GPS AMBIGUITY RESOLUTION AS A CLASSIFICATION PROBLEM

P. J. G. TEUNISSEN

*Delft Geodetic Computing Centre, Delft, The Netherlands**

Summary: *In this contribution we explore the analogy between GPS ambiguity resolution on the one hand, and the problem of statistical classification, on the other hand. As classification rules we discuss the Bayes-, the MAP-, the ML- and the Minmax-rule. It is shown to what extent ambiguity resolution may be considered a classification problem. It turns out that both problems show many similarities, although some marked differences exist as well. These similarities and differences are discussed.*

Keywords: GPS, ambiguity resolution, classification

1 INTRODUCTION

All GPS models underlying ambiguity resolution can be cast in the following conceptual frame of linear(ized) observation equations

$$y = \mathbf{A}a + \mathbf{B}b + e\,, \tag{1}$$

where y is the given data vector, a and b are the unknown parameter vectors of order n and q respectively, and where e is the noise vector of order p. Matrices **A** and **B** are the corresponding design matrices of order $p \times n$ and $p \times q$, respectively. Matrix (**A**,**B**) is assumed to be of full rank. In case of GPS, the data vector will usually consist of the 'observed minus computed' single- or dual-frequency double-differenced (DD) phase and/or pseudo range (code) observations, accumulated over all observation epochs. The entries of vector a are the DD carrier phase ambiguities, expressed in units of cycles rather than range. They are known to be integers. The entries of vector b consist of the remaining unknown parameters, such as for instance baseline components (coordinates) and possibly atmospheric delay parameters (troposphere, ionosphere).

The problem of GPS ambiguity resolution is to estimate and validate the DD carrier phase ambiguities as *integers*. For that purpose, the computation of the solution of (1) is usually divided into three different steps, see, e.g., *Teunissen (1993,1997a)*. In the first step, one simply disregards the integer constraints on the ambiguities and performs a standard adjustment. As a result one obtains the (real-valued) estimates of **a** and **b**, together with their variance-covariance matrix

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}, \begin{bmatrix} Q_{\hat{a}} & Q_{\hat{a}\hat{b}} \\ Q_{\hat{b}\hat{a}} & Q_{\hat{b}} \end{bmatrix}. \tag{2}$$

This solution is often referred to as the 'float' solution. In the second step, the 'float' ambiguity estimate $\hat{a}$ is used to compute the corresponding integer ambiguity estimate. This implies that one has to introduce a mapping from R^n, the space of reals, to Z^n, the space of integers. Examples of

* Address: Faculty of Geodesy, Delft University of Technology, Thijsseweg 11, 2629 JA Delft, The Netherlands (Fax: +31-15-278 3711; e-mail: P.J.G.Teunissen@geo.tudelft.nl)

such maps are 'integer rounding', 'integer bootstrapping' or 'integer least-squares', see, e.g., *Teunissen (1997b,1998)*. In case of the least-squares map, one solves the minimization problem

$$\min_{a \in Z^n} (\hat{a}-a)^T Q_{\hat{a}}^{-1} (\hat{a}-a) \tag{3}$$

Once the integer solution is computed, it is finally used in the third step to correct the 'float' estimate of b. As a result one obtains the 'fixed' solution

$$\breve{b} = \hat{b} - Q_{\hat{b}\hat{a}} Q_{\hat{a}}^{-1} (\hat{a}-\breve{a}) \tag{4}$$

where $\breve{a}$ denotes the integer ambiguity solution.

The purpose of this contribution is to explore the similarities and differences, if any, between the problem of GPS ambiguity resolution on the one hand, and the statistical classification problem, on the other hand. In Sect. 2 we discuss the statistical classification problem. It is presented as a multi-hypotheses testing problem in which all hypotheses are treated symmetrically. Within the classification framework, we show the relation between different classification rules. They are the Bayes-, the MAP-, the ML- and the Minmax-rule. All these rules boil down to a specific partitioning of the observation space, such that subsets are created which act as the classification regions. The shape and size of these subsets are determined by minimizing the cost of the classification rules.

In Sect. 3 we specialize the classification problem to the Gaussian case. The corresponding classification regions are presented and their geometry is discussed. Since probabilistic inferences of GPS data analysis are usually based on the normal distribution, it is the Gaussian case with which ambiguity resolution can be compared. In this comparison we consider both the estimation and validation part of GPS ambiguity resolution.

2 THE CLASSIFICATION FRAMEWORK

In this section we present a brief review of the statistical framework of classification. For more details we refer to, e.g., *Anderson (1958)*, *McLachlan (1992)* or *Huberty (1994)*. Let $\hat{a} \in R^n$ be a continuous random n-vector with probability density function (pdf) $p_{\hat{a}}(a \mid z)$. The notation is chosen such that $\hat{a}$ may be thought of as the ambiguity 'float' solution. Its pdf is assumed known apart from z, which is assumed to be an element of the discrete set $\Omega = \{a_1, \ldots, a_m\} \subset R^n$. This set may be thought of as containing all likely integer ambiguity vectors. To each element a_i of set Ω we may assign a corresponding hypothesis. Hence, we have m hypotheses H_i, $i = 1, \ldots, m$,

$$H_1 : p_{\hat{a}}(a \mid a_1), \ldots, H_m : p_{\hat{a}}(a \mid a_m) \tag{5}$$

Only one of these hypotheses can be true and the problem of classification is to decide which hypothesis to choose based on a given sample or observation of $\hat{a}$. In the classification problem, we thus want to decide from which pdf the sample was drawn.

This classification problem is essentially a multi-hypotheses testing problem in which all hypotheses are treated symmetrically. Often in problems of hypothesis testing we are used to identifying one hypothesis as the null hypothesis and the remaining hypotheses as the alternatives. This is then also reflected in how one tries to control the error probabilities. For instance, by setting the probability of the type I error to a fixed value and choosing procedures which have a small probability for the type II error. Such an

approach is clearly asymmetric in the hypotheses. For the above classification problem there is no reason to call one hypothesis the null hypothesis. All hypotheses are, therefore, treated symmetrically.

2.1 Minimizing the cost of classification

In order to pick one of the hypotheses, we need a classification rule that assigns $\hat{a} \in R^n$ to one of the H_i. We, therefore, partition the observation space R^n into disjoint sets S_i, $\cup_{i=1}^m S_i = R^n$ and $S_i \cap S_j = \{0\}$ for $i \neq j$, and decide to choose H_i when $\hat{a} \in S_i$. This classification rule, or S-rule, depends on the chosen partitioning. Hence, there are as many rules as there are ways of partitioning observation space R^n, and some of them may be considered better than others.

In order to qualify a classification rule, we need the classification probabilities and the costs involved of making a wrong decision. The probability of choosing hypothesis H_i when in fact H_j is true, reads

$$P_S(a_i \mid a_j) = \int_{S_i} p_{\hat{a}}(a_i \mid a_j)\, da, \tag{6}$$

where subscript "S" is used to show the dependence on the chosen S-rule. If we denote the cost of choosing H_i while H_j is true as

$$C(a_i \mid a_j) = \begin{cases} 0 & \text{if} \quad a_i = a_j \\ >0 & \text{if} \quad a_i \neq a_j \end{cases}, \tag{7}$$

the average cost when H_j is true becomes

$$C_S(a_j) = \sum_{i=1, i\neq j}^{m} C(a_i \mid a_j) P_S(a_i \mid a_j). \tag{8}$$

This average cost reduces to the probability of misclassifying H_j, $C_S(a_j) = 1 - P_S(a_j \mid a_j)$, when the 0 – 1 cost function, $C(a_i \mid a_j) = 1 - \delta_{ij}$, is used.

Instead of (6), we may also consider the probability of choosing H_i when H_j in fact occurs. This requires knowledge of the a priori probabilities of the hypotheses. These probabilities will be denoted as $\pi(a_i)$; $i = 1, \ldots, m$, with $\sum_{i=1}^m \pi(a_i) = 1$. The probability of choosing H_i while H_j occurs, then reads

$$P_S(a_i, a_j) = P_S(a_i \mid a_j)\pi(a_j). \tag{9}$$

The overall average cost of the classification rule then becomes

$$C_S = \sum_{\substack{i=1 \\ i\neq j}}^{m} \sum_{j=1}^{m} C(a_i \mid a_j) P_S(a_i \mid a_j) = \sum_{j=1}^{m} C(a_j)\pi(a_j). \tag{10}$$

This cost may now be used to compare and judge various classification rules. That is, we can now infer what a particular partitioning of observation space R^n does to the overall average cost. To emphasize the dependence of the cost on the chosen partitioning, we put

$$C_S = \sum_{i=1}^{m} \int_{S_i} f_i(a)da \text{, with } f_i(a) = \sum_{\substack{j=1 \\ j\neq i}}^{m} C(a_i|a_j) p_{\hat{a}}(a|a_j) \pi(a_j). \tag{11}$$

Thus the cost equals a sum of integrals of nonnegative functions $f_i(a)$, having subsets S_i as their domain of integration. With this representation we are now in the position to solve the classification problem. That is, we are now able to identify the partitioning of R^n which results in a minimal overall average cost. The cost is minimal when the subsets are chosen so that each of the nonnegative entries in the sum of (11) is minimal. But this requires each subset to correspond with the domain for which one of the m functions $f_i(a)$ is minimal. Hence, the subsets that minimize the overall average cost are given as

$$B_i = \left\{ a \in R^n \middle| f_i(a) = \min_j f_j(a) \right\}, \; i = 1, \ldots, m \,. \tag{12}$$

We thus have

$$C_B = \sum_{i=1}^{m} \int_{B_i} f_i(a)da \leq C_S = \sum_{i=1}^{m} \int_{S_i} f_i(a)da \tag{13}$$

for all S-rules. The classification rule that uses the above subsets B_i for its partitioning, is known as the Bayes-rule (*Anderson, 1958*).

2.2 Different classification rules

The Bayes-rule is a very general rule. Other rules can be derived from it when particular choices are made for the cost function and/or for the a priori probabilities. In the following we will consider the $0-1$ cost function, the situation in which all a priori probabilities are equal, and the situation in which the a priori probabilities are unknown. We will start with the Bayes-rule itself.

The Bayes rule: Since H_i is chosen when $\hat{a} \in B_i$, it follows from (12) that the Bayes-rule can be expressed as

$$\text{Choose } H_i \text{ when } f_i(\hat{a}) \leq f_j(\hat{a}), \quad \forall j \neq i \,, \tag{14}$$

or alternatively, when we substitute $f_i(a)$ of (11) into (14),

Choose H_i when

$$\sum_{\substack{k=1 \\ k\neq i}}^{m} C(a_i|a_k) p_{\hat{a}}(\hat{a}|a_k) \pi(a_k) \leq \sum_{\substack{k=1 \\ k\neq j}}^{m} C(a_j|a_k) p_{\hat{a}}(\hat{a}|a_k) \pi(a_k), \; \forall j \neq i \tag{15}$$

Note, that for executing this rule, only the ratio of the costs needs to be known.

The MAP rule: The Bayes-rule simplifes considerably if the $0-1$ cost function is used. With $C(a_i|a_j)=1-\delta_{ij}$, it follows from (15) that H_i is chosen if $p_{\hat{a}}(\hat{a}|a_i)\pi(a_i) \geq p_{\hat{a}}(\hat{a}|a_j)\pi(a_j)$, $\forall j \neq i$. Since $p(a_i|\hat{a}) \propto p_{\hat{a}}(\hat{a}|a_i)\pi(a_i)$, it follows that the rule may be expressed as

$$\text{Choose } H_i \text{ when } p(a_j|\hat{a}) \geq p(a_i|\hat{a}), \quad \forall j \neq i. \tag{16}$$

Note that $p(a_i|\hat{a})$ is the *a posteriori* distribution of a_i, i.e., it is the distribution of a_i given the data $\hat{a}$. This rule is, therefore, known as the *maximum a posteriori* or *MAP*-rule.

The ML rule: A further simplification is obtained if all a priori probabilities are the same. In that case, one can divide by the a priori probabilities in (16) so as to obtain the rule

$$\text{Choose } H_i \text{ when } p_{\hat{a}}(\hat{a}|a_i) \geq p_{\hat{a}}(\hat{a}|a_j), \quad \forall j \neq i. \tag{17}$$

This rule is known as the *maximum likelihood* or *ML*-rule.

Up to this point we have been working under the assumption that the a priori probabilities $\pi(a_i)$, $i = 1, \ldots, m$, are known. However, if they are not known, we can no longer work with the overall average cost C_S. In that case we are forced to fall back on the individual average costs $C_S(a_i)$. But now it becomes more difficult to compare two different classification rules, say S and R, since we no longer have a single criterion for comparison. It could well be that $C_S(a_i) \geq C_R(a_i)$ for some i, but that $C_S(a_j) \geq C_R(a_j)$ for $j \neq i$. In that case it is not clear which of the two rules, S or R, one should prefer. A way out of this dilemma is to make use of the so-called minmax criterion. With this criterion the maximum of the individual costs of a rule is minimized. Hence, rule R is said to be a *minmax*-rule when

$$\max_i C_R(a_i) \leq \max_i C_S(a_i) \tag{18}$$

for all S-rules. This definition though, does not yet make clear how a minmax-rule can be constructed. Fortunately it can be shown that a minmax-rule can be constructed by making use of the structure of the Bayes-rule. This is due to the following important property of the Bayes-rule: a Bayes-rule for which $\pi(a_i) > 0$, $\forall i$, and $C_B(a_i) = \text{constant}$; $\forall i$, is a minmax-rule. We will prove this property in two steps. For the first step we will show that no rule S exists for which

$$C_S(a_i) < C_B(a_i) \text{ for some } i \text{ and } C_S(a_j) \leq C_B(a_j) \text{ for } j \neq i \tag{19}$$

when all $\pi(a_i)$ are positive. For the Bayes-rule we have $C_B \leq C_S$ and thus $\sum_{i=1}^{m} C_B(a_i)\pi(a_i) \leq \sum_{i=1}^{m} C_S(a_i)\pi(a_i)$. Now suppose that $\pi(a_1) > 0$ and $C_S(a_i) \leq C_B(a_i)$ for

$i = 2, \ldots, m$. Then $\pi(a_1)[C_B(a_1) - C_S(a_1)] \leq \sum_{i=2}^{m} \pi(a_i)[C_S(a_i) - C_B(a_i)] \leq 0$. Hence, since $\pi(a_1) > 0$, it follows that $C_B(a_1) \leq C_S(a_1)$ and not that $C_S(a_1) < C_B(a_1)$. Repeating this argument for the other $\pi(a_i)$ too, shows that one can never find a rule for which (19) holds true when all $\pi(a_i)$ are positive. The second step of the proof is by contradiction. Assume that the Bayes-rule with $\pi(a_i) > 0$; $\forall i$, and $C_B(a_i) = \text{constant}$; $\forall i$, is not a minmax-rule. Then an S-rule exists with $\max_i C_S(a_i) \leq \max_i C_B(a_i)$ and thus, when not all $C_S(a_i)$ are the same, $C_S(a_i) < \max_j C_B(a_j)$ for some i, from which it follows, since $C_B(a_j) = \text{constant}$, $\forall j$, that $C_S(a_i) < C_B(a_i)$ for some i. But this contradicts the fact that no S-rule exists for which (19) holds true when all $\pi(a_i)$ are positive.

The Minmax rule: We are now in the position to express the minmax-rule. It is (15) with the constraint that all $\pi(a_i)$ are positive and chosen such that the $C_B(a_i)$ are the same for all i. The minmax subsets read, therefore,

$$B_i^{\pi} = \left\{ a \in R^n \,\middle|\, f_i^{\pi}(a) \leq f_j^{\pi}(a), \quad \forall j \neq i \right\}, \; i = 1, \ldots, m \tag{20}$$

with

$$f_i^{\pi}(a) = \sum_{\substack{j=1 \\ j \neq i}}^{m} C(a_i | a_j) p_{\hat{a}}(a | a_j) \pi(a_j)$$

and appropriate chosen $\pi(a_i) > 0$. Note that in this case the $\pi(a_i)$ are not to be interpreted as a priori probabilities anymore. They have become positive weights which are needed to enforce the minmax property.

2.3 Costs of the classification rules

Now that we have determined the different classification regions, we may consider the corresponding costs. The costs for each of the above four classification rules can be specified as follows. For the Bayes-rule we have

$$C_B = \sum_{i=1}^{m} \int_{B_i} f_i(a) da \,. \tag{21}$$

The MAP-rule follows when using the $0-1$ cost function. The corresponding cost then reads

$$C_{MAP} = 1 - \sum_{i=1}^{m} \pi(a_i) \int_{B_i^{MAP}} p_{\hat{a}}(a | a_i) da \,. \tag{22}$$

The subset B_i^{MAP} follows from using the $0-1$ cost function in (21). Note that C_{MAP} equals one minus the probability of making a correct decision. It, therefore, equals the

error probability. The ML-rule follows if we assume $\pi(a_i) = 1/m$. The cost reads, therefore,

$$C_{ML} = 1 - \frac{1}{m}\sum_{i=1}^{m} \int_{B_i^{ML}} p_{\hat{a}}(a|a_i)\,da \,. \tag{23}$$

The Minmax-rule follows from using the Bayes-rule with positive weights $\pi(a_i)$ chosen so that the individual costs are all the same. The corresponding cost then reads

$$C_{Minmax} = \sum_{\substack{i=1 \\ i \neq j}}^{m} C(a_i|a_j) \int_{B_j^{\pi}} p_{\hat{a}}(a|a_j)\,da \,. \tag{24}$$

In case of the $0-1$ cost function it reduces to

$$C_{Minmax} = 1 - \int_{B_i^{\pi}} p_{\hat{a}}(a|a_i)\,da \,. \tag{25}$$

3. GAUSSIAN CLASSIFICATION

The classification results obtained so far are all applicable to any probability density function that $\hat{a}$ might have. Having the problem of ambiguity resolution in mind though, we will now specialize to the Gaussian case. If we assume y of model (1) to be normally (Gaussian) distributed as $y \approx N(Aa + Bb, Q_y)$, then $\hat{a}$ is normally distributed too. Its probability density function then reads

$$p_{\hat{a}}(a|z) = \frac{1}{\sqrt{\det(Q_{\hat{a}})}(2\pi)^{\frac{1}{2}n}} \exp\left\{-\frac{1}{2}\|a - z\|^2_{Q_{\hat{a}}}\right\} \tag{26}$$

with the squared weighted norm $\|\cdot\|^2_{Q_{\hat{a}}} = (\cdot)^T Q_{\hat{a}}^{-1}(\cdot)$. With this Gaussian pdf, the various classification rules of the previous section simplify considerably, in particular if the $0-1$ cost function is used. We will first discuss the geometry of the classification regions and then make the connection with GPS ambiguity resolution.

3.1 Geometry of the classification regions

For the $0-1$ cost function we obtain for the Gaussian case, using the pdf of (26), the following classification rules

$$\text{Choose } H_i \text{ when } \hat{a} \in S_i \,, \tag{27}$$

where the subsets S_i, $i = 1, \ldots, m$, are given as

$$S_i = \left\{ a \in R^n \middle| \|a - a_i\|^2_{Q_{\hat{a}}} \leq \|a - a_j\|^2_{Q_{\hat{a}}} + 2k_i - 2k_j, \quad \forall j \neq i \right\}, \tag{28}$$

with MAP: $k_i = \ln \pi(a_i)$ (a priori probability),

ML: $k_i = 0$,

Minmax: $k_i = \ln \pi(a_i)$ (minmax weight).

This shows that, in this case, the MAP-, the ML- and the Minmax-rule may be considered as belonging to one family. The definition of subsets S_i changes only when constants k_i change.

In order to describe the geometry of the above classification regions, we will first rewrite (28) in a more convenient form. By using the decomposition $\|a - a_j\|^2_{Q_{\hat{a}}} = \|a - a_i\|^2_{Q_{\hat{a}}} + \|a_j - a_i\|^2_{Q_{\hat{a}}} - 2(a_j - a_i)^T Q^{-1}_{\hat{a}}(a - a_i)$, the inequality of (28) can be rewritten as an inequality which is linear in a. This yields

$$S_i = \left\{ a \in R^n \middle| v_{ji}(a) \leq k_i - k_j, \quad \forall j \neq i \right\}, \quad i = 1, \ldots, m, \tag{29}$$

with

$$v_{ji}(a) = (a_j - a_i)^T Q^{-1}_{\hat{a}} \left(a - \frac{1}{2}(a_i + a_j) \right).$$

Note that each of the inequalities in (29) describes a halfspace. Thus S_i equals the intersection of $m - 1$ such halfspaces. Since a halfspace is convex and since the intersection of convex subsets is also convex, it follows that subsets S_i are convex as well. They are referred to as *convex polytopes*.

In order to evaluate the costs of the classification rules, we need to evaluate the integrals of the probability density function over the above subsets. That is, we need to determine the probability that $\hat{a}$ falls in S_i, or equivalently, we need to determine the probability that $v_{ji}(\hat{a})$ satisfies the inequalities of (29). This shows that the distribution of $v_{ji}(\hat{a})$ is needed.

Since $\hat{a}$ is normally distributed, $v_{ji}(\hat{a})$ is normally distributed as well. It is distributed under hypotheses H_i and H_j as

$$\begin{cases} H_i : v_{ji}(\hat{a}) \approx N\left(-\frac{1}{2}\|a_j - a_i\|^2_{Q_{\hat{a}}}, \|a_j - a_i\|^2_{Q_{\hat{a}}} \right) \\ H_j : v_{ji}(\hat{a}) \approx N\left(+\frac{1}{2}\|a_j - a_i\|^2_{Q_{\hat{a}}}, \|a_j - a_i\|^2_{Q_{\hat{a}}} \right) \end{cases} \tag{30}$$

The v-statistics are correlated with the covariance

$$\mathrm{Cov}\left(v_{ji}(\hat{a}), v_{ki}(\hat{a})\right) = \left(a_j - a_i\right)^T Q_{\hat{a}}^{-1}\left(a_k - a_i\right).$$

If constants k_i, $i = 1, \ldots, m$, are all equal, an alternative way of describing the subsets (29) is

$$S_i = \left\{ a \in R^n \middle| w_{ji}(a) \le \frac{1}{2}\left\|a_j - a_i\right\|_{Q_{\hat{a}}}, \quad \forall j \ne i \right\}, \quad i = 1, \ldots, m, \tag{31}$$

with $$w_{ji}(a) = \frac{\left(a_j - a_i\right)^T Q_{\hat{a}}^{-1}\left(a - a_i\right)}{\sqrt{\left(a_j - a_i\right)^T Q_{\hat{a}}^{-1}\left(a_j - a_i\right)}}.$$

The statistic $w_{ji}(\hat{a})$ is also normally distributed. It is distributed under hypotheses H_i and H_j as

$$\begin{cases} H_i: & w_{ji}(\hat{a}) \approx N(0,1) \\ H_j: & w_{ji}(\hat{a}) \approx N\left(\left\|a_j - a_i\right\|_{Q_{\hat{a}}}, 1\right) \end{cases} \tag{32}$$

with the covariance

$$\mathrm{Cov}\left(w_{ji}(\hat{a}), w_{ki}(\hat{a})\right) = \frac{\left(a_j - a_i\right)^T Q_{\hat{a}}^{-1}\left(a_k - a_i\right)}{\left\|a_j - a_i\right\|_{Q_{\hat{a}}}\left\|a_k - a_i\right\|_{Q_{\hat{a}}}}.$$

Note that the $w_{ji}(\hat{a})$ are the well-known w-test statistics for testing one-dimensional alternative hypotheses (*Baarda, 1968*; *Teunissen, 1985*). Geometrically, they can be interpreted as orthogonal projectors which project $(\hat{a} - a_i)$ onto direction vectors $a_j - a_i$. Orthogonality is hereby measured in the metric of $Q_{\hat{a}}$. Thus S_i equals the intersection of $m - 1$ halfspaces, with the border of the j-th halfspace positioned halfway $a_j - a_i$ and orthogonal to this direction. Also note that the covariance between $w_{ji}(\hat{a})$ and $w_{ki}(\hat{a})$ can be interpreted as the cosine of the angle between a_j, a_i and a_k.

3.2 Ambiguity resolution

With the above results we are now in a position to discuss the link with the GPS ambiguity resolution. As mentioned earlier, the ambiguity resolution consists of an estimation part and a validation part. The estimation part is concerned with the problem of using $\hat{a}$, the "float" solution, to determine the corresponding integer ambiguity estimate. In classification terminology this corresponds to the decision which hypothesis to pick. The map from $\hat{a}$ to integer vector $\breve{a}$ is thus determined by the classification rule.

As an example consider (28) for the case that the k_i are all constant. The corresponding classification region is then the one that would follow from applying the least-squares principle

$$\min_{a\in\Omega}(\hat{a}-a)^T Q_{\hat{a}}^{-1}(\hat{a}-a). \tag{33}$$

Compare this minimization problem with the integer least-squares principle (3). Two things are worth mentioning. First, note that the ambiguity resolution is concerned with the determination of *integers*, whereas for the classification problem this need not be the case per se. Although the classification problem is *discrete*, the entries of set Ω need not be integer vectors. Secondly we observe that the least-squares problem (33) is constrained in the sense that set Ω is *finite*. In this sense the integer least-squares problem (3) is unconstrained, since the minimization is taken over the whole infinite space of integers Z^n. Although this difference is essential in principle, it may be less relevant for practical applications. After all, one could try to define the finite set Ω large enough to contain all candidate integer ambiguity vectors.

The validation part of ambiguity resolution is concerned with the probabilistic characteristics of the integer ambiguity solution. It is related to the optimality of the integer solution and in particular it should give an answer to the question how successful can ambiguity resolution be expected to be. It should thus provide the probability of successful integer ambiguity estimation. In the terminology of classification this is related to the costs of the classification rules. As shown previously all these costs depend on integrals of the probability density function over the respective classification regions. The evaluation of these integrals is usually very difficult, in particular when certain symmetries are absent in the shape of the classification regions. This is for instance the case when constants k_l differ. But also when all k_l are equal, the classification regions may turn out to be asymmetric. This is the case when Ω is finite and when the entries of this subset are irregularly distributed over space R^n. Even a symmetric distribution of the entries of Ω may not necessarily result in symmetric classification regions. This is due to Ω being finite. As a result the classification regions will change in shape the closer one gets to the boundaries of set Ω.

When constants k_l are all equal, the situation is fortunately somewhat simpler to handle for the case of ambiguity resolution. The integer grid points are symmetrically distributed and set Z^n is infinite. As a result all classification regions S_l will have the same shape. In fact they will be symmetric about their own gridpoint a_l. We may, therefore, concentrate on the single integral

$$\int_{S_l} p_{\hat{a}}(a|a_l)da\,. \tag{34}$$

In general this integral is still difficult to evaluate. The integral is only easy to evaluate when the ambiguity variance-covariance matrix becomes diagonal. In that case subset S_l reduces to a cube, while the pdf can be expressed as a product of one-dimensional normal distributions. As a result the integral itself also reduces to a product of one-dimensional integrals. In the case of GPS, however, the ambiguity variance-covariance matrix is far from diagonal. This simple and rigorous evaluation of (34) is, therefore, not applicable. Instead one may try to obtain approximations along this line. For that purpose one should first try to decorrelate the ambiguities as much as possible, in order to get a close to diagonal ambiguity variance-covariance matrix. For the ambiguity decorrelation process,

we refer to, e.g., *Teunissen (1993)*. Alternatively one may use the geometry of subset S_i to formulate bounds for the above integral. As an example consider

$$P\left(\left\|\hat{a}-a_i\right\|_{Q_{\hat{a}}} \le l\right) \le \int_{S_i} p_{\hat{a}}(a|a_i)da \le P\left(\left|w_{ji}(\hat{a})\right| \le r\right). \tag{35}$$

These bounds can be understood as follows. When the a_j are elements of the space of integers, subset S_i of (31) equals the intersection of pairs of halfspaces, with each pair being symmetric about a_i. The region covered by each pair is, therefore, always larger than the intersection itself. This shows that the above upper bound is applicable for $r = \frac{1}{2}\left\|a_j - a_i\right\|_{Q_{\hat{a}}}$. This bound becomes tighter when a_j is chosen as the grid point nearest to a_i. For the lower bound this same value of r is taken for l. The lower bound then describes the probability of $\hat{a}$ falling in the ellipsoidal region just contained in S_i. Note that both bounds are easy to evaluate. For the first bound we need the Chi-squared distribution and for the second we need the normal distribution.

4. SUMMARY

In this contribution we have explored the similarities and differences between, on the one hand, the statistical classification problem and, on the other hand, the problem of GPS ambiguity resolution. Both problems are very similar. For the classification problem a review was given of different classification rules. These rules may be rather complicated, as is the case with the general Bayes-rule, or relatively simple such as the maximum likelihood rule. The similarity between the two problems is in that, in both cases, the observation space is partitioned into mutually exclusive subsets, with the ruling that the grid point of one of these subsets is chosen once sample $\hat{a}$ lies in this subset. In the case of classification, the grid point is said to identify the hypothesis, while in the case of ambiguity resolution, the grid point identifies the estimated integer ambiguity vector. Thus it is the partitioning itself which defines either the classification rule, or the integer ambiguity estimator.

For the performance of the different rules, costs or error probabilities are used. They depend on the computation of integrals over the various classification regions. The same holds true for the ambiguity resolution. In the case of ambiguity resolution, the probability of successful integer estimation is particularly of relevance. It equals the integral of the multivariate normal distribution over the classification region centred at the mean of $\hat{a}$.

Apart from the similarities, we also discussed a subtle difference between the two problems. The classification problem is usually an optimization problem over a finite set of hypotheses, whereas in the case of ambiguity resolution we need to optimize over the whole space of integers. For practical use one could decide to ignore this difference as long as the finite set is chosen large enough, or alternatively, when it is permitted to replace the global integer minimization problem by a local one. For validation purposes however, the global integer minimization problem has the advantage that the geometry of the classification regions simplifies due to the presence of symmetry.

Manuscript received: 16 August 1998; Revisions accepted: 21 November 1998

References

Anderson T.W., 1958: *Introduction to Multivariate Statistical Analysis.* John Wiley and Sons.

Baarda W., 1968: A testing procedure for use in geodetic networks. Netherlands Geodetic Commission, Publications on Geodesy, New Series, Vol. 2, No. 5.

Huberty C.J., 1994: *Applied Discriminant Analysis.* John Wiley and Sons.

McLachlan G.J., 1992: *Discriminant Analysis and Statistical Pattern Recognition.* John Wiley and Sons.

Teunissen P.J.G., 1985: Quality control in geodetic networks. In: E. Grafarend and F. Sanso (Eds.), *Optimization and Design of Geodetic Networks*, Springer Verlag.

Teunissen P.J.G., 1993: Least-squares estimation of the integer GPS ambiguities. Invited Lecture, Section IV Theory and Methodology, IAG General Meeting, Beijing, China, August 1993. Also in: LGR Series, No. 6, Delft Geodetic Computing Centre.

Teunissen P.J.G., 1997a: GPS ambiguity resolution: impact of time correlation, cross correlation and satellite elevation dependence. *Stud. geophys. geod.*, **41**, 181-195.

Teunissen P.J.G., 1997b: Some remarks on GPS ambiguity resolution. *Artificial Satellites*, **32**, 119-130.

Teunissen P.J.G., 1998: Success probability of integer GPS ambiguity rounding and bootstrapping. *J. Geodesy*, in print.