward $\mu = 1 - u$ gets larger, which then, as expected, also increases the size of the undecided region $\mathcal{P}_{k+1}$.

With the above partitioning, the testing would proceed as follows. First one would execute the detection-step by

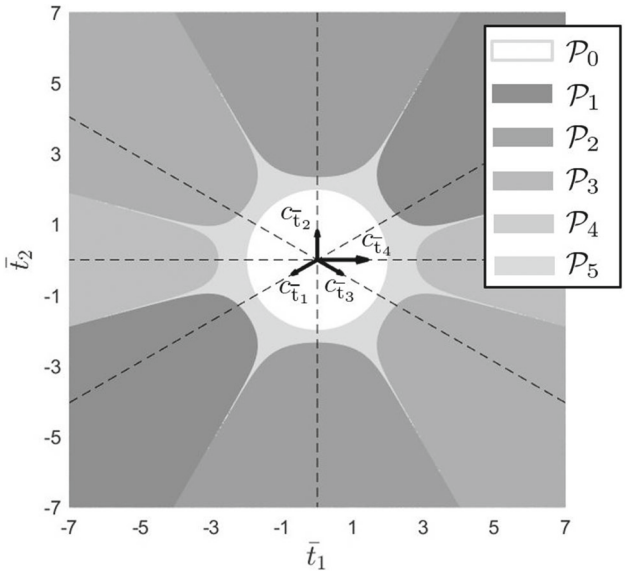

Fig. 7 Misclosure space partitioning in $\mathbb{R}^{r=2}$ for single-outlier data snooping, with an undecided region $\mathcal{P}_5$ included (cf. Example 12)

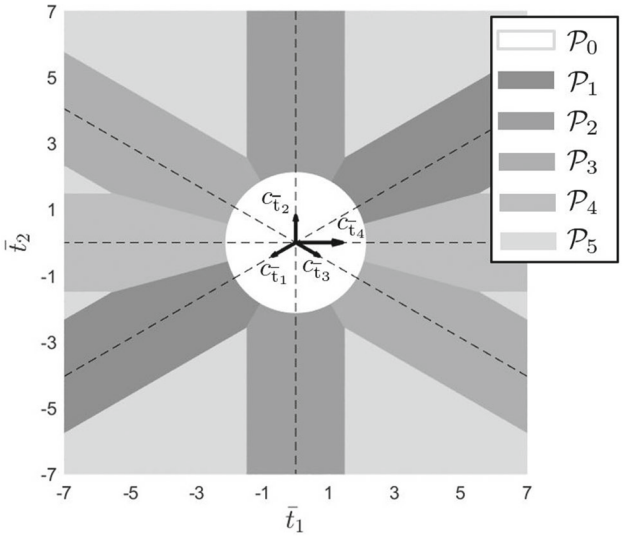

Fig. 8 Misclosure space partitioning in $\mathbb{R}^{r=2}$ for single-outlier data snooping, with an alternative undecided region $\mathcal{P}_5$ included (cf. Example 13)

checking whether or not $t \in \mathcal{P}_0$. If so, then $\mathcal{H}_0$ would be accepted. If not, then one would proceed to check whether or not $t \in \mathcal{P}_{k+1}$. This is done by computing the largest $w_\alpha^2(t)$, $\alpha = 1, \ldots, k$, say $w_i^2(t)$, and checking whether it is less than $h(t)$. If not, then $\mathcal{H}_i$ is the identified hypothesis. If so, then $t \in \mathcal{P}_{k+1}$ and the decision is made that no parameter solution can be provided.

Note that the ith alternative hypothesis is identified if $w_i^2(t) \geq w_\alpha^2(t)$, $\forall\alpha$, while $w_i^2(t) \geq h(t)$ and $w_i^2(t) > \tau^2 - ||P_{c_{t_i}}^{\perp} t||_{Q_{tt}}^2$. The latter two conditions ensure that such identification only happens if the in absolute value largest w-statistic is also sufficiently large.

The above partitioning is shown in Fig. 7 for the same model and hypotheses as used in Example 9. Compare this geometry with that of Fig. 5 and note how the undecided region $\mathcal{P}_5$ separates the regions $\mathcal{P}_{i\in[0,\ldots,4]}$ when the biases are large enough to be detected, but yet too small to be identified. □

Example 13 (*Datasnooping with $\mathcal{P}_0$ given and alternative undecided region included*) This example is to illustrate that the liberal definition of penalty functions in Sect. 3.1 allows one to interpret existing testing procedures in terms of assigned penalties. In Teunissen (2018), p.66, the following partitioning was considered,

$$\begin{aligned}
\mathcal{P}_0 &= \{t \in \mathbb{R}^r \,|\, ||t||_{Q_{tt}}^2 < \tau^2\} \\
\mathcal{P}_{k+1} &= \{t \in \mathbb{R}^r \backslash \mathcal{P}_0 |\; ||t||_{Q_{tt}}^2 - \max_{\alpha\in[1,\ldots,k]} w_\alpha^2(t) > \bar{\tau}^2\} \\
\mathcal{P}_{i\in[1,\ldots,k]} &= \{t \in \mathbb{R}^r \backslash \{\mathcal{P}_0 \cup \mathcal{P}_{k+1}\} | i = \arg \max_{\alpha\in[1,\ldots,k]} w_\alpha^2(t)\}
\end{aligned} \tag{81}$$

Its geometry is shown in Fig. 8 for the same model and hypotheses as used in Example 9. The idea behind this chosen undecided region $\mathcal{P}_{k+1}$ is that the sample of $\underline{t}$ should lie close enough to a fault line $c_{t_\alpha} b_\alpha$ for that hypothesis to be identifiable.

We can now use the results of Corollary 2c, cf. (70), to show the penalties that result in partitioning (81). If we assume $\rho_\alpha = 0$ and take the undecided rewards as

$$\mu_\alpha(t) = \frac{1}{k+1} \exp\{+\tfrac{1}{2}\left(||P_{C_{t_\alpha}}^{\perp} t||_{Q_{tt}}^2 - \ln \pi_\alpha^2 - \bar{\tau}'^2\right)\} \tag{82}$$

it follows with $||P_{C_{t_\alpha}}^{\perp} t||_{Q_{tt}}^2 = ||t||_{Q_{tt}}^2 - T_\alpha(t) + \ln \pi_\alpha^2$, from (70) that

$$\mathcal{P}_{k+1} = \{t \in \mathbb{R}^r \backslash \mathcal{P}_0 |\; ||t||_{Q_{tt}}^2 - \max_{\alpha\in[1,\ldots,k]} T_\alpha(t) > \bar{\tau}'^2\} \tag{83}$$

which indeed reduces to that of (81) if $q_\alpha = 1$, $\pi_\alpha = (1 - \pi_0)/k$ and $\bar{\tau}'^2 = \tau^2 - \ln \frac{1-\pi_0}{k}$. Hence, to obtain the linearly structured undecided region of (81), its $k+1$ reward functions need to be chosen as exponentially increasing functions of the squared-distances $||P_{c_{t_\alpha}}^{\perp} t||_{Q_{tt}}^2$ to the respective fault lines. □

Example 14 (*Datasnooping with undecided included*) The same assumptions are made as in Example 12, with (80), except that now $\mathcal{P}_0$ is not assumed to be a-priori given. Then, the misclosure space partitioning of $\mathbb{R}^r$ follows from Corollary 2c as

$$\mathcal{P}_{k+1} = \{t \in \mathbb{R}^r |\; \max_{\alpha\in[1,\ldots,k]} w_\alpha^2(t) < h(t) \wedge a < h(t)\}$$

$$\mathcal{P}_0 = \{t \in \mathbb{R}^r \backslash \mathcal{P}_{k+1} | \max_{\alpha\in[1,\ldots,k]} w_\alpha^2(t) < a\}$$
$$\mathcal{P}_{i\in[1,\ldots,k]} = \{t \in \mathbb{R}^r \backslash \{\mathcal{P}_0 \cup \mathcal{P}_{k+1}\} | i = \arg \max_{\alpha\in[1,\ldots,k]} w_\alpha^2(t)\} \quad (84)$$

Note that this partitioning reduces to that of Example 10 in case the undecided region $\mathcal{P}_{k+1}$ would be empty. This happens when the undecided reward is equal to zero, $\mu = 1-u = 0$.

With the above ordering of the partitioning, one would first check whether a parameter solution would be available by checking whether or not $t \in \mathcal{P}_{k+1}$. Only when $t \notin \mathcal{P}_{k+1}$ would one then check on the acceptability of the null-hypothesis $\mathcal{H}_0$. As in the majority of our applications the occurrence-probability of the null-hypothesis, $\mathsf{P}[\mathcal{H}_0] = \pi_0$, will be high, as well as the probability of its correct acceptance, $\mathsf{P}[t \in \mathcal{P}_0|\mathcal{H}_0]$, it is more advantageous to seek an ordering in the partitioning that starts with $\mathcal{P}_0$ rather than with $\mathcal{P}_{k+1}$. This can be achieved by noting that the complements of $\mathcal{P}_{k+1}$ and $\mathcal{P}_0$ are given as:

$$\mathbb{R}^r \backslash \mathcal{P}_{k+1} = \{t \in \mathbb{R}^r | \max_{\alpha\in[1,\ldots,k]} w_\alpha^2(t) > h(t) \vee a > h(t)\}$$
$$\mathbb{R}^r \backslash \mathcal{P}_0 = \{t \in \mathbb{R}^r | \max_{\alpha\in[1,\ldots,k]} w_\alpha^2(t) > a \vee h(t) > a\} \quad (85)$$

Combining this result with that of (84) allows us to write the partitioning in the following order:

$$\mathcal{P}_0 = \{t \in \mathbb{R}^r | \max_{\alpha\in[1,\ldots,k]} w_\alpha^2(t) < a \wedge h(t) < a\}$$
$$\mathcal{P}_{k+1} = \{t \in \mathbb{R}^r \backslash \mathcal{P}_0 | \max_{\alpha\in[1,\ldots,k]} w_\alpha^2(t) < h(t)\}$$
$$\mathcal{P}_{i\in[1,\ldots,k]} = \{t \in \mathbb{R}^r \backslash \{\mathcal{P}_0 \cup \mathcal{P}_{k+1}\} | i = \arg \max_{\alpha\in[1,\ldots,k]} w_\alpha^2(t)\} \quad (86)$$

With this ordering, we can now also compare the partitioning directly with that of Example 12, cf. (79), thus clearly showing how they differ in their definition of the detection region $\mathcal{P}_0$.

The above partitioning is shown in Fig. 9 for the same model and hypotheses as used in Example 9. □

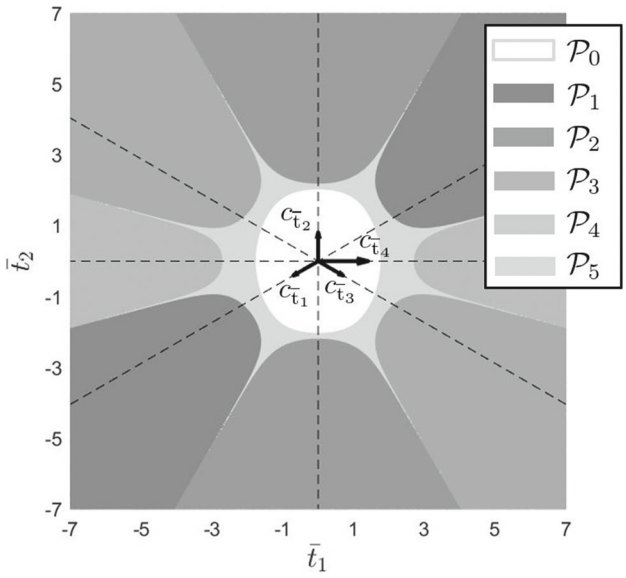

Fig. 9 Misclosure space partitioning in $\mathbb{R}^{r=2}$ for single-outlier data snooping, with an undecided region $\mathcal{P}_5$ included (cf. Example 14)

4 Maximum probability estimators

In the previous section, we have shown how the choice of penalty functions leads to corresponding minimum mean penalty partitionings of misclosure space. One such choice leads to a partitioning maximizing the probability of correct decisions. Although this property is attractive indeed from the perspective of *testing*, it may not be sufficient from the viewpoint of *estimation*. Afterall, having a maximized probability of correct hypothesis identification does not necessarily imply good performance of the DIA-estimator. The first is driven by the misclosure vector $\underline{t}$, while the second is also driven by the $\hat{\underline{x}}_i$'s. Thus instead of focussing on correct decisions, one would be better off focussing on the *consequences* of the decisions made. In this section we will therefore introduce penalty functions that penalize unwanted outcomes of the DIA-estimator. As a result two new estimators with corresponding misclosure space partitionings are identified and derived. They are the optimal DIA-estimator and the optimal WSS-estimator.

4.1 The optimal DIA-estimator

To determine an appropriate penalty function for the DIA-estimator $\bar{\underline{x}}$, we should think of one that penalizes its unwanted outcomes when a decision i is made under hypothesis $\mathcal{H}_\alpha$. As decision i corresponds with an outcome of $\hat{\underline{x}}_i$ and since such outcome is unwanted when it lies in the complement of the tolerance or safety region, $\Omega_x^c = \mathbb{R}^n \backslash \Omega_x$, we would like the probability of such outcomes happening under $\mathcal{H}_\alpha$ to be small. We therefore introduce this probability, for a given misclosure vector t, as our penalty function. Then, if the probability of such an unwanted outcome is large, the penalty will be large as well.

Definition (DIA-penalty function): The penalizing function of the DIA-estimator $\bar{\underline{x}}_{\text{DIA}} = \sum_{i=0}^k \hat{\underline{x}}_i\, p_i(\underline{t})$ is defined as:

$$\bar{\mathrm{r}}_{i\alpha}(t) = \mathsf{P}[\hat{\underline{x}}_i \in \Omega_x^c \mid t, \mathcal{H}_\alpha] \text{ for } i, \alpha \in [0, \ldots, k] \quad (87)$$

We will now show how this penalty function can be used to find the *optimal* DIA-estimator, i.e. the DIA-estimator that within its class has the largest probability of lying inside its safety region, $\mathsf{P}[\underline{\bar{x}}_{\mathrm{DIA}} \in \Omega_x]$, or equivalently, has the smallest integrity risk, $\mathsf{P}[\underline{\bar{x}}_{\mathrm{DIA}} \in \Omega_x^c]$. We have the following result.

Theorem 3 (Optimal DIA-estimator) *Let $\bar{\mathcal{P}}_{i\in[0,\ldots,k]} \subset \mathbb{R}^r$ denote the misclosure space partitioning that of all such partitionings maximizes the DIA-estimator's probability $\mathsf{P}[\underline{\bar{x}}_{\mathrm{DIA}} \in \Omega_x]$. Then,*

$$\bar{\mathcal{P}}_{i\in[0,\ldots,k]} := \arg \max_{\mathcal{P}_{i\in[0,\ldots,k]}} \mathsf{P}[\underline{\bar{x}}_{\mathrm{DIA}} \in \Omega_x] \tag{88}$$

where

$$\bar{\mathcal{P}}_{i\in[0,\ldots,k]} = \{t \in \mathbb{R}^r \mid i = \arg \min_{j\in[0,\ldots,k]} \sum_{\alpha=0}^{k} \bar{\mathtt{r}}_{j\alpha}(t)\, f_{\underline{t}}(t|\mathcal{H}_\alpha)\mathsf{P}[\mathcal{H}_\alpha]\} \tag{89}$$

□

Proof We first proof that the mean penalty (30) becomes identical to the integrity risk if the penalty functions are chosen as (87),

$$\mathsf{E}(\underline{\mathtt{r}}) = \mathsf{P}[\underline{\bar{x}}_{\mathrm{DIA}} \in \Omega_x^c] \;\text{ if }\; \mathtt{r}_{i\alpha}(t) = \bar{\mathtt{r}}_{i\alpha}(t) \tag{90}$$

We have

$$\begin{aligned}
\mathsf{E}(\underline{\mathtt{r}}) &= \sum_{i=0}^{k} \int_{\mathcal{P}_i} \sum_{\alpha=0}^{k} \mathtt{r}_{i\alpha}(t)\, f_{\underline{t}}(t|\mathcal{H}_\alpha)\mathsf{P}[\mathcal{H}_\alpha]dt \\
&\overset{(a)}{=} \sum_{i=0}^{k} \int_{\mathcal{P}_i} \sum_{\alpha=0}^{k} \mathsf{P}[\underline{\hat{x}}_i \in \Omega_x^c \mid t, \mathcal{H}_\alpha]\mathsf{P}[\mathcal{H}_\alpha|t]\, f_{\underline{t}}(t)dt \\
&\overset{(b)}{=} \sum_{i=0}^{k} \int_{\mathcal{P}_i} \mathsf{P}[\underline{\hat{x}}_i \in \Omega_x^c \mid t]\, f_{\underline{t}}(t)dt \\
&\overset{(c)}{=} \sum_{i=0}^{k} \int_{\mathbb{R}^r} \mathsf{P}[\underline{\hat{x}}_i \in \Omega_x^c \mid t]\, p_i(t)\, f_{\underline{t}}(t)dt \\
&\overset{(d)}{=} \sum_{i=0}^{k} \int_{\mathbb{R}^r} \mathsf{P}[\underline{\bar{x}}_{\mathrm{DIA}} \in \Omega_x^c \mid t]\, p_i(t)\, f_{\underline{t}}(t)dt \\
&\overset{(e)}{=} \int_{\mathbb{R}^r} \mathsf{P}[\underline{\bar{x}}_{\mathrm{DIA}} \in \Omega_x^c \mid t]\, f_{\underline{t}}(t)dt \\
&\overset{(f)}{=} \mathsf{P}[\underline{\bar{x}}_{\mathrm{DIA}} \in \Omega_x^c]
\end{aligned}$$

Step (a) follows from substituting $\mathtt{r}_{i\alpha} = \bar{\mathtt{r}}_{i\alpha}$, cf. (87), and recognizing that $f_{\underline{t}}(t|\mathcal{H}_\alpha)\mathsf{P}[\mathcal{H}_\alpha] = \mathsf{P}[\mathcal{H}_\alpha|t]\, f_{\underline{t}}(t)$. Step (b) follows from recognizing that $\mathsf{P}[\underline{\hat{x}}_i \in \Omega_x^c \mid t] = \sum_{\alpha=0}^{k} \mathsf{P}[\underline{\hat{x}}_i \in \Omega_x^c \mid t, \mathcal{H}_\alpha]\mathsf{P}[\mathcal{H}_\alpha|t]$. Step (c) introduces the indicator function $p_i(t)$ of $\mathcal{P}_i$. Step (d) recognizes, since $\underline{\bar{x}}_{\mathrm{DIA}} = \sum_{i=1}^{k} \underline{\hat{x}}_i\, p_i(\underline{t})$, the conditional probability equality $\mathsf{P}[\underline{\hat{x}}_i \in \Omega_x^c \mid t] = \mathsf{P}[\underline{\bar{x}}_{\mathrm{DIA}} \in \Omega_x^c \mid t]$ for $t \in \mathcal{P}_i$. Step (e) follows from using $\sum_{i=1}^{k} p_i(t) = 1$ and step (f) from the continuous version of the total probability rule.

Having established (90), the result (88) follows from applying Theorem 2a, cf. (31). □

In analogy with (33), also the partitioning (89) can be given an insightful probabilistic interpretation. As $\mathsf{P}[\underline{\hat{x}}_i \in \Omega_x^c|t]\, f_{\underline{t}}(t) = \sum_{\alpha=0}^{k} \mathsf{P}[\underline{\hat{x}}_i \in \Omega_x^c|t, \mathcal{H}_\alpha]\, f_{\underline{t}}(t|\mathcal{H}_\alpha)\mathsf{P}[\mathcal{H}_\alpha]$, we have

$$\bar{\mathcal{P}}_{i\in[0,\ldots,k]} = \{t \in \mathbb{R}^r \mid i = \arg \min_{j\in[0,\ldots,k]} \mathsf{P}[\underline{\hat{x}}_j \in \Omega_x^c|t]\} \tag{91}$$

thus showing that each of the defining regions $\bar{\mathcal{P}}_i$ of the optimal DIA-estimator is characterized by having the smallest misclosure-conditioned integrity risk.

If we also include a no-identification or undecided region for which the DIA-estimator is said to be unavailable, cf. (22), then we have, for the case of a constant unavailability penalty $\bar{\mathtt{r}}_{(k+1)\alpha}(t) = u$, in analogy with Corollary 1, cf. (34), the following optimal partitioning.

Corollary 3 (Unavailability included) *Let the DIA-penalty function (87) be extended with the unavailability penalty $\bar{\mathtt{r}}_{(k+1)\alpha}(t) = u$. Then the minimum mean penalty partitioning follows from (31) as*

$$\begin{aligned}
\bar{\mathcal{P}}_{k+1} &= \{t \in \mathbb{R}^r \mid u < \min_{j\in[0,\ldots,k]} \mathsf{P}[\underline{\hat{x}}_j \in \Omega_x^c|t]\} \\
\bar{\mathcal{P}}_{i\in[0,..,k]} &= \{t \in \mathbb{R}^r \backslash \bar{\mathcal{P}}_{k+1} \mid i = \arg \min_{j\in[0,..,k]} \mathsf{P}[\underline{\hat{x}}_j \in \Omega_x^c|t]\}
\end{aligned} \tag{92}$$

□

Hence, a no-identification or unavailability decision is made when the smallest misclosure-conditioned integrity risk is still considered too large.

4.2 DIA-penalty function and the choice for Ω_x

The DIA-penalty function $\bar{\mathtt{r}}_{i\alpha}(t)$ is defined in (87) as a conditional probability of $\underline{\hat{x}}_i \in \Omega_x^c$ under the alternative hypothesis $\mathcal{H}_\alpha$. The following Lemma shows how this probability can be computed directly from the distribution of $\underline{\hat{x}}_0$ under $\mathcal{H}_0$.

Lemma 3 (DIA-penalty function)*: The required probability for the DIA-penalty function (87) can be computed under $\mathcal{H}_0$ for a general x-centred region $\Omega_x \subset \mathbb{R}^n$ as*

$$\begin{aligned}
\bar{\mathtt{r}}_{i\alpha}(t) &= \mathsf{P}[\underline{\hat{x}}_0 \in \Omega_{x+\Delta x_{i\alpha}(t)}^c|\mathcal{H}_0] \\
\Delta x_{i\alpha}(t) &= A^{+}[C_i \hat{b}_i(t) - C_\alpha b_\alpha]
\end{aligned} \tag{93}$$

and specifically for the ellipsoidal region

$$\Omega_x = \{v \in \mathbb{R}^n \mid ||v - x||^2_{Q_{\hat{x}_0\hat{x}_0}} < \tau^2\} \tag{94}$$

as

$$\begin{aligned} \bar{\mathrm{r}}_{i\alpha}(t) &= 1 - \mathsf{P}[\underline{\chi}^2(n, \lambda_{i\alpha}(t) < \tau^2] \\ \lambda_{i\alpha}(t) &= ||A^+[C_i\hat{b}_i(t) - C_\alpha b_\alpha]||^2_{Q_{\hat{x}_0\hat{x}_0}} \end{aligned} \tag{95}$$

□

Proof We first prove (93). We have

$$\begin{aligned} &\bar{\mathrm{r}}_{i\alpha}(t) = \mathsf{P}[\underline{\hat{x}}_i \in \Omega_x^c \mid t, \mathcal{H}_\alpha] = \\ &\overset{(a)}{=} \mathsf{P}[\underline{\hat{x}}_0 - A^+C_i\hat{b}_i(t) \in \Omega_x^c \mid t, \mathcal{H}_\alpha] \\ &\overset{(b)}{=} \mathsf{P}[\underline{\hat{x}}_0 - A^+C_i\hat{b}_i(t) \in \Omega_x^c \mid \mathcal{H}_\alpha] \\ &\overset{(c)}{=} \mathsf{P}[\underline{\hat{x}}_0 + A^+[C_\alpha b_\alpha - C_i\hat{b}_i(t)] \in \Omega_x^c \mid \mathcal{H}_0] \end{aligned} \tag{96}$$

from which (93) follows. Step (a) follows from substituting $\underline{\hat{x}}_i = \underline{\hat{x}}_0 - A^+C_i\underline{\hat{b}}_i(t)$ and recognizing that the conditioning on t makes $\hat{b}_i(t)$ nonrandom. Step (b) follows by recognizing that $\underline{t}$ is independent of $\underline{\hat{x}}_0$, cf. (4), and for step (c) we made use of the relation $\underline{\hat{x}}_0|\mathcal{H}_\alpha = \underline{\hat{x}}_0|\mathcal{H}_0 + A^+C_\alpha b_\alpha$.

For the special ellipsoidal case, we have $\bar{\mathrm{r}}_{i\alpha}(t) \overset{(93)}{=} 1 - \mathsf{P}[\underline{\hat{x}}_0 \in \Omega_{x+\Delta x_{i\alpha}(t)}|\mathcal{H}_0] = 1 - \mathsf{P}[||\underline{\hat{x}}_0 - x - \Delta x_{i\alpha}||^2_{Q_{\hat{x}_0\hat{x}_0}} < \tau^2|\mathcal{H}_0]$, from which, with $\underline{\hat{x}}_0 \overset{\mathcal{H}_0}{\sim} \mathcal{N}_n(x, Q_{\hat{x}_0\hat{x}_0})$, (95) follows. □

Note that the evaluation of the DIA-penalty function requires knowledge of the biases. Although there are important applications for which such biases are known (e.g. when the biases form a set of a-priori known corrections), the case for which they are unknown is treated further in Sect. 5.

One may wonder whether the above penalty computation would become simpler if instead of using (87), the mean of $\bar{\mathrm{r}}_{i\alpha}(\underline{t})$ would be used, thereby eliminating its dependence on t. However, since $\mathsf{E}(\bar{\mathrm{r}}_{i\alpha}(\underline{t})|\mathcal{H}_\alpha) = \mathsf{P}[\underline{\hat{x}}_i \in \Omega_x^c|\mathcal{H}_\alpha]$ and $\underline{\hat{x}}_i$ has a variance matrix different from $Q_{\hat{x}_0\hat{x}_0}$, already the evaluation with the ellipsoid (94) would fail to reduce to straightforward Chi-square distributions, but instead would require the usage of more complicated distributions of general quadratic forms in normal random variables (Mathai and Provost 1992).

Lemma 3 shows how the penalty function $\bar{\mathrm{r}}_{i\alpha}(t)$ can be computed for any arbitrary safety region Ω_x, as well as for the special case when this region is ellipsoidal and defined through the variance matrix of $\underline{\hat{x}}_0$, cf. (94). With this latter choice the computations simplify to evaluations of noncentral Chi-square distributions. Although the choice of Ω_x is user-driven and may vary from application to application, the ellipsoidal choice (94) is relevant for applications in which one wants to judge the DIA-performance relative to the precision of $\underline{\bar{x}}_{\mathrm{DIA}}|\mathcal{H}_0 = \underline{\hat{x}}_0$, for instance, when the working hypothesis $\mathcal{H}_0$ has been specifically designed to meet certain precision requirements on $\underline{\hat{x}}_0$.

Note that the penalty function (95) is an increasing function in $\lambda_{i\alpha}(t) = ||\Delta x_{i\alpha}(t)||^2_{Q_{\hat{x}_0\hat{x}_0}}$, i.e. the penalties get larger as the noncentrality parameter gets larger. With reference to Anderson's theorem (Anderson 1955), this property remains true in general for the penalty function of (93), provided the Ω_x's are chosen as convex sets symmetric about x. Depending on the decision made and on which hypothesis is valid, the noncentrality parameter can be further discriminated as:

$$\lambda_{i\alpha}(t) = \begin{cases} ||A^+C_\alpha b_\alpha||^2_{Q_{\hat{x}_0\hat{x}_0}}, & i = 0, \alpha \neq 0 \\ ||A^+C_i\hat{b}_i(t)||^2_{Q_{\hat{x}_0\hat{x}_0}}, & i \neq 0, \alpha = 0 \\ ||A^+[C_\alpha b_\alpha - C_i\hat{b}_i(t)]||^2_{Q_{\hat{x}_0\hat{x}_0}}, & i \neq 0, \alpha \neq 0 \end{cases} \tag{97}$$

This shows how the noncentrality parameter, and thus its corresponding penalty, is driven by the actual and estimated biases. When the null-hypothesis $\mathcal{H}_{i=0}$ is accepted and thus $\underline{\hat{x}}_{i=0}$ is selected, it are the *actual* bias vectors b_α of the hypotheses that drive the penalty $\bar{\mathrm{r}}_{0\alpha}(t)$. However, when the null-hypothesis $\mathcal{H}_{\alpha=0}$ is true, it are the *estimated* bias vectors $\hat{b}_i(t)$ that drive the penalty $\bar{\mathrm{r}}_{i0}$ when $\underline{\hat{x}}_i$ is selected. And in case $\mathcal{H}_0$ is neither selected nor true ($i \neq 0, \alpha \neq 0$), the difference of the actual and estimated biases drives the penalty.

Note that the probability of the Chi-square distribution in (95) is a monotonous decreasing function in the noncentrality parameter $\lambda_{i\alpha}(t)$. Hence, in order to avoid the required calculation of the probability, one may also decide to choose the penalty function equal to the noncentrality parameter itself, $\mathrm{r}_{i\alpha}(t) = \lambda_{i\alpha}(t)$. Although this will of course negate the optimality property of Theorem 3, it will still provide a minimum mean penalty misclosure space partitioning that is based on penalizing incorrect parameter solutions.

4.3 The optimal WSS-estimator

As with the optimality in the DIA-class, one may wonder which estimator would be optimal in the WSS-class. A natural estimator in this class would be one where one would choose the weight $\omega_i(t)$ as the conditional probability $\mathsf{P}[\underline{\mathcal{H}} = \mathcal{H}_i|t]$ (cf. 26), i.e. the probability of $\mathcal{H}_i$-occurrence given the outcome of the misclosure vector being t. Although perhaps a natural choice, this choice of weighting is not one that is directly derived from the impact the weighting has on the probabilistic properties of the WSS-estimator $\underline{\bar{x}}_{\mathrm{WSS}}$. In order to achieve that, we will therefore again, just as in Theorem 3, aim for an estimator that maximizes the probability $\mathsf{P}[\underline{\bar{x}}_{\mathrm{WSS}} \in \Omega_x]$. This time, however, the maximization is not done with respect to indicator functions, as was the case with the optimal DIA-estimator (cf. 88), but instead with respect to weighting functions satisfying $\omega_i(t) \geq 0$, $i = 1, \ldots, k$ and $\sum_{i=0}^k \omega_i(t) = 1$. To simplify and to make a direct comparison with (95) of Lemma 3 possible, we will assume Ω_x given

as in (94). Assuming estimable weights, the resulting maximum probability estimator within the WSS-class is given as follows.

Theorem 4 (Optimal WSS estimator) *Let the weight vector* $\bar{\omega}(t) = (\bar{\omega}_0(t), \ldots, \bar{\omega}_k(t))^T \in \mathbb{R}^{k+1}$ *be the vector of misclosure weight functions, that of all such vector functions maximizes the WSS-estimator's probability* $\mathsf{P}[\bar{\underline{x}}_{\mathrm{WSS}} \in \Omega_x]$ *for* $\Omega_x = \{u \in \mathbb{R}^n | \; ||u - x||^2_{Q_{\hat{x}_0\hat{x}_0}} \leq \tau^2\}$. *Then*

$$\bar{\omega}(t) := \arg \max_{e_{k+1}^T \omega = 1, \omega \in \mathbb{R}^{k+1}_{\geq 0}} \mathsf{P}[\bar{\underline{x}}_{\mathrm{WSS}} \in \Omega_x] \tag{98}$$

where

$$\bar{\omega}(t) = \arg \max_{e_{k+1}^T \omega = 1, \omega \in \mathbb{R}^{k+1}_{\geq 0}} \sum_{\alpha=0}^{k} \Pi(\lambda_\alpha(t, \omega)) F_\alpha(t) \tag{99}$$

with $\Pi(\lambda) = \mathsf{P}[\chi^2(n, \lambda) \leq \tau^2]$, $F_\alpha(t) = f_{\underline{t}}(t|\mathcal{H}_\alpha)\mathsf{P}[\mathcal{H}_\alpha]$, *and*

$$\lambda_\alpha(t, \omega) = ||A^+[C_\alpha b_\alpha - \sum_{i=1}^{k} C_i \hat{b}_i(t)\omega_i]||^2_{Q_{\hat{x}_0\hat{x}_0}} \tag{100}$$

□

Proof From $\mathsf{P}[\bar{\underline{x}}_{\mathrm{WSS}} \in \Omega_x | t, \mathcal{H}_\alpha] = \mathsf{P}[\chi^2(n, \lambda_\alpha(t, \omega(t))) \leq \tau^2]$ and $\mathsf{P}[\bar{\underline{x}}_{\mathrm{WSS}} \in \Omega_x] = \int_{\mathbb{R}^r} \sum_{\alpha=0}^{k} \mathsf{P}[\bar{\underline{x}}_{\mathrm{WSS}} \in \Omega_x | t, \mathcal{H}_\alpha] F_\alpha(t) dt$, it follows that

$$\mathsf{P}[\bar{\underline{x}}_{\mathrm{WSS}} \in \Omega_x] = \int_{\mathbb{R}^r} \sum_{\alpha=0}^{k} \Pi(\lambda_\alpha(t, \omega(t))) F_\alpha(t) dt \tag{101}$$

As the $k+1$ functions $\Pi(\lambda_\alpha(t, \omega(t)))F_\alpha(t)$ are non-negative for every $t \in \mathbb{R}^r$, the maximum of (101) is obtained if for every $t \in \mathbb{R}^r$, a feasible $\omega \in \mathbb{R}^{k+1}$ is chosen such that the sum $\sum_{\alpha=0}^{k} \Pi(\lambda_\alpha(t, \omega)) F_\alpha(t)$ is maximized. This proves that the sought for maximizing vectorial weight function is given by (99). □

Note that the above estimator can indeed be seen to be a generalization of the optimal DIA-estimator. If the weights ω_i are restricted to be only 1 or 0, the noncentrality parameter $\lambda_\alpha(t, \omega)$ (cf. 100) becomes equal to $\lambda_{j\alpha}(t) = ||A^+[C_\alpha b_\alpha - C_j \hat{b}_j(t)]||^2_{Q_{\hat{x}_0\hat{x}_0}}$ for some $j \in [0, \ldots, k]$, and (99) can be rewritten as (89) with (95), thus leading to a recovering of the optimal DIA-estimator. A complication of this generalization is, however, that the optimal WSS-estimator is far more difficult to compute than the optimal DIA-estimator. The objective function of (99), which needs to be maximized over the feasible set of ω, is, as a linear combination of log-concave functions $\Pi(\lambda_\alpha)$ in the inhomogeneous quadratic forms (100), a *multimodal* function. This implies that gradient ascent algorithms will only converge to the required maximum if the chosen initial value of the weight vector ω is already close enough to the sought for maximizer. If that is the case, one can show that a relatively simple fixed-point algorithm can be devised that converges to the maximum. However, to achieve convergence independent of the feasible initialization's quality, methods of global maximization need to be employed. As the algorithmic details of how this can be achieved are not the focus of the current contribution, we will address the required methodology for the numerical solution of the above maximization problem in a separate forthcoming contribution.

5 Operational DIA-estimators

As mentioned earlier, the computation of the DIA-penalty functions, (93) and (95), requires knowledge of the noncentrality parameters and therefore of the biases b_α. Under $\mathcal{H}_0$ these are known, but this is generally not the case under the alternative hypotheses. Does this mean that if this information is lacking usage of these penalty functions becomes obsolete? No, certainly not. First we have to remember that whatever operational choice is made for the penalty functions, their corresponding mean penalty, or its sharp upperbound, will be minimized through the optimal misclosure partitionings of (31) and (49), respectively. Thus to any choice of penalty functions belongs an optimal testing procedure. Second, somewhat in analogy with the conservative approach of Lemma 2, one may take a 'minmax' approach by taking the bias-values that minimize the probability $\mathsf{P}[\bar{x}_{\mathrm{DIA}} \in \Omega_x]$ of the optimal DIA-estimator. Third, the results of Lemma 3 and the structure of the DIA-penalty functions, cf. (93) and (95), also provide a *guide* for helping to formulate operational penalty functions. For instance, even if the evaluation of the DIA-estimator would be done with a nonellipsoidal shaped safety region, the choice of the testing-defining penalty functions could, instead of (93), fall on the cheaper-to-compute functions (95). And also with respect to the handling of the bias dependency, the structure of the above DIA-penalty functions provides a guide. In the next subsections, different such proposals will be made.

5.1 Estimated DIA-penalty functions

If the bias vectors b_α, $\alpha = 1, \ldots, k$, are unknown, one may decide to estimate them so as to obtain an approximation to the optimal DIA-estimator. When we replace the unknown biases in (93) and (95) by their estimates $\hat{b}_\alpha(t)$, we obtain

the *estimated* DIA-penalty functions as

$$\begin{aligned}\hat{\hat{\mathrm{r}}}_{i\alpha}(t) &= \mathsf{P}[\underline{\hat{x}}_0 \in \Omega^c_{x+\Delta\hat{x}_{i\alpha}(t)}|\mathcal{H}_0]\\ \Delta\hat{x}_{i\alpha}(t) &= A^+[C_i\hat{b}_i(t) - C_\alpha\hat{b}_\alpha(t)]\end{aligned} \tag{102}$$

Note that $\Delta\hat{x}_{i\alpha}(t)$ is now anti-symmetric in its indices, $\Delta\hat{x}_{i\alpha}(t) = -\Delta\hat{x}_{\alpha i}(t)$. This implies, when Ω_x is convex symmetric about x, i.e. $x \in \Omega_0 \Leftrightarrow -x \in \Omega_0$, that the estimated penalty function is *symmetric* in its indices, $\hat{\hat{\mathrm{r}}}_{i\alpha}(t) = \hat{\hat{\mathrm{r}}}_{\alpha i}(t)$. This can be seen as follows:

$$\begin{aligned}\hat{\hat{\mathrm{r}}}_{i\alpha}(t) &= \mathsf{P}[(\underline{\hat{x}}_0 - x) - \Delta\hat{x}_{i\alpha}(t) \in \Omega_0]\\ &\overset{(a)}{=} \mathsf{P}[(x - \underline{\hat{x}}_0) + \Delta\hat{x}_{i\alpha}(t) \in \Omega_0]\\ &\overset{(b)}{=} \mathsf{P}[(x - \underline{\hat{x}}_0) - \Delta\hat{x}_{\alpha i}(t) \in \Omega_0]\\ &\overset{(c)}{=} \mathsf{P}[(\underline{\hat{x}}_0 - x) - \Delta\hat{x}_{\alpha i}(t) \in \Omega_0]\\ &= \hat{\hat{\mathrm{r}}}_{\alpha i}(t)\end{aligned} \tag{103}$$

where (a) is due to the symmetry with respect to origin of Ω_0, (b) due to the anti-symmetry of $\Delta\hat{x}_{i\alpha} = -\Delta\hat{x}_{\alpha i}$, and (c) due to $\underline{\hat{x}}_0 - x \overset{\mathcal{H}_0}{\sim} x - \underline{\hat{x}}_0$.

Note that $\Delta\hat{x}_{i\alpha}(t)$ of (102) can also be written as the difference of the parameter solutions under $\mathcal{H}_\alpha$ and $\mathcal{H}_i$: $\Delta\hat{x}_{i\alpha}(t) = \hat{x}_\alpha(t) - \hat{x}_i(t)$. Thus, in this case it will be the *solution separations* of the hypothesized models that drive the penalty functions. The corresponding estimated noncentrality parameters, cf. (97), read then

$$\hat{\lambda}_{i\alpha}(t) = \begin{cases} ||\hat{x}_0(t) - \hat{x}_\alpha(t)||^2_{Q_{\hat{x}_0\hat{x}_0}}, i = 0, \alpha \neq 0\\ ||\hat{x}_i(t) - \hat{x}_0(t)||^2_{Q_{\hat{x}_0\hat{x}_0}}, i \neq 0, \alpha = 0\\ ||\hat{x}_i(t) - \hat{x}_\alpha(t)||^2_{Q_{\hat{x}_0\hat{x}_0}}, i \neq 0, \alpha \neq 0\end{cases} \tag{104}$$

This shows, for instance, the closer $\hat{x}_\alpha(t)$ is to $\hat{x}_0(t)$, the smaller the penalty $\hat{\hat{\mathrm{r}}}_{0\alpha}(t)$ is.

5.2 Influential bias driven DIA-penalty functions

Any observational bias $C_\alpha b_\alpha$ can be decomposed into its *influential* and *testable* component (Teunissen 2018),

$$C_\alpha b_\alpha = \underbrace{P_A C_\alpha b_\alpha}_{\text{influential}} + \underbrace{P_A^\perp C_\alpha b_\alpha}_{\text{testable}} \tag{105}$$

where $P_A = AA^+$ and $P_A^\perp = I_m - AA^+$. The component $P_A^\perp C_\alpha b_\alpha$ is referred to as testable as precisely this component propagates into the mean of the misclosure vector under $\mathcal{H}_\alpha$: $\mathsf{E}(\underline{t}|\mathcal{H}_\alpha) = B^T C_\alpha b_\alpha = B^T(P_A^\perp C_\alpha b_\alpha)$, since $B^T A = 0$. The influential component $P_A C_\alpha b_\alpha = A(A^+ C_\alpha b_\alpha)$, on the other hand, is non-testable as it lies in the range space of A and propagates directly into the parameter solution $\underline{\hat{x}}_0$. This implies that for the performance of the DIA-estimator it is of importance that one can keep the influences of the non-testable biases $P_A C_\alpha b_\alpha$ at bay. It is therefore fitting

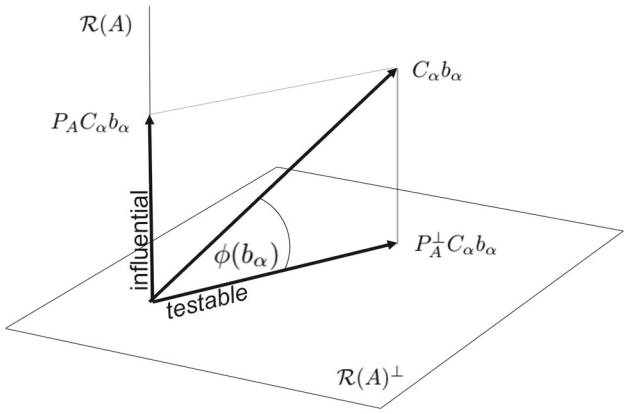

Fig. 10 Ratio of the influential and testable bias: $\tan\phi(b_\alpha) = ||P_A C_\alpha b_\alpha||_{Q_{yy}} / ||P_A^\perp C_\alpha b_\alpha||_{Q_{yy}}$

to recognize that the assigned penalties of the DIA-penalty functions are indeed driven by the influential biases, cf. (97).

Although the required penalties for the DIA-estimator to become optimal cannot be computed if the biases are unknown, it is possible to aim for a protection against influential biases that may slip testing unnoticed. To determine the sizes of such influential biases, we can of course not use the minimum mean penalty testing that we are aiming to design. But what we can do is to determine the influential biases as if a global overall model test would be executed. Since

$$||\underline{t}||^2_{Q_{tt}} \overset{\mathcal{H}_\alpha}{\sim} \chi^2(r, ||P_A^\perp C_\alpha b_\alpha||^2_{Q_{yy}}) \tag{106}$$

it follows that $\mathsf{P}[||\underline{t}||^2_{Q_{tt}} < c|\mathcal{H}_\alpha] = \text{constant}$ for all $\alpha \in [1, \ldots, k]$ if $||P_A^\perp C_\alpha b_\alpha||^2_{Q_{yy}} = \text{constant} \overset{\text{say}}{=} \lambda_0$ for all $\alpha \in [1, \ldots, k]$. Hence, by using the same yardstick λ_0 for all alternative hypotheses, we can now compute for each individual alternative hypothesis the bias vector that would have the largest influence,

$$\max_{b_\alpha \in \mathbb{R}^{q_\alpha}} ||P_A C_\alpha b_\alpha||^2_{Q_{yy}} \; s.t. \; ||P_A^\perp C_\alpha b_\alpha||^2_{Q_{yy}} = \lambda_0 \tag{107}$$

As this is the maximization of a quadratically constrained quadratic form, its solution is provided by solving a generalized eigenvalue problem:

$$\max_{||P_A^\perp C_\alpha b_\alpha||^2_{Q_{yy}} = \lambda_0} ||P_A C_\alpha b_\alpha||^2_{Q_{yy}} = \lambda_0 \lambda_{\alpha,\max} \tag{108}$$

with

$$\lambda_{\alpha,\max} = \frac{||P_A C_\alpha b_{\alpha,\max}||^2_{Q_{yy}}}{||P_A^\perp C_\alpha b_{\alpha,\max}||^2_{Q_{yy}}} \tag{109}$$

in which $b_{\alpha,\max}$ is the eigenvector that corresponds with the largest eigenvalue of the generalized eigenvalue problem $Mb = \lambda Nb$, with $M = C_\alpha^T Q_{yy}^{-1} P_A C_\alpha$ and $N =$

Table 1 Penalty matrices: maximized probability of correct decisions (left); influential bias dominated (middle); influential bias dominated with undecided penalties u_α included (right)

$$
\underset{(k+1)\times(k+1)}{[\mathtt{r}_{i\alpha}]} = \left[\begin{array}{c|cccc} r_0 & 1 & 1 & \dots & 1 & 1 \\ \hline 1 & r_1 & 1 & \dots & 1 & 1 \\ 1 & 1 & r_2 & \dots & 1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & 1 & \dots & r_{k-1} & 1 \\ 1 & 1 & 1 & \dots & 1 & r_k \end{array}\right], \quad \underset{(k+1)\times(k+1)}{[\mathtt{r}_{i\alpha}]} = \left[\begin{array}{c|cccc} 0 & \mathtt{r}_{01} & \mathtt{r}_{02} & \dots & \mathtt{r}_{0(k-1)} & \mathtt{r}_{0k} \\ \hline 0 & 0 & r & \dots & r & r \\ 0 & r & 0 & \dots & r & r \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & r & r & \dots & 0 & r \\ 0 & r & r & \dots & r & 0 \end{array}\right], \quad \underset{(k+2)\times(k+1)}{[\mathtt{r}_{i\alpha}]} = \left[\begin{array}{c|cccc} 0 & \mathtt{r}_{01} & \mathtt{r}_{02} & \dots & \mathtt{r}_{0(k-1)} & \mathtt{r}_{0k} \\ \hline 0 & 0 & r & \dots & r & r \\ 0 & r & 0 & \dots & r & r \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & r & r & \dots & 0 & r \\ 0 & r & r & \dots & r & 0 \\ \hline u_0 & u_1 & u_1 & \dots & u_{k-1} & u_k \end{array}\right]
$$

$C_\alpha^T Q_{yy}^{-1} P_A^\perp C_\alpha$. Figure 10 provides a geometric interpretation of (109). Note, since $Q^{-1}_{\hat{b}_\alpha \hat{b}_\alpha} = C_\alpha^T Q_{yy}^{-1} P_A^\perp C_\alpha$ and $Q^{-1}_{\hat{b}_\alpha(x)\hat{b}_\alpha(x)} = C_\alpha^T Q_{yy}^{-1} C_\alpha$, that (109) can also be expressed as

$$
\lambda_{\alpha,\max} = \left[\frac{||b_{\alpha,\max}||^2_{Q_{\hat{b}_\alpha(x)\hat{b}_\alpha(x)}}}{||b_{\alpha,\max}||^2_{Q_{\hat{b}_\alpha\hat{b}_\alpha}}} - 1\right] \tag{110}
$$

thus showing how the precision of $\underline{\hat{b}}_\alpha$, and the precision of its x-constrained version, $\underline{\hat{b}}_\alpha(x)$, contribute to the influential bias.

Once the generalized eigenvectors $b_{\alpha,\max}$, $\alpha = 1, \dots, k$, are known, they can be used to construct the influential bias protecting penalty functions by replacing the b_α's in (93) and (95).

5.3 Simplified DIA-penalty functions

Instead of using the full structure of the penalty functions, one can also decide to use a simplified structure, thereby simplifying the evaluation of the corresponding misclosure space partitionings. One such simplification we already met in Sect. 3.3, when discussing the maximization of the probability of correct decisions. Its corresponding penalty structure is shown in the form of a penalty matrix in Table 1 (left). This structure, however, is not really suited for our DIA-estimator as it penalizes and rewards decisions irrespective of their consequences. The structure that we propose as simplification is given as (see Table 1, middle):

$$
\left\{\begin{array}{l} (a)\ \text{under } \mathcal{H}_0,\ \text{no penalties} \\ (b)\ \text{for correct decisions, no penalties} \\ (c)\ \text{for } i = 0,\ \text{influential bias related penalties} \\ (d)\ r(t)\ \text{penalty for remaining decisions} \end{array}\right. \tag{111}
$$

This choice is motivated as follows. Giving zero penalties to correct decisions is clear. But we also give zero penalties to incorrect decisions under the null hypothesis. The rationale for this choice is that in those cases estimation takes place under larger models than that of $\mathcal{H}_0$, $\mathcal{R}([A, C_{i\neq 0}]) \supset \mathcal{R}(A)$. As a consequence, the DIA-output will be conditionally distributed as $\underline{\hat{x}}_i \overset{\mathcal{H}_0}{\sim} \mathcal{N}_n(x, Q_{\hat{x}_i,\hat{x}_i})$. Hence, the solution will then still be unbiased, albeit with a poorer precision, $Q_{\hat{x}_i\hat{x}_i} > Q_{\hat{x}_0\hat{x}_0}$. In contrast to the zero penalties for incorrect decisions under $\mathcal{H}_0$, we find the incorrect acceptance of the null-hypothesis a more severe mistake and therefore fully penalize it with influential-bias related penalties, indicated by $\mathtt{r}_{01}(t)$ through $\mathtt{r}_{0k}(t)$ (cf. Sects. 5.1 and 5.2). Finally the remaining incorrect decisions are all given the same penalty function.

The following theorem shows how the above proposal works out for the optimal misclosure space partitionings.

Theorem 5a (A proposed penalty structure) *Let the penalty function $\mathtt{r}_{i\alpha}(t)$ be structured as*

$$
\left\{\begin{array}{lll} \mathtt{r}_{i0}(t) = & 0 & \text{for } i \in [0, \dots, k] \\ \mathtt{r}_{0\alpha}(t) = & \mathtt{r}_{0\alpha}(t) & \text{for } \alpha \in [1, \dots, k] \\ \mathtt{r}_{i\alpha}(t) = & (1 - \delta_{i\alpha})r(t) & \text{for } i, \alpha \in [1, \dots, k] \end{array}\right. \tag{112}
$$

Then (31) and (49) simplify, respectively, to

$$
\left\{\begin{array}{ll} \mathcal{P}_0 & = \{t \in \mathbb{R}^r \mid \max\limits_{\alpha\in[1,\dots,k]} F_\alpha(t) < a_0(t)\} \\ \mathcal{P}_{i\in[1,\dots,k]} & = \{t \in \mathbb{R}^r \backslash \mathcal{P}_0 \mid i = \arg \max\limits_{\alpha\in[1,\dots,k]} F_\alpha(t)\} \end{array}\right. \tag{113}
$$

and

$$
\left\{\begin{array}{ll} \mathcal{P}_0 & = \{t \in \mathbb{R}^r \mid \max\limits_{\alpha\in[1,\dots,k]} T_\alpha(t) < b_0(t)\} \\ \mathcal{P}_{i\in[1,\dots,k]} & = \{t \in \mathbb{R}^r \backslash P_0 \mid i = \arg \max\limits_{\alpha\in[1,\dots,k]} T_\alpha(t)\} \end{array}\right. \tag{114}
$$

with

$$
\begin{aligned} a_0(t) &= \sum_{\alpha=1}^{k} \left(1 - \tfrac{\mathtt{r}_{0\alpha}(t)}{\mathtt{r}(t)}\right) F_\alpha(t) \\ b_0(t) &= \ln[\sum_{\alpha=1}^{k} \left(1 - \tfrac{\mathtt{r}_{0\alpha}(t)}{\mathtt{r}(t)}\right) \exp\{+\tfrac{1}{2} T_\alpha(t)\}]^2 \end{aligned} \tag{115}
$$

■

Proof First we determine $\mathcal{P}_0$. Application of (31), using the penalty structure (112), gives for $\mathcal{P}_0$ the k inequalities

$$F_\alpha(t) < \sum_{\beta=1}^{k}(1 - \tfrac{\mathtt{r}_{0\beta}(t)}{r(t)})F_\beta(t),\ \alpha \in [1,\ldots,k] \tag{116}$$

from which the first equation of (113) follows. Similarly, application of (31), using the penalty structure (112), gives for $\mathcal{P}_{i\in[1,\ldots,k]}$ the k inequalities

$$\begin{aligned} F_i(t) &> F_\alpha(t),\ \alpha \in [1,\ldots,k]\backslash\{i\} \\ F_i(t) &> \sum_{\beta=1}^{k}(1 - \tfrac{\mathtt{r}_{0\beta}(t)}{r(t)})F_\beta(t) \end{aligned} \tag{117}$$

As the last inequality of this inequality set is automatically satisfied for $t \in \mathbb{R}^r\backslash\mathcal{P}_0$ (cf. 28), we obtain (113). Finally, (114) is obtained by replacing $F_\alpha(t)$ in (113) by (52). □

Compare the above results with that of Corollary 2b and note that the main difference lies in the formation of the detection region $\mathcal{P}_0$. In (62) the upperbound on the maximum is a constant, whereas in the above Theorem the upperbounds vary in dependence of the misclosure vector t. In case of (113) and (114) the shape of the detection region $\mathcal{P}_0$ is driven by the influential-bias related penalties $\mathtt{r}_{0\alpha}(t)$. Would we assign for the decision to accept the null-hypothesis the maximum penalty under all alternative hypotheses, i.e. $\frac{\mathtt{r}_{0\alpha}}{r} = 1$ for all $\alpha \in [1,\ldots,k]$, then $a_0(t)$ (cf. 115) is identically zero and the detection region would be empty, $\mathcal{P}_0 = \emptyset$, i.e. one would then never accept $\mathcal{H}_0$. If on the other hand, $\frac{\mathtt{r}_{0\alpha}}{r} = 0$ for all $\alpha \in [1,\ldots,k]$, then the inequalities of (113) and (114) are trivially fulfilled and one would always accept the null-hypothesis, i.e. $\mathcal{P}_0 = \mathbb{R}^r$. For the intermediate cases, we learn from the expressions of (113) and (114) that the shape of the detection region would open up in the direction of $\mathcal{P}_\beta$ if its influential-bias based penalty $\mathtt{r}_{0\beta}$ reduces in size. This is also what one wants to achieve: the less harm there is in accepting $\mathcal{H}_0$ under $\mathcal{H}_\beta$, the larger the acceptance region of $\mathcal{H}_0$ can be for misclosure vectors originating from $\mathcal{H}_\beta$. This behaviour is illustrated in the following example.

Example 15 (*Simplified DIA-penalties*) Consider the null- and alternative hypotheses $\underline{y} \overset{\mathcal{H}_0}{\sim} \mathcal{N}_m(Ax, Q_{yy})$ and $\underline{y} \overset{\mathcal{H}_\alpha}{\sim} \mathcal{N}_m(Ax + c_\alpha b_\alpha, Q_{yy})$, with $\pi_\alpha = \frac{1-\pi_0}{k}$, $\alpha = 1,\ldots,k$, and assume the penalty functions (112). Application of (114) gives then

$$\begin{aligned} \mathcal{P}_0 &= \{t \in \mathbb{R}^r | \max_{\alpha\in[1,\ldots,k]} w_\alpha^2(t) < b_0'(t)\} \\ \mathcal{P}_{i\in[1,\ldots,k]} &= \{t \in \mathbb{R}^r\backslash\mathcal{P}_0 |\ i = \arg\max_{\alpha\in[1,\ldots,k]} w_\alpha^2(t)|\} \end{aligned} \tag{118}$$

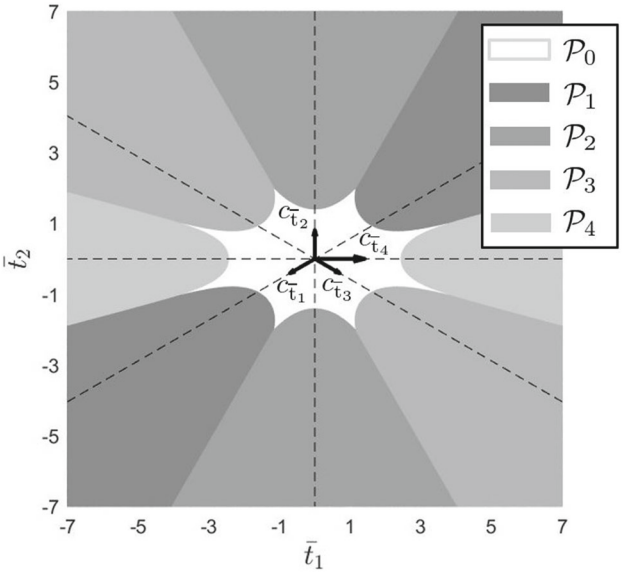

Fig. 11 Misclosure space partitioning (cf. Example 15)

with $b_0'(t) = \ln[\sum_{\alpha=1}^{k}(1 - \frac{\mathtt{r}_{0\alpha}(t)}{r(t)})\exp\{+\frac{1}{2}w_\alpha^2(t)\}]^2$. This partitioning is shown in Fig. 11 for the same model and hypotheses as used in Example 9. The functions $\mathtt{r}_{0\alpha}(t)$ and $r(t)$ were chosen to be constant. As $\mathtt{r}_{04}$ was chosen smaller than $\mathtt{r}_{01} = \mathtt{r}_{02} = \mathtt{r}_{03}$, $\mathcal{P}_0$ is elongated in the $c_{\bar{t}_4}$ direction. □

We now generalize the above results by also including the possibility of deciding that no parameter solution will be made available. Thus instead of working with the second penalty matrix of Table 1, we now work with the third.

Theorem 5b (Undecided included) *Let the penalty structure (112) be extended with the undecided penalty function* $\mathtt{r}_{(k+1)\alpha}(t) = u_\alpha(t)$, $\alpha = 0,\ldots,k$. *Then (31) and (49) simplify, respectively, to*

$$\begin{cases} \mathcal{P}_0 &= \{t \in \mathbb{R}^r \mid \max_{\alpha\in[1,\ldots,k]} F_\alpha(t) < a_0(t) \wedge a_{k+1}(t) < a_0(t)\} \\ \mathcal{P}_{k+1} &= \{t \in \mathbb{R}^r\backslash\mathcal{P}_0 \mid \max_{\alpha\in[1,\ldots,k]} F_\alpha(t) < a_{k+1}(t)\} \\ \mathcal{P}_{i\in[1,\ldots,k]} &= \{t \in \mathbb{R}^r\backslash\{P_0 \cup \mathcal{P}_{k+1}\} \mid i = \arg\max_{\alpha\in[1,\ldots,k]} F_\alpha(t)\} \end{cases} \tag{119}$$

and

$$\begin{cases} \mathcal{P}_0 &= \{t \in \mathbb{R}^r \mid \max_{\alpha\in[1,\ldots,k]} T_\alpha(t) < b_0(t) \wedge b_{k+1}(t) < b_0(t)\} \\ \mathcal{P}_{k+1} &= \{t \in \mathbb{R}^r\backslash\mathcal{P}_0 \mid \max_{\alpha\in[1,\ldots,k]} T_\alpha(t) < b_{k+1}(t)\} \\ \mathcal{P}_{i\in[1,\ldots,k]} &= \{t \in \mathbb{R}^r\backslash\{P_0 \cup \mathcal{P}_{k+1}\} \mid i = \arg\max_{\alpha\in[1,\ldots,k]} T_\alpha(t)\} \end{cases} \tag{120}$$

with

$$a_{k+1}(t) = \sum_{\alpha=1}^{k}\left(1 - \tfrac{u_\alpha(t)}{r(t)}\right)F_\alpha(t) - \tfrac{u_0(t)}{r(t)}F_0(t)$$

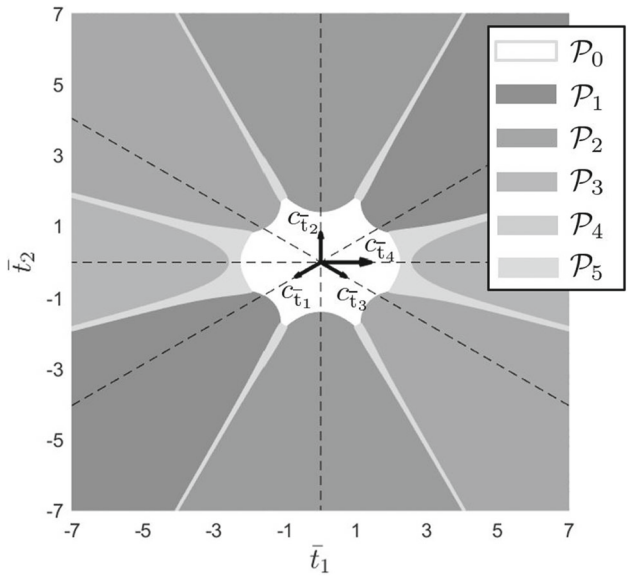

Fig. 12 Misclosure space partitioning (cf. Example 16)

$$b_{k+1}(t) = \ln\left[\sum_{\alpha=1}^{k}\left(1-\tfrac{u_\alpha(t)}{r(t)}\right)\exp\{+\tfrac{1}{2}T_\alpha(t)\} - \tfrac{u_0(t)\pi_0}{r(t)}\right]^2 \tag{121}$$

■

Proof From application of (49), using the defined functions $a_0(t)$, cf. (115), and $a_{k+1}(t)$, cf. (121), the set of inequalities for the three types of misclosure regions follow as:

$$\begin{aligned}
\mathcal{P}_0 &: \begin{cases} F_j < a_0(t) \text{ for all } j \in [1,\ldots,k] \\ a_{k+1}(t) < a_0(t) \end{cases} \\
\mathcal{P}_{k+1} &: \begin{cases} F_j(t) < a_{k+1}(t) \text{ for all } j \in [1,\ldots,k] \\ a_0(t) < a_{k+1}(t) \end{cases} \\
\mathcal{P}_{i\in[1,\ldots,k]} &: \begin{cases} F_j(t) < F_i(t) \text{ for all } j \in [1,\ldots,k]\backslash\{i\} \\ a_0(t) < F_i \\ a_{k+1}(t) < F_i \end{cases}
\end{aligned} \tag{122}$$

from which (119) follows. Finally, (120) is obtained by replacing $F_\alpha(t)$ in (119) by (52). □

One can expect that the decision of unavailability will be made if no penalties at all are set for such decision. And indeed, if $u_\alpha = 0$ for all $\alpha \in [0,\ldots,k]$, then we have for all $t \in \mathbb{R}^r$ that $a_0(t) < a_{k+1}(t)$ and $F_j < a_{k+1}(t)$ for all $j \in [1,\ldots,k]$, implying that $\mathcal{P}_{k+1} = \mathbb{R}^r$. At the other extreme, we have that $\mathcal{P}_{k+1} = \emptyset$ and the above results reduce to that of Theorem 5a if one of the following three cases hold true for all $\alpha \in [1,\ldots,k]$:

$$\begin{aligned}
&(a)\ u_\alpha = r, \\
&(b)\ u_\alpha = r_{0\alpha}, \\
&(c)\ u_\alpha > r_{0\alpha}, u_0 = 0
\end{aligned} \tag{123}$$

In case (a), we have $a_{k+1}(t) = -\frac{u_0(t)}{r(t)}F_0(t) < 0$ and therefore $\mathcal{P}_{k+1} = \emptyset$. In case (b), we have $a_{k+1}(t) = a_0(t) - \frac{u_0(t)}{r(t)}F_0(t)$ and thus $a_{k+1}(t) < a_0(t)$, which also results in $\mathcal{P}_{k+1} = \emptyset$. And in case (c), we again have $a_{k+1}(t) < a_0(t)$, thus giving $\mathcal{P}_{k+1} = \emptyset$. Under either one of the three conditions of (123) we will thus always have a parameter solution available.

Example 16 (*Simplified DIA-penalties with undecided included*) Consider the situation of the previous example, but now with an additional undecided region included. Application of (120) gives then

$$\begin{aligned}
\mathcal{P}_0 &= \{t \in \mathbb{R}^r | \max_{\alpha\in[1,\ldots,k]} w_\alpha^2(t) < b_0'(t) \wedge b_{k+1}'(t) \\
&\qquad < b_0'(t)\} \\
\mathcal{P}_{k+1} &= \{t \in \mathbb{R}^r\backslash\mathcal{P}_0 | \max_{\alpha\in[1,\ldots,k]} w_\alpha^2(t) < b_{k+1}'(t)\} \\
\mathcal{P}_{i\in[1,\ldots,k]} &= \{t \in \mathbb{R}^r\backslash\{\mathcal{P}_0 \cup \mathcal{P}_{k+1}\} | \, i \\
&\qquad = \arg\max_{\alpha\in[1,\ldots,k]} w_\alpha^2(t)|\}
\end{aligned} \tag{124}$$

with $b_{k+1}'(t) = \ln[\sum_\alpha^k\left(1-\frac{u_\alpha(t)}{r(t)}\right)\exp+\frac{1}{2}w_\alpha^2(t) - \frac{\pi_0 u_0 k}{1-\pi_0}]^2$. This partitioning is shown in Fig. 12 for the same model and hypotheses as used in Example 15, whereby the undecided penalties of the four alternative hypotheses have been chosen equal and constant, $u_1(t) = u_2(t) = u_3(t) = u_4(t) =$ constant. □

Example 17 (*An undecided region to combat poor separability*) In this example, we again work with the two partitionings (118) and (124), but now for the case in which two hypotheses are poorly separable. We use the same model and same first three alternative hypotheses as Example 9, but now with the c-vector of the fourth alternative hypothesis on purpose given as $c_4 = [0, 1, 10^{-3}]^T$. As c_2 and c_4 are almost parallel, the two hypotheses, $\mathcal{H}_2$ and $\mathcal{H}_4$, become poorly separable (Zaminpardaz and Teunissen 2019), in particular if their biases are relatively small. Figure 13 shows how inclusion of an undecided region allows one to avoid decision making between poorly separable hypotheses. In this case this was realized by having the two hypotheses $\mathcal{H}_2$ and $\mathcal{H}_4$ be assigned the smallest undecided penalties. □

Example 18 (*Poor GNSS pseudorange outlier separability*) Poor detectability and/or poor separability of pseudorange outliers also occurs with certain GNSS receiver-satellite geometries (Teunissen 1991; Almaqbile and Wang 2011;

Amiri-Simkooei et al. 2012; Teunissen 2017). In Fig. 14 (at right), one such example is shown for a six-satellite Galileo skyplot configuration (the lines of sight to satellites E01, E04, E21 and E31 all nearly lie in a plane, passing through the receiver, and the lines of sights to E09 and E19 are about perpendicular to this plane). In this case, outliers in the E09 and E19 pseudoranges are poorly separable, resulting in almost coinciding faultlines in the misclosure space. To avoid their misidentification, an example undecided region, following from Theorem 5b, is shown in Fig. 14(Left).

6 Summary and conclusions

By recognizing that members from the class of DIA-estimators are unambiguously defined by their misclosure space partitioning, one can design DIA-estimators with certain favourable properties through a proper choice of partitioning. In this contribution we introduced, in analogy to penalized integer ambiguity resolution, the concept of penalized testing with the goal of directing the performance of the DIA-estimator towards its application-dependent tolerable risk objectives. The presented theory is illustrated by means of examples, thereby also showing how it compares

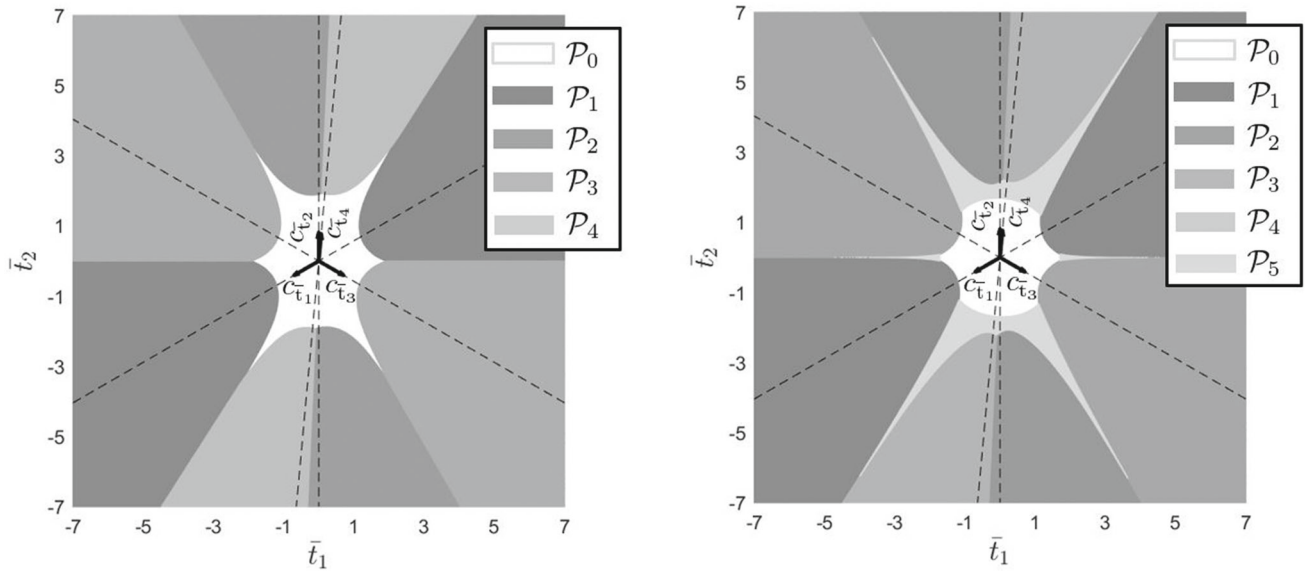

Fig. 13 Misclosure space partitioning without (left) and with (right) an undecided region (cf. Example 17)

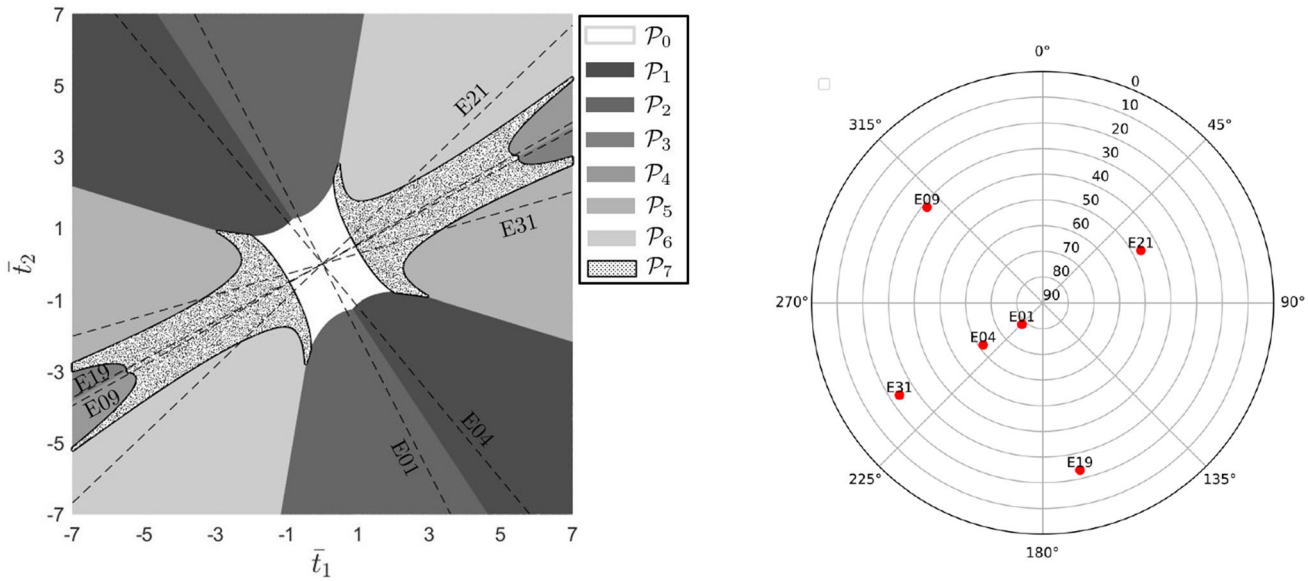

Fig. 14 Example 18: misclosure space partitioning with dashed undecided region $\mathcal{P}_7$ (left) and corresponding six-satellite Galileo skyplot (right)

and/or specializes to existing testing procedures, such as, for instance, classical data snooping.

In analogy with the aperture pull-in regions of penalized integer ambiguity resolution, we assigned *penalty functions* to each of the partitioning decision regions in misclosure space. With the use of the distribution of the misclosure vector, the mean penalty of each chosen misclosure space partitioning can then be determined and compared. By minimizing the mean penalty, the optimal partitioning for *multiple* hypothesis testing was derived, the results of which are given in Theorem 2. The results are given for different cases: the biases under the alternative hypotheses may be known or unknown, and the misclosure distribution may be normal or not. Although in most of our applications the biases are unknown, the bias-known case applies when one wants to test for certain hypothesized biases. We also included results for *constrained* minimum mean penalty misclosure partitioning. This will allow users to work with a-priori chosen decision regions, such as, for instance, the detection region of the overall model test.

As each minimum mean penalty partitioning depends on the given penalty functions, different choices can be made, in dependence of the application. So will the emphasis be on *testing* rather than *estimation*, if the data processing objective is solely to identify the correct hypothesis. In that case a logical objective is to maximize the probability of correct decisions, the results of which are given in Corollary 2, and of which classical data snooping is shown to be a special case. However, maximizing the probability of correct decisions may not be the proper objective in case the emphasis is on estimation, rather than testing.

As the quality of the DIA-estimator is not only driven by the misclosure vector $\underline{t}$, but also by the hypothesis-dependent parameter estimators $\hat{\underline{x}}_i$, one would be better off, in case parameter estimation is the objective, to focus on the *consequences* of the testing decisions rather than only on their correctness. For that purpose we introduced a special DIA-penalty function that penalizes unwanted outcomes of the DIA-estimator. It is then shown in Theorem 3 how this penalty function allows one to construct the *optimal* DIA-estimator, being the estimator that within its class has the largest probability of lying inside a user specified tolerance or safety region. By extending the analogy with integer estimation to that of integer-equivariant estimation, we also introduced and derived in Theorem 4, similar to that of the maximum probability DIA-estimator, the optimal estimator within the larger WSS-class. We indicated its computational complexities, showing that its algorithmic realization is quite more involved than that of the optimal DIA-estimator.

By a further elaboration of the DIA-penalty functions, it is shown in Lemma 3 how they are driven by the *influential* biases of the different hypotheses. This important insight then provided means for defining simplified and operational penalty functions. Two such sets were introduced. The first is based on using the BLUEs of the biases to obtain an estimate of the DIA-penalty function. The second set is constructed on the basis of the idea that the to-be-used influential biases should reflect the relative strengths of the different hypotheses, i.e. a smaller penalty should be assigned if the model is better capable to withstand the bias-propagation into the parameter solution. Using this principle, we used the largest influential biases of the overall model test when it is constrained to have an equal power for all alternative hypotheses.

For both sets of penalty functions, a further practical simplification was suggested by giving prominence to missed detections and refraining from penalizing false alarms. The resulting minimum mean penalty partitionings are given in Theorem 5. We hereby also included the option of having an additional *undecided* region to accommodate situations when it will be hard to discriminate between some of the hypotheses or when identification is unconvincing. In such situations, when one lacks confidence in the decision making, one may rather prefer to state that a solution is unavailable, than providing an actual, but possibly unreliable, parameter estimate.

Acknowledgements Sebastian Ciuban and Serge Kaplev contributed to the figures and the Delft PNT-team provided feedback and corrections to earlier versions of this contribution.

Data availability All data generated or analysed during this study are included in this contribution.

Declarations

Conflict of interest The author declares that he has no conflict of interest to this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

References

Almaqbile A, Wang J (2011) Analysis of outlier separability in integrated GPS/INS systems. In: Proceedings of the IGNSS symposium 2011, University of New South Wales, Sydney, NSW, Australia

Amiri-Simkooei AR, Asgari J, Zangeneh-Nejad F, Zaminpardaz S (2012) Basic concepts of optimization and design of geodetic networks. J Surv Eng 138(4):172–183

Anderson T (1955) The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. Proc Am Math Soc 6(2):170–176

Arnold SF (1981) The theory of linear models and multivariate analysis. Wiley, New York

Baarda W (1968a) A testing procedure for use in geodetic networks. Tech. rep., Netherlands Geodetic Commission, Publ. on Geodesy, New Series, Vol. 2(5), Delft

Baarda W (1968b) Enkele inleidende beschouwingen tot de B-methode van toetsen. Tech. rep., Laboratorium voor Geodesie, Delft

DGCC (1982) The Delft approach for the design and computation of geodetic networks. "Forty Years of Thought" In: Anniversary edition on the occasion of the 65th birthday of Professor W Baarda By staff of the Delft Geodetic Computing Centre (DGCC) 1:202–274

Gillissen I, Elema I (1996) Test results of DIA: a real-time adaptive integrity monitoring procedure, used in an integrated navigation system. Int Hydrogr Rev 73(1):75–103

Kok JJ (1984) On data snooping and multiple outlier testing. US Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, Charting and Geodetic Services

Kroese DP, Taimre T, Botev ZI (2011) Handbook of Monte Carlo methods. Wiley Series in Probability and Statistics

Lehmann R, Lösler M (2016) Multiple outlier detection: hypothesis tests versus model selection by information criteria. J Surv Eng 142(4):04016017

Lehmann R, Lösler M (2017) Congruence analysis of geodetic networks-hypothesis tests versus model selection by information criteria. J Appl Geod 11(4):271–283

Mathai AM, Provost SB (1992) Quadratic forms in random variables, vol 126. Marcel Dekker, New York

Nowel K (2020) Specification of deformation congruence models using combinatorial iterative DIA testing procedure. J Geod 94(12):1–23

Perfetti N (2006) Detection of station coordinate discontinuities within the Italian GPS fiducial network. J Geod 80(7):381–396

Robert CP, Casella G (2004) Monte Carlo statistical methods. Springer, Berlin

Salzmann M (1991) MDB: a design tool for integrated navigation systems. Bull Geod 65(2):109–115

Teunissen PJG (1989) Quality control in integrated navigation systems. IEEE Aerosp Electron Syst Mag 5(7):35–41

Teunissen PJG (1991) Differential GPS: Concepts and quality control. NIN Workshop Navstar GPS, Amsterdam (Delft Geodetic Computing Centre LGR, Delft 1991) pp 1–49

Teunissen PJG (2000) Testing theory: an introduction. Delft University Press, Series on Mathematical Geodesy and Positioning

Teunissen PJG (2003) Integer aperture GNSS ambiguity resolution. Artif Satel 38(3):79–88

Teunissen PJG (2003) Theory of integer equivariant estimation with application to GNSS. J Geod 77:402–410

Teunissen PJG (2004) Penalized GNSS ambiguity resolution. J Geod 78:235–244

Teunissen PJG (2017) Batch and recursive model validation. Chapter 24 in Springer Handbook of Global Navigation Satellite Systems, Eds Teunissen and Montenbruck

Teunissen PJG (2018) Distributional theory for the DIA method. J Geod 92:59–80

Yang L, Shen Y, Li B, Rizos C (2021) Simplified algebraic estimation for the quality control of DIA estimator. J Geod 95:1–15

Yu Y, Yang L, Shen Y, Sun N (2023) A DIA method based on maximum a posteriori estimate for multiple outliers. GPS Solut 27:199

Zaminpardaz S, Teunissen PJG (2019) DIA-datasnooping and identifiability. J Geod 93:85–101

Zaminpardaz S, Teunissen PJG (2023) Detection-only versus detection and identification of model misspecifications. J Geod 97:55

Zaminpardaz S, Teunissen PJG, Tiberius C (2020) A risk evaluation method for deformation monitoring systems. J Geod 94(3):28