Theory for the ambiguity function method probability model and global solution

Teunissen, P. J.G.; Massarweh, L.

DOI
10.1007/s00190-025-01945-7

Publication date
2025

Document Version
Final published version

Published in
Journal of Geodesy

Journal of Geodesy (2025) 99:28
https://doi.org/10.1007/s00190-025-01945-7

ORIGINAL ARTICLE

Theory for the ambiguity function method: probability model and global solution

P. J. G. Teunissen[1,2,3] · **L. Massarweh**[1]

Received: 11 October 2024 / Accepted: 10 February 2025

Abstract
In this contribution, we introduce some new theory for the classical GNSS ambiguity function (AF) method. We provide the probability model by means of which the AF-estimator becomes a maximum likelihood estimator, and we provide a globally convergent algorithm for computing the AF-estimate. The algorithm is constructed from combining the branch-and-bound principle, with a special convex relaxation of the multimodal ambiguity function, to which the projected-gradient-descent method is applied to obtain the required bounds. We also provide a systematic comparison between the AF-principle and that of integer least-squares (ILS). From this comparison, the conclusion is reached that the two principles are fundamentally different, although there are identified circumstances under which one can expect AF- and ILS-solutions to behave similarly.

Keywords GNSS · Ambiguity function (AF) method · Maximum likelihood · Integer least-squares (ILS) · Branch-and-bound (BB) · Convex relaxation · Projected-gradient-descent (PGD)

1 Introduction

The ambiguity function (AF) method is one of the earliest methods for estimating baselines from integer ambiguous GNSS carrier-phase data. The method was introduced and popularized by Counselman and Gourevitch (1981) and Remondi (1984, 1991), while its original idea of eliminating dependence on the '2π' ambiguities goes back to (Rogers et al. 1978). In fact, it was this property of invariance that contributed to the initial popularity of the method. It promised the capability of determining precise baselines, without the explicit need of having to resolve the values of integer carrier-phase ambiguities.

Although the AF-method is one of the oldest methods, its statistical and numerical evolution did not keep pace with the theoretical developments of other methods of mixed-integer inference (Teunissen 2003b, 2017; Hartman 2021). Currently, we have different classes of mixed-integer estimators, with identified optimal estimators within each class, together with numerically efficient computational algorithms (Teunissen 1995). For instance, the best integer-equivariant (BIE) estimator (Teunissen 2003a) is minimum-mean-squared-error optimal in the largest class, while the integer least-squares (ILS) estimator (Teunissen 1999) is best in maximizing the probability of correct integer estimation within the smaller integer-class.

As the maturity of the AF-method is not on par with the current methods of mixed-integer estimation, it is the goal of the present contribution to help fill in some of the theoretical gaps. In doing so, the two main innovations of this contribution are: (1) the provision of a probability model by means of which the AF-solution is given a statistical basis and (2) the provision of a global optimizer of the AF-likelihood function, having finite termination with a guaranteed epsilon tolerance.

This contribution is organized as follows: In Sect. 2, we provide a brief review of the ambiguity function method, together with examples of its use. Then, in Sect. 3 we draw attention to the possible nonuniqueness of the AF-solution. This is new, as this problem has not been addressed before in the AF-method's literature. We prove under which conditions the AF-solution is nonunique and how one can verify whether or not this nonuniqueness is problematic for the application

✉ 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] Department of Infrastructure Engineering, The University of Melbourne, Melbourne, Australia

[3] GNSS Research Centre, Curtin University of Technology, Perth, Australia

Published online: 04 April 2025

under consideration. In Sect. 4, we introduce our probability model for the AF-method. It shows what distributional assumptions need to be made in order for the AF-estimator to become a maximum likelihood estimator. It therefore provides, for the first time, a statistical basis for the AF-method of GNSS baseline determination. In doing so, we also generalize the classical expression of the AF-estimator by showing how a varying precision of the carrier-phase observables can be incorporated into the estimation scheme.

In order to describe the characteristics of the ambiguity objective function qualitatively, we provide its multivariate gradient and Hessian in Sect. 5. Their analysis shows that the ambiguity objective function is severely multimodal and that, in the absence of very accurate baseline initializations, iterative gradient descent methods will not be able to locate the sought for maximizer of the likelihood function. A global optimization method is therefore asked for, which we introduce in Sect. 6. The proposed approach is based on a branch-and-bound algorithm, which exploits a specific convex relaxation of the ambiguity function. The bounds are computed using a projected-gradient-descent (PGD) iterative method, which requires the convex lower bounding function to be continuously differentiable. Each step of the algorithm is described here, demonstrating how global optimality is guaranteed in a finite time within a user-selected epsilon tolerance.

In Sect. 7, we compare the AF estimation principle with that of integer least-squares (ILS). Although we exemplify the various marked differences between the two principles, we also show under which identified circumstances one can expect AF- and ILS-solutions to be close. We do this by making use of the primal-dual equivalence of mixed ILS theory as introduced in (Teunissen and Massarweh 2024). The presented theory is supported by means of several examples in which the workings and performance of the AF-method are numerically and graphically illustrated. Finally, Sect. 8 contains the summary and conclusions.

The following notation is used: $\mathsf{E}(.)$ and $\mathsf{D}(.)$ stand for the expectation and dispersion operators, respectively, and $\mathcal{N}_p(\mu, Q)$ denotes a p-dimensional, normally distributed random vector, with mean (expectation) μ and variance matrix (dispersion) Q. $\mathbb{R}^p$ and $\mathbb{Z}^p$ denote the p-dimensional spaces of real- and integer numbers, respectively. The Q-weighted squared norm is denoted as $||.||^2_Q = (.)^T Q^{-1}(.)$, and $\lceil x \rfloor$ denotes the rounding of x to the nearest integer. If applied to a vector, the rounding is understood to apply to each of its coordinates. $\cup$ and $\cap$ denote the union and intersection operators, and the vectorial inequality $\preceq$ denotes the all componentwise inequality $\leq$. The gradient of a function $F(b)$ is denoted as $\partial_b F(b)$, and the central Chi-square distribution with p degrees of freedom is denoted as $\chi^2(p, 0)$, with $\chi^2_\delta(p, 0)$ being its δ-percentage critical value.

2 The AF-method: a brief review

The single-baseline, k-epochs, f-frequencies, and s-satellites GNSS ambiguity function (AF) is generally defined as (Mader 1992; Lachapelle et al. 1992; Leick et al. 2015):

$$\mathrm{AF}(b) = \sum_{t=1}^{k} \sum_{j=1}^{f} \sum_{i=1}^{s} \cos\left[\tfrac{2\pi}{\lambda_j}(\phi^{ri}_{12,j}(t) - \rho^{ri}_{12}(t, b))\right] \qquad (1)$$

in which λ_j is the wavelength of the jth frequency, $\phi^{ri}_{12,j}(t) = [\phi^i_{2,j}(t) - \phi^i_{1,j}(t)] - [\phi^r_{2,j}(t) - \phi^r_{1,j}(t)]$ the double-differenced (DD) phase-observable, in units of range, on frequency j at epoch t of receivers 1, 2 and satellites r, i, and $\rho^{ri}_{12}(t, b)$ is its corresponding DD range, which depends on b, the unknown baseline vector between receivers 1 and 2.

Note that $\mathrm{AF}(b)$ is *invariant* for any perturbations of $\phi^{ri}_{12,j}(t)$ that are integer multiples of the wavelength λ_j. Hence, it is invariant for integer cycle slips in the phase data, as well as for the presence of the DD integer ambiguities $a^{ri}_{12,j} \in \mathbb{Z}$ in the observation equations $\phi^{ri}_{12,j}(t) = \lambda_j a^{ri}_{12,j} + \rho^{ri}_{12}(t, b) + \epsilon^{ri}_{12,j}(t)$. This invariance has in fact been the overarching motivation for introducing the AF-concept (Counselman and Gourevitch 1981; Remondi 1984; Mader 1992; Hofmann-Wellenhof et al. 2008; Leick et al. 2015). It promises namely of being able to resolve the unknown baseline b, without the explicit need of having to resolve the values of the integer ambiguities $a^{ri}_{12,j}$. As $\mathrm{AF}(b)$ reaches its maximum value when all DD phase errors $\epsilon^{ri}_{12,j}(t)$ are identically zero, the chosen AF-approach for resolving the unknown baseline is to *aim* for a solution that satisfies $\breve{b} = \arg\max_b \mathrm{AF}(b)$. The usual approach for doing so is by direct evaluation of the ambiguity function on the vertices of a three-dimensional rectangular grid, centred at an approximate baseline solution. The vertex that provides the largest function value is then selected as the solution sought (Rogers et al. 1978; Remondi 1984; Hofmann-Wellenhof et al. 2008). It will be clear that the numerical and statistical efficacy of such approach depends on the chosen grid spacing, grid size, and grid location.

Although the AF-principle, of working with an integer-ambiguity invariant objective function that is maximized when the errors are zero, forms the basis of all publications in the GNSS AF-literature, it is important to realize that different authors applied the principle to different objective functions. As a consequence, different baseline results will be obtained even when these authors would be using the same original data. Some authors work on single-differenced data, while others apply the principle to double-differenced data, and some work directly with the cosine function, as in (1), while other authors work with the complex phasor function, being the analytical representation of a cosine func-

tion. Remondi (1984, 1991), for instance, works with phasor norms, as a result of which his ambiguity objective function contains, in contrast to (1), both cosine and sine functions. The same phasor norms are also used by Han and Rizos (1996), but they use it on double-differenced data, instead of on single-differenced data as is done in (Remondi 1984, 1991; Remondi and Hilla 1993; Hofmann-Wellenhof et al. 2008).

In this contribution, we will use, as in (1), a 'sum of cosines' as our basis for constructing the ambiguity-invariant objective function. This is consistent with the original formulation of Rogers et al. (1978), but more importantly, it will allow us to formulate a probabilistic model for the ambiguity function method.

3 On the nonuniqueness of the AF-solution

As mentioned earlier, the attractiveness of the AF-method is the integer-ambiguity invariance of its objective function. This invariance, however, holds for any integer perturbations of its argument and thus also for those that may be generated by changes in the baseline. If such case happens, one cannot expect the baseline solution of the AF-method to be unique.

To study the possible nonuniqueness of the AF-solution, we first introduce a useful compact notation for the ambiguity function. Let $\epsilon_\phi = [\epsilon_{\phi_1}, \ldots, \epsilon_{\phi_m}]^T$ be an m-vector with its entries expressed in cycles, and let $e_m = [1, \ldots, 1]^T$ be the m-vector of ones. Then, we introduce for the AF-function the compact notation

$$\mathrm{AF} = e_m^T \cos[2\pi(\epsilon_\phi)] := \sum_{i=1}^{m} \cos[2\pi(\epsilon_{\phi_i})] \tag{2}$$

Thus, $\cos[2\pi(\epsilon_\phi)]$ is the vector that consists of the componentwise cosine values of $2\pi\epsilon_{\phi_i}$. As the general system of GNSS, carrier-phase observation equations can be written in vector–matrix cycle-form as:

$$\phi = A_\phi a + B_\phi b + \epsilon_\phi,\ a \in \mathbb{Z}^n, b \in \mathbb{R}^p, A_\phi \in \mathbb{Z}^{m\times n} \tag{3}$$

the to-be-maximized objective function of the AF-method follows upon substitution of $\epsilon_\phi = \phi - A_\phi a - B_\phi b$ into (2) as

$$\mathrm{AF}(b) = e_m^T \cos[2\pi(\phi - B_\phi b)] \tag{4}$$

Note that this formulation generalizes that of (1) in the sense that ϕ need now not be restricted to a DD-form and that b need not be restricted to a single baseline. As (4) applies to any carrier-phase system of the form (3), it holds in principle for undifferenced data and networks as well.

Also note, due to the property of the cosine function, that the ambiguity-part $A_\phi a \in \mathbb{Z}^m$ of system (3) disappeared from the objective function $\mathrm{AF}(b)$. This is also the principal attractiveness of the method as it implies that no explicit 'integer-ambiguity resolution' is required when maximizing (4).

We now show, however, what this 'invariance' does to the uniqueness of the AF-maximizer.

Theorem 1 (Nonuniqueness of AF-solution)*: Let $m\times(n+p)$ design matrix $[A_\phi, B_\phi]$ (cf. 3) be of full column rank and let $Z = [Z_1, Z_2]$, $Z_1 \in \mathbb{Z}^{m\times(m-p)}$, $Z_2 \in \mathbb{Z}^{m\times p}$, be an admissible ambiguity transformation (i.e. Z and Z^{-1} have integer entries), satisfying $B_\phi^{\perp T}[Z_1, Z_2] = [L, 0]$, where $B_\phi^\perp$ is a basis matrix of the null space of B_ϕ^T. Then, the ambiguity function (4) satisfies*

$$\mathrm{AF}(b + B_\phi^+ Z_2 \tilde{z}_2) = \mathrm{AF}(b),\ \forall \tilde{z}_2 \in \mathbb{Z}^p \tag{5}$$

in which B_ϕ^+ is a left-inverse of B_ϕ (i.e. $B_\phi^+ B_\phi = I_p$). ■

Proof From $B_\phi^{\perp T}[Z_1, Z_2] = [L, 0]$, it follows that Z_2 is an integer basis matrix of the range space of B_ϕ, i.e. $Z_2 = B_\phi X$ for some invertible $p \times p$ matrix $X = B_\phi^+ Z_2$. Therefore, using the projector property $B_\phi B_\phi^+ Z_2 = Z_2$,

$$\begin{aligned}
\mathrm{AF}(b) &= e_m^T \cos[2\pi(\phi - B_\phi b)] \\
&= e_m^T \cos[2\pi(\phi - B_\phi b - Z_2\tilde{z}_2)] \\
&= e_m^T \cos[2\pi(\phi - B_\phi(b + X\tilde{z}_2))] \\
&= e_m^T \cos[2\pi(\phi - B_\phi(b + B_\phi^+ Z_2\tilde{z}_2))] \\
&= \mathrm{AF}(b + B_\phi^+ Z_2\tilde{z}_2)
\end{aligned} \tag{6}$$

□

The important consequence of this result is that the AF-solution may not be unique even if the design matrix $[A_\phi, B_\phi]$ (cf. 3) is of full column rank. Thus, even if the phase-only system of observation equations produces a unique float solution, with corresponding integer least-squares (ILS) solution, the solution produced by the AF-method may not be unique. This possible lack of baseline uniqueness is here identified for the first time as it is not part of the deliberations in the classical AF-literature (Counselman and Gourevitch 1981; Remondi 1984; Mader 1992; Hofmann-Wellenhof et al. 2008; Leick et al. 2015).

The condition under which the above lack of uniqueness occurs is when an admissible integer matrix $Z = [Z_1, Z_2]$ can be constructed such that $B^{\perp T}[Z_1, Z_2] = [L, 0]$. This is always possible when the entries of matrix $B^\perp$ are *rational*, see Theorem 2 in (Teunissen and Khodabandeh 2022), which happens, for instance, with the geometry-free GNSS model. We hereby note, however, that even if the actual entries of $B^\perp$ are not rational, in the context of numerical computing

they could be, which then still introduces numerically the above-mentioned lack of uniqueness.

Whether or not the above-identified lack of uniqueness is problematic from a practical point of view may depend on how large the smallest perturbation $B^+ Z_2 \tilde{z}_2$ will be. If the smallest such perturbation is sufficiently large, then a local maximizer $\check{b}$ of AF(b) could still be acceptable, since the next nearest maximizer will then be far away. The smallest distance between the local maximizers is given by $D_{\min} = \min_{z \in \mathbb{Z}^p \setminus \{0\}} ||B_\phi^+ Z_2 z||_Q$ in which Q is a user-chosen positive-definite matrix, e.g. $Q = (B_\phi^T B_\phi)^{-1}$. Note that this minimization can be computed efficiently with the LAMBDA method (Teunissen 1995; Massarweh et al. 2025).

4 A probability model for the ambiguity function method

In this section, we will develop our probability model for the ambiguity function method. The starting idea is to find a probability density function (PDF) that has the solution of the ambiguity function method, $\check{b} = \arg\max_b e_m^T \cos[2\pi(\phi - B_\phi b)]$ (cf. 4), as its *maximum likelihood* estimate. In doing so, we ignore for the moment that the solution may be nonunique. We have

$$
\begin{aligned}
\check{b} &= \arg\max_b e_m^T \cos[2\pi(\phi - B_\phi b)] \\
&\overset{(a)}{=} \arg\max_b \exp\{e_m^T \cos[2\pi(\phi - B_\phi b)]\} \\
&\overset{(b)}{=} \arg\max_b \exp\{e_m^T \cos[z - \bar{z}]\},\ z = 2\pi\phi,\ \bar{z} = 2\pi B_\phi b \\
&\overset{(c)}{=} \arg\max_b \frac{\exp\{e_m^T \cos[z-\bar{z}]\}}{\int_\Omega \exp\{e_m^T \cos[z-\bar{z}]\} dz},\ \Omega = [-\pi, +\pi]^m \\
&\overset{(d)}{=} \arg\max_b \prod_{i=1}^m \frac{\exp\{\cos[z_i-\bar{z}_i]\}}{\int_{-\pi}^{+\pi} \exp\{\cos[z_i-\bar{z}_i]\} dz_i},\ z_i = c_i^T z,\ \bar{z}_i = c_i^T \bar{z} \\
&\overset{(e)}{=} \arg\max_b \prod_{i=1}^m f(z_i|\bar{z}_i, 1) \\
&\text{with } f(x|\mu,\kappa) = \frac{\exp\{\kappa \cos[x-\mu]\}}{\int_{-\pi}^{+\pi} \exp\{\kappa \cos[x-\mu]\} dx}
\end{aligned}
\tag{7}
$$

This result can be explained as follows. By taking the exponential exp in step (a), we obtain a nonnegative objective function which has the same maximizer as the original objective function. In step (b), we simplify the argument by setting $z = 2\pi\phi$ and $\bar{z} = 2\pi B_\phi b$. In step (c), we normalize the objective function such that it now can be interpreted as being a PDF. As function of z, it integrates to 1 over Ω, and as function of b, it still has $\check{b}$ as its maximizer. It is thus the multivariate PDF of the random vector z, having $\bar{z} = 2\pi B_\phi b$ as its parameter vector. As the PDF is symmetric about $\bar{z}$, $\bar{z} = 2\pi B_\phi b$ is also the mean of z. In step (d), we applied the property that the exponential of a sum can be written as a product of exponentials. As a result, the multivariate PDF is written as an m-product of univariate PDFs. Here, c_i denotes the canonical unit vector having its only nonzero entry of 1 as its ith entry; Step (e) follows by recognizing that all m PDFs are the same, except for their means $\bar{z}_i$, $i = 1, \ldots, m$. Here we also recognize that the PDFs $f(z_i|\bar{z}_i, 1)$, $i = 1, \ldots, m$, are all special cases of the well-known *circular normal* distribution $\mathcal{CN}(\mu, \kappa)$, having $f(x|\mu, \kappa)$ as its PDF (Gumbel et al. 1953). The circular normal distribution $\mathcal{CN}(\mu, \kappa)$, with mean μ and *concentration* parameter κ, is also known as the *von Mises* distribution. Its denominator is given as $2\pi I_0(\kappa) = \int_{-\pi}^{+\pi} \exp\{\kappa \cos[x - \mu]\} dx$, where $I_0(\kappa)$ is the modified Bessel function of the first kind of order 0.

The above has shown that the solution $\check{b}$ of the ambiguity function method can now be interpreted as being the maximum of the likelihood function of $2\pi\phi$, if the m random variables $2\pi\phi_i$, $i = 1, \ldots, m$, are independent and distributed as $2\pi\phi_i \sim \mathcal{CN}(2\pi c_i^T B_\phi b, 1)$. Noting that this probabilistic result is obtained with circular normal distributions having unity concentration parameters, our above derivation now also shows how to introduce a weighting scheme into the ambiguity function method. If we replace the sum of cosines, $e_m^T \cos[2\pi(\phi - B_\phi b)]$, by the weighted sum $w^T \cos[2\pi(\phi - B_\phi b)]$, $w = [w_1, \ldots, w_m]^T$, a similar derivation as above shows, with $z_i = 2\pi c_i^T \phi$ and $\bar{z}_i = 2\pi c_i^T B_\phi b$, that

$$
\arg\max_b w^T \cos[2\pi(\phi - B_\phi b)] = \arg\max_b \prod_{i=1}^m f(z_i|\bar{z}_i, w_i)
\tag{8}
$$

This shows how the classical objective function (4) needs to be changed in order to incorporate a varying precision of the GNSS carrier-phase measurements, thus allowing to include, for instance, frequency and/or elevation dependency in the precision description.

We are now in the position to summarize the above findings in a theorem. In order to do so, we will include the pseudorange data, and work, instead of with the phase-only system (3), with the extended partitioned system

$$
\begin{bmatrix} p \\ \phi \end{bmatrix} = \begin{bmatrix} 0 & B_p \\ A_\phi & B_\phi \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} \epsilon_p \\ \epsilon_\phi \end{bmatrix},\ a \in \mathbb{Z}^n,\ b \in \mathbb{R}^p,\ A_\phi \in \mathbb{Z}^{m \times n}
\tag{9}
$$

in which the pseudoranges in p are expressed in units of range, while the carrier-phases in ϕ are still expressed in cycles.

By now assuming the pseudorange observables to be normally distributed and independent of the circular normal distributed carrier-phase observables, we obtain a complete probabilistic model for the ambiguity function method and one that also eliminates the lack of uniqueness discussed in the previous section. We have the following result.

Theorem 2 (AF-ML estimator) *Let* $p \sim \mathcal{N}_m(B_p b, Q_{pp})$, *with* $\operatorname{rank} B_p = p$, $Q_{pp} > 0$, *be independent of* $2\pi\phi \sim \prod_{i=1}^{m} \mathcal{CN}(2\pi c_i^T B_\phi b, w_i)$. *Then, the* likelihood function *of* b *is given as:*

$$L(b) = \frac{\exp\{-\frac{1}{2}||p - B_p b||^2_{Q_{pp}}\}}{|2\pi Q_{pp}|^{1/2}} \frac{\exp\{w^T \cos[2\pi(\phi - B_\phi b)]\}}{\prod_{i=1}^{m} 2\pi I_0(w_i)} \tag{10}$$

and its maximizer as

$$\check{b}_{\mathrm{AF}} = \arg\max_{b\in\mathbb{R}^p} \left(-\tfrac{1}{2}||p - B_p b||^2_{Q_{pp}} + w^T \cos[2\pi(\phi - B_\phi b)]\right) \tag{11}$$

□

This result shows how the solution of the ambiguity function method can be interpreted as a maximum likelihood (ML) estimator, through which it also automatically inherits all the known properties of ML-estimators. Hence, by means of the above-identified probabilistic model, one can now apply known likelihood estimation and testing results.

Note, by making use of the trigonometric identity $\cos 2x = 1 - 2\sin^2 x$, that we may write the maximization problem (11) also as a minimization problem,

$$\check{b}_{\mathrm{AF}} = \arg\min_{b\in\mathbb{R}^p} \left(||p - B_p b||^2_{Q_{pp}} + 4w^T \sin^2[\pi(\phi - B_\phi b)]\right) \tag{12}$$

If we assume B_p to be of full column rank, we may use the orthogonal decomposition $||p - B_p b||^2_{Q_{pp}} = ||P^\perp_{B_p} p||^2_{Q_{pp}} + ||\hat{b} - b||^2_{Q_{\hat{b}\hat{b}}}$, with $P^\perp_{B_p} = I_m - B_p B_p^+$, $\hat{b} = B_p^+ p$, $B_p^+ = Q_{\hat{b}\hat{b}} B_p^T Q_{pp}^{-1}$, $Q_{\hat{b}\hat{b}} = (B_p^T Q_{pp}^{-1} B_p)^{-1}$ and write (12) also as

$$\check{b}_{\mathrm{AF}} = \arg\min_{b\in\mathbb{R}^p} \left(||\hat{b} - b||^2_{Q_{\hat{b}\hat{b}}} + 4w^T \sin^2[\pi(\phi - B_\phi b)]\right) \tag{13}$$

This formulation shows how the quadratic term on the right-hand side, and therefore the inclusion of pseudorange data through $\hat{b}$, acts as a *regularizer* on the phase-based part of the objective function. It is through this pseudorange-based regularization that the earlier mentioned nonuniqueness of the classical ambiguity function (4) is eliminated. Note in this regard that although a full rank of (9) guarantees uniqueness of the float LS-solution, that this not necessarily implies uniqueness of $\check{b}_{\mathrm{AF}}$. This is only the case if also B_p is of full rank. With single-differenced data for instance, B_p will be rank defect as the phase clock cannot be determined from code data alone. In that case, the integerness of the phase-clock coefficients implies that the phase clock can only be AF-determined up to an integer multiple of the wavelength. To be able to work with (11) (or 13), we still need to show how the weighting vector w can be chosen. For this, we rely on the properties of the circular normal distribution and the fact that GNSS carrier-phase measurements are ultra-precise. The circular normal distribution $\mathcal{CN}(\mu, \kappa)$ is a symmetric unimodular distribution having its mode at μ. The parameter κ drives its peakedness or concentration. For $\kappa \to 0$, the circular normal distribution converges to the uniform distribution, while for $\kappa \to \infty$, it converges to the point distribution $\delta(x - \mu)$. As for large, but finite values of κ, the normal distribution $\mathcal{N}(\mu, \sigma^2 = \frac{1}{\kappa})$ provides a good approximation to $\mathcal{CN}(\mu, \kappa)$ (Gumbel et al. 1953), we will use the precision of the carrier-phase observables to set the entry values of the weight vector w. For example, if σ_i denotes the phase standard deviation when expressed in units of range, i.e. $\sigma_i = \sigma_{\lambda_i \phi_i}$ with λ_i the wavelength, then the variance of $2\pi\phi_i$ can be taken as the reciprocal value of the corresponding concentration parameter, i.e. $\frac{4\pi^2\sigma_i^2}{\lambda_i^2} = \frac{1}{w_i}$.

Example 1 (*Geometry-free model*) Consider the single-epoch, multi-frequency, double-differenced (DD) geometry-free model

$$\mathsf{E}\begin{bmatrix} p \\ \phi \end{bmatrix} = \begin{bmatrix} 0 & e_f \\ I_f & \Lambda_f^{-1} e_f \end{bmatrix} \begin{bmatrix} a \\ \rho \end{bmatrix} \tag{14}$$

with $p \in \mathbb{R}^f$ the DD pseudorange vector expressed in units of range, $\phi \in \mathbb{R}^f$ the DD carrier-phase vector expressed in cycles, $e_f = (1, \ldots, 1)^T$ the f-vector of ones, $\Lambda_f = \operatorname{diag}(\lambda_1, \ldots, \lambda_f)$ the diagonal matrix of f wavelengths, $a \in \mathbb{Z}^f$ the DD integer-ambiguity vector and $\rho \in \mathbb{R}$ the scalar DD range. Then, with $\mathsf{D}(p) = \sigma_p^2 I_f$, the two parts of the objective function of (13) work out as:

$$\begin{aligned} ||\hat{b} - b||^2_{Q_{\hat{b}\hat{b}}} &:= \left(\tfrac{\hat{\rho} - \rho}{\sigma_p/\sqrt{f}}\right)^2, \ \hat{\rho} = \tfrac{1}{f}\textstyle\sum_{i=1}^{f} p_i \\ 4w^T \sin^2[\pi(\phi - B_\phi b)] &:= 4\textstyle\sum_{i=1}^{f} w_i \sin^2[\tfrac{\pi(\tilde{\phi}_i - \rho)}{\lambda_i}] \\ w_i &:= \tfrac{\lambda_i^2}{4\pi^2\sigma^2_{\tilde{\phi}_i}} \end{aligned} \tag{15}$$

with $\sigma^2_{\tilde{\phi}_i}$ being the variance of $\tilde{\phi}_i = \lambda_i \phi_i$. □

5 Multimodal ambiguity function and its optimality domain

In this section, we describe and illustrate some of the defining characteristics of the ambiguity function. In order to do so, we work from now on, for convenience sake, instead of with the objective function (11), with that of (12), i.e.

$$F(b) = ||p - B_p b||^2_{Q_{pp}} + 4w^T \sin^2[\pi(\phi - B_\phi b)] \tag{16}$$

Hence, since $F(b) = 2w^T e_m - 2\mathrm{AF}(b)$, the maximum likelihood solution of the AF-method is then computed, not as a maximizer, but as the minimizer of $F(b)$.

5.1 Minimization by iterative gradient descent

As $F(b)$ is a smooth function (in fact, it is a C^∞ function, having continuous derivatives of all orders), one may think of applying methods of *iterative gradient descent* to obtain its minimum. Such methods adhere to the following scheme (Teunissen 1990; Nocedal and Wright 2006):

$$b_{k+1} = b_k - t_k Q(b_k)\partial_b F(b_k),\ k = 0, 1, \ldots \tag{17}$$

in which $t_k > 0$ is a to-be-chosen step size and $Q(b_k) > 0$ is a to-be-chosen positive-definite matrix. The iteration is started by an externally provided initial approximation b_0 of the minimizer.

Through the choice of $Q(b_k)$, one can choose the direction of descent, and through the choice of t_k, one can enforce that $F(b_{k+1}) \leq F(b_k)$. For computing the stepsize t_k in each iteration, different line-search strategies exist, from simple to advanced (Nesterov 2018). One of the simplest starts with $t_k = 1$, followed by halving it, $t_k \leftarrow t_k/2$, until $F(b^k - t_k Q(b_k)\partial_b F(b^k)) < F(b^k)$. The simplest choice for $Q(b_k)$ would be to choose it as an identity matrix, $Q(b_k) = I_k$. As the resulting direction $-\partial_b F(b_k)$ points in the direction of steepest descent of $F(b)$ at b_k, this method is known as the *steepest descent method*. It has a local linear rate of convergence, but the iterations have the potential to zigzag when the contours of $F(b)$ are elongated at the minimizer. This is avoided when $Q(b_k)$ is chosen as the inverse-Hessian of $F(b)$, $Q(b_k) = [\partial^2_{bb} F(b_k)]^{-1}$. This gives *Newton's method*, which is known to have a local quadratic convergence. Contrary to the steepest descent method, Newton's method does not need a line-search strategy to enforce local convergence. That is, when the Hessian is positive-definite, the method has a guaranteed convergence for points sufficiently close to the solution. This is a consequence of the method being based on a linear approximation of the vanishing gradient of $F(b)$ at the minimizer: $0 = \partial_b F(\breve{b}_{\mathrm{AF}}) \approx \partial_b F(b_k) + \partial^2_{bb} F(b_k)(\breve{b}_{\mathrm{AF}} - b_k)$.

To apply the above iterative descent methods to (16) and verify whether or not a minimizer is obtained, the gradient and Hessian of $F(b)$ are needed. They are given as follows.

Lemma 1 (Ambiguity function gradient and Hessian)*: The gradient and Hessian of the objective function* $F(b) = ||p - B_p b||^2_{Q_{pp}} + 4w^T \sin^2[\pi(\phi - B_\phi b)]$ *are given as*

$$\begin{aligned} \partial_b F(b) &= -2\left(B_p^T Q_{pp}^{-1}(p - B_p b) + B_\phi^T \varphi\right) \\ \partial^2_{bb} F(b) &= 2\left(B_p^T Q_{pp}^{-1} B_p + B_\phi^T D B_\phi\right), \end{aligned} \tag{18}$$

with

$$\begin{aligned} \varphi &= \sum_{i=1}^{m} (2\pi w_i \sin[2\pi c_i^T[\phi - B_\phi b)])c_i \\ D &= \mathrm{diag}(d_1(b), \ldots, d_m(b)) \\ d_i(b) &= 4\pi^2 w_i \cos[2\pi c_i^T(\phi - B_\phi b)] \end{aligned}$$

and where c_i *is the* i*th canonical unit vector.* ■

Note, due to the presence of the diagonal matrix D, that the Hessian matrix (cf. 18) of the ambiguity function is not necessarily positive-definite. Hence, it may not be invertible, or, when it is, it may not provide a descent direction. To avoid this from happening, one may think of regularizing the Hessian as $\partial^2_{bb} F(b) - 2B_\phi^T D B_\phi$, when $\partial^2_{bb} F(b)$ fails to be positive-definite. The so-obtained regularized iteration can then again be interpreted as an iterative descent method, but now corresponding with the choice $Q(b_k) = [2(B_p^T Q_{pp}^{-1} B_p)]^{-1}$. Note that, with $t_k = 1$, this iteration boils down to

$$b_{k+1} = (B_p^T Q_{pp}^{-1} B_p)^{-1}[B_p^T Q_{pp}^{-1} p + B_\phi^T \varphi] \tag{19}$$

which can be seen to be the *fixed point iteration* of the system $\partial_b F(b) = 0$.

5.2 On the multimodality of the ambiguity function

Although the above descent methods converge to a minimum of $F(b)$, it depends on the initial approximation b_0 whether the minimum is a *local* or a *global* minimum of $F(b)$. One can have some confidence in having computed the global minimum, if b_0 would already be close enough to the global minimizer $\breve{b}_{\mathrm{AF}}$. To see whether or not one can reasonably expect this to be the case, we take the ambiguity function of the single-frequency geometry-free model as an example. For the single-frequency case, the geometry-free ambiguity function follows from (15) as:

$$F(\rho) = \tfrac{(p-\rho)^2}{\sigma_p^2} + 4w \sin^2\left(\tfrac{\pi}{\lambda}[\tilde{\phi} - \rho]\right),\ w = \tfrac{\lambda^2}{4\pi^2\sigma^2_{\tilde{\phi}}} \tag{20}$$

thus having first and second derivatives

$$\begin{aligned} d_\rho F(\rho) &= \tfrac{2}{\sigma^2_{\tilde{\phi}}}\left[\epsilon(\rho - p) - \tfrac{\lambda}{2\pi}\sin\left(\tfrac{2\pi}{\lambda}[\tilde{\phi} - \rho]\right)\right] \\ d^2_{\rho\rho} F(\rho) &= \tfrac{2}{\sigma^2_{\tilde{\phi}}}\left[\epsilon + \cos\left(\tfrac{2\pi}{\lambda}[\tilde{\phi} - \rho]\right)\right] \end{aligned} \tag{21}$$

with ϵ being the phase-code variance ratio $\epsilon = \sigma^2_{\tilde{\phi}}/\sigma^2_p$. This shows that all those points where the straight line $y = \epsilon(\rho - p)$ intersects the sine-function $y = \frac{\lambda}{2\pi}\sin(\frac{2\pi}{\lambda}(\tilde{\phi} - \rho))$, are points for which $d_\rho F(\rho) = 0$. These are therefore the points where the local minima and maxima of the function $F(\rho)$ are located. Note that their number increases, when ϵ gets

smaller, i.e. when the descending straight line $y = \epsilon(\rho - p)$ gets less tilted.

From the second derivative, we learn that of these points, only those are minima for which $\cos(\frac{2\pi}{\lambda}(\tilde{\phi} - \rho)) > -\epsilon$ holds. This shows, since $\cos x \geq -1$ for all $x \in \mathbb{R}$, that $d^2_{\rho\rho}F(\rho) > 0$ for all $\rho \in \mathbb{R}$, if $\epsilon > 1$. Hence, this is the condition for which $F(\rho)$ is convex. Thus, if $\epsilon > 1$, then $F(\rho)$ has only a single minimum. As ϵ equals the phase-code variance ratio, this would require the precision of the pseudorange (code) observables to be better than that of the phase observables, which clearly is not the case with GNSS.

In fact, in case of GNSS it is the reciprocal value of ϵ that is large, i.e. the phase-code variance ratio is very small, $\epsilon \approx 10^{-4}$. This implies that the almost horizontal line $y = \epsilon(\rho - p)$ will have a large number of intersections with $y = \frac{\lambda}{2\pi}\sin(\frac{2\pi}{\lambda}(\tilde{\phi} - \rho))$. Hence, in the typical GNSS case, there will be a large number of minima and maxima from which one would then need to select the global minimum. The ρ-values for the minima will have to satisfy $\cos(\frac{2\pi}{\lambda}(\tilde{\phi} - \rho)) > -\epsilon \approx 0$. If we define the small value δ_ϵ as satisfying $\epsilon + \cos(\frac{1}{2}\pi + 2\pi\delta_\epsilon) = 0$, then $d^2_{\rho\rho}F(\rho) > 0$ for all DD range values satisfying

$$\rho/\lambda \in \phi + z + \left(-\tfrac{1}{4} - \delta_\epsilon, +\tfrac{1}{4} + \delta_\epsilon\right), \forall z \in \mathbb{Z} \tag{22}$$

Thus, of all the solutions satisfying $d_\rho F(\rho) = 0$, those ρ/λ that lie in one of the integer translated intervals (22) will be local minima and thus candidates for a global minimum.

To see how this set of 'integer translated intervals' generalizes to the multidimensional case, we need to study the region for which the $p \times p$ Hessian matrix $\partial^2_{bb}F(b)$ is positive-definite. We have the following result.

Lemma 2 (On the ambiguity function's convexity) *Let* $\Omega_{\text{PD}} = \{b \in \mathbb{R}^p | \partial^2_{bb}F(b) > 0\}$ *be the region in which the ambiguity function is convex and define the convex polytope*

$$\Omega^\circ_{\text{PD}}(z) = \{b \in \mathbb{R}^p |\ l(z) \preceq B_\phi b \preceq r(z)\}, z \in \mathbb{Z}^n \tag{23}$$

with $l(z) = (\phi - \delta_\epsilon - \frac{1}{4}e_m) + z$, $r(z) = (\phi + \delta_\epsilon + \frac{1}{4}e_m) + z$. *Then,*

$$\Omega^\circ_{\text{PD}} = \underset{z\in\mathbb{Z}^m}{\cup} \Omega^\circ_{\text{PD}}(z) \subset \Omega_{\text{PD}} \tag{24}$$

for $\delta_\epsilon = 0$, *while if* $B_\phi = \Lambda^{-1}B_p$, *with* Λ *the diagonal wavelength matrix, and* Q_{pp} *is diagonal, relation* (24) *even holds with the entries of* $\delta_\epsilon = (\delta_{\epsilon_1}, \ldots, \delta_{\epsilon_m})^T$ *satisfying* $\epsilon_i + \cos(\frac{1}{2}\pi + 2\pi\delta_{\epsilon_i}) = 0$, $\epsilon_i = \sigma^2_{\hat{\phi}_i}/\sigma^2_{p_i} = \lambda_i^2/(4\pi^2 w_i \sigma^2_{p_i})$, $i = 1, \ldots, m$. *If then also* B_p *is invertible, we have* $\Omega^\circ_{\text{PD}} = \Omega_{\text{PD}}$. ■

Proof Relation (24), for $\delta_\epsilon = 0$, follows from the fact that $D > 0$ (cf. 18) implies $2(B_p^T Q_{pp}^{-1} B_p + B_\phi^T D B_\phi) = \partial^2_{bb}F(b) > 0$. Similarly, if $B_\phi = \Lambda^{-1}B_p$, then $Q_{pp}^{-1} + \Lambda^{-1}D\Lambda^{-1} > 0$ implies $2(B_p^T Q_{pp}^{-1} B_p + B_\phi^T D B_\phi) = \partial^2_{bb}F(b) > 0$, which gives (24) for the given δ_ϵ when Q_{pp} is diagonal. If also B_p is invertible, then $Q_{pp}^{-1} + \Lambda^{-1}D\Lambda^{-1} > 0$ is equivalent to $2(B_p^T Q_{pp}^{-1} B_p + B_\phi^T D B_\phi) = \partial^2_{bb}F(b) > 0$, which gives the equality $\Omega^\circ_{\text{PD}} = \Omega_{\text{PD}}$. □

This result shows how the region over which the ambiguity function is convex can be *inscribed* by a set of translated convex polytopes. This set is thus not guaranteed to contain the global minimizer. This is only true in the special case when B_p is invertible.

It will be clear from the above that for the typical GNSS case, i.e. when the phase-code variance ratio is small, the ambiguity function will show many local minima and maxima and thus exhibit a pronounced multimodal variability. This is illustrated in Fig. 1 for the single-frequency, geometry-free model and in Fig. 2 for the dual-frequency, geometry-free model.

Due to the multimodality of $F(b)$ and the difficulty of knowing a priori whether our initial approximation b_0 resides in the same convexity region as the global minimizer, it will generally not be possible to solve for $\check{b}_{\text{AF}}$ by using one of the iterative gradient descent methods from the start. Fortunately, we do know how we can provide a convex bound to the region in which the search for the global minimizer can be conducted. We have the following result:

Lemma 3 (Optimality domain) *Let* $\hat{b} = \arg\min_{b\in\mathbb{R}^p} ||p - B_p b||^2_{Q_{pp}}$. *Then*

$$\begin{aligned} &\check{b}_{\text{AF}} = \arg\min_{b\in\mathbb{R}^p} F(b) \in \Omega, \text{ with} \\ &\Omega = \{b \in \mathbb{R}^p |\ ||p - B_p b||^2_{Q_{pp}} \leq F(\hat{b})\} \end{aligned} \tag{25}$$

■

Proof Clearly, $\check{b}_{\text{AF}} \in \Omega_{\text{F}} = \{b \in \mathbb{R}^p |\ F(b) \leq F(b_0)\}$ for any $b_0 \in \mathbb{R}^p$ and thus also for $b_0 = \hat{b}$. Then, the result follows by recognizing that $||p - B_p b||^2_{Q_{pp}} \leq F(b)$ for all $b \in \mathbb{R}^p$. □

Geometrically, the region Ω of (25) captures all those $b \in \mathbb{R}^p$ of which the vectors $B_p b$ have a weighted squared distance to p that is not greater than $F(\hat{b})$. With B_p being of full column rank and by using the orthogonal decomposition $||p - B_p b||^2_{Q_{pp}} = ||P^\perp_{B_p} p||^2_{Q_{pp}} + ||\hat{b} - b||^2_{Q_{\hat{b}\hat{b}}}$, in which $Q_{\hat{b}\hat{b}} = (B_p^T Q_{pp}^{-1} B_p)^{-1}$, the region can also be written in ellipsoidal form,

$$\Omega = \{b \in \mathbb{R}^p |\ ||\hat{b} - b||^2_{Q_{\hat{b}\hat{b}}} \leq r^2\} \tag{26}$$

with $r^2 = F(\hat{b}) - ||P^\perp_{B_p} p||^2_{Q_{pp}}$. Note, since $\hat{b} \sim \mathcal{N}_p(b, Q_{\hat{b}\hat{b}})$, that Ω would become a confidence region with confidence level $1 - \alpha$, if r^2 would be chosen as $r^2 = \chi^2_\alpha(p, 0)$.

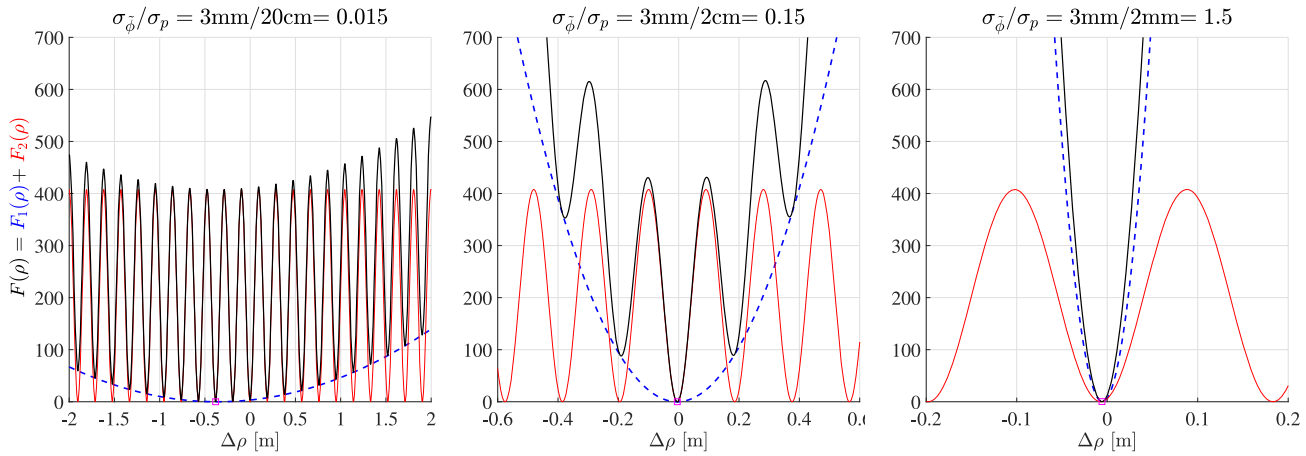

Fig. 1 The $L1$ single-frequency, geometry-free ambiguity function $F(\rho) = F_1(\rho) + F_2(\rho)$ (cf. 20) is illustrated for different standard deviation values of the code and phase data. As the parabolic term $F_1(\rho)$ (in blue) rapidly increases for larger values $\epsilon = \sigma^2_{\tilde{\phi}}/\sigma^2_p$, the ambiguity function $F(\rho)$ becomes strictly convex for $\epsilon > 1$ (note: $\Delta\rho$ is the difference between the variable ρ and the range model value)

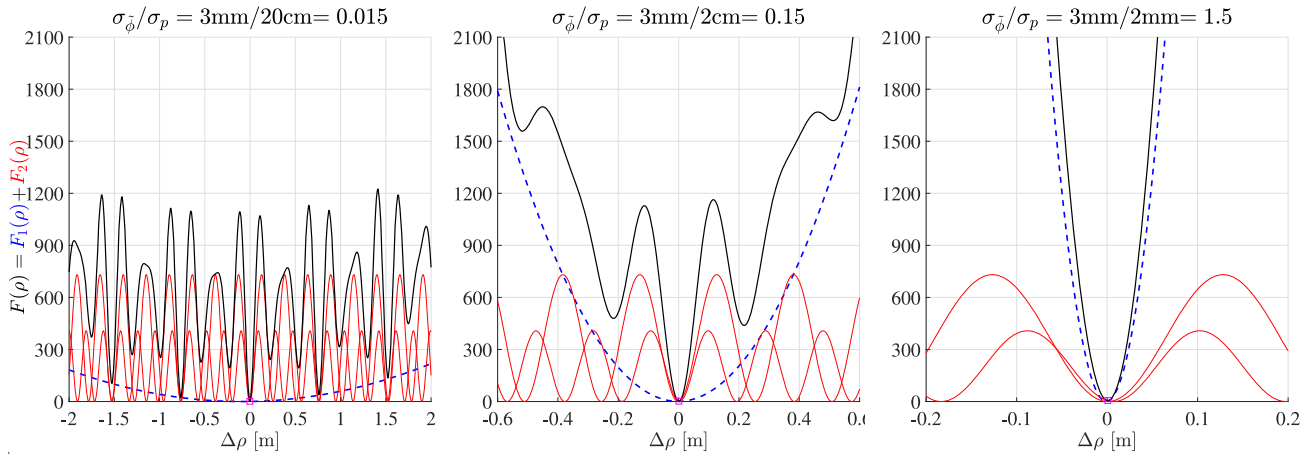

Fig. 2 The $L1$-$L2$ dual-frequency, geometry-free ambiguity function $F(\rho) = F_1(\rho) + F_2(\rho)$ is illustrated for different standard deviation values of the code and phase data as presented in Fig. 1. The term $F_2(\rho)$ is a sum of two sine squared functions, illustrated individually in the plots (red curves), each one having different amplitudes and periods

6 Global minimization of the ambiguity function

In this section, we present our proposed method for finding the global minimizer $\check{b}_{\text{AF}}$ of the ambiguity function $F(b)$. The method is based on that of Teunissen and Massarweh (2024), be it that a different convex relaxing lower bounding function needs to be constructed. We therefore first provide a brief review of the characteristic components of the algorithm, followed by our construction of the required convex relaxation of the ambiguity function.

6.1 Branch-and-bound-based minimization

As $\check{b}_{\text{AF}} \in \Omega$ (cf. 26), we can reformulate our original minimization problem $\min_{b\in\mathbb{R}^p} F(b)$ as the minimization of $F(b)$ over a bounded convex region, $\min_{b\in\Omega} F(b)$. This can be further simplified if we replace Ω by the ellipsoid-circumscribing box $\mathcal{C} = \{b \in \mathbb{R}^p | \, |b_\alpha - \hat{b}_\alpha| \le r\sigma_{\hat{b}_\alpha}, \alpha = 1, \ldots, p\} \supset \Omega$. Hence, this brings our task to solving

$$\check{b}_{\text{AF}} = \arg \min_{b\in\mathcal{C}\subset\mathbb{R}^p} F(b) \tag{27}$$

The challenge in solving (27) is due to the multimodality of the ambiguity function. It is not convex, and it has a multitude of local minima over $\mathcal{C}$. As this challenge is similar to the ones of dual mixed-integer least-squares computations, the same algorithmic components will be used as in the method of Teunissen and Massarweh (2024), i.e. a branch-and-bound-driven minimization, for which the required lower bounds are constructed from projected-gradient-descent solutions of a convex relaxed objective function.

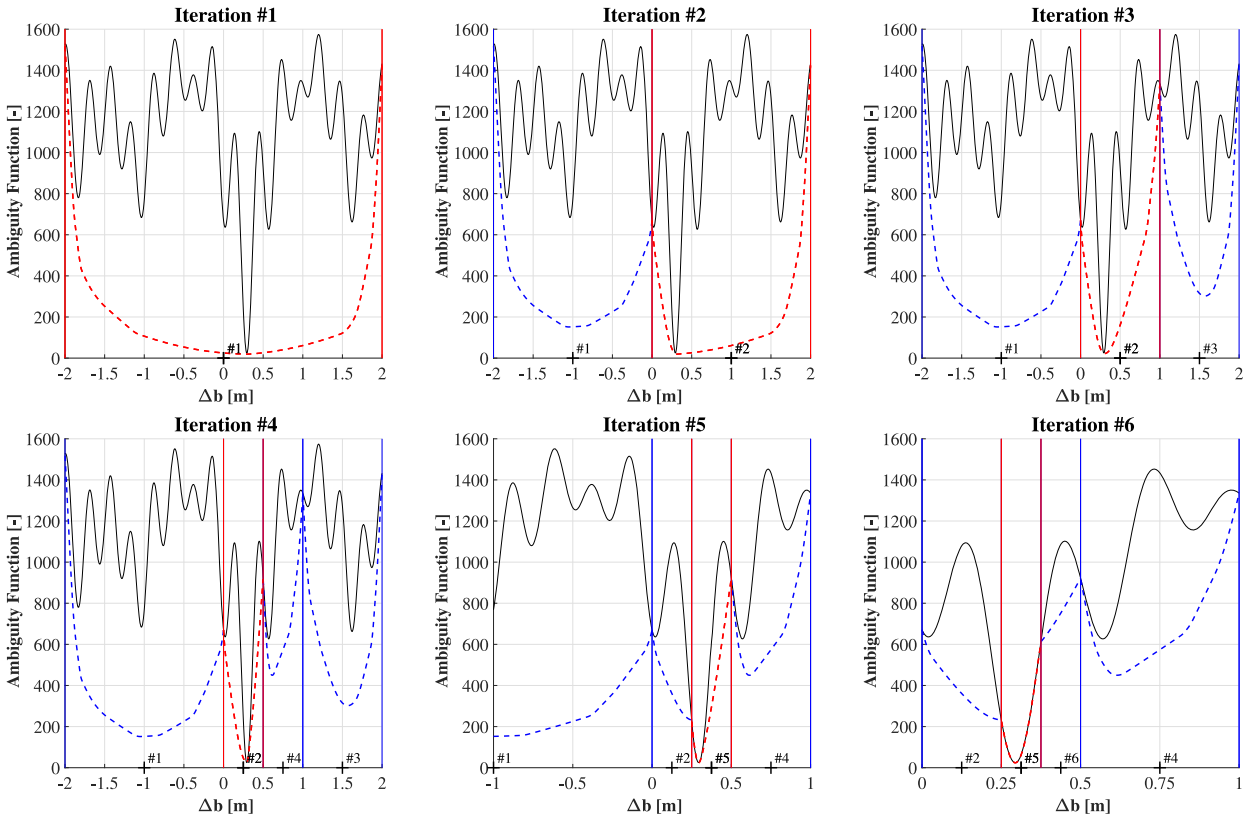

Fig. 3 An illustrative example for $p = 1$, showing the branch-and-bound iterations for computing the global minimum of the ambiguity function (in black). In the first iteration, the interval $[-2, +2]$ is defined as search domain for the global optimum, where Δb refers to the initial value. The convex lower bound is represented by dashed lines for each interval, where at each iteration the most promising one (in red colour) is halved, thus further isolating the global optimum

Branch and Bound (BB)

The branch-and-bound algorithms (Lawler and Wood 1966; Balakrishnan et al. 1991) represent a general approach to the global minimization solution of nonconvex problems. In fact, they ensure a provable upper and lower bound of the global minimum is maintained, thus providing a predefined accuracy for the computed solution, see Guida (2015). The BB algorithms involve partitioning a problem into subproblems (branching) and solving these subproblems to the optimal level, using bounds to eliminate the need to consider suboptimal solutions (bounding). Different mechanizations exist to implement BB solutions; nonetheless, we focus here on a simple approach where we start at the first level with an initial box $\mathcal{C} \subset \mathbb{R}^p$, and then, we proceed by an iterative halving of the boxes.

Hence, we can compute lower $L_1 = L(\mathcal{C})$ and upper $U_1 = U(\mathcal{C})$ bounds, here being global and local at the same time, such that

$$L_1 = L(\mathcal{C}) \leq F(\check{b}_{\mathrm{AF}}) \leq U_1 = U(\mathcal{C}) \tag{28}$$

and the algorithm terminates if $U_1 - L_1 \leq \epsilon$ (note: here the user-defined stop-criterion ϵ should not be confused with the phase-code variance ratio). If $U_1 - L_1 > \epsilon$, then we move to the second level where $\mathcal{C}$ is partitioned in two boxes, i.e. $\mathcal{C} = \mathcal{B}_1 \cup \mathcal{B}_2$, and for each new partition $i = 1, 2$ we compute both lower $L(\mathcal{B}_i)$ and upper $U(\mathcal{B}_i)$ bounds. Notice that these are generally local bounds, whereas we can then construct the new global lower and upper bounds as follows

$$\min(L(\mathcal{B}_1), L(\mathcal{B}_2)) \leq F(\check{b}_{\mathrm{AF}}) \leq \min(U(\mathcal{B}_1), U(\mathcal{B}_2)) \tag{29}$$

with $L_2 = \min(L(\mathcal{B}_1), L(\mathcal{B}_2))$ and $U_2 = \min(U(\mathcal{B}_1), U(\mathcal{B}_2))$. As both $\mathcal{B}_1$ and $\mathcal{B}_2$ are 'smaller' than $\mathcal{C}$ (i.e. they are its partition), one can generally expect the local bounds for $\mathcal{B}_i$ to get sharper in subsequent BB-iterations and the difference between upper and lower bounds to converge uniformly to zero, see Balakrishnan et al. (1991).

The algorithm terminates if $U_2 - L_2 < \epsilon$, otherwise we keep partitioning one of the two boxes, thus ending up with three boxes. For each one, we compute again local lower and upper bounds. Then, we update the global bounds, i.e. taken as minimum over local bounds found for the partitions of $\mathcal{C}$. The value of local bounds determines which box between $\mathcal{B}_1$ and $\mathcal{B}_2$ is split, i.e. the one with the smallest value, so this is equivalent to the global lower bound.

Table 1 Overview of different cases and conditions for the definition of convex lower-bound terms $g_{i,L}$

Case	Condition	Function	Plot
$z_U = 0$	$[l_i, u_i] \in \left[0, \frac{1}{4}\right]$	$g_{i,L} = \sin^2(\pi x_i)$	Figure 4a
	$[l_i, u_i] \in \left[\frac{3}{4}, 1\right]$		
	$[l_i, u_i] \in \left[\frac{1}{4}, \frac{3}{4}\right]$	$g_{i,L} = \sin^2(\pi l_i) + (x_i - l_i)\frac{\sin^2(\pi u_i) - \sin^2(\pi l_i)}{u_i - l_i}$	Figure 4b
	$l_i \in \left[0, \frac{1}{4}\right], u_i \in \left[\frac{1}{4}, 1\right]$	$g_{i,L} = \text{funL}(l_i, u_i, x_i)$	Figure 4c
	$l_i \in \left[0, \frac{3}{4}\right], u_i \in \left[\frac{3}{4}, 1\right]$	$g_{i,L} = \text{funU}(l_i, u_i, x_i)$	Figure 4d
$z_U > 0$	$x_i < 1$	$g_{i,L} = \begin{cases} \text{funU}(l_i, 1, x_i), & l_i < \frac{3}{4} \\ \sin^2(\pi x_i), & \text{otherwise} \end{cases}$	
	$x_i \in [1, z_U]$	$g_{i,L} = 0$	Figure 5
	$x_i > z_U$	$g_{i,L} = \begin{cases} \text{funL}(0, u_i - z_U, x_i - z_U), & u_i > z_U + \frac{1}{4} \\ \sin^2(\pi x_i), & \text{otherwise} \end{cases}$	

After k iterations, we end up with a partitioning of the form $\mathcal{C} = \cup_{i=1}^{k} \mathcal{B}_i$, where the global lower and upper bounds are such that

$$L_k \leq \min_{b \in \mathcal{C}} F(b) \leq U_k \tag{30}$$

where $L_k = \min_{i=1,\ldots,k} L(\mathcal{B}_i)$ is nondecreasing, while $U_k = \min_{i=1,\ldots,k} U(\mathcal{B}_i)$ replaces U_{k-1} only if $U_k < U_{k-1}$, thus assuring that the global upper bound is nonincreasing (i.e. possibly speeding up the BB algorithm). Partitioning terminates if the difference of these bounds is small enough, $U_k - L_k \leq \epsilon$.

With reference to Lemma 3 (cf. 25), we note that the above procedure can be aided by applying box-shrinking, i.e. any b^* that has a function value $F(b^*)$ smaller than the previously used can be used to shrink the set Ω.

Lower and Upper Bounds

For any box partition $\mathcal{B}$, we shall compute lower and upper bounds of $\min_{b \in \mathcal{B}} F(b)$, where the *local* lower bound $L(\mathcal{B})$ is certainly the most challenging one to compute. In fact, the standard gradient-based methods unfortunately cannot ensure global convergence, given that $F(b)$ is not necessarily convex.

However, we can find a differentiable convex lower bounding function $F_L(b) \leq F(b), \forall b \in \mathcal{B}$, whose minimizer is easily found as

$$L(\mathcal{B}) = \min_{b \in \mathcal{B}} F_L(b) \leq \min_{b \in \mathcal{B}} F(b) \tag{31}$$

where $\mathcal{B} = \{b \in \mathbb{R}^p |\ b_L \preceq b \preceq b_U\}$. We refer to Sect. 6.2 for more details on the $F_L(b)$ construction.

The local *upper* bounds $U(\mathcal{B})$ are trivial, given that any $b_* \in \mathcal{B}$ can in principle be used, i.e.

$$U(\mathcal{B}) = F(b_*) \geq \min_{b \in \mathcal{B}} F(b) \tag{32}$$

where the 'centre of gravity' of the box, i.e. $b_* = \frac{1}{2}(b_L + b_U)$ for box $\mathcal{B}$ bounded as $b_L \preceq b \preceq b_U$, represents one simple choice.

Another good choice would be to take b_* as the minimizer of the lower bounding function, $b_* = \arg\min_{b \in \mathcal{B}} F_L(b)$ (cf. 31), the idea being that if $F_L(b)$ approximates $F(b)$ well in $\mathcal{B}$, then $F_L(b_*)$ should not differ too much from $F(b_*)$. A third option is to exploit the smoothness of the AF-function and compute b_* as a local minimizer of $F(b)$ over $\mathcal{B}$, thereby using, for instance, the projected-gradient-descent method (see below). This local minimizer becomes then automatically the global minimizer over $\mathcal{B}$, once the branching has reached the stage that $F(b)$ is truely convex over $\mathcal{B}$.

Projected Gradient Descent (PGD)

As our constructed lower bounding convex function $F_L(b)$ is only continuously differentiable, we use the PGD-method to solve for the lower bound (31). The algorithmic steps for doing so are as follows (Bertsekas 1999; Nocedal and Wright 2006):

1. Initialize: Start with a feasible solution, $b^0 \in \mathcal{B}$ and then loop for $k = 0, \ldots$ until the stop criterium $||b^{k+1} - b^k|| \leq \delta_{\text{PGD}}$ is satisfied, given a user-selected small threshold $\delta_{\text{PGD}} \ll 1$.
2. PGD step: Compute stepsize $t_k > 0$ and projected gradient descent

$$b^{k+1} = \text{P}_{\mathcal{B}}(b^k - t_k \partial_b F_L(b^k)) \tag{33}$$

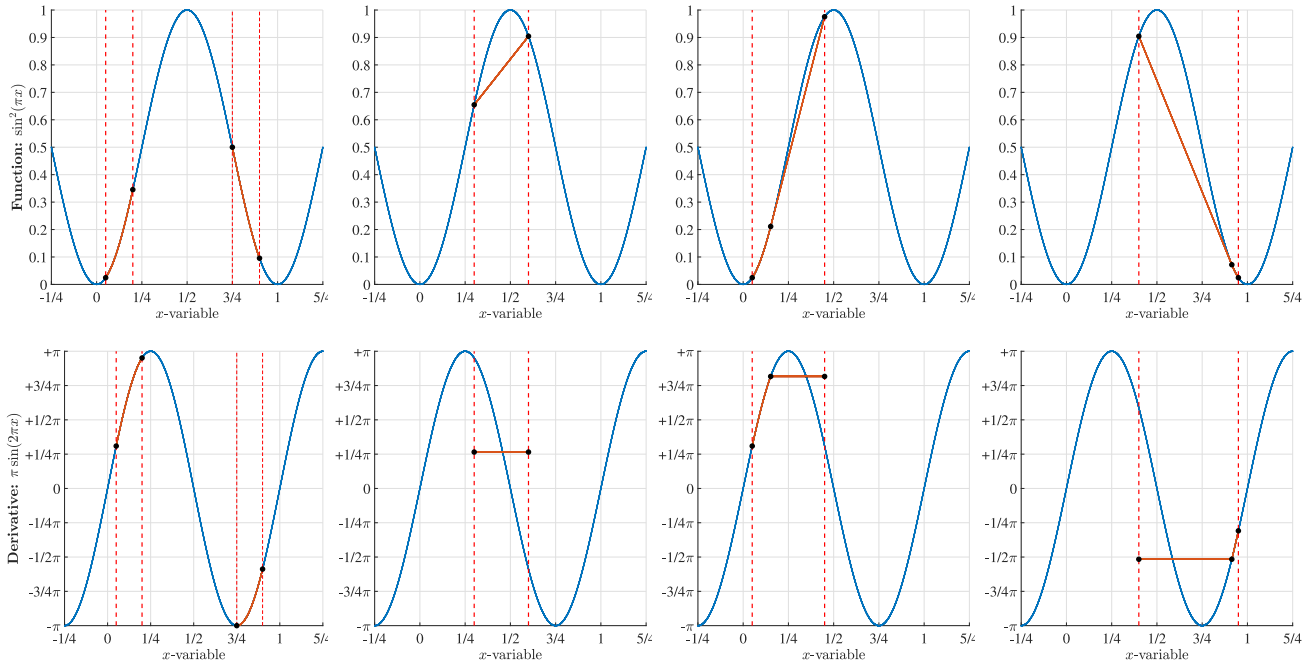

Fig. 4 Illustration of the different convex lower-bound terms $g_{i,L}$ (top) and their derivatives $g'_{i,L}$ (bottom) for the case $z_U = 0$ in Table 1. We consider $\{l = 0.05, u = 0.20\}$ and $\{l = 0.75, u = 0.90\}$ in the first column, $\{l = 0.30, u = 0.60\}$ in the second column, $\{l = 0.05, u = 0.45\}$ in the third column, and $\{l = 0.40, u = 0.95\}$ in the fourth column

such that $F_L(b^{k+1}) \leq F_L(b^k)$, where

$$\mathrm{P}_{\mathcal{B}}(y) = \arg\min_{b \in \mathcal{B}} ||y - b||^2 \tag{34}$$

As $\mathcal{B}$ is a box, the orthogonal projection $\mathrm{P}_{\mathcal{B}} : \mathbb{R}^p \to \mathcal{B}$ can be computed very efficiently. We have for $\alpha = 1, \ldots, p$,

$$\mathrm{P}_{\mathcal{B}}(y)_\alpha = \begin{cases} b_{\alpha,L} & \text{if } y_\alpha \leq b_{\alpha,L} \\ y_\alpha & \text{if } b_{\alpha,L} \leq y_\alpha \leq b_{\alpha,U} \\ b_{\alpha,U} & \text{if } b_{\alpha,U} \leq y_\alpha \end{cases} \tag{35}$$

in which $b_{\alpha,L}$, y_α, and $b_{\alpha,U}$ denote the α-coordinates of b_L, y, and b_U, respectively.

Example 2 (*BB height-determination*) Fig. 3, first panel, shows, over an initial interval $[-2, +2]$ metres, the ambiguity function (16) of an $L1$ single-frequency, DD short-baseline model for which only the height difference is assumed unknown, i.e. $p = 1$. The true, simulated, height difference is 30.0 cm, while the AF-minimizer is found to be 30.3 cm, using a pseudorange (code) and phase measurement precision of 20 cm and 2 mm, respectively (note: Δb is shown with respect to initial value).

In the first iteration, the initial search interval $[-2, +2]$ is taken and the convex lower bounding function (red dashed line) is evaluated. This interval represents our initial box $C \in \mathbb{R}^p$ for $p = 1$, which is therefore split in two for the second iteration. For both intervals, we compute the respective

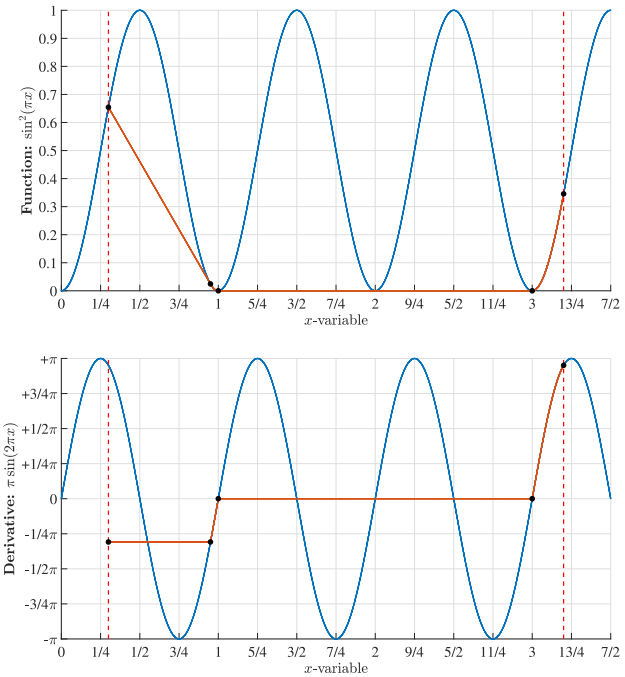

Fig. 5 Illustration of the convex lower-bound term $g_{i,L}$ given the case $z_U > 0$ described in Table 1, along with the associated derivative terms $g'_{i,L}$, where we consider for this example a particular interval $\{l = 0.3, u = 3.2\}$

convex lower bounding functions (blue and red dashed lines) and their minimum value, i.e. convex lower bound (CLB). The one with lowest CLB value is taken as most promising interval, here marked in red colour. For the upper bound computations, we simply consider the centre of current intervals, i.e. $b_* = (b_L + b_U)/2$, and compute $F(b_*)$. As the intervals get shorter, the difference between upper and lower local bounds get smaller as well.

The process is repeated, and with further iterations the global minimum is isolated in smaller intervals where the difference between $F(b)$ and $F_L(b)$ also gets smaller, until convergence. Note that in the last iteration, we have isolated the global minimum in an interval where the ambiguity function is convex and iterative gradient descent could have been adopted, see Sect. 5.1. □

6.2 Convex relaxation of the ambiguity function

We now determine our differentiable convex lower bounding (CLB) function of:

$$\begin{aligned} F(b) &= ||p - B_p b||^2_{Q_{pp}} + 4w^T \sin^2[\pi(\phi - B_\phi b)] \\ &= ||p - B_p b||^2_{Q_{pp}} + \underbrace{4\sum_{i=1}^{m} w_i \sin^2[\pi c_i^T(\phi - B_\phi b)]}_{G(b)} \end{aligned} \tag{36}$$

As the first term on the right-hand side is already convex, we concentrate on finding a convex relaxation of the second term $G(b)$. We thus aim to find a differentiable convex function $G_L : \mathcal{B} \to \mathbb{R}$, such that

$$G_L(b) \le G(b), \ \forall b \in \mathcal{B} \tag{37}$$

so that we obtain the convex lower bounding (CLB) function

$$F_L(b) = ||p - B_p b||^2_{Q_{pp}} + G_L(b) \le F(b), \ \forall b \in \mathcal{B} \tag{38}$$

with its gradient given by

$$\partial_b F_L(b) = -2B_p^T Q_{pp}^{-1}(p - B_p b) + \partial_b G_L(b) \tag{39}$$

Given that $G(b)$ is a summation of scalar terms,

$$G(b) = 4\sum_{i=1}^{m} w_i g_i(x_i(b)), \ \ g_i(x_i(b)) = \sin^2(\pi x_i(b)) \tag{40}$$

where $x_i(b) = c_i^T(\phi - B_\phi b) \in \mathbb{R}$, we may seek a CLB function of similar structure,

$$G_L(b) = 4\sum_{i=1}^{m} w_i g_{i,L}(x_i(b)) \tag{41}$$

such that $g_{i,L}(x_i(b)) \le g_i(x_i(b)), \forall b \in \mathcal{B}$, thus satisfying (37). The box constraint implies that also $x_i(b)$ is bounded in an interval $[l_i, u_i]$ for which $g_{i,L}$ is required to be convex. These intervals can be found based on the projection-lemma described in Teunissen and Massarweh (2024).

At this point, we provide the definition of $g_{i,L}$ for $x_i \in [l_i, u_i]$, and different cases should be distinguished as summarized in Table 1. Notice that we define $z_L = \lfloor l_i \rfloor$ and $z_U = \lfloor u_i \rfloor$, where $\lfloor \cdot \rfloor$ refers to the floor function. Hence, we subtract z_L from the aforementioned quantities, so they are re-defined as $x_i := x_i - z_L$, $l_i := l_i - z_L$, $u_i := u_i - z_L$, and $z_U := z_U - z_L$. Thus, we always have $l_i \in [0, 1]$ with $u_i \in (l_i, 1)$ if $z_U = 0$ or $u_i > 1$ if $z_U > 0$. These two cases are separated in Table 1 and are, respectively, illustrated in Figs. 4 and 5.

When constructing these CLB terms, we make use of two auxiliary functions 'funL' and 'funU' that are defined as:

Auxiliary Function #1:

$$\text{funL}(l, u, x) =$$

$$\begin{cases} \sin^2(\pi l) + (x - l)\frac{\sin^2(\pi u) - \sin^2(\pi l)}{u - l}, & l \ge x_T \\ \sin^2(\pi x), & l < x_T, x \le x_T \\ \sin^2(\pi x_T) + \pi(x - x_T)\sin(2\pi x_T), & l < x_T, x > x_T \end{cases} \tag{42}$$

where $x_T \in [0, \frac{1}{4}]$ is obtained from

$$\sin^2(\pi u) = \sin^2(\pi x_T) + \pi(u - x_T)\sin(2\pi x_T) \tag{43}$$

Auxiliary Function #2:

$$\text{funU}(l, u, x) =$$

$$\begin{cases} \sin^2(\pi l) + (x - l)\frac{\sin^2(\pi u) - \sin^2(\pi l)}{u - l}, & u \le x_T \\ \sin^2(\pi x), & u > x_T, x \ge x_T \\ \sin^2(\pi x_T) + \pi(x - x_T)\sin(2\pi x_T), & u > x_T, x < x_T \end{cases} \tag{44}$$

where $x_T \in \left[\frac{3}{4}, 1\right]$ is obtained from

$$\sin^2(\pi l) = \sin^2(\pi x_T) + \pi(l - x_T)\sin(2\pi x_T) \tag{45}$$

At the same time, the derivative in $b \in \mathbb{R}^p$ of each CLB term $g_{i,L}$ follows as

$$\partial_b g_{i,L}(x_i(b)) = g'_{i,L}(x_i)\partial_b x_i(b) = -g'_{i,L}(x_i)B_\phi^T c_i \tag{46}$$

given $x_i \in [l_i, u_i]$, while $g'_{i,L}(x_i) = \partial_{x_i} g_{i,L}(x_i)$ can be easily computed based on the elementary expressions provided in Table 1 and has been illustrated for the two cases $z_U = 0$ and

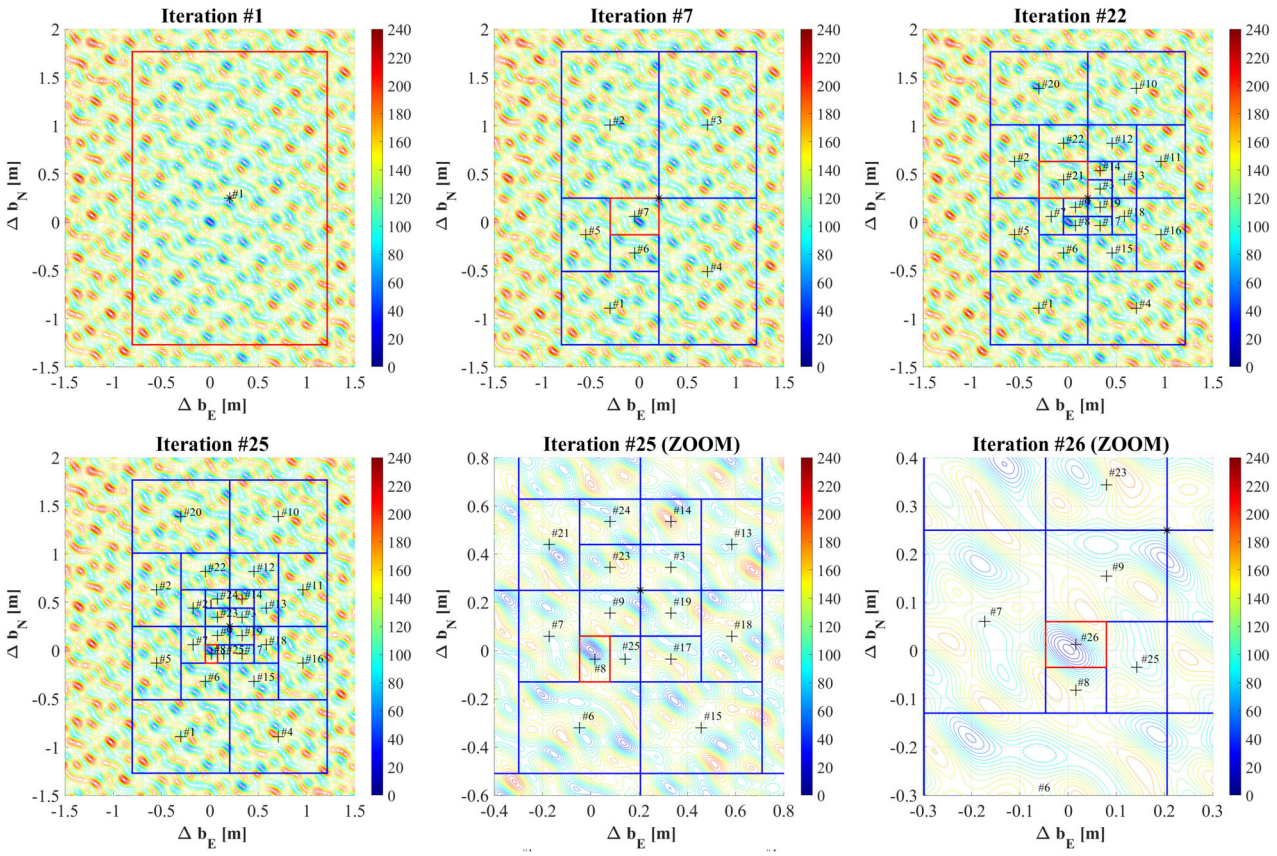

Fig. 6 A branch-and-bound iteration towards the global minimum of the ambiguity function $F(b)$ (cf. 16) for an $L1$ single-frequency, DD short-baseline model with the horizontal position coordinates assumed unknown, where at each iteration the most promising box (in red colour) is halved

$z_U > 0$, respectively, in Figs. 4 and 5. The gradient $\partial_b G_L(b)$ is ultimately computed as:

$$\partial_b G_L(b) = -4 \sum_{i=1}^{m} w_i g'_{i,L}(x_i) B_\phi^T c_i \tag{47}$$

and $\partial_b F_L(b)$ will resemble the expression shown in (18), after substituting $\pi \sin(2\pi x_i)$ with $g'_{i,L}(x_i)$ since we are considering now the convex lower bound of $F(b)$.

Example 3 *(BB position determination)* In analogy with example 2, this example illustrates the BB-iterations towards the global minimum of the ambiguity function $F(b)$ (cf. 16) in case of an $L1$ single-frequency, DD short-baseline model for which only the horizontal position coordinates are assumed unknown. The panels of Fig. 6 show the contourlines of $F(b)$, together with the per iteration-step increasing box-densification. Shown are the results for iterations #1, #7, #22, #25, together with a zoom-in of the last two iteration steps, #25 and #26. The red box is every time the most promising box to be split. It is the box with lowest CLB value, computed from the convex lower bounding functions that we introduced in Sect. 6.2. □

7 The AF- and LS-principle compared

The fact that the ambiguity objective function (16) has a 'sum-of-squares' structure and that the approximation $\sin^2(x) \approx x^2$ holds for small x, has led some authors to link the AF-principle to that of least-squares, e.g. (Rogers et al. 1978; Lachapelle et al. 1992; Leick et al. 2015), with some even stating that the two are 'fundamentally equivalent'. Although we show in this section that this statement is incorrect, we also show under which identified circumstances one can expect AF- and ILS-solutions to have similar behaviour.

7.1 A least-squares relation

To make a strict comparison between the AF- and ILS-objective functions possible, we rely on the primal-dual equivalence of mixed ILS theory as introduced in (Teunissen and Massarweh 2024). We therefore first summarize the for current purposes relevant material in the following theorem (see Sect. 4 of (ibid)).

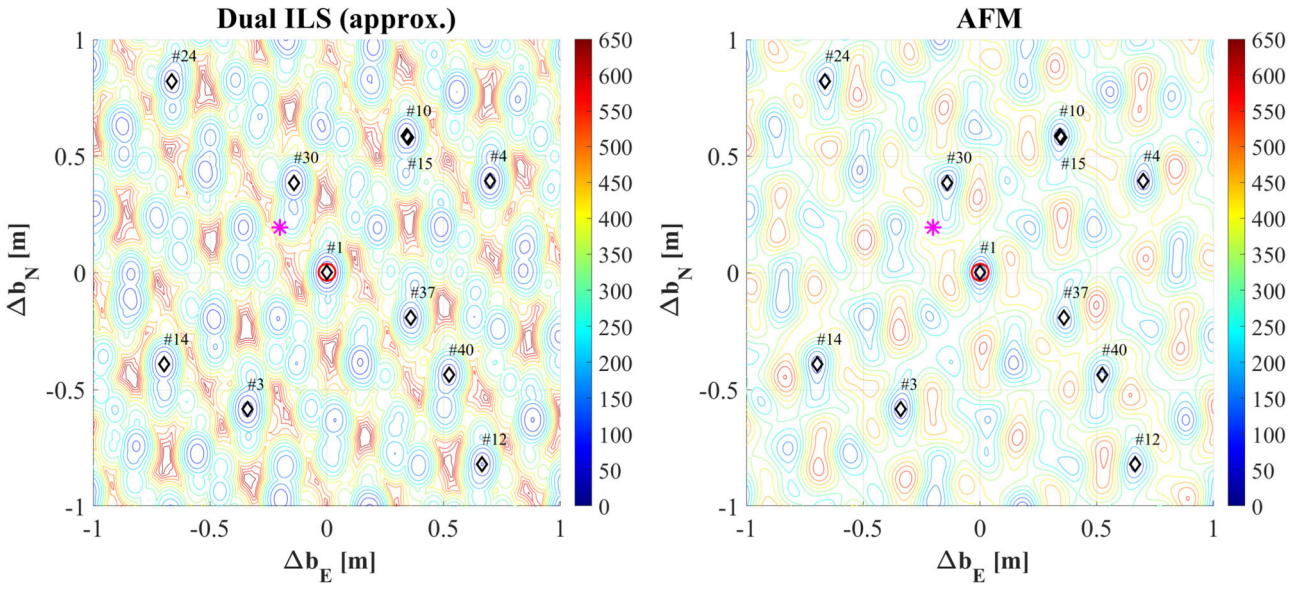

Fig. 7 Contourlines zoom-in of the dual-ILS function $\mathcal{D}^\circ(b)$ (left) and ambiguity function $F(b)$ (right) for the single-epoch model of example 3. The colourbar scale of $\mathcal{D}^\circ(b)$ is half that of $F(b)$. The float solution $\hat{b}$ is shown as a magenta asterisk and the AF-solution $\check{b}_{\mathrm{AF}} = \arg\min_b F(b)$ as a red circle. The ten black diamonds show the locations of 10 out of the 40 smallest local minima of $\mathcal{D}^\circ(b)$, with diamond #1 showing the location of the ILS-solution $\check{b}^\circ = \arg\min_b \mathcal{D}^\circ(b) = \hat{b}^\circ(\check{a}^\circ)$, with $\check{a}^\circ = \arg\min_{a\in\mathbb{Z}^n} ||\hat{a}-a||^2_{Q^\circ_{\hat{a}\hat{a}}}$

Theorem 3 (Primal-Dual mixed ILS)*: Let the dispersion of the float ambiguity and baseline estimators, $\hat{a} \in \mathbb{R}^n$ and $\hat{b} \in \mathbb{R}^p$, be given as*

$$\mathrm{D}\begin{bmatrix}\hat{a}\\ \hat{b}\end{bmatrix} = \begin{bmatrix} Q_{\hat{a}\hat{a}} & Q_{\hat{a}\hat{b}} \\ Q_{\hat{b}\hat{a}} & Q_{\hat{b}\hat{b}}\end{bmatrix} \tag{48}$$

and define the baseline objective function as

$$\mathcal{D}^\circ(b) = ||\hat{b}-b||^2_{Q_{\hat{b}\hat{b}}} + ||\hat{a}(b)-\check{a}^\circ(b)||^2_{Q^\circ_{\hat{a}(b)\hat{a}(b)}} \tag{49}$$

with $\hat{a}(b) = \hat{a} - Q_{\hat{a}\hat{b}}Q^{-1}_{\hat{b}\hat{b}}(\hat{b}-b)$, $\check{a}^\circ(b) = \arg\min_{z\in\mathbb{Z}^n} ||\hat{a}(b)-z||^2_{Q^\circ_{\hat{a}(b)\hat{a}(b)}}$ and $Q^\circ_{\hat{a}(b)\hat{a}(b)}$ being an approximation of the actual baseline-conditioned ambiguity variance matrix $Q_{\hat{a}(b)\hat{a}(b)} = Q_{\hat{a}\hat{a}} - Q_{\hat{a}\hat{b}}Q^{-1}_{\hat{b}\hat{b}}Q_{\hat{b}\hat{a}}$. Then, the minimizer of $\mathcal{D}^\circ(b)$,

$$\check{b}^\circ = \arg\min_{b\in\mathbb{R}^p} \mathcal{D}^\circ(b) \tag{50}$$

is the solution of the mixed ILS problem

$$\left.\begin{matrix}\check{a}^\circ\\ \check{b}^\circ\end{matrix}\right\} = \arg\min_{a\in\mathbb{Z}^n, b\in\mathbb{R}^p}\left(||\hat{a}-a||^2_{Q^\circ_{\hat{a}\hat{a}}} + ||\hat{b}^\circ(a)-b||^2_{Q^\circ_{\hat{b}(a)\hat{b}(a)}}\right)$$

where

$$\begin{aligned} Q^\circ_{\hat{a}\hat{a}} &= Q^\circ_{\hat{a}(b)\hat{a}(b)} + Q_{\hat{a}\hat{b}}Q^{-1}_{\hat{b}\hat{b}}Q_{\hat{a}\hat{b}}\\ \hat{b}^\circ(a) &= \hat{b} - Q_{\hat{b}\hat{a}}Q^{\circ-1}_{\hat{a}\hat{a}}(\hat{a}-a)\\ Q^\circ_{\hat{b}(a)\hat{b}(a)} &= Q_{\hat{b}\hat{b}} - Q_{\hat{b}\hat{a}}Q^{\circ-1}_{\hat{a}\hat{a}}Q_{\hat{a}\hat{b}}\end{aligned}$$

■

Proof See Section 4 of (Teunissen and Massarweh 2024). □

From the above theorem, two important conclusions can be drawn. First, if $Q^\circ_{\hat{a}(b)\hat{a}(b)}$ is chosen to be equal to $Q_{\hat{a}(b)\hat{a}(b)}$, then also $Q^\circ_{\hat{a}\hat{a}} = Q_{\hat{a}\hat{a}}$ and $\hat{b}^\circ(a) = \hat{b}(a)$, from which follows that $\check{b}^\circ = \arg\min_{b\in\mathbb{R}^p} \mathcal{D}^\circ(b)$ will be identical to

$$\check{b} = \hat{b}(\check{a}) \text{ with } \check{a} = \arg\min_{a\in\mathbb{Z}^n} ||\hat{a}-a||^2_{Q_{\hat{a}\hat{a}}} \tag{51}$$

Thus, with the choice $Q^\circ_{\hat{a}(b)\hat{a}(b)} = Q_{\hat{a}(b)\hat{a}(b)}$, the baseline solution $\check{b}^\circ = \arg\min_{b\in\mathbb{R}^p} \mathcal{D}^\circ(b)$ is identical to the ILS-baseline estimator $\check{b}$, as a consequence of which it will also share its statistical optimality properties such as having a maximum ambiguity success rate (Teunissen 1999).

The second conclusion that can be drawn from the above theorem is that even if $Q^\circ_{\hat{a}(b)\hat{a}(b)} \neq Q_{\hat{a}(b)\hat{a}(b)}$, the minimizer $\check{b}^\circ = \arg\min_{b\in\mathbb{R}^p} \mathcal{D}^\circ(b)$ is still an ILS-baseline estimator, namely

$$\check{b}^\circ = \hat{b}^\circ(\check{a}^\circ) \text{ with } \check{a}^\circ = \arg\min_{a\in\mathbb{Z}^n} ||\hat{a}-a||^2_{Q^\circ_{\hat{a}\hat{a}}} \tag{52}$$

but now one which uses an incorrect ambiguity-weighting through $Q^\circ_{\hat{a}\hat{a}}$ and an incorrect baseline-mapping through $\hat{b}^\circ(a)$; compare (51) with (52). Hence, (52) will not have the optimality properties of (51). Using the statistical and distributional properties of $\check{a}^\circ$ and $\check{b}^\circ$ as given in (Teunissen and Massarweh 2024), one can study by how much these properties differ from those of the optimal estimators in (51). As the

above makes clear, this departure from optimality is driven by the difference between $Q^{\circ}_{\hat{a}(b)\hat{a}(b)}$ and $Q_{\hat{a}(b)\hat{a}(b)}$. In (ibid), we discussed some cases for which the estimators $\check{b}^{\circ}$ and $\check{b}$ performed similarly due to their small difference between $Q^{\circ}_{\hat{a}(b)\hat{a}(b)}$ and $Q_{\hat{a}(b)\hat{a}(b)}$.

For our current AF-analysis, the relevance of Theorem 3 lies now in the fact that with a special choice for $Q^{\circ}_{\hat{a}(b)\hat{a}(b)}$, the ILS-baseline producing objective function $\mathcal{D}^{\circ}(b)$ can be given a structure that closely resembles that of the ambiguity objective function $F(b)$. For the ambiguity objective function (16), we have, with $w_i = \frac{1}{4\pi^2\sigma^2_{\phi_i}}$,

$$F(b) = ||\hat{b} - b||^2_{Q_{\hat{b}\hat{b}}} + \sum_{i=1}^{m} \frac{\sin^2\left(\pi c_i^T(\phi - B_\phi b)\right)}{\pi^2\sigma^2_{\phi_i}} \tag{53}$$

A very similar structure is obtained for $\mathcal{D}^{\circ}(b)$ if $Q^{\circ}_{\hat{a}(b)\hat{a}(b)} = \text{diag}(\sigma^2_{\hat{a}_1(b)}, \ldots, \sigma^2_{\hat{a}_n(b)})$. With this choice, we have $\check{a}^{\circ} = \arg\min_{a\in\mathbb{Z}^n} ||\hat{a} - a||^2_{Q^{\circ}_{\hat{a}\hat{a}}} = (\lceil \hat{a}_1(b)\rfloor, \ldots, \lceil \hat{a}_n(b)\rfloor)^T$, and therefore $\mathcal{D}^{\circ}(b) = ||\hat{b} - b||^2_{Q_{\hat{b}\hat{b}}} + \sum_{i=1}^{n} \frac{1}{\sigma^2_{\hat{a}_i(b)}} \left(\hat{a}_i(b) - \lceil \hat{a}_i(b)\rfloor\right)^2$, which can be worked out as

$$\begin{aligned} \mathcal{D}^{\circ}(b) &= ||\hat{b} - b||^2_{Q_{\hat{b}\hat{b}}} + \sum_{i=1}^{n} \frac{d^2\left(c_i^T A^+_\phi(\phi - B_\phi b)\right)}{\sigma^2_{(A^+_\phi\phi)_i}} \\ &\text{with} \\ d(x) &= (x - \lceil x\rfloor) \\ A^+_\phi &= (A_\phi^T Q^{-1}_{\phi\phi} A_\phi)^{-1} A_\phi^T Q^{-1}_{\phi\phi} \\ \sigma^2_{(A^+_\phi\phi)_i} &= c_i^T (A_\phi^T Q^{-1}_{\phi\phi} A_\phi)^{-1} c_i \end{aligned} \tag{54}$$

The two expressions, (53) and (54), look very similar, but with the following marked differences,

1. The functions used in the two sums are $\frac{\sin^2(\pi x)}{\pi^2}$ in (53) and $(x - \lceil x\rfloor)^2$ in (54).
2. The arguments used in the two functions are $c_i^T(\phi - B_\phi b)$ in (53) and $c_i^T A^+_\phi(\phi - B_\phi b)$ in (54), while normalized with the variances $\sigma^2_{\phi_i}$ in (53) and $\sigma^2_{(A^+_\phi\phi)_i}$ in (54).
3. Furthermore, as a consequence, the sum in (53) is over m, the dimension of the phase vector ϕ, while the sum in (54) is over n, the dimension of the ambiguity vector a.

In the absence of these differences, the two functions $F(b)$ and $\mathcal{D}^{\circ}(b)$ would be identical and the ambiguity function determined baseline $\check{b}_{\text{AF}} = \arg\min_{b\in\mathbb{R}^p} F(b)$ would be a true ILS-baseline, albeit one determined from an incorrect ambiguity variance matrix.

Note that the second of the above differences (and implictly the third as well) is due to the fact that $\mathcal{D}^{\circ}(b)$ is based on a conditional least-squares ambiguity estimation, while this is not true for $F(b)$. The ambiguity function $F(b)$ works namely directly on the phase data, and it thus therefore not exploit any time-constancy in the ambiguities if such would be present. Hence, if we assume to work with DD phase data in either a single-epoch model or a multi-epoch model in which all ambiguities are disconnected in time, then $A_\phi = I_m$, from which follows that the last two of the above differences disappear, since then $A^+_\phi = I_m$ and $m = n$.

As immunity for cycle slips is considered one of the attractive features of the AF-method, we summarize the properties for the case $A_\phi = I_m$ separately in the following corollary.

Corollary 1 (AF as ILS approximation) *If $A_\phi = I_m$, replacement of $\frac{\sin^2\pi x}{\pi^2}$ in (53) by $(x - \lceil x\rfloor)^2$ turns the ambiguity function $F(b)$ into $\mathcal{D}^{\circ}(b)$, the minimizer of which is the ILS-baseline $\check{b}^{\circ}$* (cf. 52) *with $Q^{\circ}_{\hat{a}(b)\hat{a}(b)} = \text{diag}(\sigma^2_{\hat{a}_1(b)}, \ldots, \sigma^2_{\hat{a}_n(b)})$.* ■

This result shows that the extend to which the AF-solution $\check{b}_{\text{AF}} = \arg\min_{b\in\mathbb{R}^p} F(b)$ can be considered an approximate ILS-solution hinges on the approximation of $(x - \lceil x\rfloor)^2$ by $\frac{\sin^2\pi x}{\pi^2}$. Although the maxima of these two functions are quite different, their minima are identical and their local behaviour around these minima is also very similar. For $x = z + \delta$, with $z \in \mathbb{Z}$ and δ = small, we have namely $\frac{\sin^2\pi x}{\pi^2} \approx \delta^2 = (x - \lceil x\rfloor)^2$. This shows that under the condition of Corollary 1, one can indeed expect the minimizers of $F(b)$ and $\mathcal{D}^{\circ}(b)$ to be quite close. This is illustrated, for the model of example 3, in Fig. 7. To the left we have the contourlines of $\mathcal{D}^{\circ}(b)$ and to the right those of $F(b)$. Apart from their difference in scale, the two contour plots show a very similar topography, with their local minima at almost identical locations. The ten black diamonds, for instance, show the locations of 10 out of the 40 smallest local minima of $\mathcal{D}^{\circ}(b)$, but they are at the same time also very close to the corresponding local minima of $F(b)$. And this also holds true for their global minimum, with the black diamond #1 identifying the ILS-solution $\check{b}^{\circ} = \arg\min_b \mathcal{D}^{\circ}(b)$ and the red circle identifying $\check{b}_{\text{AF}} = \arg\min_b F(b)$.

As the strength of the underlying model is not in play in the properties captured by Corollary 1, the close-to-ILS behaviour of the AF-solution $\check{b}_{\text{AF}}$ will not change when varying the ambiguity success rate of the model. In Table 2, we show the formal standard deviation (σ_{UP}) and the simulated RMS values of $\hat{b}$ (float), $\check{b}^{\circ}$ (ILS), and $\check{b}_{\text{AF}}$, for three different success-rates (SR) of the model used in example 2. The success rates were reduced by reducing the number of satellites from 8 via 6 to 5. These results show the consistency between σ_{UP} and the float-RMS values and how the AF-RMS values follow those of ILS. The RMS values of AF and ILS get poorer for smaller success rates and for very small success rates even poorer than the RMS of the float solution. This shows just as for ILS; it is the ambiguity success rate that plays a decisive role in the quality of the AF-solution.

Table 2 Formal standard deviation (σ_{UP}) and simulated RMS values of $\hat{b}$ (float), $\check{b}^{\circ}$ (ILS), and $\check{b}_{\mathrm{AF}}$, for three different success rates (SR) of the model used in Example 2

SR (%)	σ_{UP} (cm)	Float (cm)	ILS (cm)	AFM (cm)
99.99	101.9	101.8	1.0	1.1
86.84	194.7	194.5	113.6	113.9
45.06	418.1	418.0	372.5	372.9

For $\check{b}^{\circ}$ (cf. 52), matrix $Q^{\circ}_{\hat{a}(b)\hat{a}(b)}$ was chosen as $Q^{\circ}_{\hat{a}(b)\hat{a}(b)} = \mathrm{diag}(\sigma^2_{\hat{a}_1(b)}, \ldots, \sigma^2_{\hat{a}_n(b)})$

Hence, although the ambiguity function is a function of b only, the statistical and probabilistic quality of its minimizer is still, like the ILS estimator of b, in a large part driven by the ambiguity success rate of the model.

As an additional remark to our comparison of the two objective functions (53) and (54), we note that the difference in their phase arguments can easily be eliminated by *re-defining* the ambiguity function such that $c_i^T(\phi - B_\phi b)$ and $\sigma^2_{\phi_i}$ in (53) are replaced by $c_i^T A_\phi^+(\phi - B_\phi b)$ and $\sigma^2_{(A_\phi^+\phi)_i}$, respectively. This, however, would eliminate the immunity-to-cycle-slip property of the ambiguity function.

7.2 What about differencing?

The above considerations have shown that the extend to which the AF-solution can be expected to be close to the statistically optimal ILS-solution $\check{b}$ depends on the differences listed for (51) and (52) and on how well $Q^{\circ}_{\hat{a}(b)\hat{a}(b)}$ approximates the actual ambiguity variance matrix $Q_{\hat{a}(b)\hat{a}(b)}$. However, in addition to this, there are also two other aspects that one should keep in mind when comparing AF with ILS. The first is their difference in solution-uniqueness as discussed in Sect. 3 and the second is their ability to handle data transformations, like e.g. phase differencing.

In our review Sect. 2 we already alluded to the fact that the AF baseline solution lacks invariance against different forms of data differencing. To understand the essence of this better, consider data differencing as a linear transformation and then first recall the property which the least-squares principle has with respect to invertible linear transformations of the data. In case of least squares, the parameter solution remains invariant provided the weighting accommodates the effect of the data transformation. Such accommodation, however, is not generally available with the AF-method. The w-vector in (11) can take care of variations in precision, but not accommodate any correlations that a linear transformation of the data may introduce.

Another way of describing this difference between the two principles is to consider the PDF transformation rule for the normal distribution and the circular normal distribution. The normal distribution is closed under linear transformations. That is, a linear transformation of a normally distributed random vector is again normally distributed. Such is, however, not the case with the circular normal distribution. A linear combination of circular normally distributed random variables is *not* circular normally distributed anymore.

This lack of being closed under linear transformations implies that one has to be careful when formulating the carrier-phase part of the AF-likelihood function. Under the assumption that the considered carrier-phase observables are circular normally distributed, usage of (11), as maximum likelihood estimator, requires the carrier-phase observables to be independent. This implies, in the context of GNSS, that only undifferenced (UD) or (between-receiver) single-differenced (SD) carrier-phase observables would qualify as potential entries of the m-vector ϕ. Double-differenced (DD) carrier-phase observables, being correlated amongst themselves, would then not qualify in principle.

But the usage of the AF-method with DD carrier-phase observables is of course not forbidden. As our numerical examples have shown, their AF-results can become quite close to the ILS-solutions and similarly to the SD-based AF-results, as a consequence of the typically high precision of the GNSS phase-observables. What one should keep in mind, however, is that the solution would then not be invariant for the in principle arbitrary way in which DD observables can be defined. Hence, from the same UD data set, one would then obtain different AF solutions when using different definitions of the DD observables. Although these solution differences can be small, it is important that this lack of invariance is understood when applying the AF-method.

8 Summary and conclusions

In this contribution, we introduced new theory for the ambiguity function method. Its two main components are (1) the provision of a probability model by means of which the AF-estimator can be identified as a maximum likelihood estimator and (2) the provision of a global optimizer of the AF-likelihood function, having finite termination with a guaranteed epsilon tolerance.

It was shown that for the AF-estimator to be a maximum likelihood estimator, the multivariate distribution of the phase data must consist of independent circular normal distributions. Although the suggestion of the circular normal distribution for phase data is not new, see (Cai et al. 2007), our linkage of the circular normal distribution to the ambiguity function method and its requirement for the AF-estimator to be a maximum likelihood estimator are new. In this context, we also showed how the ambiguity function method can be generalized so as to enable the inclusion of varying weights for the phase data.

Although the attractiveness of the ambiguity function method is the integer-ambiguity invariance of its objective function, we showed that this same invariance may also cause nonuniqueness in its solution. As this possible nonuniqueness appears to be overlooked in the AF-literature, we determined in Theorem 1 the explicit conditions that need to be satisfied for such nonuniqueness to occur. This nonuniqueness is then removed by code-regularization in Theorem 2.

To better understand the challenges of computing the AF-solution, we first characterized the multimodality of the ambiguity function, provided its gradient and Hessian matrix in Theorem 2 and determined a convex region in which its minimizer is guaranteed to reside. As iterative gradient descent methods alone will not be able to ensure the determination of the global minimizer of the multimodal ambiguity function, we introduced our globally convergent algorithm which is constructed from combining the branch-and-bound principle, with a special convex relaxation of the ambiguity function, to which the projected-gradient-descent method is applied. Each of the method's three constituents was described, with special emphasis to the construction of the required continuously differentiable, convex lower bounding function of the multimodal ambiguity function. Several examples were provided in which the workings and performance of our AF-algorithm were numerically and graphically illustrated.

Finally, a further comparison between the AF-principle and that of ILS-estimation was made using the primal-dual equivalence of mixed ILS theory as introduced in (Teunissen and Massarweh 2024). Based on this equivalence, as summarized in Theorem 3, it was shown that the differences are driven by those listed for (51) and (52), as well as by the impact of neglecting the correlation between the baseline conditioned, float ambiguities. From this comparison, as well as from the identified differences in nonuniqueness and dependence on linear transformations, the conclusion is reached that the two principles are fundamentally different, although there are identified circumstances, as was shown, under which one can expect AF- and ILS-solutions to behave similarly.

Author Contributions PT devised the statistical and algorithmic concepts, developed the theory and wrote the paper, LM checked the theory and text, developed and tested the integrated algorithm and computed the numerical/graphical results.

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

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Balakrishnan V, Boyd S, Balemi S (1991) Branch and bound algorithm for computing the minimum stability degree of parameter-dependent linear systems. J Robust Nonlinear Control 1(4):295–317

Bertsekas D (1999) Nonlinear programming, 2nd edn. Athena Scientific, New York

Cai J, Grafarend EW, Hu C (2007) The statistical property of the GNSS carrier phase observations and its effects on the hypothesis testing of the related estimators. In: Proceedings of ION GNSS 20th international technical meeting of the satellite division, pp 331–338

Counselman CC, Gourevitch SA (1981) Miniature interferometric terminals for earth surveying: ambiguity and multipath with global positioning system. IEEE Trans Geosci Remote Sens GE 19(4):244–252

Guida A (2015) A branch and bound algorithm for the global optimization and its improvements. Ph.D. thesis, University of Florence pp 1–89

Gumbel EJ, Greenwood JA, Durand D (1953) The circular normal distribution: theory and tables. J Amer Statist Assoc 48(261):131–152

Han S, Rizos C (1996) Improving the computational efficiency of the ambiguity function algorithm. J Geodesy 70:330–341

Hartman S (2021) The history of RTK—part 3: everybody wanted it. The American Surveyor, pp 1–9

Hofmann-Wellenhof B, Lichtenegger H, Wasle E (2008) GNSS: global navigation satellite systems: GPS, Glonass, Galileo, and more. Springer, New York

Lachapelle G, Canon ME, Erickson C, Falkenberg W (1992) High-precision C/A code technology for rapid-static DGPS surveys. In: Proceedings of the 6th international geodetic symposium on satellite positioning, DMA/OSU, pp 1–6

Lawler EL, Wood DE (1966) Branch-and-bound methods: a survey. Oper Res 14(4):699–719

Leick A, Rapoport L, Tatarnikov D (2015) GPS satellite surveying, 4th edn. Wiley, New York

Mader GL (1992) Rapid static and kinematic global positioning system solutions using the ambiguity function technique. J Geophys Res 97(B3):2371–2383

Massarweh L, Verhagen S, Teunissen PJG (2025) New LAMBDA toolbox for mixed-integer models: estimation and evaluation. GPS Solut 29:14. https://doi.org/10.1007/s10291-024-01738-z

Nesterov Y (2018) Lectures on convex optimization. Springer optimization and its applications 137

Nocedal J, Wright S (2006) Numerical optimization, 2nd edn. Springer, New York

Remondi BW (1984) Using the global positioning system phase observable for relative geodesy: modelling, processing and results. Ph.D. Dissertation, University of Texas

Remondi BW (1991) Pseudo-kinematic GPS results using the ambiguity function method. Navig J Inst Navig 38(1):17–36

Remondi BW, Hilla SA (1993) Pseudo-kinematic surveying based on full-wavelength dual-frequency GPS observations. NOAA Tech Memorandum NOS NGS 56:24p

Rogers AEE, Knight CA, Hintegregger HF, Witney AR (1978) Geodesy by radio interferometry: determination of a 124-km base line vector with ∼5-mm repeatability. J Geophys Res 83(B1):325–333

Teunissen PJG (1990) Nonlinear least squares. Manuscr Geodaet 15(3):137–150

Teunissen PJG (1995) The least-squares ambiguity decorrelation adjustment: a method for fast GPS integer ambiguity estimation. J Geod 70(1–2):65–82

Teunissen PJG (1999) An optimality property of the integer least-squares estimator. J Geod 73(11):587–593

Teunissen PJG (2003a) Theory of integer equivariant estimation with application to GNSS. J Geodesy 77(7–8):402–410

Teunissen PJG (2003) Towards a unified theory of GNSS ambiguity resolution. J Glob Position Syst 2(1):1–12

Teunissen PJG (2017) Carrier-phase integer ambiguity resolution. In: Chapter 23 in Handbook of GNSS, Teunissen and Montenbruck (Eds), pp 661–685

Teunissen PJG, Khodabandeh A (2022) PPP-RTK theory for varying transmitter frequencies with satellite and terrestrial positioning applications. J Geodesy 96:84

Teunissen PJG, Massarweh L (2024) Primal and dual mixed-integer least-squares: distributional statistics and global algorithm. J Geodesy 98(63):1–26