An Algorithm for Resolving 2π Ambiguities in Interferometric Measurements by use of Multiple Wavelengths

Mats G. Löfdahl[a] and Henrik Eriksson[b]
[a]Royal Swedish Academy of Sciences
Stockholm Observatory, SE–133 36 Saltsjöbaden, Sweden
phone: +46-8-164480, fax: +46-8-164228, email: mats@astro.su.se
[b]NADA, KTH, SE–100 44 Stockholm, Sweden
email: henrik@nada.kth.se

Submitted to Optical Engineering 15 September 2000
Revised version 2 January 2001

Abstract

Measurements of differences in optical path length in monochromatic light with any interferometric method are insensitive to errors that are a whole number of waves. If measurements are performed in several wavelengths, this ambiguity can be resolved. We present a general algorithm for finding the correct distance post facto, given multiple measurements in different wavelengths. Applied e.g. to piston measurements of a segmented mirror, the capture range of a wavefront sensor can be extended from ± half a wave to several waves. The extended capture range can be calculated and depends on the selection of wavelengths used for measurements and on the expected accuracy of the method used.

1 Introduction

With any interferometric distance measurement method using monochromatic or narrow-band light, there is a modulo 2π ambiguity in the measured optical path lengths (OPLs). This is caused by the repetitive nature of the wavefronts or fringes involved.

One application where this is a problem is wavefront sensing for telescope optics. When the wavefront is sensed for every pixel in a pupil map, phase unwrapping methods are used to recover the true shape of the wavefront[1]. A problem for monolithic mirrors, the situation is even worse for segmented mirrors and sparse array telescopes. Relying on continuity, the phase unwrapping works within segment boundaries. But even when the 2π ambiguity has been resolved within a segment, it persists for the step height between neighboring segments. Although not noticeable in the wavefront sensor wavelength, such inter-segment phase errors do degrade the optical performance in broadband light and in wavelengths other than the measurement wavelength.

If the OPL measurements are performed in several wavelengths, the 2π ambiguity can be resolved. One way of measuring OPLs beyond the ambiguity limit are methods based on two-wavelength interferometry[2]. The capture range, i.e. the number of waves of OPL that can be handled, of such measurements is extended because they are equivalent to measurements with a larger, equivalent wavelength, that can be written (for the case of two wavelengths, λ_1 and λ_2) as $\lambda_{\text{eq}} = \lambda_1\lambda_2/|\lambda_1-\lambda_2|$. In principle, λ_{eq} (and thereby the capture range) can be extended to any large number by choosing closely spaced wavelengths. However, in practice λ_{eq} is limited by other factors, such as fringe finesse and amplification of errors. For a general method, useful in post-processing of data from any interferometric method, like phase diversity and curvature sensing, a direct way of using knowledge about measurement errors in the individual methods is needed.

Such a method was proposed in the context of phasing the segments of the Keck telescopes [3, 4]. In that method, the OPL is given by a χ^2 minimization involving measurements in different wavelengths together with the associated statistical uncertainties. However, the solution returned by this method is sometimes in error by much more than the measurement uncertainties and the capture range is not clear. For two wavelengths, these errors occur when the wavelengths are too closely spaced or too far apart. It is not a problem in the given context, partly because the algorithm is used together with a broadband method that is less accurate but does not suffer from the ambiguity problem, but it is clearly a drawback in a general method.

In this paper we present a straightforward algorithm, that incorporates measurement errors and is quite general in that it can be applied to multiple wavelength measurements with any of a variety of methods. It works by establishing sets of candidate solution intervals for each measurement channel, limited by the maximum errors of the method used and regularly spaced along the distance axis with the wavelength of each channel, and then rejecting all solutions that are not in the intersection of these sets. We will refer to this algorithm as the candidate-solution rejection algorithm (CSRA). Using the CSRA, the capture range for unambiguous measurements can be calculated. This cap-

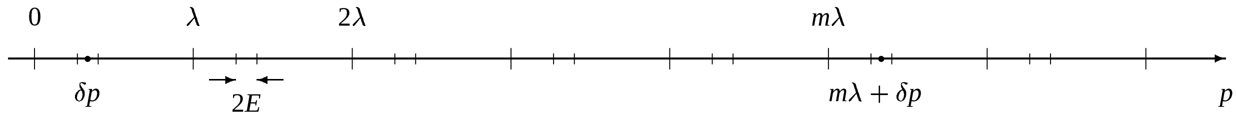


Figure 1: Periodic candidate solution intervals. We measure the ambiguous OPD δp within a $2E$ wide error interval. The wanted OPL $p = m\lambda + \delta p$. The unknown m yields the ambiguity in the form of a sequence of intervals where the solution might be.

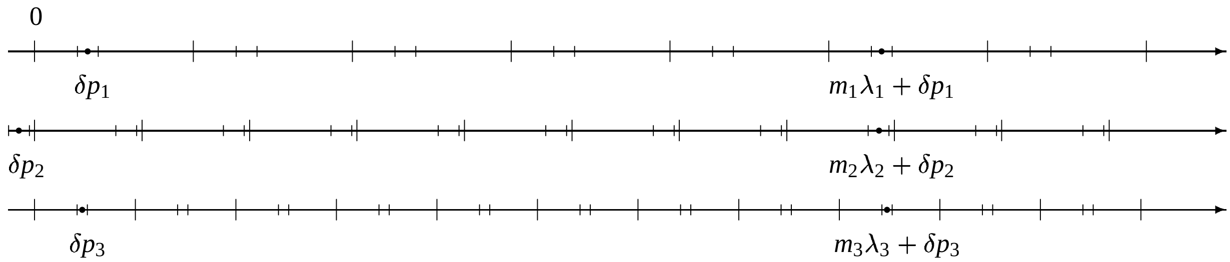


Figure 2: Periodic candidate solution intervals in multiple wavelengths. Compare with Fig. 1. Note the ambiguity is resolved because the only intersection of all three sets of intervals is with the intervals marked $m_k\lambda_k + \delta p_k$.

ture range can be extended from $\pm$ half a wave by tens or hundreds of times, depending on the choice of wavelengths and the associated measurement accuracies. The CSRA is presented in a slightly different form in Ref. [5].

The paper is organized as follows. In Section 2 we introduce our notation and state the problems of finding the extended capture range and the correct distance within that capture range. We look at the two-wavelength capture range analytically in Section 3, discussing the optimal choice of wavelengths. In Section 4 we derive the CSRA, an algorithm for any number of wavelengths, and show how it can be used to find the correct solution as well as the capture range. We end with a short discussion in Section 5.

2 The 2π ambiguity problem

Consider a monochromatic interferometry method working in a wavelength λ. Attempts to measure an OPL p are really measurements of δp, the optical path difference (OPD) to the nearest whole number of wavelengths, so that

$$p = m\lambda + \delta p + \epsilon, \tag{1}$$

where m is an unknown integer number and ϵ is an unknown measurement error with a maximum absolute value E, $|\epsilon| < E$. In addition to the ordinary measurement uncertainty E, there is a 2π ambiguity that comes from the fact that solutions are possible for every value of m. The result is a sequence of $2E$ wide candidate solution intervals centered on each of the $m\lambda + \delta p$, see Fig. 1. Because we are using the maximum error rather than e.g. the standard deviation, the correct OPL must be in one of the intervals – but we don't know which one.

If we have measurements for a set of wavelengths, $\{\lambda_k\}_{k=1}^{K}$, with corresponding maximum absolute errors, $\{E_k\}_{k=1}^{K}$, we have more information, because the candidate solution periodicity is different in the K channels. We get

$$p = m_k\lambda_k + \delta p_k + \epsilon_k; \qquad k \in \{1, \ldots, K\}, \tag{2}$$

where most m_k can be excluded because we know that the error intervals for all wavelength channels should overlap at the correct measurement, see Fig. 2. However, we cannot exclude extra overlaps sufficiently far from the correct solution. If the distance between the correct solution and the first other possible overlap is denoted by $2P$, we can define the capture range, $-P < p < P$, in which there can be only one overlap – at the correct solution.

3 Capture range analysis of the two-wavelength case

For the sake of simplicity, we first consider the case of two wavelengths, $\lambda_1 > \lambda_2$. The capture range is characterized by the relation

$$m_1\lambda_1 \approx m_2\lambda_2 \tag{3}$$

for some multiples of the wavelengths. (Note that these are not the measurement related m_k of the previous section.) Figure 3 depicts one critical case, where $m_1\lambda_1 - 2E_1 = m_2\lambda_2 + 2E_2$ is the distance between two pin-pointed solutions where the error intervals touch. In a less extreme case, where $m_1\lambda_1 - 2E_1 > m_2\lambda_2 + 2E_2$, there will be two solution intervals $m_1\lambda_1 - 2E_1$ apart, so this is the appropriate value for the capture interval $2P$. In the opposite situation, where $m_2\lambda_2 - 2E_2 > m_1\lambda_1 + 2E_1$, the capture interval will of course be $2P = m_2\lambda_2 - 2E_2$. But for which m_1 and m_2 does the first overlap occur?

The important parameter is the ratio λ_1/λ_2. If this ratio has a rational approximation, $\lambda_1/\lambda_2 \approx m_2/m_1$, for some small integers m_1 and m_2, then $m_1\lambda_1 \approx m_2\lambda_2$ making overlaps this critical distance apart possible. By use of continued fractions, we can define a sequence of progressively

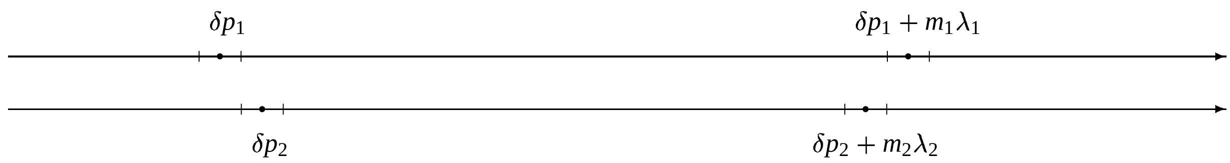


Figure 3: One critical case for the capture range with two wavelengths. The distance between the two intersection points is $m_1\lambda_1 - 2E_1 = m_2\lambda_2 + 2E_2$.

better, alternately smaller and larger, rational approximations, m_2/m_1 (called approximants), to λ_1/λ_2. We will refer to approximants in the beginning of the sequence as "early" and the ones following as "later". We refer to the appendix for a review of used terms and properties of continued fractions needed in deriving the following theorem.

Theorem 1 *For two wavelengths, the capture range $2P$ is the largest of the two numbers $m_1\lambda_1 - 2E_1$ and $m_2\lambda_2 - 2E_2$, where m_2/m_1 is the first approximant in the regular continued fraction for λ_1/λ_2 satisfying $|m_1\lambda_1 - m_2\lambda_2| < 2(E_1 + E_2)$. If a/b and c/d are consecutive approximants and $b < \lambda_2/2(E_1+E_2) < d$, then the approximant m_2/m_1 defined above is a/b.*

Proof Consider the continued fractions expansion of λ_1/λ_2. According to property P1 of the appendix, we get a sequence of rational approximants m_2/m_1, alternately smaller and greater than λ_1/λ_2. The case $m_2/m_1 < \lambda_1/\lambda_2$ corresponds to Fig. 3 and the case $m_2/m_1 > \lambda_1/\lambda_2$ to the opposite situation. Property P2 tells us that the smallest integers m_1 and m_2 for which $|m_2(\lambda_1/\lambda_2) - m_1| < 2(E_1 + E_2)/\lambda_2$ are denominator and numerator of one of these approximants, so the first part of the theorem follows.

For the second part we consider the case of two successive approximants a/b and c/d. Property P3 implies that $|b\lambda_1 - a\lambda_2| < \lambda_2/d$ and from the assumption $d > \lambda_2/2(E_1+E_2)$ we conclude that a/b is indeed the approximant m_2/m_1 determining $2P$. ∎

By examination of Theorem 1, we note that the largest capture range is obtained when m_1 and m_2 are large, so later approximants are better than early ones. But as $|m_1\lambda_1 - m_2\lambda_2|$ is smaller for a late approximant, large maximum errors tend to select early approximants, resulting in small capture ranges. If the error intervals are small, late approximants are obtained when λ_1/λ_2 is not close to any rational fraction with reasonably small denominator and the resulting capture range is large. In musical terms this means that the two frequencies do not harmonize – which is bad for the music lover but good for our purpose – and that the disharmony criterion is set by the error intervals.

When the maximum errors are small, the capture range depends critically on the later approximants. We can get some insight by considering a case when the error intervals are independent of wavelength, and we are free to vary the wavelengths. This may correspond to a situation where we have a tunable laser and want to maximize the capture range. In order to demonstrate how and why a small change in wavelength can affect the capture range, we consider two cases that only differ by 10 nm in one of the wavelengths,

$$\text{Case 1:} \quad \begin{cases} \lambda_1 = 1.640\ \mu\text{m}; & E_1 = 7.5\ \text{nm}, \\ \lambda_2 = 1.100\ \mu\text{m}; & E_2 = 7.5\ \text{nm}, \end{cases} \tag{4}$$

and

$$\text{Case 2:} \quad \begin{cases} \lambda_1 = 1.630\ \mu\text{m}; & E_1 = 7.5\ \text{nm}, \\ \lambda_2 = 1.100\ \mu\text{m}; & E_2 = 7.5\ \text{nm}. \end{cases} \tag{5}$$

We first note that for both cases $2(E_1 + E_2) = 30$ nm. The first three approximants of λ_1/λ_2 are 1, $3/2$, and $79/53$ for Case 1 and 1, $3/2$, and $40/27$ for Case 2, respectively. For Case 1 the first approximant that satisfies the inequality in Theorem 1 is $3/2$, giving us a capture range of $P = 1.66\ \mu$m. For Case 2, however, the inequality is not satisfied by $3/2$ due to the smaller difference between the wavelengths, so we need to use $40/27$. That makes the capture range significantly longer, $P = 22.19\ \mu$m. Alternatively, we could use the inequality $27 < \lambda_2/2(E_1+E_2) \approx 36.7 < 53$ to see that the capture range is given by the second approximant in Case 1, while we have to use a later approximant in Case 2, resulting in a larger capture range. Because of this sensitivity to small changes in the wavelengths, it is a good idea to calculate the capture range for appropriate perturbations of the actual wavelengths, particularly when using white light and filters rather than laser wavelengths.

For two wavelengths, finding the capture range is thus a simple matter that can be handled analytically. What can be said if there are more wavelengths? By pairwise applications of Theorem 1, one obtains a lower bound on the capture range, but in general this is not the best capture range possible. We need a way to find out how all the wavelengths combine to extend the capture range. We also have the problem of actually finding the correct solution within the capture range. The CSRA, given in the next section, is a numerical algorithm that can be used for both purposes.

4 An algorithm for any number of wavelengths

4.1 The candidate-solution rejection algorithm

The CSRA should find a solution p that is consistent with the measurements and error margins in all wavelength channels. The idea is to find the intersection of the error intervals in all wavelengths within the capture range. It

```
algorithm CSRA(λ, δp, E, Q_0)
  K ← length(λ)
  for k ← 1 to K do begin
    Get λ_k, δp_k and E_k as the kth elements
      in λ, δp, and E
    N_{k−1} ← length(Q_{k−1})
    Q_k ← {}
    for n ← 1 to N_{k−1} do begin
      Get q_n and ℰ_n as the nth element in Q_{k−1}
      Calculate m_lo and m_hi from Eqs. (9)–(10)
      for integer m : m_lo < m < m_hi do begin
        Calculate q_new and ℰ_new from
          Eqs. (11)–(14)
        Q_k ← Q_k ∪ {q_new ± ℰ_new}
      end
    end
  end
  return Q_K
end
```

Figure 4: The CSRA as pseudo code. The CSRA can be used both when when determining the capture range (see Section 4.2) and searching for a solution (see Section 4.3). The length($\mathbf{x}$) function should return the number of elements in the array $\mathbf{x}$.

can be implemented as a straight-forward search algorithm, shown as pseudo-code in Fig. 4.

The key is that Eq. (2) should hold for all k simultaneously, to within the specified error limits. This is equivalent to an overlap between the error intervals such that p is within the intersection of the intervals. This real-solution intersection can be found by starting from the entire capture range and successively rejecting intervals that do not overlap for all the wavelengths. We keep the candidate solutions that are consistent with the first k wavelengths in a list $\mathbf{Q}_k = \{q_{kn} \pm \mathcal{E}_{kn}\}_{n=1}^{N_k}$, where each element contains both q_{kn}, the nth candidate OPL for p, and the associated maximum error, $\mathcal{E}_{kn}$. $\mathbf{Q}_k$ could be implemented as two arrays with N_k elements each or a single $2 \times N_k$ array but we use the $q_{kn} \pm \mathcal{E}_{kn}$ notation for brevity and compactness of notation. The CSRA has to be initialized with a list $\mathbf{Q}_0$, usually containing a single element but in general the CSRA can be initialized with any number of candidate solution intervals. In $\mathbf{Q}_1$ we keep candidate solutions that are consistent with the measurements using λ_1, etc. Any solution that is consistent with all the wavelengths can be found in the final list, $p = q_{Kn} \pm \mathcal{E}_{Kn} \in \mathbf{Q}_K$ for some n, where $2\mathcal{E}_{Kn}$ is the width of the intersection of the overlapping error intervals. With a correct capture range, $N_K = 1$. Note that we also gain accuracy, since $\mathcal{E}_{Kn} \leq \min_k E_k$ and often much smaller.

For each candidate solution in $\mathbf{Q}_{k-1}$, intervals are included in $\mathbf{Q}_k$ only if they are consistent with any of the candidate solutions provided by the wavelength λ_k, i.e. if there is an overlap between the intervals $q_{(k-1)n} \pm \mathcal{E}_{(k-1)n}$ and $\delta p_k + m\lambda_k \pm E_k$ for some integer m. For this to be the case,

$$q_{(k-1)n} - \mathcal{E}_{(k-1)n} < \delta p_k + m\lambda_k + E_k \tag{6}$$

and

$$q_{(k-1)n} + \mathcal{E}_{(k-1)n} > \delta p_k + m\lambda_k - E_k. \tag{7}$$

This can be expressed as an interval for m,

$$m_{\text{lo}} < m < m_{\text{hi}}, \tag{8}$$

where

$$m_{\text{lo}} = \frac{q_{(k-1)n} - \mathcal{E}_{(k-1)n} - \delta p_k - E_k}{\lambda_k} \tag{9}$$

and

$$m_{\text{hi}} = \frac{q_{(k-1)n} + \mathcal{E}_{(k-1)n} - \delta p_k + E_k}{\lambda_k}. \tag{10}$$

Candidates that do not meet this requirement for m are rejected. For each m that does meet the requirement, we include an interval, $q_{\text{new}} \pm \mathcal{E}_{\text{new}}$, where

$$q_{\text{new}} = \frac{q_{\text{hi}} + q_{\text{lo}}}{2} \tag{11}$$

and

$$\mathcal{E}_{\text{new}} = \frac{q_{\text{hi}} - q_{\text{lo}}}{2} \tag{12}$$

and the limits of the intersection are

$$q_{\text{lo}} = \max\{q_{(k-1)n} - \mathcal{E}_{(k-1)n}, \delta p_k + m\lambda_k - E_k\} \tag{13}$$

and

$$q_{\text{hi}} = \min\{q_{(k-1)n} + \mathcal{E}_{(k-1)n}, \delta p_k + m\lambda_k + E_k\}. \tag{14}$$

Except for the initial interval, equal to the capture range, Eq. (8) will most often be satisfied by at most a single integer value of m. Therefore the CSRA is a linear algorithm.

4.2 Using the CSRA to calculate the capture range

We start by demonstrating how the CSRA can be used to find the capture range $|p| < P$ for a given set of wavelengths, $\{\lambda_k\}$, and the corresponding maximum errors, $\{E_k\}$.

The idea is to examine the smallest possible distance between consecutive intersections involving all wavelengths. We force a match at the origin by setting all $\delta p_k = 0$ and then step m_1 (where λ_1 is the longest wavelength) through $m_1 = 1, 2, \ldots$ until we find such an intersection. For each m_1, we call CSRA($\boldsymbol{\lambda}$, $\delta\mathbf{p}$, $\mathbf{E}$, $\mathbf{Q}_0$) with $\mathbf{Q}_0 = \{m_1\lambda_1 \pm 2E_1\}$ as the initial solution interval. We only have to look for intersections with the other wavelengths, so we exclude $k = 1$ in the call so that $\boldsymbol{\lambda} = \{\lambda_k\}_2^K$ and $\mathbf{E} = 2\{E_k\}_2^K$. Note that we have to use twice the real maximum errors to account for errors at both ends of the capture range. This corresponds to the factor 2 in front of the error term in Theorem 1. In short, the procedure is to repeatedly make the call

$$\mathbf{Q}_K \leftarrow \text{CSRA}(\{\lambda_k\}_2^K, \mathbf{0}, 2\{E_k\}_2^K, \{m_1\lambda_1 \pm 2E_1\}), \tag{15}$$

while increasing m_1 until a non-empty list $\mathbf{Q}_K \neq \{\}$ is returned. The capture range is then equal to the returned

Table 1: Capture ranges $\pm P$ calculated for subsets of $\{\lambda_k\}_{k=1}^{5} = \{1.152, 0.633, 0.612, 0.594, 0.543\}$ μm, a set of commercially available HeNe laser wavelengths. $E_k = \lambda_k/20$.

Ranking	$P/1\ \mu$m	Subset size	Subset indices k
1	9.20	5	$\{1,2,3,4,5\}$
	9.20	4	$\{1,3,4,5\}$
	9.20	4	$\{1,2,4,5\}$
	9.20	4	$\{1,2,3,5\}$
5	4.59	3	$\{1,3,5\}$
6	3.50	3	$\{1,2,5\}$
7	0.60	4	$\{1,2,3,4\}$
	0.60	3	$\{1,2,4\}$
	0.60	3	$\{1,2,3\}$
	0.60	2	$\{1,2\}$
11	0.58	3	$\{1,3,4\}$
	0.58	2	$\{1,3\}$
13	0.56	3	$\{1,4,5\}$
	0.56	2	$\{1,4\}$
15	0.52	2	$\{1,5\}$
16	0.28	4	$\{2,3,4,5\}$
	0.28	3	$\{3,4,5\}$
	0.28	3	$\{2,4,5\}$
	0.28	3	$\{2,3,5\}$
	0.28	3	$\{2,3,4\}$
	0.28	2	$\{3,5\}$
	0.28	2	$\{3,4\}$
	0.28	2	$\{2,5\}$
	0.28	2	$\{2,4\}$
	0.28	2	$\{2,3\}$
26	0.27	2	$\{4,5\}$

OPL, corrected with the error of the non-rejected solution, or

$$2P = q_{Kn} - \mathcal{E}_{Kn}, \quad (16)$$

where n is the index of the smallest q_{Kn} in the unlikely circumstance that $\mathbf{Q}_K$ is returned with more than a single element.

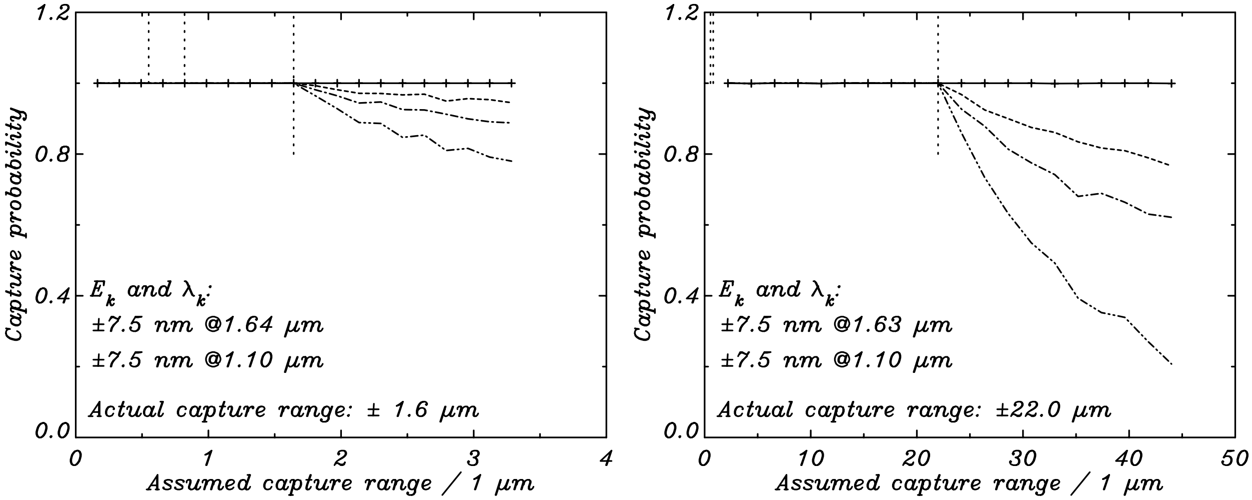
Application of the CSRA to the example in Section 2, Eqs. (4)–(5), yields the capture ranges $P = 1.6$ μm for Case 1 and $P = 22.0$ μm for Case 2, in excellent agreement with the analytical result.

Finding the capture range is a fast process with the CSRA. It is easy to do it for all combinations of available wavelengths and selecting the smallest set that gives a good enough capture range. This can help us e.g. selecting which lasers to purchase. As an example, we show in Table 1 the results for a set of five commercially available HeNe laser wavelengths, $\lambda_k \in \{1.152, 0.633, 0.612, 0.594, 0.543\}$ μm. We first note the three groups of subsets that are separated with horizontal lines in the table. The two bottom groups do not give any extension of the capture range over approximately $\pm\lambda/2$ at all – the difference between the two is that the longest wavelength, λ_1, is not present in the very bottom group. We note then that λ_1 and λ_5 are present in all subsets in the top ranking group and that it does not matter much which two other wavelengths are combined with λ_1 and λ_5. Excluding λ_4 may be the best choice, because it gives us the full factor 16 extension with only four wavelengths, while also allowing a fair amount of extension with only three wavelengths as in the 5th or 6th ranking subsets.

Table 2: Sample capture ranges $\pm P$ calculated for subsets of $\{\lambda_k\}_{k=1}^{10} = \{5.870, 5.855, 4.403, 3.499, 3.349, 3.254, 2.951, 1.917, 1.479, 1.133\}$ μm, a random set of wavelengths. $E_k = \lambda_k/10$. An asterisk in the third column indicates the best subset of a particular size.

Ranking	$P/1\ \mu$m	Subset size	Subset indices k
1	281.61	10*	$\{1,2,3,4,5,6,7,8,9,10\}$
	281.61	9*	$\{1,3,4,5,6,7,8,9,10\}$
3	281.59	9	$\{1,2,3,4,5,6,7,9,10\}$
4	255.10	9	$\{1,2,3,4,6,7,8,9,10\}$
5	237.52	9	$\{1,2,3,5,6,7,8,9,10\}$
	237.52	8*	$\{1,3,5,6,7,8,9,10\}$
7	169.47	9	$\{2,3,4,5,6,7,8,9,10\}$
	169.47	8	$\{2,3,5,6,7,8,9,10\}$
⋮			
14	161.06	7*	$\{2,3,4,5,6,7,9\}$
15	160.85	9	$\{1,2,3,4,5,6,7,8,10\}$
	160.85	8	$\{1,2,3,5,6,7,8,10\}$
⋮			
44	76.70	6*	$\{2,3,4,5,6,7\}$
45	70.15	9	$\{1,2,3,4,5,6,8,9,10\}$
46	70.15	8	$\{2,3,4,5,6,8,9,10\}$
⋮			
134	40.52	5*	$\{2,4,5,6,9\}$
	40.52	5*	$\{1,4,5,6,9\}$
136	40.07	6	$\{3,4,5,7,8,9\}$
⋮			
261	19.96	4*	$\{1,3,5,9\}$
262	19.94	6	$\{2,3,5,8,9,10\}$
	19.94	5	$\{2,3,8,9,10\}$
⋮			
541	8.43	3*	$\{2,3,8\}$
	8.43	3*	$\{1,3,8\}$
543	8.40	7	$\{1,2,3,4,5,6,10\}$
⋮			
689	3.15	2*	$\{2,4\}$
	3.15	2	$\{1,4\}$
691	3.01	7	$\{1,2,5,6,7,8,9\}$
⋮			
1013	0.59	2	$\{9,10\}$

Greater extensions of the capture range are possible with smaller E_k, other choices of λ_k, and with more available wavelength channels. See Table 2 for a large set of random wavelengths (one realization of a uniform distribution be-



(a) (b)

Figure 5: Simulation results for two similar cases with very different capture range, see Section 4.3. (a) Case 1, (b) Case 2. The solid line shows the 100% capture probability for different assumed capture ranges (tic marks). The dash-dot-dotted line shows how the fraction of unique solutions decreases outside the actual capture range. The two remaining curves represent the fractions of correct solutions found with two different strategies for selecting one of multiple solutions. Dash-dotted: pick the smallest solution. Dashed: pick the solution with the largest $\mathcal{E}_{Kn}$. The vertical dotted lines show the actual capture range P (long line) and the individual $\lambda_k/2$ capture ranges (shorter lines) for comparison.

tween 1 μm and 6 μm) and Ref. [5] for examples involving filter wavelengths available on the Next Generation Space Telescope (NGST).

Another type of analysis possible with the CSRA is calculating the capture range for a certain set of available wavelengths as a function of the E_k. This way one can find out what accuracy is needed for the capture range to jump to a useful value.

4.3 Using the CSRA for resolving the ambiguity

We initialize the CSRA with a single-element list, $\mathbf{Q}_0$, whose only element is the capture range expressed as zero plus/minus half the width of the capture range and call it with the full sets of wavelengths, measurements and maximum errors,

$$\mathbf{Q}_K \leftarrow \mathrm{CSRA}(\{\lambda_k\}_1^K, \{\delta p_k\}_1^K, \{E_k\}_1^K, \{0 \pm P\}). \quad (17)$$

The correct solution is guaranteed to be in the returned list $\mathbf{Q}_K$, unless the measurements were off by more than the assumed E_k.

If we use a larger capture range than the correct one, the CSRA may return more than a single solution. We may want to run the CSRA in this mode if we have calculated a certain capture range but find that we cannot guarantee that the real OPL is within that range. The CSRA can then be called with a P given by some maximum error that we can guarantee with the understanding that extra solutions will be returned in addition to the correct one.

Different strategies can be used to select one of the remaining candidate solutions. Picking the smallest solution may appear to be a conservative strategy when we can move e.g. a mirror segment and make a new set of measurements. However, such a strategy could diverge with small steps in the wrong direction. An almost rejected solution would have a small $\mathcal{E}_{Kn}$ (for a rejected solution, it is zero), so we may want to select the solution that maximizes $\mathcal{E}_{Kn}$. For imaging with a segmented telescope, the best method is probably to move the mirror segment to each of the candidate positions and compare the contrasts or Strehl ratios in broadband image data. Finally, we may also try to improve the result by making another set of measurements using the same wavelengths. This new set of data is then used together with the old data, in hope that the measurement errors are different enough (not necessarily smaller) the second time to exclude more candidates.

In Fig. 5 we show simulation results where we have used the CSRA to find the capture range for the example in Section 3, Eqs. (4)–(5). Piston measurements were simulated by randomly selecting 2000 numbers from a uniform distribution between zero and the capture range and reduced to measurements within $\pm\lambda_k/2$ for each k. The CSRA was then used to recover the correct solution. This was done not only for the actual capture range, but also for larger and smaller capture ranges, always generating measurements within the assumed capture range. The correct phase was indeed always recovered, although outside the actual capture range not always as a unique solution. The rate of single solutions drop outside the actual capture range, quicker for Case 2 than for Case 1. The strategy of selecting for large $\mathcal{E}_{Kn}$ outside the actual capture range seems quite promising, particularly in Case 1, where this may be a way of regaining some of the capture range lost with respect to Case 2. See Ref. [5] for examples involving filter wavelengths available on the NGST.

5 Conclusion

We conclude that the CSRA can resolve the modulo 2π ambiguity of monochromatic interferometric OPD measurements uniquely within an extended capture range. The capture range can be predicted by using the CSRA or, for two wavelengths, by using Theorem 1. It can be demonstrated that the capture range can be extended hundreds of times with mild assumptions on the measurement accuracy. This procedure can relax the demands on coarse initial positioning of e.g. segments of a mirror telescope. The CSRA can also be used for selecting a subset of available filters.

If the measurement wavelengths, $\{\lambda_k\}$, can be selected more freely, like when tunable lasers or filters are available, the optimization problem is no longer discrete. Finding the optimal $\{\lambda_k\}$ would require dense sampling because the capture range is a discontinuous function of $\{\lambda_k\}$.

The CSRA can be used outside of its capture range, where it will yield false positives but at least significantly reduce the list of candidate solutions. One can then examine the remaining candidate solutions one by one or try to use additional measurements to reject more of the remaining candidate solutions.

Uncertainties in $\{\lambda_k\}$ could be partly handled by calculating capture ranges for $\{\lambda_k\}$ with appropriate perturbations. However, it would be an improvement of the CSRA if these uncertainties were incorporated into the method, just like the measurement errors.

Generalizations of the CSRA are possible. Because it does not depend heavily on periodicity, it may be useful for broadband WFS methods with quasi-periodic candidate solutions. However, if no bracketing like the one in Eq. (8) can be found, the algorithm will no longer be linear.

Another generalization is possible for applications where the error distributions are known. Rather than forming intersections of maximum error intervals, error distributions for measurements in different wavelengths could be combined by convolution. This would allow ranking of candidate solutions in terms of likelihood. Such a method would resemble the χ^2 method of Refs. [3, 4], although one should retain the CSRA's ability to calculate a capture range.

Acknowledgments

M.G.L. gratefully acknowledges stimulating discussions on the subject of this paper with Gopal Vasudevan of Lockheed Martin Space Systems and Gary Chanan of the University of California, Irvine. We also thank the anonymous referees for their valuable suggestions and comments, that helped us improve the paper. This research was supported in part by Independent Research and Development funds at Lockheed Martin Space Systems, Advanced Technology Center, Palo Alto, California.

A Continued Fractions

The *regular continued fraction* representing any real number can be found through a simple process of repeatedly separating the integer part of a real number and inverting the decimals. As an example, we can write $\pi = 3 + 0.14159\ldots = 3+1/7.06251\ldots = 3+1/(7+0.06251\ldots)$, etc. This results in a unique expansion of the following type,

$$\pi = 3 + \cfrac{1}{7 + \cfrac{1}{15 + \cfrac{1}{1 + \cdots}}},$$

finite for rational numbers and infinite for irrational numbers. The expansion can be truncated at any plus-sign, giving us a sequence of successively better rational approximations (in our example 3, 22/7, 333/106, 355/113, ...) called the *approximants*.

See e.g. Ref. [6] for the theory of continued fractions and proofs of the following properties needed in Section 3 to prove Theorem 1.

P1. The approximants of a real number α are alternately smaller and greater than α. For example,

$$3 < \frac{333}{106} < \cdots < \pi < \cdots < \frac{355}{113} < \frac{22}{7}.$$

P2. For a given $r > 0$, the smallest positive integers a and b for which $|b\alpha - a| < r$ are numerator and denominator of an approximant. For example, $|113\pi - 355| < 0.0001$.

P3. The degree of approximation is given by $|\alpha - a/b| < 1/bd$, where a/b and c/d are consecutive approximants to α. For example, $|\pi - 22/7| = 0.00126\ldots < 1/(7\cdot 106) = 0.00134\ldots$.

References

[1] D. C. Ghiglia and M. D. Pritt, *Two-Dimensional Phase Unwrapping*, Wiley, New York (1998).

[2] C. Polhemus, "Two-wavelength interferometry," *Applied Optics* **12**(9), 2071–2074 (1973).

[3] G. Chanan, M. Troy, F. Dekens, S. Michaels, J. Nelson, T. Mast and D. Kirkman, "Phasing the mirror segments of the Keck telescopes: the broadband phasing algorithm," *Applied Optics* **37**(1), 140–155 (1998).

[4] G. Chanan, C. Ohara and M. Troy, "Phasing the mirror segments of the Keck telescopes II: the narrow-band phasing algorithm," *Applied Optics* **39**(25), 4706–4714 (2000).

[5] M. G. Löfdahl and H. Eriksson, "Resolving piston ambiguities when phasing a segmented mirror," in *UV, Optical, and IR Space Telescopes and Instruments VI*, J. B. Breckinridge and P. Jacobsen, eds., vol. 4013 of *Proc. SPIE*, pp. 774–782 (2000), URL `www.astro.su.se/~mats/offprints/lofdahl00resolving.pdf`.

[6] A. Khintchine, *Continued Fractions*, P. Noordhoff, Groningen, The Netherlands (1963), translated by Peter Wynn.

Mats Löfdahl received a MSc in engineering physics and applied mathematics from the Royal Institute of Technology, KTH (Stockholm, Sweden) in 1989. He completed his PhD in astrophysics in 1996 at Stockholm University. His main research interests are wavefront sensing and image restoration, particularly by use of phase diversity. His current research includes imaging and adaptive optics for solar telescopes as well as phasing of sparse and segmented aperture telescopes.

Henrik Eriksson, after graduating from KTH as an engineering physicist, wrote his thesis in pure mathematics 1967 but soon turned to applied maths and combinatorics. He is now professor in Computer Science at KTH and, which is more important, Mats Löfdahl's father-in-law.