

Resolving Piston Ambiguities when Phasing a Segmented Mirror

Mats G. Löfdahl^a, Henrik Eriksson^b

^aRoyal Swedish Academy of Sciences
Stockholm Observatory, SE-133 36 Saltsjöbaden, Sweden

^bNADA, KTH, SE-100 44 Stockholm, Sweden

ABSTRACT

Wavefront sensing in monochromatic light is insensitive to segment piston errors that are a whole number of waves. If the wavefront sensing is performed in several wavelengths, this ambiguity can be resolved. We give an algorithm for finding the correct phase, given multiple measurements in different wavelengths. Using this algorithm, the capture range of a wavefront sensor can be extended from on the order of $\pm\lambda/2$ in piston to several waves. This relaxes the demands on an initial, coarse alignment method. The extended capture range depends on the selection of wavelengths available for phase measurements and the expected accuracy of the wavefront sensing method used.

Keywords: Segmented mirrors, phasing, piston, phase ambiguity

1. INTRODUCTION

In the literature it is noted that with any monochromatic wavefront sensing (WFS) method, the phase is recovered modulo 2π . When the phase is sensed for every pixel in a pupil map, powerful phase unwrapping methods are used to recover the true shape of the wavefront.¹⁻³ 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 *between* the segments. Although not sensed in the wavefront sensor wavelength, such inter-segment phase errors do degrade the performance when observing in wide-band light or in other wavelengths.

If the WFS is performed in several wavelengths, this inter-segment 2π ambiguity can be resolved using the straightforward algorithm presented in this paper, which we will refer to as the candidate-solution rejection algorithm (CSRA). Using the CSRA, the capture range of the wavefront sensor used can be extended from on the order of half a wave to several waves. We will discuss how this range depends on the selection of wavelengths available for piston measurements and the expected accuracy of the WFS method used. Extending the capture range should translate directly to relaxed demands on the initial coarse alignment method.

The method proposed here should not be confused with multiple color differential interferometry, where two or more wavelengths are combined optically. The CSRA allows you to use your favorite WFS method and combine the measurements in post processing.

The paper is organized as follows. In Section 2 we introduce notation, state the problems of finding the extended capture range and finding the correct piston within that capture range. We also analyze a simple case analytically discussing how the wavelengths should be selected for best capture range and we point out the difficulties in treating the general case. In Section 3 we derive the CSRA and show how it can be used to find the correct solution given a set of measurements and the capture range given a set of filters and error limits. We demonstrate the performance with simulations. In Section 4 we discuss the results.

Email address: mats@astro.su.se.

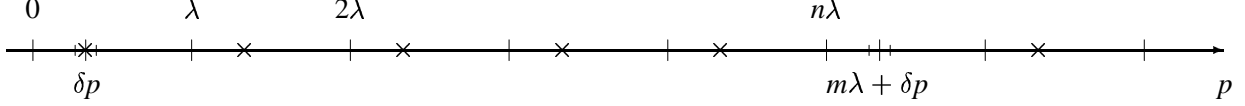


Figure 1. Periodic candidate solutions. The measured piston is δp . The correct solution, $p = m\lambda + \delta p$, is shown within a symmetric error interval. Extra candidate solutions are marked with \times .

2. CAPTURE RANGE EXTENSION PROBLEM

We are phasing a segmented mirror. We use one segment as a reference and measure the relative piston of the others. A measurement can be a single accurate measurement or the sum of consecutive corrections in a closed loop that is allowed to converge.

What we want to measure is a piston p , which is how much a segment is off from the reference. We are using a monochromatic WFS method with a maximum error $|\epsilon| < E$. However, we are really measuring δp , the distance to the nearest whole number of wavelengths off from the reference, so that

$$p = m\lambda + \delta p + \epsilon, \quad (1)$$

where m is an unknown integer number. Figure 1 illustrates the situation, with an error interval centered on the correct solution.

If we have measurements for a set of wavelengths, $\{\lambda_k\}_{k=1}^K$, the situation is different, because the candidate solutions have different periodicity in the different wavelengths. We get

$$p = m_k \lambda_k + \delta p_k + \epsilon_k; \quad \forall k. \quad (2)$$

Because the error intervals for all wavelength channels have to overlap at the correct piston, some candidate n_k can be excluded. Ideally, one would want overlap only at the correct solution but we cannot exclude extraneous 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 we can be sure there is only one solution. So, if the piston is within this capture range, the true piston can be determined unambiguously from the modulo 2π measurements.

For the sake of simplicity we first consider the case of two wavelengths, $\lambda_1 > \lambda_2$. The important parameter is the ratio λ_1/λ_2 . If this ratio has a rational approximation,

$$\frac{\lambda_1}{\lambda_2} \approx \frac{m_2}{m_1}, \quad (3)$$

for some small integers m_1 and m_2 , then $m_1\lambda_1 \approx m_2\lambda_2$ making overlaps this distance apart possible. As illustrated in Fig. 2, the necessary condition is

$$|m_1\lambda_1 - m_2\lambda_2| < 2(E_1 + E_2). \quad (4)$$

For given E_1 and E_2 , we need to find the smallest integers m_1 and m_2 satisfying Eq. (4). The theory of continued fractions⁵ (see Appendix A) can be used to derive a general expression:

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 $|m_1\lambda_1 - m_2\lambda_2| < 2(E_1 + E_2)$.*



Figure 2. The critical case for two wavelengths.

Remark 1 *The expression for the approximation degree tells us that if a/b and c/d are consecutive approximants and $d > \lambda_2/2(E_1 + E_2)$, then the capture range $2P$ will be at least $a\lambda_2 - 2E_2$.*

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 late approximants are better approximations, so $|m_1\lambda_1 - m_2\lambda_2|$ are smaller for a late approximant. Thus, large maximum errors tend to select early approximants, resulting in small capture ranges.

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 want to choose wavelengths in a way that maximizes 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,

$$\text{Case 1: } \begin{cases} \lambda_1 = 1.640 \mu\text{m}; & E_1 = 7.5 \text{ nm} \\ \lambda_2 = 1.100 \mu\text{m}; & E_2 = 7.5 \text{ nm} \end{cases} \quad \text{Case 2: } \begin{cases} \lambda_1 = 1.630 \mu\text{m}; & E_1 = 7.5 \text{ nm} \\ \lambda_2 = 1.100 \mu\text{m}; & E_2 = 7.5 \text{ nm}. \end{cases} \quad (5)$$

We first note that for both cases $2(E_1 + E_2) = 30 \text{ nm}$. 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 $P = 1.66 \mu\text{m}$. 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, $P = 22.19 \mu\text{m}$. We can also use Remark 1: $27 < \lambda_2/2(E_1 + E_2) \approx 36.7 < 53$, so the capture range is given by the second approximant in Case 1 but we have to use a later approximant in Case 2, resulting in a better capture range.

For every pair of wavelengths, finding the capture range is thus a simple matter. 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 too poor to be useful. 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 can be used for both.

3. CAPTURE RANGE EXTENSION SOLUTION

3.1. Candidate-solution rejection algorithm

The CSRA should find a solution p that is consistent with the measurements and error margins in all wavelength channels. It can be implemented as a straight-forward search algorithm, shown as pseudo-code in Fig. 3.

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

-
1. Initialize \mathbf{Q}_0 . Declare k , m , and n as integers.
 2. **for** $k \leftarrow 1$ **to** K **do begin**
 3. $N_{k-1} \leftarrow N(\mathbf{Q}_{k-1})$
 4. $\mathbf{Q}_k \leftarrow \{\}$
 5. **for** $n \leftarrow 1$ **to** N_{k-1} **do begin**
 6. Get q_n and E_n as the n th element in \mathbf{Q}_{k-1} .
 7. Calculate m_{low} and m_{high} from Eq. (7).
 8. **for** $m : m_{\text{low}} < m < m_{\text{high}}$ **do begin**
 9. Calculate q_{new} and E_{new} from Eqs. (8)–(10).
 10. $\mathbf{Q}_k \leftarrow \mathbf{Q}_k + \{q_{\text{new}} \pm E_{\text{new}}\}$
 11. **end**
 12. **end**
 13. **end**
 14. The non-rejected solutions are in \mathbf{Q}_K .
-

Figure 3. The CSRA in pseudo code. The initialization and procedure are different when searching for a solution (see Section 3.1) and when searching for the capture range (see Section 3.2).

a list $\mathbf{Q}_k = \{q_{kn} \pm \mathcal{E}_{kn}\}_{n=1}^{N_k}$, where q_{kn} is the n th candidate piston with a maximum error \mathcal{E}_{kn} . 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 solutions. 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 \mathcal{E}_{Kn} is the intersection of the overlapping error intervals. If we used the correct capture range, $N_K = 1$.

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 m . For this to be the case, we require that the low limit of one interval be smaller than the upper limit of the other and vice versa, i.e.

$$(q_{(k-1)n} - \mathcal{E}_{(k-1)n} \leq \delta p_k + m\lambda_k + E_k) \wedge (q_{(k-1)n} + \mathcal{E}_{(k-1)n} \geq \delta p_k + m\lambda_k - E_k) \quad (6)$$

or, expressed as an interval for m ,

$$\frac{q_{(k-1)n} - \mathcal{E}_{(k-1)n} - \delta p_k - E_k}{\lambda_k} = m_{\text{low}} < m < m_{\text{high}} = \frac{q_{(k-1)n} + \mathcal{E}_{(k-1)n} - \delta p_k + E_k}{\lambda_k}. \quad (7)$$

For each integer m that meets this requirement, there is a non-rejected interval, $q_{\text{new}} \pm \mathcal{E}_{\text{new}}$, where

$$q_{\text{new}} = \frac{q_{\text{high}} + q_{\text{low}}}{2} \quad \text{and} \quad \mathcal{E}_{\text{new}} = \frac{q_{\text{high}} - q_{\text{low}}}{2} \quad (8)$$

and the low and high limits of the overlap interval are

$$q_{\text{low}} = \max\{q_{(k-1)n} - \mathcal{E}_{(k-1)n}, \delta p_k + m\lambda_k - E_k\}, \quad (9)$$

$$q_{\text{high}} = \min\{q_{(k-1)n} + \mathcal{E}_{(k-1)n}, \delta p_k + m\lambda_k + E_k\}. \quad (10)$$

Except for the initial interval, Eq. (7) will usually be satisfied by zero or one integer value of m . Therefore the CSRA is a linear algorithm.

3.2. Extended 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, $\{E_k\}$.

We force a match at the origin by setting all $\delta p_k = 0$. We then step m_1 (where λ_1 is the longest wavelength) through $m_1 = 1, 2, \dots$ until we have overlap for all wavelengths. That is, for each value of m_1 , we initialize $\mathbf{Q}_0 \leftarrow \{m_1 \lambda_1 \pm 2E_1\}$ and run the CSRA through the wavelengths $\{\lambda_k\}_{k=2}^K$ until we find an m_1 for which \mathbf{Q}_K is not empty. Note that we use twice the real maximum errors for this step, because otherwise we implicitly assume that the centers of the matching intervals have to be perfectly aligned at the origin. To avoid inventing more notation at this point, we note that the CSRA should be called with the following substitutions: $E_k \leftarrow 2E_k$ and $\{\lambda_k\}_{k=1}^K \leftarrow \{\lambda_k\}_{k=2}^K$.

The capture range is then equal to this distance, corrected with the error of the non-rejected solution, or

$$2P = q_{K_n} - \mathcal{E}_{K_n}. \quad (11)$$

We have used the CSRA to find the capture range for the example in Section 2, Eq. (5). The capture ranges found were $P = 1.6 \mu\text{m}$ for Case 1 and $P = 22.0 \mu\text{m}$ for Case 2, in excellent agreement with the mathematical result.

Finding the capture range is such a fast process with the CSRA, that it is easy to select the best subset of available filters. Just run the above process for all combinations of filter wavelengths and list them in order of capture range. Pick the smallest set of filters that give a good enough capture range. In Table 1(a) we show the results for the NGST broadband filters known to us at the time of preparing this presentation. These filters are centered at 4.8, 4.2, 3.4, 2.1, and 1.6 μm , respectively, and the FWHM $\approx 20\%$ of the central wavelength.⁴ Note that 4.2 is a multiple of 2.1 so you do not gain anything by using both. In fact, 3 filters are enough to extend the capture range to almost $\pm 17 \mu\text{m}$, an improvement by a factor 7 from using only the longest wavelength.

In Table 1(b) we show a larger set of filter wavelengths, that was drawn from a random uniform distribution between 1 and 6 μm . Lacking information on the full set of NGST filters, we want to demonstrate what we can expect from wavelengths that were not selected for good performance with the CSRA. In this example, it is possible to get a factor 95 improvement in capture range, compared to the longest wavelength. With a list like this, it is easy to search through the best results for each subset size, N , and pick the smallest subset that has a good enough capture range.

3.3. Finding the correct piston

We initialize the CSRA with a single-element list whose only element is the capture range expressed as zero plus/minus half the width of the capture range, $\mathbf{Q}_0 = \{0 \pm P\}$. After running the CSRA, the correct solution has to remain in the list, or the measurements were off by more than the error ranges.

If we use a larger capture range than the correct one, we may end up with 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 piston is within that range. We can then run the CSRA with a P given by some maximum error that we *can* guarantee and accept that we will be given false solution *in addition* to the correct one. Different strategies can be used to select one of the remaining candidate solutions. Picking the smallest phase may be a conservative strategy but can still give a correction in the wrong direction. An almost rejected solution would have a small remaining error limit so, we may want to select the solution that maximizes \mathcal{E}_{K_n} . We test both in the simulations below. If possible, the best method is probably to move to each of the candidate pistons and compare wide-band image data; PSF or MTF if we have access to a point source or contrast comparison if not. 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, hoping that the measurement errors are different enough (not necessarily smaller) the second time to exclude more candidates.

Table 1. Capture ranges $\pm P$ calculated for all N -element subsets of $\{\lambda_k\}_{k=1}^K$. The best result for each N is indicated with an asterisk.

(a) $\{\lambda_k\}_{k=1}^5 = \{4.8, 4.2, 3.4, 2.1, 1.6\}$ μm . Maximum errors $\pm\lambda_k/14$.

Ranking	$P/1 \mu\text{m}$	N	Subset indices, $\{k\}$
1	16.76	5*	{1, 2, 3, 4, 5}
	16.76	4*	{1, 3, 4, 5}
	16.76	4*	{1, 2, 3, 5}
	16.76	4*	{1, 2, 3, 4}
	16.76	3*	{1, 2, 3}
6	16.69	4	{1, 2, 4, 5}
7	10.35	4	{2, 3, 4, 5}
8	8.69	3	{2, 3, 5}
9	8.26	3	{2, 3, 4}
10	7.20	3	{1, 4, 5}
11	5.10	3	{1, 3, 4}
12	4.86	3	{1, 3, 5}
	4.86	2*	{1, 3}
14	4.05	3	{2, 4, 5}
15	3.16	3	{3, 4, 5}
	3.16	2	{3, 4}
17	2.29	3	{1, 2, 5}
	2.29	2	{2, 5}
	2.29	2	{1, 5}
20	2.06	3	{1, 2, 4}
	2.06	2	{1, 4}
	2.06	2	{1, 2}
23	1.95	2	{2, 4}
24	1.80	2	{2, 3}
25	1.49	2	{3, 5}
26	0.90	2	{4, 5}

(b) $\{\lambda_k\}_{k=1}^{10} = \{5.870, 5.855, 4.403, 3.499, 3.349, 3.254, 2.951, 1.917, 1.479, 1.133\}$ μm . Maximum errors $\pm\lambda_k/10$.

Ranking	$P/1 \mu\text{m}$	N	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}

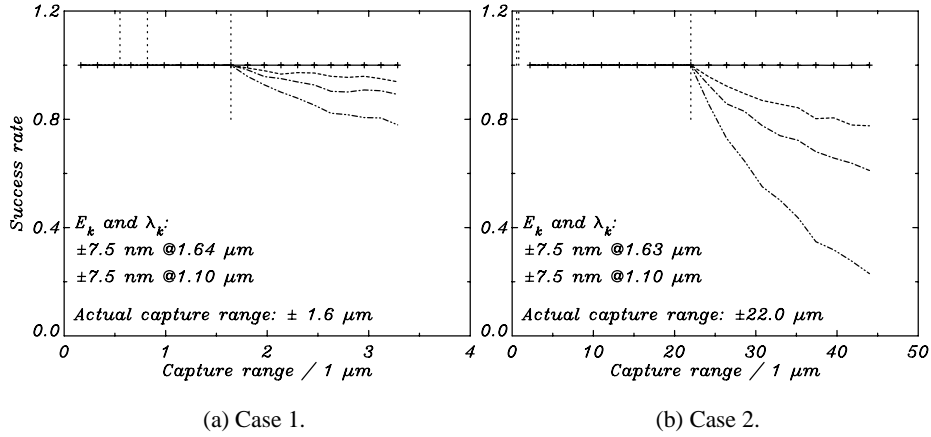


Figure 4. Success rate for two similar cases with very different capture range, see Eq. (5) in Section 2. See Section 3.3 for an explanation of the diagrams.

In Figs. 4–6 we show simulation results where we have used the CSRA to find the capture range for different sets of filters and maximum errors. Then we simulated piston measurements by randomly selecting 2000 numbers from a uniform distribution between zero and the capture range, reducing them to measurements within $\pm\lambda_k/2$ for each k , and then used the CSRA 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. In each diagram, a solid line (with tic marks indicating assumed capture range) shows the rate of success in the sense that the correct phase was indeed recovered, although not necessarily as a unique solution; a dash-dot-dotted line shows the rate of correct unique solutions. The dash-dotted line represents cases when the smallest solution is correct and the dashed line represents cases where selecting the largest error interval gives the right answer. Also, there are vertical dotted lines that show the capture range P (long line at unity) and the individual $\lambda_k/2$ capture ranges (shorter lines) for comparison.

First, we again use the example in Section 2, Eq. (5). The results are shown in Fig. 4. Note the 100% success rate within the actual capture range and the quicker drop outside the capture range for Case 2. The strategy of selecting for large error intervals 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.

With the known broad-band filter wavelengths of the NGST, we demonstrate the effect of changing the maximum errors in Fig. 5. It is evident that this affects both the capture range and the performance outside of the capture range. In Fig. 6 we show the results using subsets of the NGST filters. Compare with Table 1.

4. DISCUSSION

We conclude that the CSRA can resolve the modulo 2π ambiguity of monochromatic piston measurements uniquely within an extended capture range. The capture range can be predicted by using the CSRA. Using this procedure can relax the demands on coarse initial phasing. The CSRA can also be used to make recommendations for filter subset selection, minimizing the number of filters that have to be used.

The CSRA can be used outside of its capture range, where it significantly reduces 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.

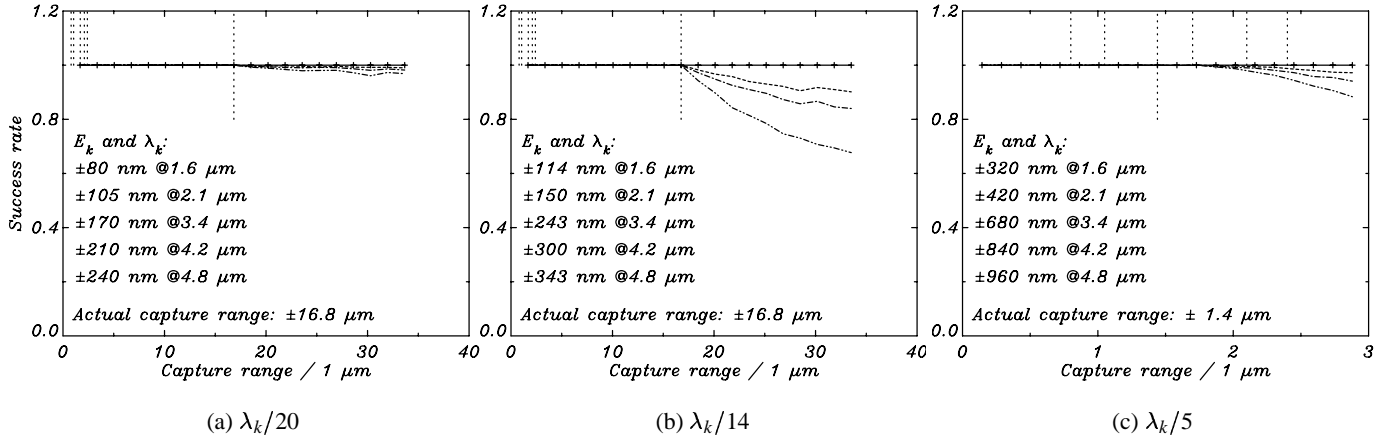


Figure 5. Success rate for NGST broadband wavelengths. Different maximum errors. See Section 3.3.

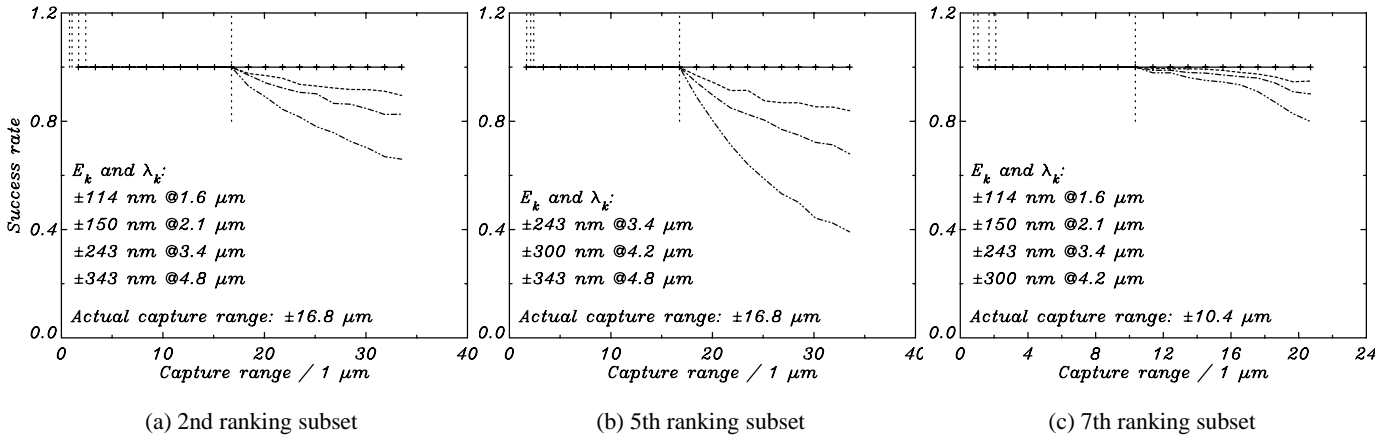


Figure 6. Success rate for NGST broadband filter wavelengths; $E_k = \lambda_k/14$. Using subsets of the NGST broad-band filter with different ranking, see Table 1. Compare with 1st ranking full set in Fig. 5(b). See Section 3.3.

Generalizations of the CSRA are possible. E.g., it does not depend heavily on periodicity. As long as we can make lists of candidate solutions for each wavelength, we can use the CSRA to narrow down the list. This may be useful for wide-band WFS methods, that may result in quasi-periodic candidate solutions, however the CSRA will no longer be linear. For some applications it could make sense to use a formulation based on expected error distributions rather than maximum error intervals. This may allow ranking of candidate solutions in terms of likelihood.

ACKNOWLEDGMENTS

This research was supported in part by Independent Research and Development funds at Lockheed Martin Space Systems, Advanced Technology Center, Palo Alto, California

REFERENCES

1. R. G. Lyon, “DCATT wavefront sensing and optical control study,” Tech. Rep. WFSC-0001, NASA/GSFC, February 22 1999.
2. M. G. Löfdahl, R. L. Kendrick, A. Harwit, K. E. Mitchell, A. L. Duncan, J. H. Seldin, R. G. Paxman, and D. S. Acton, “A phase diversity experiment to measure piston misalignment on the segmented primary mirror of the Keck II telescope,” in *Space Telescopes and Instruments V*, P. Y. Bely and J. B. Breckinridge, eds., vol. 3356 of *Proc. SPIE*, pp. 1190–1201, 1998.
3. D. C. Ghiglia and M. D. Pritt, *Two-Dimensional Phase Unwrapping*, Wiley, New York, 1998.
4. J. Trauger, K. Stapelfeldt, R. Sahai, D. Backman, C. Beichman, G. Djorgovski, M. Ealey, E. Gaidos, C. Grillmair, Y. Gürsel, S. Kulkarni, R. Lyon, S. Macenka, C. Martin, M. Ressler, M. Shao, and M. Werner, “High-contrast origins science for NGST,” Pre-phase A science instrumentation concept study report for the Next Generation Space Telescope (NGST): in response to NRA 98-GSFC-1, NASA/GSFC, September 1 1999.
5. A. Khintchine, *Continued Fractions*, P. Noordhoff, Groningen, The Netherlands, 1963. Translated by Peter Wynn.

APPENDIX A. CONTINUED FRACTIONS

We refer to Ref. 5 for the theory of continued fractions and proofs to the elementary results that we quote in this appendix.

The *regular continued fraction* representing a real number is an expression of the following type,

$$\pi = 3 + \frac{1}{7 + \frac{1}{15 + \frac{1}{1 + \dots}}} \quad (12)$$

Every real number has a unique expansion of this kind, finite for rational numbers and infinite for irrational numbers. The expansion can be truncated at any plus-sign, giving us successively better rational approximations (in our example 3, 22/7, 333/106, 355/113, ...) called the *approximants*. The following facts will be important to us:

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

$$3 < \frac{333}{106} < \dots < \pi < \dots < \frac{355}{113} < \frac{22}{7}. \quad (13)$$

- For two consecutive approximants a/b and c/d , the relation $|ad - bc| = 1$ holds. For example, $22 \cdot 106 - 7 \cdot 333 = 1$.
- 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, $|\pi - 22/7| = 0.00126\dots < 1/(7 \cdot 106) = 0.00134\dots$
- For a given $r > 0$, the smallest positive integers m_1 and m_2 for which $|m_2\alpha - m_1| < r$ are numerator and denominator of an approximant. For example, $|113\pi - 355| < 0.0001$.

The algorithm for finding the continued fraction for any real number is simple. Again, we take π as the example:

Subtract integer part:	$\pi = 3 + 0.14159\dots$
Invert:	$\pi = 3 + \frac{1}{7.06251\dots}$
Subtract integer part:	$\pi = 3 + \frac{1}{7 + 0.06251\dots}$
Invert:	etc.