1999

The separation of multiple mutually interfering FM signals and the use of predistortion to compensate for nonlinear RF amplifiers

Sven Anders Mattsson
Iowa State University

The separation of multiple mutually interfering FM signals and the use of predistortion to compensate for nonlinear RF amplifiers

by

Sven Anders Mattsson

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Electrical Engineering (Communications and Signal Processing)

Major Professor: Steve F. Russell

Iowa State University

Ames, Iowa

1999

UMI Number: 9977378

Graduate College
Iowa State University

This is to certify that the Doctoral dissertation of

Sven Anders Mattsson

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program

Signature was redacted for privacy.

For the Graduate College

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

My dissertation research covers two signal processing problems originally presented by Rockwell-Collins in Cedar Rapids, the research sponsor. Both problems are aimed at improving the performance of communication systems. The first problem is separation of three FM signals, and the second problem is pre-distortion for compensation of nonlinear amplifiers. My first project at Iowa State University was about cross coupled phase locked loops (CCPLLs) for separation of two FM signals. That project had me intrigued with the possibility of separating three FM signals, which is the emphasis of this dissertation. This naturally extends to the possibility of separating more than three FM signals. Most results presented in this dissertation also apply to the case of an arbitrary, but finite, number of FM signals. The second problem also results from research done on behalf of Rockwell-Collins and, simply put, deals with pre-shaping the signal before it is sent to the amplifier so that the amplifier output is a perfect replica of the original signal. The need for pre-distorting the signal arises mainly from the fact that *high efficiency* and *linearity* do not go well together. This work has been highly successful because it has shown a way of implementing high-order pre-distorters for a broad class of amplifiers without having to directly work with the usual, but complicated approach, Volterra series.

The work on separation of three FM signals is a continuation of my Masters's thesis, which dealt with the separation of two signals. At first glance, the separation of three or more mutually interfering signals might be perceived as a small extension to my work done on separating two FM signals, [22], [23], this is a significantly more difficult problem because the shortcut that can be used when separating two signals fails when the number of signals increases beyond this limit. Also, the sufficient conditions that have been found are given in the terms of maximum constraints, i.e., the maximum phase error must be suppressed by a certain factor, while the filtering of the received phase only guarantees a suppression in the mean-squared sense. Put differently, the requirements are given in the L^∞ norm while the known properties are in the L^2 norm, leading to a problem because the L^∞ norm is not easily related to the L^2 norm. Note that convergence in the L^∞ norm and L^2 norm are equivalent only if one works with functions that are continuous and "live" on compact sets, i.e., [1] $f \in C_c$ and the convergence is to a continuous function. In other words, the only converging sequences that are considered are those that converge to a continuous function, i.e., only working with the common subset C_c of the L^∞ and L^2 spaces. However the following is false $||f||_2 < ||g||_2 \Leftrightarrow ||f||_\infty < ||g||_\infty$, which can be formulated in electrical engineering terms as saying that the maximum value of a signal does not tell us anything about the energy of the signal and vice versa. Still, the research has shown for finite number of FM signals that a stable solution exists, and the research gives a way to find this solution. It also presents a strong case for the uniqueness of the solution although a thorough proof is not presented for the general case. The conclusion is supported by the fact that the simulations never found any other solution than the known one. To show this, some intermediate results were needed. The first was to show how the amplitudes of the signals can be found. The second one was to derive an expression for the received phase that shows that there is a capture effect, which was needed when showing stability. Because the way to find the three signals that are guaranteed to work is unpractical because of computational complexity, a simpler method was derived. This simpler method to separate three FM signals works extremely well, even when the weakest signal is 80 dB below the strongest signal, but it is not absolutely guaranteed to always separate the signals. Fortunately it is possible to verify if the method was successful by checking the residual signal.

Pre-distortion is actually a theoretically well researched area, although it is not widely known. The key element in the theory is the Volterra series representation of nonlinear time invariant systems, including linear systems. A result of this theory is that, given the Volterra series representation, the inverse (if it exists) can be found from the Volterra series representation. This inverse, which will work as a pre-distorter, is given as yet another Volterra series. To get a different perspective, this is the same as a linear systems where the inverse of the linear filter is yet another linear filter. There are unfortunately two major problems associated with Volterra series. The first problem is that although it is straight forward to measure them, the measurements are extremely complicated. The second problem is that implementing a Volterra series based pre-distorter is numerically intense. As the order of the pre-distorter increases, the numerical problems become overwhelming. The reason that the high-order terms are hard to implement is that the n'th term in the Volterra series is an n dimensional function. This function is then convolved n times with the input to produce an output. Assuming that one wants to store an n dimensional function in a computer using, say 100 samples per dimension, then the required storage space is 10^{2n} samples. In other words, the required memory grows exponentially.

[1] The notation $f \in C_c$ implies that f is continous and $f(x) = 0$ when $|x| > a$ for some $a < \infty$.

To further complicate things, for each output at each time, there is an nth-order integration that must be carried out. For these reasons, it is not reasonable to implement anything larger than third-order pre-distorters.

In the pre-distorter research, high order pre-distorters have been found for a fairly wide class of amplifiers by circumventing the Volterra series approach. Because the implementation of these pre-distorters is very simple, there are no practical limitations on the order. For example, an 11th-order pre-distorter is easily implemented. This work resulted in a paper submitted to the *IEEE Transactions on Communications* [31]; I have decided to include it in its entirety in the dissertation instead of presenting the material "side-by-side" with the work on separation of FM signals.

Dissertation organization

Most material related to pre-distortion is in Chapter 10 on page 76; besides this, the organization is fairly traditional. The dissertation starts with an introduction followed by a literature survey, chapter 2, which mainly deals with signal separation. After this comes chapter 3 with the problem statements. This is followed by chapters 4-8 that contain the main theoretical results that I have found. Chapter 4 shows how the amplitudes of the three signals are found from the received signal. Chapter 5 finds a series expansion for the received phase. Chapter 6 uses these results to show that the problem has a unique solution and that this solution is stable. Chapter 7 derives and discusses some sufficient and necessary conditions for successful separation. Finally, Chapter8 shows a method that will always separate any three, or more, FM signals in a finite amount of time. Unfortunately it is too computationally intensive to be of practical use, at least in its present state. Chapter 9 presents an alternative way of separating the three signals, which has the advantage of being practically feasible to implement, but it will not guarantee the separation of the three signals. Chapter 9 also presents the simulation approaches and results, and has a discussion about practical considerations. Chapter 10 about pre-distortion is a paper that Dr. Steve F. Russell, my major professor, and I submitted to IEEE Transactions on Communications about pre-distortion. It documents past research, our contribution with respect to theoretical results, and ends with simulation results. The final two chapters are future work and conclusions; both chapters deal with signal separation and pre-distortion.

2 LITERATURE SURVEY

Interference rejection has received considerable, and constant, attention during the last 20 years. The published articles can be broadly put into two categories. In the first category, the authors use multiplication and/or subtraction techniques, [1], [2], [4]. In the second category, they use cross-coupled phase-locked loops (CCPLLs), [10]-[13]. Besides these, some articles also deal with specific problems related to interference rejection, [3], [6]-[9]. One article, [14], uses extended Kalman filtering for separation. Unfortunately, no article gives firm rules for when and how two or more signals can be separated/detected. All methods presented depend on the capture effect in the case of two interfering signals; unfortunately, they all lack a thorough description of the capture effect. Figure 2.1 shows a geometrical representation of how the two FM signals S_1 and S_2 add up the received signal R.

The articles in the first category generally assume an adjacent channel interferer, and most of the proposed methods break down in the case of a co-channel interferer. The chosen method is to use a "beat detector" followed by a lowpass filter to produce a signal whose phase is the phase difference between the two signals. A second signal is produced by hardlimiting the first signal. This is where the capture effects comes in. It is assumed that the output from the hardlimiter is essentially the stronger signal with a small phase error caused by the weak signal (this is not true if the capture effect is weak). These two outputs are used for separating the two input signals, or alternatively, for enhancing one of the signals. One inherent weakness is that the beat detector will only work in the desired way when the frequency difference between the two carriers is sufficiently large (approximately the bandwidth of the signals) to avoid frequency folding. This makes these methods unusable for co-channel interference. The hardlimiting/multiplication method has one advantage; it does not depend on subtracting signals from each other. This eliminates the need for amplitude estimation, which simplifies the scheme. Furthermore, [1] mentions that the hardlimiter is a "conventional hardlimiter-discriminator," implying a capture effect that is not described in the article. It is somewhat surprising that much effort was expended deriving very long, complicated expressions for some optimal filters that were based on simplified approximations of the behavior of the hardlimiter-discriminator and beat

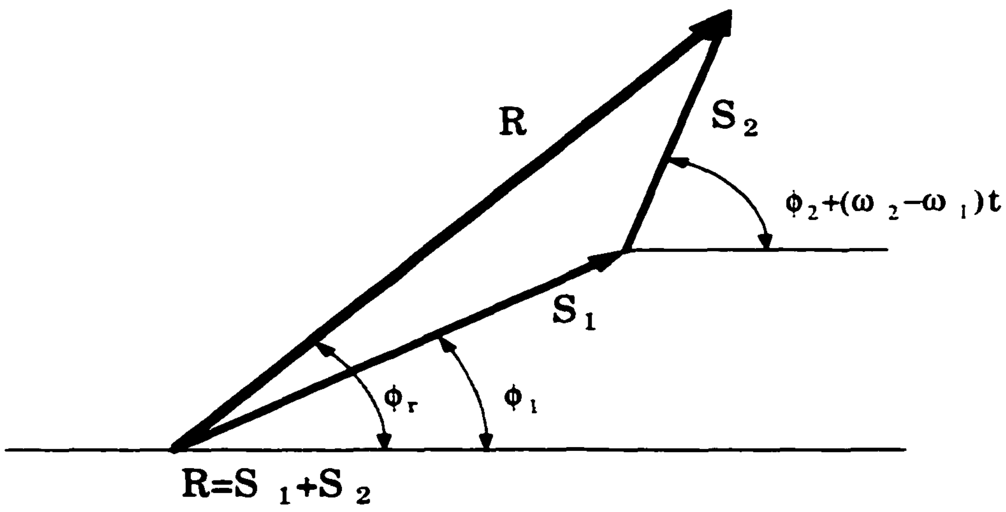


Figure 2.1 Vector representation of the received signal.

detector.

In the second category, cross-coupled phase-locked loops (CCPLLs) are used. CCPLLs are shown in Figure 2.2. S_1 and S_2 are the received signals, and θ_{1est} and θ_{2est} are the phase estimates of the two signals. The CCPLLs work by having the first PLL (PLL1) track the stronger signal (S_1). This relies on the capture effect. The output from the VCO is shifted by an additional 90^o, producing a signal estimate that is 180^o out of phase, which is added to the input. This process is equivalent to subtracting the estimate from the received signal ($S_1 + S_2$), leaving a residual signal that is supposedly dominated by S_2.

The second PLL (PLL2) tracks the weaker signal, which is subtracted at the input of the first PLL, etc. Simulations, both computer and experimental [10]-[13], show that sometimes CCPLLs can separate mutually interfering signals. If the two signals are of roughly the same magnitude, the separation fails. Moreover, if the magnitude differs by more than 20 to 30 dB, the separation fails. None of the articles explain why there is only a limited range when separation is possible.

It is well known that a PLL does give a capture effect. However, none of the published articles on CCPLLs tries to describe the capture effect, or use this to explain why and when CCPLLs succeed in separating FM signals. The term cross-coupled phase-locked loop is actually somewhat misleading. Cross-coupling two phase-locked loops will not work, but given the phase estimate from the first PLL and an estimate of the amplitude, an estimate of the stronger signal is obtained, which can be sub-

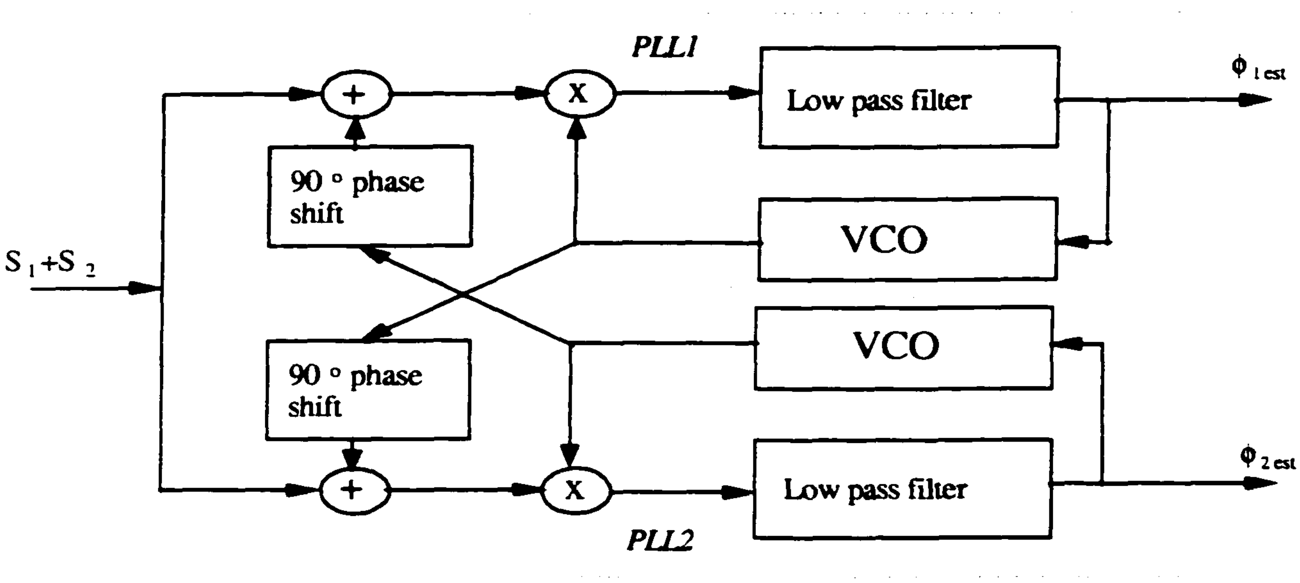


Figure 2.2 Cross-Coupled Phase-Locked Loops, (CCPLLs).

tracted from the received signal. A similar process applies for the weaker signal. This implies that the amplitude must be properly estimated for separation of the two signals to take place. In [13], CCPLLs with amplitude estimations are simulated by basically lowpass filtering the received amplitude. The simulations only show that it can work; no results showing the limitations were published.

It seems that the focus on PLLs as phase estimators is one of the reasons why only elementary descriptions of the capabilities of CCPLLs are available. Describing the CCPLLs mathematically results in nonlinear, coupled differential equations. The order of the equations is determined by the chosen filters. The regions of successful separation (stability) can be found by simulations. Only results from simulations using first and second order filters have been published. These results are of limited value. There are at least seven parameters; two for each second order filter, the modulation index for the two signals, and the amplitude ratio of the two signals. This excludes the amplitude estimation scheme that would introduce at least two more variables. Due to this complexity, it is impractical present simulation results for the huge number of parameter combinations for which two FM signals can be separated using CCPLLs.

The only article that tries to describe the capture effect is [3]. The authors' approach is very straight forward; the phase of the received signal as a function of the two signals is derived using simple geometry and expanded in a series. In finding a useful expression for the received phase, they fail to realize how the capture effect works. In addition to this, their following conclusions are incorrect ([3], page 533). "If this signal is supplied to a frequency demodulator, at the output of the demodulator only the modulation of the stronger signal will be heard...." and " The modulation of the weaker signal is only perceptible if $\omega_1 = \omega_2$." In the first quotation, the term signal refers to the phase of the received signal, and in the second quotation, ω_1 and ω_2 refers to the carrier frequencies. It will be shown that neither of these statements is true. Except for their mistaken conclusions, they do find an expression for the received phase that is very useful. The article also shows that the expected value of the received phase is that of the stronger signal. This is useful for carrier frequency estimation. No article showed that the amplitudes of the signals should be found, although one article, [13], did implement a scheme for amplitude estimation. This scheme relied on the capture effect to work, i.e., if the capture effect is strong, it will work. It takes as the amplitude estimate the part of the received signal that is in phase with the estimated phase.

It is my opinion that the reason that no theoretical results describing the behavior of CCPLLs have been published is because the actual equations governing the CCPLLs are extremely hard to work with. Considering that no theoretically complete description of a PLL with filters in the feedback loop of higher order than four are available, it is not surprising that a theoretical description of the CCPLLs is lacking. In my Master's thesis, [23], as well as in this dissertation, I look at the received phase and not at what happens in a PLL. These results can then be applied to the CCPLLs, since a PLL is, in essence, a phase estimator, albeit not the optimum one.

My Master's thesis used the expression from [3] and [19], namely:

$$\theta_R = \theta_1 - \sum_{n=1}^{\infty} \left(-\frac{A_2}{A_1}\right)^n \frac{1}{n} \sin\left(n\left(\theta_2 - \theta_1\right)\right) \qquad \left|\frac{A_2}{A_1}\right| < 1 \tag{2.1}$$

which states that in the case of two interfering signals $S_p = A_p e^{i\theta_p}$, $p \in [1, 2]$, the received phase is that of the stronger signal and the expected value of the received phase is also the expected value of the stronger signal's phase. More importantly, the error terms are in themselves FM signals which the authors failed to report in [3]. This is easily recognized by looking at the n-th term, $\left(-\frac{A_2}{A_1}\right)^n \frac{1}{n} \sin\left(n\left(\theta_2 - \theta_1\right)\right)$, and noting that $\sin\left(n\left(\theta_2 - \theta_1\right)\right)$ is an FM signal. Also note that as n increases, the modulation index increases, resulting in progressively more wideband terms. The same definition of the modulation index as in [16] will be used, namely defining it as the ratio between the 3dB bandwidths of θ and $\sin(\theta)$.

Equation 2.1 governs under what conditions CCPLLs work and do not work. If the original signals are heavily modulated, filtering the received phase will efficiently suppress the phase error, resulting in a very good estimate of the phase. Using the received phase to form a phase estimate and the known, or estimated, amplitude of the stronger signal, to form an estimate of the stronger signal and subtracting this from the received signal produces a new signal in which the second signal dominates. Equation 2.1 shows that the previous procedure can now be repeated to form a good estimate of the second signal's phase and therefore a good estimate of the second signal.

If the signals have little modulation, i.e., a low modulation index, then $\sin(n(\theta_2 - \theta_1))$ is not very spread out in frequency compared to θ_1. In this case, the capture effect is not "strong" enough to guarantee that the second signal dominates when the estimate of the stronger signal is subtracted from the received signal. For example, if the two signals are just two carriers at the same frequency but with different phases, there is now capture effect, and subsequently, the CCPLLs cannot separate the signals.

Another fact that follows from Equation 2.1, is that, as the ratio of the amplitudes $\left|\frac{A_2}{A_1}\right|$ approaches unity, more and more error terms start to significantly contribute to the error, making it increasingly harder for the CCPLLs to separate the signals. In the extreme case when $\left|\frac{A_2}{A_1}\right| = 1$, Equation 2.1 is no longer valid: instead one can use simple geometry to show that the phase is simply:

$$\theta_R = \frac{\theta_2 + \theta_1}{2} \qquad \left|\frac{A_2}{A_1}\right| = 1 \tag{2.2}$$

Where we have assumed that the signals are at the same carrier frequency. If the two phases θ_2 and θ_1 have the same bandwidths, there will obviously be no capture effect, and the two signals cannot be separated using CCPLLs. No published papers, except my own work,[22] and [23], relating to CCPLLs, have been able to demonstrate the separation of two signals when the amplitudes are equal or to explain why that happens.

Another practical problem with CCPPLs is that the feedback structure introduces delays. The way to avoid this is to cascade the estimators[1], [20]-[23]. This structure can compensate for any delays and is superior to CCPLLs (Figure 2.3).

It turns out that any two FM signals can be separated, except for a possible phase ambiguity [23] (Figure 2.4). Once more, by forgetting about CCPLLs and looking at the signals, it was shown in [23] that the amplitudes of the two FM signals can be found if the distribution of the phase difference is known (assuming the phases to be uniformly distributed on the interval $(0, 2\pi]$ which is very reasonable). The received signal can be seen as the vector sum of the two signals (essentially working in a two-dimensional signal space, corresponding to I and Q components) (Figure 2.1). Looking at the problem from a geometrical viewpoint, it involves finding the angles of a triangle when the three sides are known.

[1] To my knowledge, D. L. Abbey at Rockwell-Collins, Cedar Rapids, invented this.

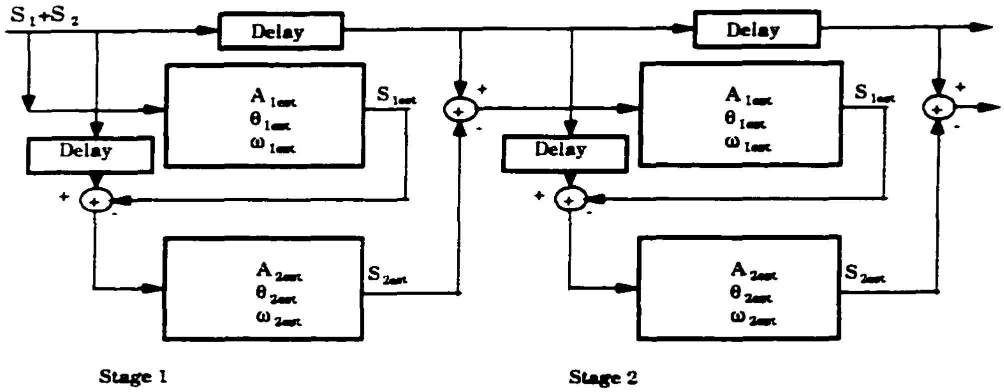


Figure 2.3 Cascaded estimators.

The angles are given by:

$$\cos(\theta_R - \theta_1) = \frac{R^2 + A_1^2 - A_2^2}{2RA_1} \quad (2.3)$$

$$\cos(\pi + \theta_1 - \theta_2 - (\omega_1 - \omega_2)t) = \frac{R^2 - A_1^2 + A_2^2}{2RA_2} \quad (2.4)$$

The ambiguity arises because we only know the cosine of the angular difference (see Figure 2.4). It

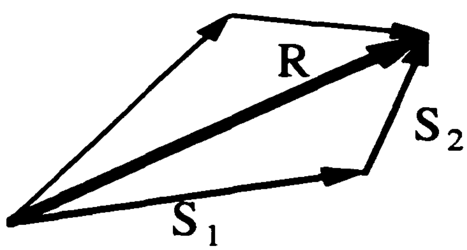


Figure 2.4 The phase ambiguity.

was not conclusively shown in my Master's thesis, [23], that this ambiguity can be overcome; however, some results from this dissertation will show that it is possible to pick the right solution by requiring that the phases be bandlimited.

The CCPLLs can be seen as estimators, so it makes intuitive sense to use an optimum estimator to find the two signals. The only article to try this approach, using extended Kalman filtering to separate the signals, is Reference [14]. It is well known that for a linear dynamical system corrupted by Gaussian noise, the optimum estimator is the Kalman filter, [24] and [25], sometimes referred to as the Kalman-Bucy filter when treated from a stochastic differential equations form, [25]-[29]. If the system is not linear, one can use the extended Kalman filter as done in [14]; however, there is no guarantee that this filter is optimum. This is an interesting approach, but it does not seem to perform too well in practice. The results shown in the article are far from ideal, for example the extended Kalman filter does not manage to properly estimate the amplitudes of the two signals (Figure 8 on page 1379 in [14]). As mentioned previously, it is possible to correctly estimate the amplitudes.

Some practical problems associated with CCPLLs and cascaded estimators were mentioned in my Master's thesis. The problem with delays in CCPLLs have already been discussed; this problem can be avoided in cascaded estimators. The fact that any practical filter will be non-ideal will cause problems. Because of, this there will be distortion on the estimate of the stronger signal even when the second signal is not present. Subtracting this estimate from the received signal will leave a residue, due to the non ideal filter, even though there theoretically should not be any. Obviously, any second signal that is weaker than this self-induced residue will be "drowned" by it, i.e., the second estimator (PLL) will lock to this residue and not to the second, weaker signal. Therefore, there is always a practical lower limit on how much weaker the second signal can be relative the stronger signal depending on the quality of the signal processing implementation. This is evident in all published results for CCPLLs, where the demodulated second signal suffers from more and more distortion as it gets weaker.

An example from [23] shows successful separation of two signals using cascaded estimators (Figure 2.5). The modulation index, defined as the ratio between the bandwidths of the modulated signal and the phase signal, is 5. The measure SDR_2 is the distortion on the modulated signal; $SDR_{2\,PM}$ is the distortion on the phase and $SDR_{2\,PM\;mod}$ is the distortion on the phase when compensated for phase slips. Finally, SDR_{2FM} is the distortion on the phase derivative, i.e., the important figure of merit for an FM signal.

Lately, much work has been focused on interference rejection in CDMA systems; this is an area that this dissertation does not directly deal with since the separation in this case is generally based on using the "built-in" correlation properties of the signal and therefore forms a special case. However, the ideas related to separation of FM signals still apply.

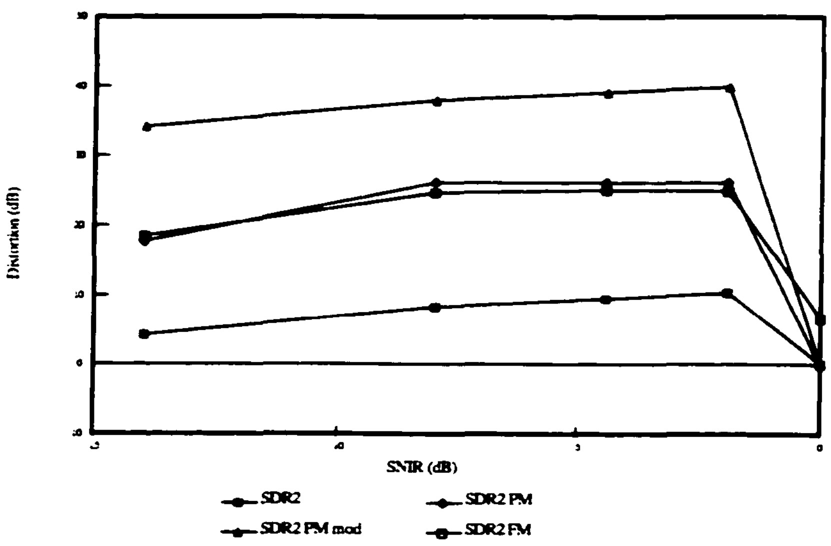


Figure 2.5 Successful separation of two signals. The figure shows various distortion measures for the weaker signal.

3 PROBLEM STATEMENT

Separation of three FM signals

When a received signal is the sum of three FM signals, when is it possible to find (separate) the three signals and how should it be done? The signals may even be at the same carrier frequency. In the reminder of this dissertation, it will be assumed that the carrier frequencies are the same or very close together because this represents a worst case. The modulation index could also be the same for all the signals. The problem is more complicated than it might appear to be, since FM modulation is a nonlinear modulation scheme. On the other hand, it is the nonlinear behavior that makes it possible to separate the signals! Introducing the following notation; the three signals are represented as follows:

$$S_i = A_i \cos(\omega_i t + \theta_i(t)) \quad i \in [1, 2, 3] \tag{3.1}$$

where $\theta_i(t)$ are assumed to be bandlimited and A_i and ω_i are unknown constants. The received signal $S_R(t)$ is simply the sum of the three signals (Figure 3.1):

$$S_R(t) = S_1(t) + S_3(t) + S_3(t) \tag{3.2}$$

Mathematically, the problem can be formulated as follows: given the received signal $S_R(t)$ and the bandwidths of $\theta_i(t)$, find $\theta_i(t)$. An equivalent way of formulating the problem is to view the signals as vectors:

$$S_i = A_i e^{i(\omega_i t - \theta_i(t))} \quad i \in [1, 2, 3] \tag{3.3}$$

From an electrical engineering point of view, this corresponds to representing the signal as inphase (I) and quadrature (Q) components. This representation has the added advantage of allowing for a geometrical interpretation. To emphasize that an quantity is estimated, it will be denoted with the subscript est, for example, S_{1est} denotes an estimate of S_1.

In the literature survey, it was seen that in the case of two interfering FM signals, the cosine of the phase difference between the signals and the received phase can always be found using the law of cosines. In the case of three FM signals, an extra degree of freedom is introduced; therefore, it is necessary to

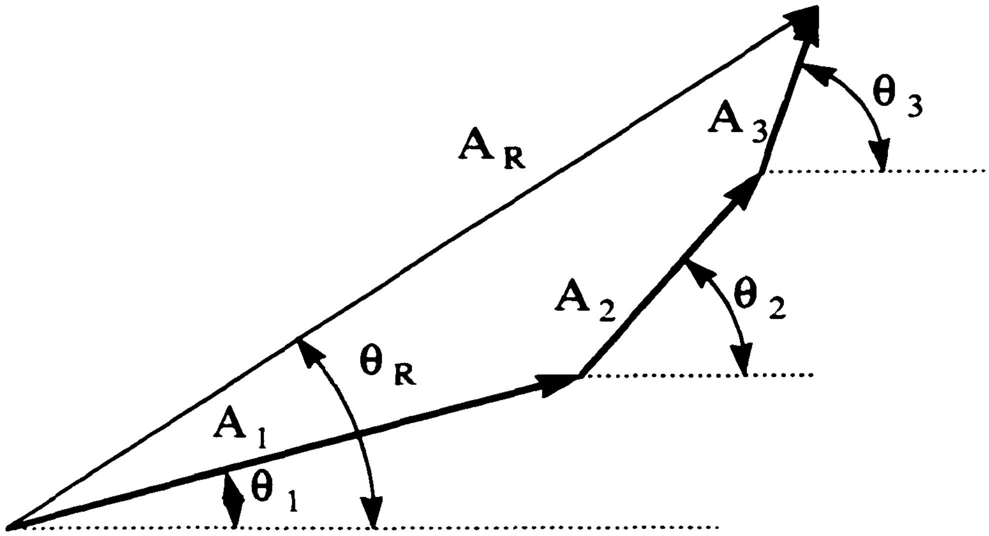


Figure 3.1 The received signal modeled as a vector.

introduce some other restriction(s). The natural choice is that the phases should be bandlimited: I will show later that this is enough to ensure uniqueness of the solution. The problem statement is easily extended to the case of more than three FM signals by letting the received signal be:

$$S_R(t) = \sum_{i=1}^{N} S_i(t) \tag{3.4}$$

where

$$S_i = A_i e^{i(\omega_i t - \theta_i(t))} \quad i \in [1, .., N] \tag{3.5}$$

Pre-distortion

The second problem, pre-distortion, is more of a practical problem. The theory of Volterra series shows how a pre-distorter should be implemented, given the Volterra series representation of the amplifier. The problem is that it is difficult to both measure the Volterra series and to implement the pre-distorter. The problem is to find an easily implemented pre-distorter for some well-specified amplifiers. See the chapter "High-Order Volterra Series Based pre-distortion for High-Efficiency RF Power Amplifiers" on page 76 for more details.

Relevance

No study or paper has been published that attempts to analyze the problem of separating three interfering FM signals, and only one has mentioned the problem [12]. Traditionally, the problem is to demodulate a signal that is corrupted by noise. As the spectra gets more utilized, the desired signal is more likely to be corrupted by other signals rather than noise. Radio systems are in many situations definitely moving away from the noise limited case and into the interference limited case. This is because the spectrum gets more crowded as more users occupy the same frequencies. Some newer systems are designed on the assumption of a certain level of interference. As an example, consider a cellular radio system. If it would be possible to separate the interfering signals, it would be possible to reuse the frequencies more often, resulting in a higher utilization. Traditionally in a cellular system, increased capacity is achieved by a so called "cell split." This is the procedure of putting the transmitters, base stations, or sites closer together. The problem is that the cells get so small that it is practically impossible to find suitable positions. The main problem is that, to avoid unacceptable interference, the base station must be within an increasingly small area, which cannot always be found. From this point of view, being able to separate the signals, and therefore eliminate the interferer, would make it possible to increase the capacity with minimum sacrifice in quality.

Another example is a car radio. Probably everyone has been sitting in a stopped car in front of a red light listening to the car radio and noticing how one radio station is fading out and another one is breaking through. A few seconds later, the process repeats itself and the radio is once again receiving the first signal. By separating the received signal, it would be possible to keep listening to either of the two stations.

A related problem is that of adjacent channels. An adjacent channel should ideally only transmit power within its allocated bandwidth, but in practice it will transmit outside this range. Therefore, the Federal Communication Commission (FCC) regulations dictate that the transmitter must be below a certain spectral level in its adjacent channels. The actual figure varies depending on frequency and system but is generally given as a number of dB below the carrier. A receiver will have problems if the received signal is very weak compared to the adjacent channel, in which case the residual power from the adjacent channel will still overpower the desired signal. Once again, by tracking and subtracting the received adjacent channel, the two signals can be separated, resulting in a cancellation of the interference caused by the adjacent channel.

The second problem, pre-distortion has started to receive renewed attention. The reason is that it is practically impossible to build a linear amplifier, especially one with high efficiency. All practical amplifiers are more or less nonlinear due to the fact both transistors and tubes are inherently nonlinear. It has been known since people started building amplifiers that they were not linear, so this is nothing new. The main reason that pre-distortion did not receive any significant attention earlier is probably because there was no practical way of implementing pre-distorters (lack of digital signal processing [DSP]) and the use of FM modulation, which is less susceptible to amplifier distortion. In existing amplifiers, high linearity is sometimes achieved by using a class A amplifier, resulting in low efficiency and is impractical for high-efficiency applications. Another thing that is done in practice is to stay away from saturating the transmitter, referred to as back-off. The idea is that as the input signal gets smaller the distortion decreases, so to stay away (back-off) from the point where one has the maximum output power from the amplifier, one gets a more linear behavior. In essence, this trades peak output power for linearity.

Another problem is that components will age, resulting in most cases in an amplifier that gets gradually more nonlinear. The fact that power amplifiers are fairly nonlinear is one reason FM is popular. Because the envelop is constant, there will be no amplitude distortion! However, many digital modulation schemes are neither FM nor AM but a combination. In these cases, any distortion caused by a non-linear amplifier will degrade the overall system performance. Another aspect is that the more linear the amplifier is, the less spurious emission in the adjacent channels.

Because it is not practical to build a linear power amplifier, the alternative is to pre-distort the signal so that the distortion of the amplifier will be cancelled out. Also, when pre-distorting the signal there is less need, or no need, to "back-off," i.e., the amplifier can be operated closer to, or at, its peak output power without sacrificing linearity and performance.

4 AMPLITUDE ESTIMATION

Finding the amplitudes

This chapter will show that just as in the case of two mutually interfering FM signals, one can always estimate the amplitude of the three signals. The method for estimating the amplitude of the two FM signals,[22], [23], is based on estimating the moments of the received signal. Through a simple extension, this also has the advantage of working in the presence of Gaussian noise. Expanding on this idea, it will be shown through direct calculations that the amplitudes, (A_1, A_2, A_3) can be estimated through the second, fourth, and sixth moments of the received signal. It is seen that this method could be extended to estimate the amplitudes for the case of more than three interfering signals by simply including higher moments. Furthermore, the calculations can be expanded to include the presence of noise by evaluating the eighth moment. To simplify the calculations, it is assumed the signals are uncorrelated and the distribution of the signals' phases are uniform $p(\theta) = \frac{1}{2\pi}$.

$$S_R = S_1 + S_2 + S_3, \tag{4.1}$$

where $S_i = A_i \cos(\omega_i t + \theta_i(t))$. Define the mean of x, $\overline{x}$ by:

$$\overline{x} = \int_{-\infty}^{\infty} x p(x) dx \tag{4.2}$$

Assuming the process to be ergodic, which can be assumed in our case, this can be written as

$$\overline{x} = \lim_{T \to \infty} \frac{1}{T} \int_0^T x(t) dt \tag{4.3}$$

The second moment of the received signal is:

$$\overline{S_R^2} = \left(\overline{S_1^2} + \overline{S_2^2} + \overline{S_3^2}\right) = \frac{1}{2}\left(A_1^2 + A_2^2 + A_3^2\right). \tag{4.4}$$

The fourth moment is:

$$\overline{S_R^4} = \frac{3}{2}A_1^2A_2^2 + \frac{3}{2}A_1^2A_3^2 + \frac{3}{2}A_2^2A_3^2 + \frac{3}{8}A_1^4 + \frac{3}{8}A_2^4 + \frac{3}{8}A_3^4, \tag{4.5}$$

and the sixth moment:

$$\overline{S_R^6} = \frac{5}{16}A_1^6 + \frac{90}{8}A_1^2A_2^2A_3^2 + \frac{45}{16}A_1^4A_2^2 + \frac{45}{16}A_1^4A_3^2 + \frac{45}{16}A_1^2A_2^4 + \frac{45}{16}A_1^2A_3^4 + \frac{5}{16}A_2^6 + \frac{45}{16}A_2^4A_3^2 + \frac{45}{16}A_2^2A_3^4 + \frac{5}{16}A_3^6. \tag{4.6}$$

The derivation of these equations are shown later in this chapter.

Define k_1, k_2, and k_3 as:

$$k_1 = 2 \cdot \overline{S_R^2} \tag{4.7}$$

$$k_2 = \left(\overline{S_R^4} - \frac{3}{8}\left(2 \cdot \overline{S_R^2}\right)^2\right)\frac{4}{3} = A_1^2A_2^2 + A_1^2A_3^2 + A_2^2A_3^2$$

In order to get k_3 in a suitable form, the following equality is handy:

$$\overline{S_R^6} - \frac{5}{16}\left(2 \cdot \overline{S_R^2}\right)^3 = \frac{75}{8}A_1^2A_2^2A_3^2 + \frac{30}{16}A_1^4A_2^2 + \frac{30}{16}A_1^4A_3^2 + \frac{30}{16}A_1^2A_2^4 + \frac{30}{16}A_1^2A_3^4 + \frac{30}{16}A_2^4A_3^2 + \frac{30}{16}A_2^2A_3^4 \Rightarrow \tag{4.8}$$

$$\begin{aligned} k_3 &= \left(\overline{S_R^6} - \frac{5}{16}\left(2 \cdot \overline{S_R^2}\right)^3 - k_1k_2\frac{30}{16}\right)\frac{8}{30} \\ &= A_1^2A_2^2A_3^2 \end{aligned} \tag{4.9}$$

The equations are of the form:

$$\begin{aligned} k_1 &= A_1^2 + A_2^2 + A_3^2 \\ k_2 &= A_1^2A_2^2 + A_1^2A_3^2 + A_2^2A_3^2 \\ k_3 &= A_1^2A_2^2A_3^2 \end{aligned} \tag{4.10}$$

Although this looks fairly innocent, it is a third-order equation system. This can be solved analytically. Because the equation system is symmetric, not only does the vector $[A_1, A_2, A_3]$ solve the equations, but so does every permutation of this vector. This is not a limitation since we know that the strongest signal is the one with the greatest amplitude etc.

Using standard techniques, [30], solve the equation for the three unknowns A_1^2, A_2^2, and A_3^2:

$$\begin{aligned} k_1 &= A_1^2 + A_2^2 + A_3^2 \\ k_2 &= A_1^2A_2^2 + A_1^2A_3^2 + A_2^2A_3^2 \\ k_3 &= A_1^2A_2^2A_3^2 \end{aligned} \tag{4.11}$$

This equation implies that:

$$0 = z^3 - k_1z^2 + k_2z - k_3 \tag{4.12}$$

Substitute $z = x + k_1/3$

$$\begin{aligned} 0 &= (x+k_1/3)^3 - k_1(x+k_1/3)^2 + k_2(x+k_1/3) - k_3 \\ &= x^3 + x^2k_1 + xk_1^2/3 + k_1^3/27 - k_1x^2 - 2xk_1^2/3 - k_1^3/9 + k_2x + k_2k_1/3 - k_3 \\ &= x^3 + x\left(-k_1^2/3 + k_2\right) - 2k_1^3/27 + k_2k_1/3 - k_3 \end{aligned} \tag{4.13}$$

Define:

$$D = \frac{\left[k_2 - \frac{k_1^2}{3}\right]^3}{27} + \frac{\left[k_3 - \frac{1}{3}k_1k_2 + \frac{2k_1^3}{27}\right]^2}{4} \tag{4.14}$$

$$\begin{aligned} s &= \left(\frac{1}{2}\left[k_3 - \frac{1}{3}k_1k_2 + \frac{2k_1^3}{27}\right] - \sqrt{D}\right)^{1/3} \\ t &= \left(\frac{1}{2}\left[k_3 - \frac{1}{3}k_1k_2 + \frac{2k_1^3}{27}\right] + \sqrt{D}\right)^{1/3} \end{aligned} \tag{4.15}$$

The three roots to 4.13 are given by:

$$\begin{aligned} z_1 &= s + t \\ z_{2,3} &= \frac{s+t}{2} \pm \frac{s-t}{2}i\sqrt{3} \end{aligned} \tag{4.16}$$

The solution to 4.11 is as follows:

$$\begin{aligned} A_1^2 &= s + t + \frac{k_1}{3} \\ A_2^2 &= \frac{s+t}{2} + \frac{s-t}{2}i\sqrt{3} + \frac{k_1}{3} \\ A_3^2 &= \frac{s+t}{2} - \frac{s-t}{2}i\sqrt{3} + \frac{k_1}{3}. \end{aligned} \tag{4.17}$$

The problem can be extended by taking the noise into consideration. Assuming that the noise is Gaussian with known variance, the three equations will be slightly different, but it will still be a third-order equation. Assuming, more realistically, that the noise variance is unknown, one would have to evaluate $\overline{S_R^8}$, which will give a fourth-order equation. It can still be solved analytically, because there is a formula for the solution of a fourth-order equation and even for a fifth order equation. For higher order equations, one is forced to solve it numerically. It would be possible to analytically find the amplitudes of five interfering FM signals or, four FM signals in Gaussian noise.

Derivation of Equations 4.4 through 4.6

Using Matlab's symbolic toolbox, the following expansions are found:

$$
\begin{aligned}
(a+b+c)^4 = {} & 12\cdot a\cdot b^2\cdot c + 4\cdot a^3\cdot b + 4\cdot a^3\cdot c + 6\cdot a^2\cdot b^2 \\
& + 6\cdot a^2\cdot c^2 + 4\cdot a\cdot b^3 + 4\cdot a\cdot c^3 + 4\cdot b^3\cdot c \\
& + 6\cdot b^2\cdot c^2 + b^4 + c^4 + a^4 + 4\cdot b\cdot c^3 \\
& + 12\cdot a^2\cdot b\cdot c + 12\cdot a\cdot b\cdot c^2
\end{aligned} \tag{4.18}
$$

$$
\begin{aligned}
(a+b+c)^6 = {} & 30\cdot a\cdot b^4\cdot c + 30\cdot a\cdot b\cdot c^4 + a^6 + 60\cdot a^3\cdot b^2\cdot c \\
& + 30\cdot a^4\cdot b\cdot c + 60\cdot a^3\cdot b\cdot c^2 + 60\cdot a^2\cdot b^3\cdot c + 90\cdot a^2\cdot b^2\cdot c^2 \\
& + 60\cdot a^2\cdot b\cdot c^3 + 60\cdot a\cdot b^3\cdot c^2 + 60\cdot a\cdot b^2\cdot c^3 + 6\cdot a\cdot b^5 + 6\cdot a\cdot c^5 \\
& + 6\cdot a^5\cdot b + 6\cdot a^5\cdot c + 15\cdot a^4\cdot b^2 + 15\cdot a^4\cdot c^2 + 20\cdot a^3\cdot b^3 + 20\cdot a^3\cdot c^3 \\
& + 15\cdot a^2\cdot b^4 + 15\cdot a^2\cdot c^4 + 6\cdot b\cdot c^5 + b^6 + 6\cdot b^5\cdot c \\
& + 15\cdot b^4\cdot c^2 + 15\cdot b^2\cdot c^4 + 20\cdot b^3\cdot c^3 + c^6
\end{aligned} \tag{4.19}
$$

This could, of course, be done without using of any computer software because it is simply an application of binomial expansion. Assuming a, b, and c to have symmetric distributions (Insuring that any term involving an odd exponent will disappear), zero mean[1], and independent gives the expected values as follows:

$$
\overline{(a+b+c)^4} = \overline{6\cdot a^2\cdot b^2} + \overline{6\cdot a^2\cdot c^2} + \overline{6\cdot b^2\cdot c^2} + \overline{b^4} + \overline{c^4} + \overline{a^4} \tag{4.20}
$$

$$
\begin{aligned}
\overline{(a+b+c)^6} = {} & \overline{a^6} + \overline{90\cdot a^2\cdot b^2\cdot c^2} + \overline{15\cdot a^4\cdot b^2} + \overline{15\cdot a^4\cdot c^2} \\
& + \overline{15\cdot a^2\cdot b^4} + \overline{15\cdot a^2\cdot c^4} + \overline{b^6} + \overline{15\cdot b^4\cdot c^2} + \overline{15\cdot b^2\cdot c^4} + \overline{c^6}
\end{aligned} \tag{4.21}
$$

Only assuming a, b,and c to be independent and to have zero mean gives the same results:

$$
\overline{(a+b+c)^4} = \overline{6\cdot a^2\cdot b^2} + \overline{6\cdot a^2\cdot c^2} + \overline{6\cdot b^2\cdot c^2} + \overline{b^4} + \overline{c^4} + \overline{a^4} \tag{4.22}
$$

$$
\begin{aligned}
\overline{(a+b+c)^6} = {} & \overline{a^6} + \overline{60\cdot a^2\cdot b^3\cdot c} + \overline{90\cdot a^2\cdot b^2\cdot c^2} + \\
& + \overline{15\cdot a^4\cdot b^2} + \overline{15\cdot a^4\cdot c^2} + \overline{20\cdot a^3\cdot b^3} + \overline{20\cdot a^3\cdot c^3}
\end{aligned} \tag{4.23}
$$

[1]This is not a limitation since any RF signal will have zero mean.

$$+\overline{15 \cdot a^2 \cdot b^4} + \overline{15 \cdot a^2 \cdot c^4} + \overline{b^6}$$
$$+\overline{15 \cdot b^4 \cdot c^2} + \overline{15 \cdot b^2 \cdot c^4} + \overline{20 \cdot b^3 \cdot c^3} + \overline{c^6}$$
$$= \overline{a^6} + \overline{90 \cdot a^2 \cdot b^2 \cdot c^2} + \overline{15 \cdot a^4 \cdot b^2} + \overline{15 \cdot a^4 \cdot c^2}$$
$$+\overline{15 \cdot a^2 \cdot b^4} + \overline{15 \cdot a^2 \cdot c^4} + \overline{b^6}$$
$$+\overline{15 \cdot b^4 \cdot c^2} + \overline{15 \cdot b^2 \cdot c^4} + \overline{c^6}$$

Because the terms $a, b,$and c, in reality, are of the form $A_i \cos(\omega_i t + \theta_i(t))$, and we can assume that the phase is evenly distributed over the interval $(0, 2\pi]$, the following relation is useful for calculating the different moments:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \sin^{2n}(x)dx = \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos^{2n}(x)dx = \frac{(2n-1)!!}{(2n)!!} \tag{4.24}$$

where $n!! \triangleq n\,(n-2)\,(n-4)\ldots$. Using this when evaluating the moments in Equations 4.20 to 4.23 gives Equations 4.4 to 4.6.

5 THE RECEIVED PHASE

The next step is to find a useful expression for the received phase. The calculations are a bit messy, but the result, Equation 5.11, is "neat" and is crucial for any further work. This result also gives an easy expression of the phase distortion in terms of the amplitudes of the three signals. For the notation see Figure 3.1. From [3] and [19], a series expansion that can be used in when working with two FM signals. Start by expanding the received phase θ_R :

$$\begin{aligned} \theta_R &= \theta_1 + \arctan\left(\frac{R_2 \sin(\theta_{R2} - \theta_1)}{A_1 + R_2 \cos(\theta_{R2} - \theta_1)}\right) \\ &= \theta_1 - \sum_{n=1}^{\infty} \left(-\frac{R_2}{A_1}\right)^n \frac{1}{n} \sin(n(\theta_{R2} - \theta_1)) \qquad \left|\frac{R_2}{A_1}\right| < 1 \end{aligned} \tag{5.1}$$

where $R_2 = \sqrt{A_2^2 + A_3^2 + 2A_2A_3\cos(\theta_3 - \theta_2)}$ is the envelop of S_2+S_3 and $\theta_{R2} = \theta_2 + \arcsin\left(\frac{A_3 \sin(\theta_3-\theta_2)}{R_2}\right)$ is the phase of $S_2 + S_3$. Note that $R_2 \leq A_2 + A_3 < A_1$, which implies that $\left|\frac{R_2}{A_1}\right| < 1$, here is where the assumption that $A_2 + A_3 < A_1$ is necessary.

Let $x = \arcsin\left(\frac{A_3/A_2 \sin(\theta_3-\theta_2)}{R_2}\right)$.

$$\begin{aligned} \theta_R &= \theta_1 - \sum_{n=1}^{\infty} \left(-\frac{R_2}{A_1}\right)^n \frac{1}{n} \sin(n(\theta_2 - \theta_1 + \arcsin(x))) = \\ & \theta_1 - \sum_{n=1}^{\infty} \left(-\frac{R_2}{A_1}\right)^n \frac{1}{n} \left[\sin(n(\theta_2 - \theta_1)) \cos(n \arcsin(x)) + \cos(n(\theta_2 - \theta_1)) \sin(n \arcsin(x))\right] \end{aligned} \tag{5.2}$$

$$\begin{aligned} \theta_R &= \theta_1 - \sum_{n=1}^{\infty} \left(-\frac{R_2}{A_1}\right)^{2n} \sin(2n(\theta_2 - \theta_1)) \times \\ & \left[\frac{1}{n} - \frac{n}{2!}x^2 + \frac{2n(4n^2 - 2^2)}{4!}x^4 + \frac{2n(4n^2 - 2^2)(4n^2 - 4^2)}{6!}x^6 - \right] \\ & - \sum_{n=1}^{\infty} \left(-\frac{R_2}{A_1}\right)^{2n} \cos(2n(\theta_2 - \theta_1)) \cos(\arcsin(x)) \times \\ & \left[x - \frac{4n^2 - 2^2}{3!}x^3 + \frac{(4n^2 - 2^2)(4n^2 - 4^2)}{5!}x^5 - \right] \\ & - \sum_{n=1}^{\infty} \left(-\frac{R_2}{A_1}\right)^{2n-1} \sin((2n-1)(\theta_2 - \theta_1)) \cos(\arcsin(x)) \times \end{aligned} \tag{5.3}$$

$$\left[\frac{1}{(2n-1)} - \frac{(2n-1)-1/(2n-1)}{2!}x^2 + \frac{[(2n-1)-1/(2n-1)]\left[(2n-1)^2-3^2\right]}{4!}x^4 - \right]$$

$$-\sum_{n=1}^{\infty}\left(-\frac{R_2}{A_1}\right)^{2n-1}\cos\left((2n-1)(\theta_2-\theta_1)\right)\times$$

$$\left[x - \frac{(2n-1)^2-1^2}{3!}x^3 + \frac{\left[(2n-1)^2-1^2\right]\left[(2n-1)^2-3^2\right]}{5!}x^5 - \right]$$

$$\theta_R = \theta_1 - \sum_{n=1}^{\infty}\left(-\frac{R_2}{A_1}\right)^{2n}\sin\left(2n(\theta_2-\theta_1)\right)\times \tag{5.4}$$

$$\left[\begin{array}{c}\frac{1}{n} - \frac{n}{2!}\left[\frac{A_3/A_2\sin(\theta_3-\theta_2)}{R_2}\right]^2 + \frac{2n\left(4n^2-2^2\right)}{4!}\left[\frac{A_3/A_2\sin(\theta_3-\theta_2)}{R_2}\right]^4 + \\ \frac{2n\left(4n^2-2^2\right)\left(4n^2-4^2\right)}{6!}\left[\frac{A_3/A_2\sin(\theta_3-\theta_2)}{R_2}\right]^6 - \dots\end{array}\right]$$

$$-\sum_{n=1}^{\infty}\left(-\frac{R_2}{A_1}\right)^{2n}\cos\left(2n(\theta_2-\theta_1)\right)\frac{1+A_3/A_2\cos(\theta_3-\theta_2)}{R_2}\times$$

$$\left[\begin{array}{c}\frac{A_3/A_2\sin(\theta_3-\theta_2)}{R_2} - \frac{4n^2-2^2}{3!}\left[\frac{A_3/A_2\sin(\theta_3-\theta_2)}{R_2}\right]^3 + \\ \frac{\left(4n^2-2^2\right)\left(4n^2-4^2\right)}{5!}\left[\frac{A_3/A_2\sin(\theta_3-\theta_2)}{R_2}\right]^5 - \dots\end{array}\right]$$

$$-\sum_{n=1}^{\infty}\left(-\frac{R_2}{A_1}\right)^{2n-1}\sin\left((2n-1)(\theta_2-\theta_1)\right)\frac{1+A_3/A_2\cos(\theta_3-\theta_2)}{R_2}\times$$

$$\left[\begin{array}{c}\frac{1}{(2n-1)} - \frac{(2n-1)-1/(2n-1)}{2!}\left[\frac{A_3/A_2\sin(\theta_3-\theta_2)}{R_2}\right]^2 + \\ \frac{[(2n-1)-1/(2n-1)]\left[(2n-1)^2-3^2\right]}{4!}\left[\frac{A_3/A_2\sin(\theta_3-\theta_2)}{R_2}\right]^4 - \dots\end{array}\right]$$

$$-\sum_{n=1}^{\infty}\left(-\frac{R_2}{A_1}\right)^{2n-1}\cos\left((2n-1)(\theta_2-\theta_1)\right)\times$$

$$\left[\begin{array}{c}\frac{A_3/A_2\sin(\theta_3-\theta_2)}{R_2} - \frac{(2n-1)^2-1^2}{3!}\left[\frac{A_3/A_2\sin(\theta_3-\theta_2)}{R_2}\right]^3 + \\ \frac{\left[(2n-1)^2-1^2\right]\left[(2n-1)^2-3^2\right]}{5!}\left[\frac{A_3/A_2\sin(\theta_3-\theta_2)}{R_2}\right]^5 - \dots\end{array}\right]$$

This expansion still contains powers of sin and cos, which is not desirable. Alternatively, the received phase can be written as follows:

$$\begin{aligned}\theta_R &= \theta_1 - \sum_{n=1}^{\infty}\left(-\frac{R_2}{A_1}\right)^n\frac{1}{n}\sin\left(n\left(\theta_2-\theta_1+\theta_{R2}\right)\right) = \\ &\theta_1 - \sum_{n=1}^{\infty}\left(-\frac{R_2}{A_1}\right)^n\frac{1}{n}\sin\left(n(\theta_2-\theta_1)\right)\left(\frac{(\cos(\theta_{R2})+i\sin(\theta_{R2}))^n+(\cos(\theta_{R2})-i\sin(\theta_{R2}))^n}{2}\right) \\ &-\sum_{n=1}^{\infty}\left(-\frac{R_2}{A_1}\right)^n\frac{1}{n}\cos\left(n(\theta_2-\theta_1)\right)\left(\frac{(\cos(\theta_{R2})+i\sin(\theta_{R2}))^n-(\cos(\theta_{R2})-i\sin(\theta_{R2}))^n}{2i}\right)\end{aligned} \tag{5.5}$$

$$\theta_R = \theta_1 - \sum_{n=1}^{\infty}\left(\frac{A_2\sqrt{1+(A_3/A_2)^2+2A_3/A_2\cos(\theta_3-\theta_2)}}{A_1}\right)^n \frac{1}{n}\sin\left(n\left(\theta_2-\theta_1\right)\right)\times \tag{5.6}$$

$$\left(\frac{\left[\begin{array}{c}(1+A_3/A_2\cos(\theta_3-\theta_2)+iA_3/A_2\sin(\theta_3-\theta_2))^n+\\(1+A_3/A_2\cos(\theta_3-\theta_2)-iA_3/A_2\sin(\theta_3-\theta_2))^n\end{array}\right]}{\left(\sqrt{1+(A_3/A_2)^2+2A_3/A_2\cos(\theta_3-\theta_2)}\right)^n 2}\right)$$

$$-\sum_{n=1}^{\infty}\left(\frac{A_2\sqrt{1+(A_3/A_2)^2+2A_3/A_2\cos(\theta_3-\theta_2)}}{A_1}\right)^n \frac{1}{n}\cos\left(n\left(\theta_2-\theta_1\right)\right)\times$$

$$\left(\frac{\left[\begin{array}{c}(1+A_3/A_2\cos(\theta_3-\theta_2)+iA_3/A_2\sin(\theta_3-\theta_2))^n-\\(1+A_3/A_2\cos(\theta_3-\theta_2)-iA_3/A_2\sin(\theta_3-\theta_2))^n\end{array}\right]}{\left(\sqrt{1+(A_3/A_2)^2+2A_3/A_2\cos(\theta_3-\theta_2)}\right)^n 2i}\right)$$

$$\theta_R = \theta_1 - \sum_{n=1}^{\infty}\left(-\frac{A_2}{A_1}\right)^n \frac{1}{n}\sin\left(n\left(\theta_2-\theta_1\right)\right)\left(\frac{\left(1+A_3/A_2e^{i(\theta_3-\theta_2)}\right)^n+\left(1+A_3/A_2e^{-i(\theta_3-\theta_2)}\right)^n}{2}\right)$$

$$-\sum_{n=1}^{\infty}\left(-\frac{A_2}{A_1}\right)^n \frac{1}{n}\cos\left(n\left(\theta_2-\theta_1\right)\right)\left(\frac{\left(1+A_3/A_2e^{i(\theta_3-\theta_2)}\right)^n-\left(1+A_3/A_2e^{-i(\theta_3-\theta_2)}\right)^n}{2i}\right) \tag{5.7}$$

Using the binomial theorem, this can be expanded as follows:

$$\theta_R = \theta_1 - \sum_{n=1}^{\infty}\left(-\frac{A_2}{A_1}\right)^n \frac{1}{n}\sin\left(n\left(\theta_2-\theta_1\right)\right)\times \tag{5.8}$$

$$\left(\frac{\sum_{k=0}^{n}\binom{n}{k}(A_3/A_2)^k e^{i(\theta_3-\theta_2)k}+\sum_{k=0}^{\infty}\binom{n}{k}(A_3/A_2)^k e^{-i(\theta_3-\theta_2)k}}{2}\right)$$

$$-\sum_{n=1}^{\infty}\left(-\frac{A_2}{A_1}\right)^n \frac{1}{n}\cos\left(n\left(\theta_2-\theta_1\right)\right)\times$$

$$\left(\frac{\sum_{k=0}^{n}\binom{n}{k}(A_3/A_2)^k e^{i(\theta_3-\theta_2)k}-\sum_{k=0}^{\infty}\binom{n}{k}(A_3/A_2)^k e^{-i(\theta_3-\theta_2)k}}{2i}\right)$$

$$\theta_R = \theta_1 - \sum_{n=1}^{\infty}\left(-\frac{A_2}{A_1}\right)^n \frac{1}{n}\sin\left(n\left(\theta_2-\theta_1\right)\right)\left(\sum_{k=0}^{n}\binom{n}{k}(A_3/A_2)^k\cos(k(\theta_3-\theta_2))\right) \tag{5.9}$$

$$-\sum_{n=1}^{\infty}\left(-\frac{A_2}{A_1}\right)^n \frac{1}{n}\cos\left(n\left(\theta_2-\theta_1\right)\right)\left(\sum_{k=0}^{n}\binom{n}{k}(A_3/A_2)^k\sin(k(\theta_3-\theta_2))\right)$$

$$\theta_R = \theta_1 - \sum_{n=1}^{\infty}\sum_{k=0}^{n}\left(-\frac{A_2}{A_1}\right)^n \frac{1}{n}\binom{n}{k}(A_3/A_2)^k\cdot \tag{5.10}$$

$$\sin\left(n\left(\theta_2-\theta_1\right)\right)\cos(k(\theta_3-\theta_2))+\cos\left(n\left(\theta_2-\theta_1\right)\right)\sin(k(\theta_3-\theta_2))$$

$$\theta_R = \theta_1 - \sum_{n=1}^{\infty}\sum_{k=0}^{n}\left(-\frac{A_2}{A_1}\right)^n \frac{1}{n}\binom{n}{k}\left(\frac{A_3}{A_2}\right)^k \sin\left((n-k)\,\theta_2 - n\theta_1 + k\theta_3\right) \tag{5.11}$$

This is the description(s) of the received phase that is most useful. First of all, it shows that there will be a capture effect on the strongest signal, S_1, since the error terms will be more wideband than the actual phase. Secondly, the equation also shows that there will be a capture effect on the second signal, provided that one can find a good enough estimate, S_{1est}, of S_1 and subtract this estimate from the received signal. To see this, consider $S_1 - S_{1est} + S_2 + S_3$: it was previously shown that one can find the amplitudes of the three signals, so it can be assumed that $|S_{1est}| = A_1$, then:

$$S_1 - S_{1est} = A_1 \sin\left(\theta_1 - \theta_{1est}\right) e^{i(\theta_1 - \theta_{1est})/2} \tag{5.12}$$

If $|A_1 \sin(\theta_1 - \theta_{1est})| < A_2 - A_3$, then one can use Equation 5.11 to write the phase of $S_1 - S_{1est} + S_2 + S_3$, denoted θ_{res}, as follows:

$$\theta_{res} = \theta_2 -$$

$$\sum_{n=1}^{\infty}\sum_{k=0}^{n}\left(-\frac{A_1 \sin(\theta_1 - \theta_{1est})}{A_2}\right)^n \frac{1}{n}\binom{n}{k}\left(\frac{A_3}{A_1 \sin(\theta_1 - \theta_{1est})}\right)^k \sin\left((n-k)\,\theta_2 - n\frac{\theta_1 + \theta_{1est}}{2} + k\theta_3\right) \tag{5.13}$$

Through an argument similar to the one above, it is seen that there will now be a capture effect on the second signal.

It is far from obvious when (and if) the series expansion converges. Fortunately, it is easy to show that it does converge.

$$\begin{aligned}&\left|\sum_{n=1}^{\infty}\sum_{k=0}^{n}\left(-\frac{A_2}{A_1}\right)^n \frac{1}{n}\binom{n}{k}(A_3/A_2)^k \sin\left((n-k)\,\theta_2 - n\theta_1 + k\theta_3\right)\right| \\ \le\ &\sum_{n=1}^{\infty}\sum_{k=0}^{n}\left(\frac{A_2}{A_1}\right)^n \frac{1}{n}\binom{n}{k}(A_3/A_2)^k = \sum_{n=1}^{\infty}\left(\frac{A_2 + A_3}{A_1}\right)^n \frac{1}{n}.\end{aligned} \tag{5.14}$$

The last sum converges by the root criteria if $A_1 < A_2 + A_3$. By the dominated covergence theorem, the first sum is absolutely convergent. It is also possible to find a closed expression for the last sum, [30]:

$$\sum_{n=1}^{\infty}\left(\frac{A_2 + A_3}{A_1}\right)^n \frac{1}{n} = \ln\left(\frac{A_1}{A_1 - A_2 - A_3}\right). \tag{5.15}$$

If $A_2 + A_3 \to A_1$, the sum approaches ∞. This sum represents an upper bound on the phase error, but it is overly pessimistic. The maximum phase distortion on S_1's phase is found from simple geometric considerations to be as follows:

$$\max|\theta_R - \theta_1| = \arccos\left(\frac{A_2 + A_3}{A_1}\right). \tag{5.16}$$

It is also possible to get an approximation of the second moment (standard deviation) of the phase error by assuming the phase difference to be uniformly distributed (this assures that the terms are orthogonal to each other). The series expansion is a two-dimensional Fourier series expansion because the three phases are independent of each other and are uniformly distributed. This assumption ensures that the inner product fulfills the usual requirements:

$$\left\langle e^{in(\theta_2-\theta_1)}, e^{-ik(\theta_3-\theta_2)} \right\rangle = 0 \tag{5.17}$$

$$\left\langle e^{in(\theta_2-\theta_1)}, e^{-ik(\theta_3-\theta_2)} \right\rangle = \begin{cases} 0 : n \neq k \\ 1 : n = k \end{cases}$$

So the functions $\sin(n(\theta_2-\theta_1))$, $\cos(k(\theta_3-\theta_2))$, $\cos(n(\theta_2-\theta_1))$, and $\sin(k(\theta_3-\theta_2)$ form a basis. Taking the coefficients of the series expansion can be can be seen as an isomorphic mapping between two Hilbert spaces. the power in the phase error can be found by squaring and adding the coefficients in the series expansion (Parseval's theorem).

In the case of one interferer only ($a = A_1/A_2$):

$$E\left((\theta_R-\theta_1)^2\right) = \int_{-\pi}^{\pi} \arctan^2\left(\frac{a\sin(\Delta\theta)}{1+a\cos(\Delta\theta)}\right)\frac{1}{2\pi}d\Delta\theta = \sum_{n=1}^{\infty}\frac{(A_1/A_2)^{2n}}{4\pi n^2}. \tag{5.18}$$

For the case of two interferers

$$E\left((\theta_R-\theta_1)^2\right) = \sum_{n=1}^{\infty}\sum_{k=0}^{n}\left(\frac{A_2}{A_1}\right)^{2n}\frac{1}{4\pi n^2}\binom{n}{k}^2\left(\frac{A_3}{A_2}\right)^{2k}. \tag{5.19}$$

It was once more assumed that all the signals' phases are independent, and the phase differences are uniformly distributed on $[0, 2\pi)$.

The case of three carriers

In this part, the special case of the phase of three carriers will be analyzed. This section will provide exact expressions for the received phase in this case. This special case is important for two reasons. First, it represents a worst case in the sense that the phase error is narrow banded because there is no modulation. Secondly, because it is possible to find the Fourier transform (and the Fourier series expansion), it is possible to tell what will happen when filtering the received phase. It will be assumed that the signals are not on the same carrier frequency and that the receiver is not narrow banded enough to directly filter out the carriers.

Beginning with the expression for the received phase, it is possible to write the phase error as follows:

$$\begin{aligned}\theta_z &= \theta_R - \theta_1 \\ &= \sum_{n=1}^{\infty}\sum_{k=0}^{n}\left(-\frac{A_2}{A_1}\right)^n \frac{1}{n}\binom{n}{k}\left(\frac{A_3}{A_2}\right)^k \sin\left((n-k)\,\theta_2 - n\theta_1 + k\theta_3\right)\end{aligned} \tag{5.20}$$

Let $\theta_1 = 0. \theta_2(t) = \theta_2 + \Delta\omega t. \theta_3(t) = \theta_3 - \Delta\omega t$. Due to geometrical considerations. it is enough to consider $\theta_1 = 0. \theta_2(t) = c + \Delta\omega t. \theta_3(t) = c - \Delta\omega t$.

$$\begin{aligned}\theta_z &= \theta_R - \theta_1 \\ &= \sum_{n=1}^{\infty}\sum_{k=0}^{n}\left(-\frac{A_2}{A_1}\right)^n \frac{1}{n}\binom{n}{k}\left(\frac{A_3}{A_2}\right)^k \sin\left((n-k)\left(c+\Delta\omega t\right) + k\left(c - \Delta\omega t\right)\right) \\ &= \sum_{n=1}^{\infty}\sum_{k=0}^{n}\left(-\frac{A_2}{A_1}\right)^n \frac{1}{n}\binom{n}{k}\left(\frac{A_3}{A_2}\right)^k \sin\left((n-2k)\,\Delta\omega t + nc\right)\end{aligned} \tag{5.21}$$

The Fourier series expansion is found by identifying terms in the above expression. For $n - 2k = 0$

$$\begin{aligned}&\sum_{\substack{n=1 \\ n \text{ even}}}^{\infty}\left(-\frac{A_2}{A_1}\right)^n \frac{1}{n}\binom{n}{n/2}\left(\frac{A_3}{A_2}\right)^{n/2} \sin(nc) \\ =\ &\sum_{n=1}^{\infty}\left(\frac{A_2}{A_1}\right)^{2n} \frac{1}{2n}\binom{2n}{n}\left(\frac{A_3}{A_2}\right)^n \sin(2nc)\end{aligned} \tag{5.22}$$

For $n - 2k = \pm 1$

$$\begin{aligned}&\sum_{n=1}^{\infty}\left(-\frac{A_2}{A_1}\right)^{2n-1} \frac{1}{2n-1}\binom{2n-1}{\frac{2n-1\pm 1}{2}}\left(\frac{A_3}{A_2}\right)^{\frac{2n-1\pm 1}{2}} \sin\left(\mp\Delta\omega t + (2n-1)\,c\right) \\ =\ &\sum_{n=1}^{\infty}\left(-\frac{A_2}{A_1}\right)^{2n-1} \frac{1}{2n-1}\binom{2n-1}{\frac{2n-1\pm 1}{2}}\left(\frac{A_3}{A_2}\right)^{\frac{2n-1\pm 1}{2}} \sin\left(\mp\Delta\omega t + (2n-1)\,c\right) \\ =\ &\sum_{n=1}^{\infty}\left(-\frac{A_2}{A_1}\right)^{2n-1} \frac{1}{2n-1}\binom{2n-1}{\frac{2n-1\pm 1}{2}}\left(\frac{A_3}{A_2}\right)^{\frac{2n-1\pm 1}{2}} \sin\left(\mp\Delta\omega t + (2n-1)\,c\right) \\ =\ &\sum_{n=1}^{\infty}\left(-\frac{A_2}{A_1}\right)^{2n-1} \frac{1}{2n-1}\binom{2n-1}{\frac{2n-1\pm 1}{2}}\left(\frac{A_3}{A_2}\right)^{\frac{2n-1\pm 1}{2}} \times \\ &\left[\sin\left(\mp\Delta\omega t\right)\cos\left((2n-1)\,c\right) + \cos\left(\mp\Delta\omega t\right)\sin\left((2n-1)\,c\right)\right] \\ =\ &\sum_{n=1}^{\infty}\left(-\frac{A_2}{A_1}\right)^{2n-1} \frac{1}{2n-1}\binom{2n-1}{\frac{2n-1\pm 1}{2}}\left(\frac{A_3}{A_2}\right)^{\frac{2n-1\pm 1}{2}} \times \\ &\left[\mp\sin\left(\Delta\omega t\right)\cos\left((2n-1)\,c\right) + \cos\left(\Delta\omega t\right)\sin\left((2n-1)\,c\right)\right]\end{aligned} \tag{5.23}$$

If $c = 0$

$$= \sin\left(\Delta\omega t\right)\sum_{n=1}^{\infty}\left(-\frac{A_2}{A_1}\right)^{2n-1} \frac{1}{2n-1}\left(\frac{A_3}{A_2}\right)^{n-1}\left[\binom{2n-1}{n}\left(\frac{A_3}{A_2}\right) - \binom{2n-1}{n-1}\right] \tag{5.24}$$

Because the expression is somewhat messy, it would be desirable to directly evaluate the Fourier series coefficients. The standard way is to use the methods of residues. Unfortunately, this turns out to be hard since the phase error is not a rational function of $\sin(\theta_2-\theta_1)$, $\sin(\theta_3-\theta_1)$, $\cos(\theta_2-\theta_1)$, and $\cos(\theta_3-\theta_1)$, leading to great difficulties in finding the residues. Fortunately, the derivative of the phase error is a rational function of $\sin(\theta_2-\theta_1)$, $\sin(\theta_3-\theta_1)$, $\cos(\theta_2-\theta_1)$, and $\cos(\theta_3-\theta_1)$, so it is much easier to find the Fourier series expansion of the derivative. Once this is found, one can integrate term by term and find the DC term to get the phase error. There is no way of knowing the average of the signal by looking at its derivative: therefore, the extra step of finding the average error is needed.

$$
\begin{aligned}
&\partial\left[\begin{array}{c} -\sum_{n=1}^{\infty}\sum_{k=0}^{n}\left(-\frac{A_2}{A_1}\right)^n \frac{1}{n}\binom{n}{k}(A_3/A_2)^k \times \\ \sin(n(\theta_2-\theta_1))\cos(k(\theta_3-\theta_2))+\cos(n(\theta_2-\theta_1))\sin(k(\theta_3-\theta_2)) \end{array}\right] \qquad (5.25)\\
=\;& \partial \arctan\left(\frac{A_2\sin(\theta_2-\theta_1)+A_3\sin(\theta_3-\theta_1)}{A_1+A_2\cos(\theta_2-\theta_1)+A_3\cos(\theta_3-\theta_1)}\right)\\
=\;& \frac{1}{1+\left(\frac{A_2\sin(\theta_2-\theta_1)+A_3\sin(\theta_3-\theta_1)}{A_1+A_2\cos(\theta_2-\theta_1)+A_3\cos(\theta_3-\theta_1)}\right)^2}\times\\
& \frac{\left[\begin{array}{c} \left\{\begin{array}{c} [A_2\cos(\theta_2-\theta_1)\partial(\theta_2-\theta_1)+A_3\cos(\theta_3-\theta_1)\partial(\theta_3-\theta_1)] \\ \times[A_1+A_2\cos(\theta_2-\theta_1)+A_3\cos(\theta_3-\theta_1)] \end{array}\right\}- \\ \left\{\begin{array}{c} [A_2\sin(\theta_2-\theta_1)+A_3\sin(\theta_3-\theta_1)] \\ \times[-A_2\sin(\theta_2-\theta_1)\partial(\theta_2-\theta_1)-A_3\sin(\theta_3-\theta_1)\partial(\theta_3-\theta_1)] \end{array}\right\} \end{array}\right]}{[A_1+A_2\cos(\theta_2-\theta_1)+A_3\cos(\theta_3-\theta_1)]^2}\\
=\;& \frac{\left(\begin{array}{c} [A_2\cos(\theta_2-\theta_1)\partial(\theta_2-\theta_1)][A_1+A_2\cos(\theta_2-\theta_1)+A_3\cos(\theta_3-\theta_1)] \\ +[A_2\sin(\theta_2-\theta_1)+A_3\sin(\theta_3-\theta_1)][A_2\sin(\theta_2-\theta_1)\partial(\theta_2-\theta_1)] \end{array}\right)}{[A_1+A_2\cos(\theta_2-\theta_1)+A_3\cos(\theta_3-\theta_1)]^2+[A_2\sin(\theta_2-\theta_1)+A_3\sin(\theta_3-\theta_1)]^2}+\\
& \frac{[\partial(\theta_3-\theta_1)]\left[A_1A_3\cos(\theta_3-\theta_1)+A_3A_2\cos(\theta_2-\theta_3)+A_3^2\right]}{[A_1+A_2\cos(\theta_2-\theta_1)+A_3\cos(\theta_3-\theta_1)]^2+[A_2\sin(\theta_2-\theta_1)+A_3\sin(\theta_3-\theta_1)]^2}
\end{aligned}
$$

$$
\begin{aligned}
=\;& \frac{\left[\begin{array}{c} \partial(\theta_2-\theta_1)\left[A_1A_2\cos(\theta_2-\theta_1)+A_2^2+A_3A_2\cos(\theta_2-\theta_3)\right]+ \\ \partial(\theta_3-\theta_1)\left[A_1A_3\cos(\theta_3-\theta_1)+A_3A_2\cos(\theta_2-\theta_3)+A_3^2\right] \end{array}\right]}{[A_1+A_2\cos(\theta_2-\theta_1)+A_3\cos(\theta_3-\theta_1)]^2+[A_2\sin(\theta_2-\theta_1)+A_3\sin(\theta_3-\theta_1)]^2}\\
=\;& \frac{\left[\begin{array}{c} \partial(\theta_2-\theta_1)\left[A_1A_2\cos(\theta_2-\theta_1)+A_2^2+A_3A_2\cos(\theta_2-\theta_3)\right]+ \\ \partial(\theta_3-\theta_1)\left[A_1A_3\cos(\theta_3-\theta_1)+A_3A_2\cos(\theta_2-\theta_3)+A_3^2\right] \end{array}\right]}{\left[\begin{array}{c} A_1^2+A_2^2+A_3^2+2A_1A_2\cos(\theta_2-\theta_1)+2A_1A_3\cos(\theta_3-\theta_1) \\ +2A_2A_3\cos(\theta_2-\theta_1)\cos(\theta_3-\theta_1)+2A_2A_3\sin(\theta_2-\theta_1)\sin(\theta_3-\theta_1) \end{array}\right]}
\end{aligned}
$$

$$= \frac{\left[\begin{array}{c}\partial(\theta_2-\theta_1)\left[A_1A_2\cos(\theta_2-\theta_1)+A_2^2+A_3A_2\cos(\theta_2-\theta_3)\right]+\\ \partial(\theta_3-\theta_1)\left[A_1A_3\cos(\theta_3-\theta_1)+A_3A_2\cos(\theta_2-\theta_3)+A_3^2\right]\end{array}\right]}{A_1^2+A_2^2+A_3^2+2A_1A_2\cos(\theta_2-\theta_1)+2A_1A_3\cos(\theta_3-\theta_1)+2A_2A_3\cos(\theta_2-\theta_3)}$$

To proceed, some more assumptions about the phases are needed. One special case worth noting is when $\theta_2-\theta_1= -(\theta_3-\theta_1)$, which corresponds to a situation when the frequency spacing between the carriers is the same and they are all in phase at $t=0$. In this case,

$$\begin{aligned}
&\frac{\partial}{\partial(\theta_2-\theta_1)}\arctan\left(\frac{(A_2-A_3)\sin(\theta_2-\theta_1)}{A_1+(A_2+A_3)\cos(\theta_2-\theta_1)}\right) \qquad (5.26)\\
&= \frac{\left(\frac{(A_2-A_3)\cos(\theta_2-\theta_1)}{A_1+(A_2+A_3)\cos(\theta_2-\theta_1)}-\frac{(A_2-A_3)\sin(\theta_2-\theta_1)(-(A_2+A_3)\sin(\theta_2-\theta_1))}{(A_1+(A_2+A_3)\cos(\theta_2-\theta_1))^2}\right)}{1+\left(\frac{(A_2-A_3)\sin(\theta_2-\theta_1)}{A_1+(A_2+A_3)\cos(\theta_2-\theta_1)}\right)^2}\\
&= \frac{\left(\begin{array}{c}(A_2-A_3)\cos(\theta_2-\theta_1)(A_1+(A_2+A_3)\cos(\theta_2-\theta_1))-\\ (A_2-A_3)\sin(\theta_2-\theta_1)(-(A_2+A_3)\sin(\theta_2-\theta_1))\end{array}\right)}{(A_1+(A_2+A_3)\cos(\theta_2-\theta_1))^2+((A_2-A_3)\sin(\theta_2-\theta_1))^2}\\
&= \frac{\left(\begin{array}{c}A_1(A_2-A_3)\cos(\theta_2-\theta_1)+(A_2-A_3)(A_2+A_3)\cos^2(\theta_2-\theta_1)+\\ (A_2-A_3)(A_2+A_3)\sin^2(\theta_2-\theta_1)\end{array}\right)}{A_1^2+(A_2+A_3)^2\cos^2(\theta_2-\theta_1)+2A_1(A_2+A_3)\cos(\theta_2-\theta_1)+(A_2-A_3)^2\sin^2(\theta_2-\theta_1)}\\
&= \frac{\left(\begin{array}{c}A_1(A_2-A_3)\cos(\theta_2-\theta_1)+(A_2-A_3)(A_2+A_3)\cos^2(\theta_2-\theta_1)+\\ (A_2-A_3)(A_2+A_3)\sin^2(\theta_2-\theta_1)\end{array}\right)}{A_1^2+A_2^2+A_3^2+2A_2A_3\cos^2(\theta_2-\theta_1)+2A_1(A_2+A_3)\cos(\theta_2-\theta_1)-2A_2A_3\sin^2(\theta_2-\theta_1)}\\
&= \frac{A_1(A_2-A_3)\cos(\theta_2-\theta_1)+A_2^2-A_3^2}{A_1^2+A_2^2+A_3^2+2A_2A_3\cos[2(\theta_2-\theta_1)]+2A_1(A_2+A_3)\cos(\theta_2-\theta_1)}\\
&= \frac{A_1(A_2-A_3)\cos(\theta_2-\theta_1)+A_2^2-A_3^2}{A_1^2+A_2^2+A_3^2+4A_2A_3\cos^2(\theta_2-\theta_1)-2A_2A_3+2A_1(A_2+A_3)\cos(\theta_2-\theta_1)}
\end{aligned}$$

Factorize the denominator by setting $\cos(\theta_2-\theta_1)=t$

$$\begin{aligned}
A_1^2+A_2^2+A_3^2+4A_2A_3\cos^2(\theta_2-\theta_1)-2A_2A_3+2A_1(A_2+A_3)\cos(\theta_2-\theta_1) &= 0 \Leftrightarrow \qquad (5.27)\\
A_1^2+A_2^2+A_3^2+4A_2A_3t^2-2A_2A_3+2A_1(A_2+A_3)t &= 0\\
t^2+\frac{A_1(A_2+A_3)}{2A_2A_3}t+\frac{A_1^2+A_2^2+A_3^2-2A_2A_3}{4A_2A_3} &= 0\\
\left(t+\frac{A_1(A_2+A_3)}{4A_2A_3}\right)^2-\left(\frac{A_1(A_2+A_3)}{4A_2A_3}\right)^2+\frac{A_1^2+A_2^2+A_3^2-2A_2A_3}{4A_2A_3} &= 0
\end{aligned}$$

$$t+\frac{A_1(A_2+A_3)}{4A_2A_3} = \pm\sqrt{\left(\frac{A_1(A_2+A_3)}{4A_2A_3}\right)^2-\frac{A_1^2+A_2^2+A_3^2-2A_2A_3}{4A_2A_3}} \qquad (5.28)$$

$$t+\frac{A_1(A_2+A_3)}{4A_2A_3} = \pm\sqrt{\frac{A_1^2A_2^2+A_1^2A_3^2+2A_1^2A_2A_3}{(4A_2A_3)^2}+\frac{-4A_2A_3A_1^2-4A_3A_2^3-4A_2A_3^3+8A_2^2A_3^2}{4A_2A_3}}$$

$$t+\frac{A_1(A_2+A_3)}{4A_2A_3} = \pm\frac{\sqrt{(A_1(A_2-A_3))^2-4A_2A_3(A_2-A_3)^2}}{4A_2A_3}$$

$$t = -\frac{A_1(A_2+A_3)}{4A_2A_3}\pm\frac{(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}$$

Is this always greater than A_1? It obviously is since $A_2/(4A_2A_3) > 1/(2A_3) > 1$ because $A_1 < A_2 + A_3$. It is also seen that $t/A_1 > 1$ from plotting the expressions, so:

$$\begin{aligned}&A_1^2+A_2^2+A_3^2+4A_2A_3\cos^2(\theta_2-\theta_1)-2A_2A_3+2A_1(A_2+A_3)\cos(\theta_2-\theta_1) \\ =\ & 4A_2A_3\left(\cos(\theta_2-\theta_1)+\frac{A_1(A_2+A_3)}{4A_2A_3}+\frac{(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right)\cdot \\ &\left(\cos(\theta_2-\theta_1)+\frac{A_1(A_2+A_3)}{4A_2A_3}-\frac{(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right)\end{aligned} \tag{5.29}$$

Returning to the expression

$$\frac{A_1(A_2-A_3)\cos(\theta_2-\theta_1)+A_2^2-A_3^2}{A_1^2+A_2^2+A_3^2+4A_2A_3\cos^2(\theta_2-\theta_1)-2A_2A_3+2A_1(A_2+A_3)\cos(\theta_2-\theta_1)} \tag{5.30}$$

for which we want to find the Fourier series expansion, forces us to evaluate the integral:

$$\begin{aligned}c_n &= \frac{1}{2\pi}\int_0^{2\pi}\frac{(A_1(A_2-A_3)\cos(\theta_2-\theta_1)+A_2^2-A_3^2)\cos(n(\theta_2-\theta_1))\,d(\theta_2-\theta_1)}{A_1^2+A_2^2+A_3^2+4A_2A_3\cos^2(\theta_2-\theta_1)-2A_2A_3+2A_1(A_2+A_3)\cos(\theta_2-\theta_1)} \\ &= \frac{1}{2\pi}\int_0^{2\pi}\frac{[A_1(A_2-A_3)\cos(\theta_2-\theta_1)+A_2^2-A_3^2]\cos(n(\theta_2-\theta_1))\,d(\theta_2-\theta_1)}{\left[\begin{array}{c}4A_2A_3\left(\cos(\theta_2-\theta_1)+\frac{A_1(A_2+A_3)}{4A_2A_3}+\frac{(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right)\times \\ \left(\cos(\theta_2-\theta_1)+\frac{A_1(A_2+A_3)}{4A_2A_3}-\frac{(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right)\end{array}\right]}\end{aligned} \tag{5.31}$$

This can be done using residues

$$c_n = \oint_{|z|=1}\frac{\frac{1}{4A_2A_3}\left(A_1(A_2-A_3)\left(\frac{z+z^{-1}}{2}\right)+A_2^2-A_3^2\right)\left(\frac{z^n+z^{-n}}{2}\right)}{\left\{\begin{array}{c}\left(\left(\frac{z+z^{-1}}{2}\right)+\frac{A_1(A_2+A_3)+(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right)\times \\ \left(\left(\frac{z+z^{-1}}{2}\right)+\frac{A_1(A_2+A_3)-(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right)\end{array}\right\}}\frac{dz}{iz} \tag{5.32}$$

$$= \oint_{|z|=1}\frac{\left[A_1(A_2-A_3)\left(\frac{z^2+1}{2}\right)+(A_2^2-A_3^2)z\right]\left(\frac{z^n+z^{-n}}{2}\right)}{\left[\begin{array}{c}4A_2A_3\left(\left(\frac{z^2+1}{2}\right)+\left(\frac{A_1(A_2+A_3)}{4A_2A_3}+\frac{(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right)z\right)\times \\ \left(\left(\frac{z^2+1}{2}\right)+\left(\frac{A_1(A_2+A_3)}{4A_2A_3}-\frac{(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right)z\right)\end{array}\right]}\frac{dz}{i} \tag{5.33}$$

The expression implies:

$$0 = \left(\frac{z^2+1}{2}\right) + \left(\frac{A_1(A_2+A_3)}{4A_2A_3} \pm \frac{(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right) z \Rightarrow \tag{5.34}$$

$$z_1 = -\left(\frac{A_1(A_2+A_3)}{4A_2A_3} \pm \frac{(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right) + \tag{5.35}$$
$$\sqrt{\left(\frac{A_1(A_2+A_3)}{4A_2A_3} \pm \frac{(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right)^2 - 1}$$

$$z_1 = -\left(\frac{A_1(A_2+A_3)}{4A_2A_3} \pm \frac{(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right) - \tag{5.36}$$
$$\sqrt{\left(\frac{A_1(A_2+A_3)}{4A_2A_3} \pm \frac{(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right)^2 - 1}$$

since

$$\frac{A_1(A_2+A_3)}{4A_2A_3} \pm \frac{(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3} > 1 \tag{5.37}$$

Define the following parameters

$$r_1 = \frac{A_1(A_2+A_3)}{4A_2A_3} + \frac{(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3} \tag{5.38}$$
$$r_2 = \frac{A_1(A_2+A_3)}{4A_2A_3} - \frac{(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}$$

The poles of Equation 5.32 are:

$$z_1 = -\left(\frac{A_1(A_2+A_3)+(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right) \tag{5.39}$$
$$+\sqrt{\left(\frac{A_1(A_2+A_3)+(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right)^2 - 1} \tag{5.40}$$
$$= -r_1 + \sqrt{r_1^2-1} \quad \text{of order 1, inside } |z|=1$$

$$z_2 = -\left(\frac{A_1(A_2+A_3)+(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right) - \tag{5.41}$$
$$\sqrt{\left(\frac{A_1(A_2+A_3)+(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right)^2 - 1} \tag{5.42}$$
$$= -r_1 - \sqrt{r_1^2-1} \quad \text{of order 1, outside } |z|=1$$

$$z_3 = -\left(\frac{A_1(A_2+A_3)-(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right) + \tag{5.43}$$
$$\sqrt{\left(\frac{A_1(A_2+A_3)-(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right)^2 - 1} \tag{5.44}$$
$$= -r_2 + \sqrt{r_2^2-1} \quad \text{of order 1, inside } |z|=1$$

$$\begin{aligned} z_3 &= -\left(\frac{A_1(A_2+A_3)-(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right)- \\ &\quad \sqrt{\left(\frac{A_1(A_2+A_3)-(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right)^2-1} \qquad (5.45) \\ &= -r_2-\sqrt{r_2^2-1} \quad \text{of order 1, outside } |z|=1 \\ z_5 &= 0 \text{ of order } n \end{aligned}$$

therefore,

$$\oint_{|z|=1} \frac{\left[A_1(A_2-A_3)\left(\frac{z^2+1}{2}\right)+(A_2^2-A_3^2)z\right]\left(\frac{z^n+z^{-n}}{2}\right)}{\left[\begin{array}{c} A_2A_3\left(\left(\frac{z^2+1}{2}\right)+\left(\frac{A_1(A_2+A_3)}{4A_2A_3}+\frac{(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right)z\right)\times \\ \left(\left(\frac{z^2+1}{2}\right)+\left(\frac{A_1(A_2+A_3)}{4A_2A_3}-\frac{(A_2-A_3)\sqrt{A_1^2-4A_2A_3}}{4A_2A_3}\right)z\right)\end{array}\right]}\frac{dz}{i}$$

$$= \oint_{|z|=1} \frac{\left[A_1(A_2-A_3)\left(\frac{z^2+1}{2}\right)+(A_2^2-A_3^2)z\right]\left(\frac{z^n+z^{-n}}{2}\right)}{A_2A_3(z-z_1)(z-z_2)(z-z_3)(z-z_4)}\frac{dz}{i} \qquad (5.46)$$

$$\begin{aligned} = \oint_{|z|=1} & \frac{\left[A_1(A_2-A_3)\left(\frac{z^{n+2}+z^n}{2}\right)+(A_2^2-A_3^2)z^{n+1}\right]+\left[A_1(A_2-A_3)\left(\frac{z^{-n+2}+z^{-n}}{2}\right)+(A_2^2-A_3^2)z^{-n+1}\right]}{A_2A_3}\times \\ & [\frac{1}{(z-z_1)(z_1-z_2)(z_1-z_3)(z_1-z_4)}+\frac{1}{(z_2-z_1)(z-z_2)(z_2-z_3)(z_2-z_4)} \\ & +\frac{1}{(z_3-z_1)(z_3-z_2)(z-z_3)(z_3-z_4)}+\frac{1}{(z_4-z_1)(z_4-z_2)(z_4-z_3)(z-z_4)}]\frac{dz}{i} \end{aligned}$$

Using the following series expansions that converges for $|z|=1$

$$\frac{1}{z-z_1} = \frac{z^{-1}}{1-z_1/z} = z^{-1}\sum_{i=0}^{\infty}(z_1/z)^i \qquad |z_1/z|<1 \Leftrightarrow |z_1|<|z| \text{ Pole inside unit circle} \qquad (5.47)$$

$$\frac{1}{z-z_1} = \frac{-z_1^{-1}}{1-z/z_1} = -z_1^{-1}\sum_{i=0}^{\infty}(z/z_1)^i \qquad |z/z_1|>1 \Leftrightarrow |z|>|z_1| \text{ Pole outside unit circle}$$

the series expansion that converge for $|z|=1$ is as follows:

$$\begin{aligned} & \frac{1}{(z-z_1)(z_1-z_2)(z_1-z_3)(z_1-z_4)}+\frac{1}{(z_2-z_1)(z-z_2)(z_2-z_3)(z_2-z_4)}+ \qquad (5.48) \\ & \frac{1}{(z_3-z_1)(z_3-z_2)(z-z_3)(z_3-z_4)}+\frac{1}{(z_4-z_1)(z_4-z_2)(z_4-z_3)(z-z_4)} \\ = & \frac{z^{-1}\sum_{i=0}^{\infty}(z_1z^{-1})^i}{(z_1-z_2)(z_1-z_3)(z_1-z_4)}+\frac{-z_2^{-1}\sum_{i=0}^{\infty}(z/z_2)^i}{(z_2-z_1)(z_2-z_3)(z_2-z_4)} \\ & +\frac{z^{-1}\sum_{i=0}^{\infty}(z_3z^{-1})^i}{(z_3-z_1)(z_3-z_2)(z_3-z_4)}+\frac{-z_4^{-1}\sum_{i=0}^{\infty}(z/z_4)^i}{(z_4-z_1)(z_4-z_2)(z_4-z_3)} \end{aligned}$$

To find the residues, it is enough to find the z^{-1} term in the series expansion of:

$$\frac{\left\{\begin{array}{l}\left[A_1(A_2-A_3)\left(\frac{z^{n-2}+z^n}{2}\right)+(A_2^2-A_3^2)z^{n-1}\right]+\\ \left[A_1(A_2-A_3)\left(\frac{z^{-n-2}+z^{-n}}{2}\right)+(A_2^2-A_3^2)z^{-n-1}\right]\end{array}\right\}}{2A_2A_3}\times \tag{5.49}$$

$$\left[\frac{z^{-1}\sum_{i=0}^{\infty}(z_1z^{-1})^i}{(z_1-z_2)(z_1-z_3)(z_1-z_4)}+\frac{-z_2^{-1}\sum_{i=0}^{\infty}(z/z_2)^i}{(z_2-z_1)(z_2-z_3)(z_2-z_4)}\right.$$
$$\left.+\frac{z^{-1}\sum_{i=0}^{\infty}(z_3z^{-1})^i}{(z_3-z_1)(z_3-z_2)(z_3-z_4)}+\frac{-z_4^{-1}\sum_{i=0}^{\infty}(z/z_4)^i}{(z_4-z_1)(z_4-z_2)(z_4-z_3)}\right] \tag{5.50}$$

Depending on $n=0, n=1, n=2$, or $n\geq 3$, the residues are as follows:

for $n = 0$, the z^{-1} term is as folows: (5.51)

$$\frac{1}{A_2A_3}\times\frac{A_1(A_2-A_3)\frac{z_1^2+1}{2}+(A_2^2-A_3^2)z_1}{(z_1-z_2)(z_1-z_3)(z_1-z_4)}+\frac{A_1(A_2-A_3)\frac{z_3^2+1}{2}+(A_2^2-A_3^2)z_3}{(z_3-z_1)(z_3-z_2)(z_3-z_4)}$$

for $n = 1$ (5.52)

$$\frac{\left[A_1(A_2-A_3)z+A_1(A_2-A_3)\frac{z^3}{2}+(A_2^2-A_3^2)z^2\right]+\left[A_1(A_2-A_3)\frac{z^{-1}}{2}+(A_2^2-A_3^2)\right]}{2A_2A_3}\times$$

$$\left[\frac{z^{-1}\sum_{i=0}^{\infty}(z_1z^{-1})^i}{(z_1-z_2)(z_1-z_3)(z_1-z_4)}+\frac{-z_2^{-1}\sum_{i=0}^{\infty}(z/z_2)^i}{(z_2-z_1)(z_2-z_3)(z_2-z_4)}\right.$$
$$\left.+\frac{z^{-1}\sum_{i=0}^{\infty}(z_3z^{-1})^i}{(z_3-z_1)(z_3-z_2)(z_3-z_4)}+\frac{-z_4^{-1}\sum_{i=0}^{\infty}(z/z_4)^i}{(z_4-z_1)(z_4-z_2)(z_4-z_3)}\right]$$

$\Rightarrow$ The z^{-1} term is as follows:

$$\begin{aligned} c_1 = {} & \frac{\frac{A_1(A_2-A_3)}{2}z_1^3+(A_2^2-A_3^2)z_1^2+A_1(A_2-A_3)z_1+(A_2^2-A_3^2)}{2A_2A_3(z_1-z_2)(z_1-z_3)(z_1-z_4)}+ \\ & \frac{\frac{A_1(A_2-A_3)}{2}z_3^3+(A_2^2-A_3^2)z_3^2+A_1(A_2-A_3)z_3+(A_2^2-A_3^2)}{2A_2A_3(z_3-z_1)(z_3-z_2)(z_3-z_4)}+ \\ & -\frac{A_1(A_2-A_3)z_2^{-1}}{4A_2A_3(z_2-z_1)(z_2-z_3)(z_2-z_4)}-\frac{A_1(A_2-A_3)z_4^{-1}}{4A_2A_3(z_4-z_1)(z_4-z_2)(z_4-z_3)} \end{aligned}$$

for $n = 2$ (5.53)

$$\frac{\left[A_1(A_2-A_3)\left(\frac{z^4+z^2}{2}\right)+(A_2^2-A_3^2)z^3\right]+\left[A_1(A_2-A_3)\left(\frac{1+z^{-2}}{2}\right)+(A_2^2-A_3^2)z^{-1}\right]}{2A_2A_3}\times$$

$$\left[\frac{z^{-1}\sum_{i=0}^{\infty}(z_1 z^{-1})^i}{(z_1-z_2)(z_1-z_3)(z_1-z_4)}+\frac{-z_2^{-1}\sum_{i=0}^{\infty}(z/z_2)^i}{(z_2-z_1)(z_2-z_3)(z_2-z_4)}+\frac{z^{-1}\sum_{i=0}^{\infty}(z_3 z^{-1})^i}{(z_3-z_1)(z_3-z_2)(z_3-z_4)}+\frac{-z_4^{-1}\sum_{i=0}^{\infty}(z/z_4)^i}{(z_4-z_1)(z_4-z_2)(z_4-z_3)}\right]$$

the z^{-1} term is as follows

$$c_2=\frac{A_1(A_2-A_3)\frac{z_1^4+z_1^2+1}{2}+(A_2^2-A_3^2)z_1^3}{2A_2A_3(z_1-z_2)(z_1-z_3)(z_1-z_4)}-\frac{A_1(A_2-A_3)\frac{z_2^{-2}}{2}+(A_2^2-A_3^2)z_2^{-1}}{2A_2A_3(z_2-z_1)(z_2-z_3)(z_2-z_4)}+\frac{A_1(A_2-A_3)\frac{z_3^4+z_3^2+1}{2}+(A_2^2-A_3^2)z_3^3}{2A_2A_3(z_3-z_1)(z_3-z_2)(z_3-z_4)}-\frac{A_1(A_2-A_3)\frac{z_4^{-2}}{2}+(A_2^2-A_3^2)z_4^{-1}}{2A_2A_3(z_4-z_1)(z_4-z_2)(z_4-z_3)}$$

$$n \geq 3 \Rightarrow \tag{5.54}$$

$$\frac{\left\{\begin{array}{l}\left[A_1(A_2-A_3)\left(\frac{z^{n-2}+z^n}{2}\right)+(A_2^2-A_3^2)z^{n-1}\right]+\\ \left[A_1(A_2-A_3)\left(\frac{z^{-n-2}+z^{-n}}{2}\right)+(A_2^2-A_3^2)z^{-n-1}\right]\end{array}\right\}}{2A_2A_3}\times$$

$$\frac{z^{-1}\sum_{i=0}^{\infty}(z_1 z^{-1})^i}{(z_1-z_2)(z_1-z_3)(z_1-z_4)}+\frac{-z_2^{-1}\sum_{i=0}^{\infty}(z/z_2)^i}{(z_2-z_1)(z_2-z_3)(z_2-z_4)}+\frac{z^{-1}\sum_{i=0}^{\infty}(z_3 z^{-1})^i}{(z_3-z_1)(z_3-z_2)(z_3-z_4)}+\frac{-z_4^{-1}\sum_{i=0}^{\infty}(z/z_4)^i}{(z_4-z_1)(z_4-z_2)(z_4-z_3)}$$

the z^{-1} term is as follows

$$\underset{n\geq 3}{c_n}=\frac{A_1(A_2-A_3)\frac{z_1^{n-2}+z_1^n}{2}+(A_2^2-A_3^2)z_1^{n-1}}{(z_1-z_2)(z_1-z_3)(z_1-z_4)}-\frac{A_1(A_2-A_3)\frac{z_2^{-n-2}+z_2^{-n}}{2}+(A_2^2-A_3^2)z_2^{-n-1}}{(z_2-z_1)(z_2-z_3)(z_2-z_4)}+\frac{A_1(A_2-A_3)\frac{z_3^{n-2}+z_3^n}{2}+(A_2^2-A_3^2)z_3^{n-1}}{(z_3-z_1)(z_3-z_2)(z_3-z_4)}-\frac{A_1(A_2-A_3)\frac{z_4^{-n-2}+z_4^{-n}}{2}+(A_2^2-A_3^2)z_4^{-n-1}}{(z_4-z_1)(z_4-z_2)(z_4-z_3)}$$

A slight, but important, generalization of the above case is to find the Fourier expansion of the expression $\arctan\left(\frac{C\sin(\theta_2-\theta_1-\phi_1)}{A+B\cos(\theta_2-\theta_1-\phi_2)}\right)$ This expression will make it possible to handle the case when the carriers have equal frequency spacing but different phase shifts. The approach is the same as before: find the derivative and its Fourier series expansion, and then integrate term by term.

Beginning with the partial derivative,

$$
\begin{aligned}
&\frac{\partial}{\partial(\theta_2-\theta_1)}\arctan\left(\frac{C\sin(\theta_2-\theta_1+\phi_3)}{A+B\cos(\theta_2-\theta_1+\phi_2)}\right) \qquad (5.55)\\
&= \frac{\dfrac{C\cos(\theta_2-\theta_1+\phi_3)}{A+B\cos(\theta_2-\theta_1+\phi_2)}-\dfrac{C\sin(\theta_2-\theta_1+\phi_3)\left(-B\sin(\theta_2-\theta_1+\phi_2)\right)}{\left(A+B\cos(\theta_2-\theta_1+\phi_2)\right)^2}}{1+\left(\dfrac{C\sin(\theta_2-\theta_1+\phi_3)}{A+B\cos(\theta_2-\theta_1+\phi_2)}\right)^2}\\
&= \frac{C\cos(\theta_2-\theta_1+\phi_3)\left(A+B\cos(\theta_2-\theta_1+\phi_2)\right)+C\sin(\theta_2-\theta_1+\phi_3)\left(B\sin(\theta_2-\theta_1+\phi_2)\right)}{\left(A+B\cos(\theta_2-\theta_1+\phi_2)\right)^2+\left(C\sin(\theta_2-\theta_1+\phi_3)\right)^2}\\
&= \frac{\left[\begin{array}{c}AC\cos(\theta_2-\theta_1+\phi_3)+C\cos(\theta_2-\theta_1+\phi_3)\,B\cos(\theta_2-\theta_1+\phi_2)+\\ C\sin(\theta_2-\theta_1+\phi_3)\left(B\sin(\theta_2-\theta_1+\phi_2)\right)\end{array}\right]}{A^2+B\cos^2(\theta_2-\theta_1+\phi_2)+2AB\cos(\theta_2-\theta_1+\phi_2)+C^2\sin^2(\theta_2-\theta_1+\phi_3)}\\
&= \frac{AC\cos(\theta_2-\theta_1+\phi_3)+BC\cos(\phi_2-\phi_3)}{A^2+B^2\cos^2(\theta_2-\theta_1+\phi_2)+2AB\cos(\theta_2-\theta_1+\phi_2)+C^2\sin^2(\theta_2-\theta_1+\phi_3)}\\
&= \frac{AC\cos(\theta_2-\theta_1+\phi_3)+BC\cos(\phi_2-\phi_3)}{A^2+B^2\dfrac{1+\cos[2(\theta_2-\theta_1+\phi_2)]}{2}+2AB\cos(\theta_2-\theta_1+\phi_2)+C^2\dfrac{1-\cos[2(\theta_2-\theta_1+\phi_3)]}{2}}
\end{aligned}
$$

and finding the zeros of the denominator:

$$
\begin{aligned}
\left(A+B\cos(\theta_2-\theta_1+\phi_2)\right)^2+\left(C\sin(\theta_2-\theta_1+\phi_3)\right)^2 &= 0 \qquad (5.56)\\
A+B\cos(\theta_2-\theta_1+\phi_2) &= \pm iC\sin(\theta_2-\theta_1+\phi_3)\\
z &= e^{i(\theta_2-\theta_1)}\\
2A+B\left[ze^{i\phi_2}+z^{-1}e^{-i\phi_2}\right]\mp C\left[ze^{i\phi_3}-z^{-1}e^{-i\phi_3}\right] &= 0\\
2Az+B\left[z^2e^{i\phi_2}+e^{-i\phi_2}\right]\mp C\left[z^2e^{i\phi_3}-e^{-i\phi_3}\right] &= 0\\
2Az+z^2\left[Be^{i\phi_2}\mp Ce^{i\phi_3}\right]+Be^{-i\phi_2}\pm Ce^{-i\phi_3} &= 0\\
z^2+\frac{2Az}{Be^{i\phi_2}\mp Ce^{i\phi_3}}+\frac{Be^{-i\phi_2}\pm Ce^{-i\phi_3}}{Be^{i\phi_2}\mp Ce^{i\phi_3}} &= 0\\
\left(z+\frac{A}{Be^{i\phi_2}\mp Ce^{i\phi_3}}\right)-\left(\frac{A}{Be^{i\phi_2}\mp Ce^{i\phi_3}}\right)^2 &= -\frac{Be^{-i\phi_2}\pm Ce^{-i\phi_3}}{Be^{i\phi_2}\mp Ce^{i\phi_3}}
\end{aligned}
$$

$$
\begin{aligned}
z &= -\frac{A}{Be^{i\phi_2}\mp Ce^{i\phi_3}}\pm\frac{\sqrt{A^2-\left(Be^{-i\phi_2}\pm Ce^{-i\phi_3}\right)\left(Be^{i\phi_2}\mp Ce^{i\phi_3}\right)}}{Be^{i\phi_2}\mp Ce^{i\phi_3}} \qquad (5.57)\\
z &= -\frac{A}{Be^{i\phi_2}\mp Ce^{i\phi_3}}\pm\frac{\sqrt{A^2-\left(B^2-C^2\mp i2BC\sin(\phi_3-\phi_2)\right)}}{Be^{i\phi_2}\mp Ce^{i\phi_3}}\\
z_{1,2} &= -\frac{A+\sqrt{A^2-B^2+C^2\pm i2BC\sin(\phi_3-\phi_2)}}{Be^{i\phi_2}\mp Ce^{i\phi_3}}
\end{aligned}
$$

$$z_{3,4} = -\frac{A - \sqrt{A^2 - B^2 + C^2 \pm i2BC\sin(\phi_3 - \phi_2)}}{Be^{i\phi_2} \mp Ce^{i\phi_3}}$$

produces,

$$\begin{aligned} &(A + B\cos(\theta_2 - \theta_1 + \phi_2))^2 + (C\sin(\theta_2 - \theta_1 + \phi_3))^2 \\ = \; & z^{-2}(z - z_1)(z - z_2)(z - z_3)(z - z_4)\frac{\left(Be^{i\phi_2} + Ce^{i\phi_3}\right)\left(Be^{i\phi_2} - Ce^{i\phi_3}\right)}{4} \\ z = \; & e^{i(\theta_2 - \theta_1)} \end{aligned} \tag{5.58}$$

and

$$\begin{aligned} &\frac{\partial}{\partial(\theta_2 - \theta_1)}\arctan\left(\frac{C\sin(\theta_2 - \theta_1 + \phi_3)}{A + B\cos(\theta_2 - \theta_1 + \phi_2)}\right) \\ = \; & \frac{AC\cos(\theta_2 - \theta_1 + \phi_3) + BC\cos(\phi_2 - \phi_3)}{A^2 + B^2\cos^2(\theta_2 - \theta_1 + \phi_2) + 2AB\cos(\theta_2 - \theta_1 + \phi_2) + C^2\sin^2(\theta_2 - \theta_1 + \phi_3)} \\ = \; & \frac{AC\frac{ze^{i\phi_3} + z^{-1}e^{-i\phi_3}}{2} + BC\cos(\phi_2 - \phi_3)}{z^{-2}(z - z_1)(z - z_2)(z - z_3)(z - z_4)\frac{\left(Be^{i\phi_2} + Ce^{i\phi_3}\right)\left(Be^{i\phi_2} - Ce^{i\phi_3}\right)}{4}} \\ = \; & \frac{2AC\left(ze^{i\phi_3} + z^{-1}e^{-i\phi_3}\right) + 4BC\cos(\phi_2 - \phi_3)}{z^{-2}(z - z_1)(z - z_2)(z - z_3)(z - z_4)\left(Be^{i\phi_2} + Ce^{i\phi_3}\right)\left(Be^{i\phi_2} - Ce^{i\phi_3}\right)} \\ z = \; & e^{i(\theta_2 - \theta_1)} \end{aligned} \tag{5.59}$$

Finding the Fourier coefficients (in this case, the complex Fourier series)

$$c_n = \tag{5.60}$$

$$\begin{aligned} & \frac{1}{2\pi}\int_0^{2\pi}\frac{[AC\cos(\theta_2 - \theta_1 + \phi_3) + BC\cos(\phi_2 - \phi_3)]\, e^{in(\theta_2 - \theta_1)}d(\theta_2 - \theta_1)}{A^2 + B^2\cos^2(\theta_2 - \theta_1 + \phi_2) + 2AB\cos(\theta_2 - \theta_1 + \phi_2) + C^2\sin^2(\theta_2 - \theta_1 + \phi_3)} \\ = \; & \frac{1}{2\pi}\oint_{|z|=1}\frac{2AC\left(ze^{i\phi_3} + z^{-1}e^{-i\phi_3}\right) + 4BC\cos(\phi_2 - \phi_3)}{z^{-2}(z - z_1)(z - z_2)(z - z_3)(z - z_4)\left(Be^{i\phi_2} + Ce^{i\phi_3}\right)\left(Be^{i\phi_2} - Ce^{i\phi_3}\right)}\frac{z^n dz}{iz} \\ = \; & \frac{1}{2\pi}\oint_{|z|=1}\frac{2AC\left(z^2e^{i\phi_3} + e^{-i\phi_3}\right) + 4zBC\cos(\phi_2 - \phi_3)}{(z - z_1)(z - z_2)(z - z_3)(z - z_4)\left(Be^{i\phi_2} + Ce^{i\phi_3}\right)\left(Be^{i\phi_2} - Ce^{i\phi_3}\right)}\frac{z^n dz}{i} \\ = \; & \frac{1}{2\pi}\oint_{|z|=1}\frac{2AB\left(z^{2-n}e^{i\phi_3} + z^n e^{-i\phi_3}\right) + 4z^{1-n}BC\cos(\phi_2 - \phi_3)}{\left(Be^{i\phi_2} + Ce^{i\phi_3}\right)\left(Be^{i\phi_2} - Ce^{i\phi_3}\right)} \times \\ & \left[\begin{array}{c} \frac{1}{(z - z_1)(z_1 - z_2)(z_1 - z_3)(z_1 - z_4)} + \frac{1}{(z - z_2)(z_2 - z_1)(z_2 - z_3)(z_2 - z_4)} + \\ \frac{1}{(z - z3)(z_3 - z_1)(z_3 - z_2)(z_3 - z_4)} + \frac{1}{(z - z_4)(z_4 - z_1)(z_4 - z_2)(z_4 - z_3)} \end{array}\right]\frac{dz}{i} \end{aligned} \tag{5.61}$$

One would use residues to solve this, but first it is necessary to know whether the poles are inside the unit circle. Before this is done, a way of relating the expression $\frac{C\sin(\theta_2 - \theta_1 + \phi_3)}{A + B\cos(\theta_2 - \theta_1 + \phi_2)}$ to $\frac{A_2\sin(\theta_2 - \theta_1 + \varphi_2) + A_3\sin(-(\theta_2 - \theta_1) + \varphi_3)}{A_1 - A_2\cos(\theta_2 - \theta_1 + \varphi_2) + A_3\cos(-(\theta_2 - \theta_1) + \varphi_3)}$ is needed. The following equalities will provide the necessary results:

$$
\begin{aligned}
& A_2\sin(\theta_2-\theta_1+\varphi_2)+A_3\sin(-(\theta_2-\theta_1)+\varphi_3) \qquad (5.62)\\
=\; & \sqrt{A_2^2+A_3^2-2A_2A_3\cos(\varphi_2+\varphi_3)}\sin\left(\theta_2-\theta_1+\varphi_2+\arctan\left(\frac{A_3\sin(\varphi_2+\varphi_3)}{A_2-A_3\cos(\varphi_2+\varphi_3)}\right)\right)\\
& A_2\cos(\theta_2-\theta_1+\varphi_2)+A_3\cos(-(\theta_2-\theta_1)+\varphi_3) \qquad (5.63)\\
=\; & A_2\sin\left(\theta_2-\theta_1+\varphi_2+\frac{\pi}{2}\right)+A_3\sin\left(-(\theta_2-\theta_1)+\varphi_3+\frac{\pi}{2}\right)\times\\
=\; & \sqrt{A_2^2+A_3^2-2A_2A_3\cos(\varphi_2+\varphi_3+\pi)}\times\\
& \sin\left(\theta_2-\theta_1+\varphi_2+\frac{\pi}{2}+\arctan\left(\frac{A_3\sin(\varphi_2+\varphi_3+\pi)}{A_2-A_3\cos(\varphi_2+\varphi_3+\pi)}\right)\right)\\
& \sqrt{A_2^2+A_3^2+2A_2A_3\cos(\varphi_2+\varphi_3)}\cos\left(\theta_2-\theta_1+\varphi_2-\arctan\left(\frac{A_3\sin(\varphi_2+\varphi_3)}{A_2+A_3\cos(\varphi_2+\varphi_3)}\right)\right)
\end{aligned}
$$

So, one should use

$$
\begin{aligned}
A &= A_1 \\
B &= \sqrt{A_2^2+A_3^2+2A_2A_3\cos(\varphi_2+\varphi_3)}\\
C &= \sqrt{A_2^2+A_3^2-2A_2A_3\cos(\varphi_2+\varphi_3)}\\
\phi_2 &= \varphi_2-\arctan\left(\frac{A_3\sin(\varphi_2+\varphi_3)}{A_2+A_3\cos(\varphi_2+\varphi_3)}\right)\\
\phi_3 &= \varphi_3+\arctan\left(\frac{A_3\sin(\varphi_2+\varphi_3)}{A_2-A_3\cos(\varphi_2+\varphi_3)}\right)
\end{aligned} \tag{5.64}
$$

$$
\begin{aligned}
z_{1,2} &= -\frac{A+\sqrt{A^2-B^2+C^2\pm i2BC\sin(\phi_3-\phi_2)}}{Be^{i\phi_2}\mp Ce^{i\phi_3}}\\
z_{3,4} &= -\frac{A-\sqrt{A^2-B^2+C^2\pm i2BC\sin(\phi_3-\phi_2)}}{Be^{i\phi_2}\mp Ce^{i\phi_3}}
\end{aligned} \tag{5.65}
$$

$$
z_{1,2} = -\frac{A_1+\sqrt{\begin{array}{c} A_1^2-4A_2A_3\cos(\varphi_2+\varphi_3)\pm \\ \left[\begin{array}{c} i2\sqrt{\left(A_2^2+A_3^2\right)^2-4A_2^2A_3^2\cos^2(\varphi_2+\varphi_2)}\times \\ \sin\left(\arctan\dfrac{A_3\sin(\varphi_2+\varphi_3)}{A_2+A_3\cos(\varphi_2+\varphi_3)}+\arctan\dfrac{A_3\sin(\varphi_2+\varphi_3)}{A_2-A_3\cos(\varphi_2+\varphi_3)}\right)\end{array}\right]\end{array}}}{\left[\begin{array}{c}\left\{\sqrt{A_2^2+A_3^2+2A_2A_3\cos(\varphi_2+\varphi_3)}\times\exp\left\{i\left[\varphi_2-\arctan\left(\dfrac{A_3\sin(\varphi_2+\varphi_3)}{A_2+A_3\cos(\varphi_2+\varphi_3)}\right)\right]\right\}\right\}\mp \\ \left\{\sqrt{A_2^2+A_3^2-2A_2A_3\cos(\varphi_2+\varphi_3)}\times\exp\left\{i\left[\varphi_2+\arctan\left(\dfrac{A_3\sin(\varphi_2+\varphi_3)}{A_2-A_3\cos(\varphi_2+\varphi_3)}\right)\right]\right\}\right\}\end{array}\right]} \tag{5.66}
$$

$$= -\frac{A_1 + \sqrt{\begin{array}{c} A_1^2 - 4A_2A_3\cos(\varphi_2+\varphi_3) \pm i2\sqrt{(A_2^2+A_3^2)^2 - 4A_2^2A_3^2\cos^2(\varphi_2+\varphi_2)}\times \\ \left\{ \dfrac{\left[\begin{array}{c} A_3\sin(\varphi_2+\varphi_3)(A_2 - A_3\cos(\varphi_2+\varphi_3)) + \\ A_3\sin(\varphi_2+\varphi_3)(A_2 + A_3\cos(\varphi_2+\varphi_3)) \end{array}\right]}{\left[\begin{array}{c} \sqrt{(A_2+A_3\cos(\varphi_2+\varphi_3))^2 + (A_3\sin(\varphi_2+\varphi_3))^2} \\ \times\sqrt{(A_2-A_3\cos(\varphi_2+\varphi_3))^2 + (A_3\sin(\varphi_2+\varphi_3))^2} \end{array}\right]} \right\} \end{array}}}{\left\{ \dfrac{\sqrt{A_2^2+A_3^2+2A_2A_3\cos(\varphi_2+\varphi_3)}e^{i\varphi_2}\left[A_2 + A_3\cos(\varphi_2+\varphi_3) - iA_3\sin(\varphi_2+\varphi_3)\right]}{\sqrt{(A_2+A_3\cos(\varphi_2+\varphi_3))^2 + (A_3\sin(\varphi_2+\varphi_3))^2}} \right\} \mp \left\{ \dfrac{\sqrt{A_2^2+A_3^2-2A_2A_3\cos(\varphi_2+\varphi_3)}e^{i\varphi_2}\left[A_2 - A_3\cos(\varphi_2+\varphi_3) + iA_3\sin(\varphi_2+\varphi_3)\right]}{\sqrt{(A_2-A_3\cos(\varphi_2+\varphi_3))^2 + (A_3\sin(\varphi_2+\varphi_3))^2}} \right\}} \tag{5.67}$$

$$= \left\{ -A_1 + \sqrt{\begin{array}{c} A_1^2 - 4A_2A_3\cos(\varphi_2+\varphi_3) \pm \\ i2\sqrt{(A_2^2+A_3^2)^2 - 4A_2^2A_3^2\cos^2(\varphi_2+\varphi_2)}2A_2A_3\sin(\varphi_2+\varphi_3)\times \\ \left[\sqrt{A_2^2+A_3^2+2A_2A_3\cos(\varphi_2+\varphi_3)}\sqrt{A_2^2+A_3^2-2A_2A_3\cos(\varphi_2+\varphi_3)}\right]^{-1} \end{array}} \right\} \times \tag{5.68}$$

$$\left\{ \begin{array}{c} \sqrt{A_2^2+A_3^2+2A_2A_3\cos(\varphi_2+\varphi_3)}e^{i\varphi_2}\left[A_2 + A_3\cos(\varphi_2+\varphi_3) - iA_3\sin(\varphi_2+\varphi_3)\right]\times \\ 1/\sqrt{(A_2+A_3\cos(\varphi_2+\varphi_3))^2 + (A_3\sin(\varphi_2+\varphi_3))^2} \\ \mp\sqrt{A_2^2+A_3^2-2A_2A_3\cos(\varphi_2+\varphi_3)}e^{i\varphi_2}\left[A_2 - A_3\cos(\varphi_2+\varphi_3) + iA_3\sin(\varphi_2+\varphi_3)\right]\times \\ 1/\sqrt{(A_2-A_3\cos(\varphi_2+\varphi_3))^2 + (A_3\sin(\varphi_2+\varphi_3))^2} \end{array} \right\}^{-1}$$

$$= -\frac{A_1 + \sqrt{A_1^2 - 4A_2A_3\cos(\varphi_2+\varphi_3) \pm i4A_2A_3\sin(\varphi_2+\varphi_3)}}{e^{i\varphi_2}\left[A_2 + A_3\cos(\varphi_2+\varphi_3) - iA_3\sin(\varphi_2+\varphi_3) \mp (A_2 - A_3\cos(\varphi_2+\varphi_3) + iA_3\sin(\varphi_2+\varphi_3))\right]}$$

$$\begin{aligned} \Rightarrow\quad z_1 &= -\frac{A_1 + \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}} = -\frac{A_1 + \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{2A_3e^{-i\varphi_3}} && (5.69) \\ z_2 &= -\frac{A_1 + \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_2} = -\frac{A_1 + \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_2} \\ z_3 &= -\frac{A_1 - \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}} = -\frac{A_1 - \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{2A_3e^{-i\varphi_3}} \\ z_4 &= -\frac{A_1 - \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_2} = -\frac{A_1 - \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_2} \end{aligned}$$

Since $A_1^2 - 4A_2A_3 \geq 0$, this follows from the assumption that $A_1 > A_2 + A_3$. Clearly $z_{1,2}$ are both outside the unit circle. Are the other two inside the unit circle? That is,

$$\begin{aligned} \left|A_1 - \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}\right| &\gtrless \left|2A_3e^{-i\varphi_3}\right| = 2A_3 && (5.70) \\ \left|A_1 - \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}\right| &\leq -A_1 + \sqrt{A_1^2 + 4A_2A_3} \end{aligned}$$

It is enough to show that $0 < -A_1 + \sqrt{A_1^2 + 4A_2A_3} < 2A_3$

$$-A_1 + \sqrt{A_1^2 + 4A_2A_3} \gtrless 2A_3 \Leftrightarrow \tag{5.71}$$

$$\begin{aligned}
\sqrt{A_1^2+4A_2A_3} &\gtrless 2A_3+A_1 \Leftrightarrow \\
A_1^2+4A_2A_3 &\gtrless A_1^2+4A_3^2+4A_1A_3 \Leftrightarrow \\
A_2 &\gtrless A_3+A_1 \Leftrightarrow \\
A_2-A_3 &\gtrless A_1 \Leftrightarrow \\
\text{since } A_1 &> A_2+A_3 \Leftrightarrow \\
A_2-A_3 &< A_1 \Rightarrow \\
-A_1+\sqrt{A_1^2+4A_2A_3} &< 2A_3 \text{ Q.E.D}
\end{aligned}$$

So, the zeros $z_{3,4}$ are both inside the unit circle.

$$\begin{aligned}
B &= \sqrt{A_2^2+A_3^2+2A_2A_3\cos(\varphi_2+\varphi_3)} \\
C &= \sqrt{A_2^2+A_3^2-2A_2A_3\cos(\varphi_2+\varphi_3)} \\
\phi_2 &= \varphi_2-\arctan\left(\frac{A_3\sin(\varphi_2+\varphi_3)}{A_2+A_3\cos(\varphi_2+\varphi_3)}\right) \\
\phi_3 &= \varphi_3+\arctan\left(\frac{A_3\sin(\varphi_2+\varphi_3)}{A_2-A_3\cos(\varphi_2+\varphi_3)}\right)
\end{aligned} \tag{5.72}$$

$$\begin{aligned}
&\sqrt{A_2^2+A_3^2+2A_2A_3\cos(\varphi_2+\varphi_3)}\exp\left\{i\left[\varphi_2-\arctan\left(\frac{A_3\sin(\varphi_2+\varphi_3)}{A_2+A_3\cos(\varphi_2+\varphi_3)}\right)\right]\right\} \\
&= e^{i\varphi_2}\left[A_2+A_3\cos(\varphi_2+\varphi_3)-iA_3\sin(\varphi_2+\varphi_3)\right] \\
&\sqrt{A_2^2+A_3^2-2A_2A_3\cos(\varphi_2+\varphi_3)}\exp\left\{i\left[\varphi_2+\arctan\left(\frac{A_3\sin(\varphi_2+\varphi_3)}{A_2-A_3\cos(\varphi_2+\varphi_3)}\right)\right]\right\} \\
&= e^{i\varphi_2}\left[A_2-A_3\cos(\varphi_2+\varphi_3)+iA_3\sin(\varphi_2+\varphi_3)\right]
\end{aligned} \tag{5.73}$$

$$\begin{aligned}
\left(Be^{i\phi_2}+Ce^{i\phi_3}\right)\left(Be^{i\phi_2}-Ce^{i\phi_3}\right) &\Leftrightarrow e^{i2\varphi_2}4A_2\left[A_3\cos(\varphi_2+\varphi_3)-iA_3\sin(\varphi_2+\varphi_3)\right] \\
&= e^{i2\varphi_2}4A_2A_3e^{-i(\varphi_2-\varphi_3)} \\
&= 4A_2A_3e^{i(\varphi_2-\varphi_3)}
\end{aligned} \tag{5.74}$$

Finally, we have

$$\begin{aligned}
&(A_1+A_2\cos(\theta_2-\theta_1+\varphi_2)+A_3\cos(-(\theta_2-\theta_1)+\varphi_3))^2 \\
&+(A_2\sin(\theta_2-\theta_1+\varphi_2)+A_3\sin(-(\theta_2-\theta_1)+\varphi_3))^2 \\
&= z^{-2}\Big(z+\frac{A_1+\sqrt{A_1^2-4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{2A_3e^{-i\varphi_3}}\Big)\Big(z+\frac{A_1+\sqrt{A_1^2-4A_2A_3e^{+i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_2}\Big)\times \\
&\Big(z+\frac{A_1-\sqrt{A_1^2-4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{2A_3e^{-i\varphi_3}}\Big)\Big(z+\frac{A_1-\sqrt{A_1^2-4A_2A_3e^{+i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_2}\Big)A_2A_3e^{i(\varphi_2-\varphi_3)}
\end{aligned} \tag{5.75}$$

The expression we want to work with is as follows:

$$\frac{\partial}{\partial\omega t}\arctan\left(\frac{A_2\sin(\omega t+\phi_2)+A_3\sin(-\omega t+\phi_3)}{A_1+A_2\cos(\omega t+\phi_2)+A_3\cos(-\omega t+\phi_3)}\right) \tag{5.76}$$

$$= \frac{A_2^2-A_3^2+A_1A_2\cos(\omega t+\phi_2)-A_1A_3\cos(-\omega t+\phi_3)}{(A_2\sin(\omega t+\phi_2)+A_3\sin(-\omega t+\phi_3))^2+(A_1+A_2\cos(\omega t+\phi_2)+A_3\cos(-\omega t+\phi_3))^2}$$

therefore,

$$\int_0^{2\pi}\frac{\frac{1}{2\pi}\left[A_2^2-A_3^2+A_1A_2\cos(\omega t+\phi_2)-A_1A_3\cos(-\omega t+\phi_3)\right]e^{in(\theta_2-\theta_1)}d(\theta_2-\theta_1)}{\left[\begin{array}{c}(A_1+A_2\cos(\theta_2-\theta_1+\varphi_2)+A_3\cos(-(\theta_2-\theta_1)+\varphi_3))^2\\ +(A_2\sin(\theta_2-\theta_1+\varphi_2)+A_3\sin(-(\theta_2-\theta_1)+\varphi_3))^2\end{array}\right]} \tag{5.77}$$

$$= \oint_{|z|=1}\frac{\frac{1}{2\pi}\left[A_2^2-A_3^2+A_1A_2\left(ze^{i\varphi_2}+z^{-1}e^{-i\varphi_2}\right)\frac{1}{2}-A_1A_3\left(z^{-1}e^{i\varphi_3}+ze^{-i\varphi_3}\right)\frac{1}{2}\right]z^n dz/(iz)}{\begin{array}{c}z^{-2}(z+\frac{A_1+\sqrt{A_1^2-4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{2A_3e^{-i\varphi_3}})(z+\frac{A_1+\sqrt{A_1^2-4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_2})\times\\ (z+\frac{A_1-\sqrt{A_1^2-4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{2A_3e^{-i\varphi_3}})(z+\frac{A_1-\sqrt{A_1^2-4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_2})A_2A_3e^{i(\varphi_2-\varphi_3)}\end{array}}$$

$$= \oint_{|z|=1}\frac{\frac{1}{2\pi i}\left[\left(A_2^2-A_3^2\right)z+A_1A_2\frac{z^2e^{i\varphi_2}+e^{-i\varphi_2}}{2}-A_1A_3\frac{e^{i\varphi_3}+z^2e^{-i\varphi_3}}{2}\right]z^n dz}{\begin{array}{c}(z+\frac{A_1+\sqrt{A_1^2-4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{2A_3e^{-i\varphi_3}})(z+\frac{A_1+\sqrt{A_1^2-4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_2})\times\\ (z+\frac{A_1-\sqrt{A_1^2-4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{2A_3e^{-i\varphi_3}})(z+\frac{A_1-\sqrt{A_1^2-4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_2})A_2A_3e^{i(\varphi_2-\varphi_3)}\end{array}}$$

$$= \oint_{|z|=1}\frac{\frac{1}{2\pi i}\left[\left(A_2^2-A_3^2\right)z+z^2A_1\frac{A_2e^{i\varphi_2}-A_3e^{-i\varphi_3}}{2}-A_1\frac{A_3e^{i\varphi_3}-A_2e^{-i\varphi_2}}{2}\right]z^n dz}{\begin{array}{c}(z+\frac{A_1+\sqrt{A_1^2-4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{2A_3e^{-i\varphi_3}})(z+\frac{A_1+\sqrt{A_1^2-4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_2})\times\\ (z+\frac{A_1-\sqrt{A_1^2-4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{2A_3e^{-i\varphi_3}})(z+\frac{A_1-\sqrt{A_1^2-4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_2})A_2A_3e^{i(\varphi_2-\varphi_3)}\end{array}} \tag{5.78}$$

$$\begin{aligned} c_n = {} & \oint_{|z|=1}\frac{\frac{1}{2\pi i}\left[\left(A_2^2-A_3^2\right)z+z^2A_1\frac{A_2e^{i\varphi_2}-A_3e^{-i\varphi_3}}{2}-A_1\frac{A_3e^{i\varphi_3}-A_2e^{-i\varphi_2}}{2}\right]z^n dz}{(z-z_1)(z_1-z_2)(z_1-z_3)(z_1-z_4)A_2A_3e^{i(\varphi_2-\varphi_3)}} \\ & +\oint_{|z|=1}\frac{\frac{1}{2\pi i}\left[\left(A_2^2-A_3^2\right)z+z^2A_1\frac{A_2e^{i\varphi_2}-A_3e^{-i\varphi_3}}{2}-A_1\frac{A_3e^{i\varphi_3}-A_2e^{-i\varphi_2}}{2}\right]z^n dz}{(z_2-z_1)(z-z_2)(z_2-z_3)(z_2-z_4)A_2A_3e^{i(\varphi_2-\varphi_3)}} \\ & +\oint_{|z|=1}\frac{\frac{1}{2\pi i}\left[\left(A_2^2-A_3^2\right)z+z^2A_1\frac{A_2e^{i\varphi_2}-A_3e^{-i\varphi_3}}{2}-A_1\frac{A_3e^{i\varphi_3}-A_2e^{-i\varphi_2}}{2}\right]z^n dz}{(z_3-z_1)(z_3-z_2)(z-z_3)(z_3-z_4)A_2A_3e^{i(\varphi_2-\varphi_3)}} \\ & +\oint_{|z|=1}\frac{\frac{1}{2\pi i}\left[\left(A_2^2-A_3^2\right)z+z^2A_1\frac{A_2e^{i\varphi_2}-A_3e^{-i\varphi_3}}{2}-A_1\frac{A_3e^{i\varphi_3}-A_2e^{-i\varphi_2}}{2}\right]z^n dz}{(z_4-z_1)(z_4-z_2)(z_4-z_3)(z-z_4)A_2A_3e^{i(\varphi_2-\varphi_3)}} \end{aligned} \tag{5.79}$$

$$= \tag{5.80}$$

$$-\oint_{z=1}\frac{\left[\sum_{i=0}^{\infty}\left(\frac{z}{z_1}\right)^i\right]\left[(A_2^2-A_3^2)z+z^2A_1\frac{A_2e^{i\varphi_2}-A_3e^{-i\varphi_3}}{2}-A_1\frac{A_3e^{i\varphi_3}-A_2e^{-i\varphi_2}}{2}\right]z^n dz}{2\pi i z_1(z_1-z_2)(z_1-z_3)(z_1-z_4)A_2A_3e^{i(\varphi_2-\varphi_3)}}$$

$$-\oint_{z=1}\frac{\left[\sum_{i=0}^{\infty}\left(\frac{z}{z_2}\right)^i\right]\left[(A_2^2-A_3^2)z+z^2A_1\frac{A_2e^{i\varphi_2}-A_3e^{-i\varphi_3}}{2}-A_1\frac{A_3e^{i\varphi_3}-A_2e^{-i\varphi_2}}{2}\right]z^n dz}{2\pi i z_2(z_2-z_1)(z_2-z_3)(z_2-z_4)A_2A_3e^{i(\varphi_2-\varphi_3)}}$$

$$+\oint_{z=1}\frac{\sum_{i=0}^{\infty}\left(\frac{z_3}{z}\right)^i\left[(A_2^2-A_3^2)z+z^2A_1\frac{A_2e^{i\varphi_2}-A_3e^{-i\varphi_3}}{2}-A_1\frac{A_3e^{i\varphi_3}-A_2e^{-i\varphi_2}}{2}\right]z^n dz}{2\pi i z(z_3-z_1)(z_3-z_2)(z_3-z_4)A_2A_3e^{i(\varphi_2-\varphi_3)}}$$

$$+\oint_{z=1}\frac{\sum_{i=0}^{\infty}\left(\frac{z_4}{z}\right)^i\left[(A_2^2-A_3^2)z+z^2A_1\frac{A_2e^{i\varphi_2}-A_3e^{-i\varphi_3}}{2}-A_1\frac{A_3e^{i\varphi_3}-A_2e^{-i\varphi_2}}{2}\right]z^n dz}{2\pi i z(z_4-z_1)(z_4-z_2)(z_4-z_3)A_2A_3e^{i(\varphi_2-\varphi_3)}}$$

It is now an easy task to evaluate this integral using residues for different values of n

$$\begin{aligned} c_0 &= \frac{\left[z_3^2A_1\left(A_2e^{i\varphi_2}-A_3e^{-i\varphi_3}\right)\frac{1}{2}+\left(A_2^2-A_3^2\right)z_3-A_1\left(A_3e^{i\varphi_3}-A_2e^{-i\varphi_2}\right)\frac{1}{2}\right]}{(z_3-z_1)(z_3-z_2)(z_3-z_4)A_2A_3e^{i(\varphi_2-\varphi_3)}} \\ &+\frac{\left[z_4^2A_1\left(A_2e^{i\varphi_2}-A_3e^{-i\varphi_3}\right)\frac{1}{2}+\left(A_2^2-A_3^2\right)z_4-A_1\left(A_3e^{i\varphi_3}-A_2e^{-i\varphi_2}\right)\frac{1}{2}\right]}{(z_4-z_1)(z_4-z_2)(z_4-z_3)A_2A_3e^{i(\varphi_2-\varphi_3)}} \\ c_1 &= \frac{\left[z_3^3A_1\left(A_2e^{i\varphi_2}-A_3e^{-i\varphi_3}\right)\frac{1}{2}+\left(A_2^2-A_3^2\right)z_3^2-z_3A_1\left(A_3e^{i\varphi_3}-A_2e^{-i\varphi_2}\right)\frac{1}{2}\right]}{(z_3-z_1)(z_3-z_2)(z_3-z_4)A_2A_3e^{i(\varphi_2-\varphi_3)}} \\ &+\frac{\left[z_4^3A_1\left(A_2e^{i\varphi_2}-A_3e^{-i\varphi_3}\right)\frac{1}{2}+\left(A_2^2-A_3^2\right)z_4^2-z_4A_1\left(A_3e^{i\varphi_3}-A_2e^{-i\varphi_2}\right)\frac{1}{2}\right]}{(z_4-z_1)(z_4-z_2)(z_4-z_3)A_2A_3e^{i(\varphi_2-\varphi_3)}} \end{aligned} \tag{5.81}$$

In general.

$$\begin{aligned} \underset{n\geq 0}{c_n} &= \frac{z_3^{n-2}A_1\left(A_2e^{i\varphi_2}-A_3e^{-i\varphi_3}\right)\frac{1}{2}+\left(A_2^2-A_3^2\right)z_3^{n-1}-z_3^nA_1\left(A_3e^{i\varphi_3}-A_2e^{-i\varphi_2}\right)\frac{1}{2}}{(z_3-z_1)(z_3-z_2)(z_3-z_4)A_2A_3e^{i(\varphi_2-\varphi_3)}} \\ &+\frac{z_4^{n-2}A_1\left(A_2e^{i\varphi_2}-A_3e^{-i\varphi_3}\right)\frac{1}{2}+\left(A_2^2-A_3^2\right)z_4^{n-1}-z_4^nA_1\left(A_3e^{i\varphi_3}-A_2e^{-i\varphi_2}\right)\frac{1}{2}}{(z_4-z_1)(z_4-z_2)(z_4-z_3)A_2A_3e^{i(\varphi_2-\varphi_3)}} \end{aligned} \tag{5.82}$$

For negative values. it is slightly different for the first two coefficients. Remember that since the original function is real. its Fourier transform. and hence its Fourier series expansion. must be complex conjugate symmetric. $c_n = \overline{c_{-n}}$. For this reason. we do not need to evaluate the c_n coefficients for negative values of n.

$$\begin{aligned} &\frac{\partial}{\partial\omega t}\arctan\left(\frac{A_2\sin(\omega t+\varphi_2)+A_3\sin(-\omega t+\varphi_3)}{A_1+A_2\cos(\omega t+\varphi_2)+A_3\cos(-\omega t+\varphi_3)}\right) \\ &= \sum_{n=-\infty}^{\infty}c_ne^{-in\omega t}=c_0+\sum_{n=1}^{\infty}c_ne^{-in\omega t}+c_{-n}e^{in\omega t} \\ &= c_0+\sum_{n=1}^{\infty}c_ne^{-in\omega t}+\overline{c_ne^{-in\omega t}}=c_0+2\sum_{n=1}^{\infty}\mathrm{real}\left\{c_ne^{-in\omega t}\right\} \\ &= c_0+2\sum_{n=1}^{\infty}|c_n|\cos\left(-n\omega t+\angle c_n\right) \end{aligned} \tag{5.83}$$

$$\Rightarrow \tag{5.84}$$

$$\begin{aligned}
& \arctan\left(\frac{A_2 \sin(\omega t + \phi_2) + A_3 \sin(-\omega t + \phi_3)}{A_1 + A_2 \cos(\omega t + \phi_2) + A_3 \cos(-\omega t + \phi_3)}\right) \\
= \quad & c_0 \omega t - \frac{2}{n} \sum_{n=1}^{\infty} |c_n| \sin(-n\omega t + \angle c_n) + c \qquad 0 \le \omega t < 2\pi \Rightarrow \\
& \arctan\left(\frac{A_2 \sin(\omega t + \phi_2) + A_3 \sin(-\omega t + \phi_3)}{A_1 + A_2 \cos(\omega t + \phi_2) + A_3 \cos(-\omega t + \phi_3)}\right) \\
= \quad & c + \frac{2}{n} \sum_{n=1}^{\infty} |c_n| \sin(n\omega t - \angle c_n) - c_0 \cos(n\pi) \sin(n\omega t) \\
= \quad & c + \frac{2}{n} \sum_{n=1}^{\infty} |c_n| \sin(n\omega t) \cos(\angle c_n) - |c_n| \cos(n\omega t) \sin(\angle c_n) - c_0 \cos(n\pi) \sin(n\omega t) \\
= \quad & c + \frac{2}{n} \sum_{n=1}^{\infty} [\mathrm{real}(c_n) - c_0 (-1)^n] \sin(n\omega t) - imag(c_n) \cos(n\omega t)
\end{aligned}$$

Using the expressions for the poles:

$$z_1 = -\frac{A_1 + \sqrt{A_1^2 - 4A_2 A_3 e^{-i(\varphi_2 - \varphi_3)}}}{e^{i\varphi_2} 2A_3 e^{-i(\varphi_2 - \varphi_3)}} = -\frac{A_1 + \sqrt{A_1^2 - 4A_2 A_3 e^{-i(\varphi_2 - \varphi_3)}}}{2A_3 e^{-i\varphi_3}} \tag{5.85}$$

$$z_2 = -\frac{A_1 + \sqrt{A_1^2 - 4A_2 A_3 e^{-i(\varphi_2 - \varphi_3)}}}{e^{i\varphi_2} 2A_2} = -\frac{A_1 + \sqrt{A_1^2 - 4A_2 A_3 e^{-i(\varphi_2 - \varphi_3)}}}{e^{i\varphi_2} 2A_2} \tag{5.86}$$

$$z_3 = -\frac{A_1 - \sqrt{A_1^2 - 4A_2 A_3 e^{-i(\varphi_2 - \varphi_3)}}}{e^{i\varphi_2} 2A_3 e^{-i(\varphi_2 - \varphi_3)}} = -\frac{A_1 - \sqrt{A_1^2 - 4A_2 A_3 e^{-i(\varphi_2 - \varphi_3)}}}{2A_3 e^{-i\varphi_3}}$$

$$z_4 = -\frac{A_1 - \sqrt{A_1^2 - 4A_2 A_3 e^{-i(\varphi_2 - \varphi_3)}}}{e^{i\varphi_2} 2A_2} = -\frac{A_1 - \sqrt{A_1^2 - 4A_2 A_3 e^{-i(\varphi_2 - \varphi_3)}}}{e^{i\varphi_2} 2A_2} \tag{5.87}$$

To find the value of c, one can either solve for some suitable value of ωt or use the series expansion and find its DC term

$$\theta_R = \theta_1 - \sum_{n=1}^{\infty} \sum_{k=0}^{n} \left(-\frac{A_2}{A_1}\right)^n \frac{1}{n} \binom{n}{k} (A_3/A_2)^k \sin((n-k)\theta_2 - n\theta_1 + k\theta_3) \tag{5.88}$$

with

$$\begin{aligned}
\theta_1 &= 0 \\
\theta_2 &= \omega t + \varphi_2 \\
\theta_3 &= -\omega t + \varphi_3
\end{aligned} \tag{5.89}$$

$$\begin{aligned}
\theta_R &= -\sum_{n=1}^{\infty} \sum_{k=0}^{n} \left(-\frac{A_2}{A_1}\right)^n \frac{1}{n} \binom{n}{k} (A_3/A_2)^k \sin((n-k)(\omega t + \varphi_2) + k(-\omega t + \varphi_3)) \\
&= -\sum_{n=1}^{\infty} \sum_{k=0}^{n} \left(-\frac{A_2}{A_1}\right)^n \frac{1}{n} \binom{n}{k} (A_3/A_2)^k \sin((n-2k)\omega t + (n-k)\varphi_2 + k\varphi_3) \\
&= -\sum_{n=1}^{\infty} \sum_{k=0}^{n} \left(-\frac{A_2}{A_1}\right)^n \frac{1}{n} \binom{n}{k} (A_3/A_2)^k \times \\
& \quad [\sin((n-2k)\omega t) \cos((n-k)\varphi_2 + k\varphi_3) + \cos((n-2k)\omega t) \sin((n-k)\varphi_2 + k\varphi_3)]
\end{aligned} \tag{5.90}$$

The DC term is found by summing all the of terms for which $n - 2k = 0$.

$$\begin{aligned} c &= -\sum_{n=1}^{\infty} \left(-\frac{A_2}{A_1}\right)^{2n} \frac{1}{2n} \binom{2n}{n} (A_3/A_2)^n \sin\left(n\left(\varphi_2 + \varphi_3\right)\right) \\ &= -\sum_{n=1}^{\infty} \left(\frac{A_2 A_3}{A_1^2}\right)^{n} \frac{1}{2n} \binom{2n}{n} \sin\left(n\left(\varphi_2 + \varphi_3\right)\right) \end{aligned} \tag{5.91}$$

This is equivalent to

$$c = \frac{1}{2\pi} \int_0^{2\pi} \arctan\left(\frac{A_2 \sin(\omega t + \phi_2) + A_3 \sin(-\omega t + \phi_3)}{A_1 + A_2 \cos(\omega t + \phi_2) + A_3 \cos(-\omega t + \phi_3)}\right) d\omega t \tag{5.92}$$

Returning to:

$$\begin{aligned} \underset{n\geq 0}{c_n} &= \frac{z_3^{n-2} A_1 \left(A_2 e^{i\varphi_2} - A_3 e^{-i\varphi_3}\right)\frac{1}{2} + \left(A_2^2 - A_3^2\right) z_3^{n-1} - z_3^n A_1 \left(A_3 e^{i\varphi_3} - A_2 e^{-i\varphi_2}\right)\frac{1}{2}}{(z_3 - z_1)(z_3 - z_2)(z_3 - z_4) A_2 A_3 e^{i(\varphi_2 - \varphi_3)}} \\ &+ \frac{z_4^{n-2} A_1 \left(A_2 e^{i\varphi_2} - A_3 e^{-i\varphi_3}\right)\frac{1}{2} + \left(A_2^2 - A_3^2\right) z_4^{n-1} - z_4^n A_1 \left(A_3 e^{i\varphi_3} - A_2 e^{-i\varphi_2}\right)\frac{1}{2}}{(z_4 - z_1)(z_4 - z_2)(z_4 - z_3) A_2 A_3 e^{i(\varphi_2 - \varphi_3)}} \\ &= \frac{z_3^n \left[z_3^2 A_1 \left(A_2 e^{i\varphi_2} - A_3 e^{-i\varphi_3}\right)\frac{1}{2} + \left(A_2^2 - A_3^2\right) z_3 - A_1 \left(A_3 e^{i\varphi_3} - A_2 e^{-i\varphi_2}\right)\frac{1}{2}\right]}{AA} \\ &+ \frac{z_4^n \left[z_4^2 A_1 \left(A_2 e^{i\varphi_2} - A_3 e^{-i\varphi_3}\right)\frac{1}{2} + \left(A_2^2 - A_3^2\right) z_4 - A_1 \left(A_3 e^{i\varphi_3} - A_2 e^{-i\varphi_2}\right)\frac{1}{2}\right]}{BB} \end{aligned} \tag{5.93}$$

Where AA and BB are given by:

$$\begin{aligned} AA &= \left(\frac{A_1 + \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}} - \frac{A_1 - \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}}\right) \times \\ &\left(\frac{A_1 + \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_2} - \frac{A_1 - \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}}\right) \times \\ &\left(\frac{A_1 - \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_2} - \frac{A_1 - \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}}\right) A_2A_3e^{i(\varphi_2-\varphi_3)} \\ BB &= \left(\frac{A_1 + \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}} - \frac{A_1 - \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_2}\right) \times \\ &\left(\frac{A_1 + \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_2} - \frac{A_1 - \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_2}\right) \times \\ &\left(\frac{A_1 - \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}} \frac{A_1 - \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_2}\right) A_2A_3e^{i(\varphi_2-\varphi_3)} \end{aligned}$$

$$\begin{aligned} \underset{n\geq 0}{c_n} &= +\frac{z_3^n \left[z_3^2 A_1 \left(A_2 e^{i\varphi_2} - A_3 e^{-i\varphi_3}\right)\frac{1}{2} + \left(A_2^2 - A_3^2\right) z_3 - A_1 \left(A_3 e^{i\varphi_3} - A_2 e^{-i\varphi_2}\right)\frac{1}{2}\right]}{CC} \\ &+ \frac{z_4^n \left[z_4^2 A_1 \left(A_2 e^{i\varphi_2} - A_3 e^{-i\varphi_3}\right)\frac{1}{2} + \left(A_2^2 - A_3^2\right) z_4 - A_1 \left(A_3 e^{i\varphi_3} - A_2 e^{-i\varphi_2}\right)\frac{1}{2}\right]}{DD} \end{aligned} \tag{5.94}$$

Where CC and DD are given by:

$$\begin{aligned} CC &= \left(\frac{2\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2+\varphi_3)}}}{e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}}\right) \times \\ &\left(\frac{A_1 + \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2+\varphi_3)}}}{e^{i\varphi_2}2A_2} - \frac{A_1 - \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2+\varphi_3)}}}{e^{i\varphi_2}2A_3e^{-i(\varphi_2+\varphi_3)}}\right) \times \\ &\left(\frac{A_1 - \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2+\varphi_3)}}}{e^{i\varphi_2}2A_2} - \frac{A_1 - \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2+\varphi_3)}}}{e^{i\varphi_2}2A_3e^{-i(\varphi_2+\varphi_3)}}\right) A_2A_3e^{i(\varphi_2-\varphi_3)} \end{aligned}$$

$$DD = \left(\frac{A_1 + \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}} - \frac{A_1 - \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_2}\right) \times$$
$$\left(\frac{2\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_2}\right) \times$$
$$\left(\frac{A_1 - \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}} - \frac{A_1 - \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_2}\right)A_2A_3e^{i(\varphi_2-\varphi_3)}$$

$$\underset{n\geq 0}{c_n} = \frac{z_3^n\left[z_3^2A_1\frac{A_2e^{i\varphi_2} - A_3e^{-i\varphi_3}}{2} + \left(A_2^2 - A_3^2\right)z_3 - A_1\frac{A_3e^{i\varphi_3} - A_2e^{-i\varphi_2}}{2}\right]}{EE} + \frac{z_4^n\left[z_4^2A_1\frac{A_2e^{i\varphi_2} - A_3e^{-i\varphi_3}}{2} + \left(A_2^2 - A_3^2\right)z_4 - A_1\frac{A_3e^{i\varphi_3} - A_2e^{-i\varphi_2}}{2}\right]}{FF} \tag{5.95}$$

Where EE and FF are given by:

$$EE = \left[\begin{array}{c} e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}A_1 + e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}} \\ -e^{i\varphi_2}2A_2A_1 + e^{i\varphi_2}2A_2\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}} \end{array}\right] \times$$
$$\frac{A_2A_3e^{i(\varphi_2-\varphi_3)}}{e^{i\varphi_2}2A_2e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}} \times \frac{2\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}e^{i\varphi_2}2A_2e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}} \times$$
$$\left[\begin{array}{c} e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}A_1 - e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}} \\ -e^{i\varphi_2}2A_2A_1 + e^{i\varphi_2}2A_2\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}} \end{array}\right]$$

$$FF = \left[\begin{array}{c} e^{i\varphi_2}2A_2\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}} - e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}A_1 \\ +A_1e^{i\varphi_2}2A_2 + e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}} \end{array}\right] \times$$
$$\frac{2\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}e^{i\varphi_2}2A_2e^{i\varphi_2}2A_2} \times \frac{A_2A_3e^{i(\varphi_2-\varphi_3)}}{e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}e^{i\varphi_2}2A_2}$$
$$\left[\begin{array}{c} e^{i\varphi_2}2A_2A_1 - e^{i\varphi_2}2A_2\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}} - e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}A_1 \\ +e^{i\varphi_2}2A_3e^{-i(\varphi_2-\varphi_3)}\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}} \end{array}\right]$$

$$\underset{n\geq 0}{c_n} = \frac{z_3^n\left[z_3^2A_1\frac{A_2e^{i\varphi_2} - A_3e^{-i\varphi_3}}{2} + \left(A_2^2 - A_3^2\right)z_3 - A_1\frac{A_3e^{i\varphi_3} - A_2e^{-i\varphi_2}}{2}\right]}{GG} + \frac{z_4^n\left[z_4^2A_1\frac{A_2e^{i\varphi_2} - A_3e^{-i\varphi_3}}{2} + \left(A_2^2 - A_3^2\right)z_4 - A_1\frac{A_3e^{i\varphi_3} - A_2e^{-i\varphi_2}}{2}\right]}{HH} \tag{5.96}$$

Where GG and HH are given by:

$$GG = \left\{\begin{array}{c} A_1\left[A_3e^{-i\varphi_3} - A_2e^{i\varphi_2}\right] + A_3e^{-i\varphi_3}\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2+\varphi_3)}} \\ +e^{i\varphi_2}A_2\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2+\varphi_3)}} \end{array}\right\} \times$$
$$\frac{2\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2+\varphi_3)}}}{e^{i\varphi_2}2A_3e^{-i(\varphi_2+\varphi_3)}2A_3e^{i(\varphi_2-\varphi_3)}A_2} \times \frac{A_2A_3e^{i(\varphi_2-\varphi_3)}}{2A_3e^{i(\varphi_2-\varphi_3)}A_2}$$
$$\left(A_1\left[A_3e^{-i\varphi_3} - A_2e^{i\varphi_2}\right] - A_3e^{-i\varphi_3}\sqrt{A_1^2 - 4A_2A_3e^{+i(\varphi_2+\varphi_3)}} + e^{i\varphi_2}A_2\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2+\varphi_3)}}\right)$$

$$HH = \left\{ \begin{array}{c} A_1 \left[e^{i\varphi_2} A_2 - A_3 e^{-i\varphi_3}\right] + e^{i\varphi_2} A_2 \sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}} \\ +A_3 e^{-i\varphi_3}\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}} \end{array} \right\} \times$$

$$\frac{2\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}{2A_3e^{i(\varphi_2-\varphi_3)}A_2e^{i\varphi_2}2A_2}\frac{A_2A_3e^{i(\varphi_2-\varphi_3)}}{2A_3e^{i(\varphi_2-\varphi_3)}A_2}\times$$

$$\left(A_1\left[e^{i\varphi_2}A_2 - A_3e^{-i\varphi_3}\right] - e^{i\varphi_2}A_2\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}} + A_3e^{-i\varphi_3}\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}\right)$$

$$\underset{n\geq 0}{c_n} = \frac{\left\{ \begin{array}{c} z_3^n\left[z_3^2A_1\frac{A_2e^{i\varphi_2} - A_3e^{-i\varphi_3}}{2} + (A_2^2 - A_3^2)\,z_3 - A_1\frac{A_3e^{i\varphi_3} - A_2e^{-i\varphi_2}}{2}\right]\times \\ \left[\frac{4e^{i\varphi_2}A_2A_3^2e^{-i2\varphi_3}}{\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}\right] \end{array}\right\}}{\left\{ \begin{array}{c} \left[A_1\left[A_3e^{-i\varphi_3} - A_2e^{i\varphi_2}\right] + e^{i\varphi_2}A_2\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}\right]^2 - \\ \left[A_3e^{-i\varphi_3}\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}\right]^2 \end{array}\right\}} + \tag{5.97}$$

$$\frac{\left\{ \begin{array}{c} z_4^n\left[z_4^2A_1\frac{A_2e^{i\varphi_2} - A_3e^{-i\varphi_3}}{2} + (A_2^2 - A_3^2)\,z_4 - A_1\frac{A_3e^{i\varphi_3} - A_2e^{-i\varphi_2}}{2}\right]\times \\ \left[\frac{4e^{i\varphi_2}A_2^2A_3e^{i(\varphi_2-\varphi_3)}}{\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}}\right] \end{array}\right\}}{\left\{ \begin{array}{c} \left[A_1\left[e^{i\varphi_2}A_2 - A_3e^{-i\varphi_3}\right] + A_3e^{-i\varphi_3}\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}\right]^2 - \\ \left[e^{i\varphi_2}A_2\sqrt{A_1^2 - 4A_2A_3e^{-i(\varphi_2-\varphi_3)}}\right]^2 \end{array}\right\}}$$

After going through these tedious exercises, it becomes apparent it is much easier to take the Fast Fourier Transform of the received phase and simply check what the terms are. On the other hand, it is useful to have some analytical expressions, which are fairly easy to calculate on a computer, that show the complex relation between the amplitudes and the phase offsets.

6 UNIQUENESS AND STABILITY OF THE SOLUTION

Uniqueness of the solution

Before working on a problem, it is good to know if a unique solution exists, i.e., is the problem well posed? If there are multiple solutions or perhaps infinitely many solutions, it might be futile to look for a solution. This Chapter will argue that there is a unique solution to the problem, provided the phases are bandlimited. The argument is based on the assumption that the nonlinear operations given by Equations 6.5 and 6.6 will indicate an increase in bandwidth. The way the problem is formulated, the received signal S_R is the sum of three signals, S_1, S_2, and S_3, which all have the same phase bandwidth. It is therefore obvious that a solution exists. But is the solution unique?

Assume that there are two sets of solutions S_1, S_2, S_3, and $\widetilde{S_1}, \widetilde{S_2}, \widetilde{S_3}$ such that

$$\begin{aligned} S_R &= S_1 + S_2 + S_3 \\ S_R &= \widetilde{S_1} + \widetilde{S_2} + \widetilde{S_3} \end{aligned} \tag{6.1}$$

This implies that

$$0 = S_1 - \widetilde{S_1} + S_2 - \widetilde{S_2} + S_3 - \widetilde{S_3}. \tag{6.2}$$

Since it has been shown that the amplitudes can be found and there is only one unique solution to that problem, the amplitudes of S_i and $\widetilde{S_i}$ are the same. This implies that

$$S_j - \widetilde{S_j} = 2A_i \sin\left(\frac{\theta_j - \widetilde{\theta_j}}{2}\right) e^{i(\theta_j + \tilde{\theta}_j)/2 + i\pi/2} \tag{6.3}$$

Note that the bandwidth of the *phase* of $S_j - \widetilde{S_j}$ is the same as the phases of S_j, or $\widetilde{S_j}$. Rewrite Equation 6.2 as follows:

$$S_1 + S_2 - \widetilde{S_2} + S_3 - \widetilde{S_3} = \widetilde{S_1} \tag{6.4}$$

and it is clear that $\widetilde{S_1}$ is just S_1 plus two extra terms, namely, $S_2 - \widetilde{S_2}$ and $S_3 - \widetilde{S_3}$. Then,

$$\widetilde{\theta_1} = \theta_1 + \arctan \frac{A_2 \sin\left(\frac{\theta_2 - \tilde{\theta}_2}{2}\right) \sin\left(\frac{\theta_2 + \tilde{\theta}_2 + \pi}{2} - \theta_1\right) + A_3 \sin\left(\frac{\theta_3 - \tilde{\theta}_3}{2}\right) \sin\left(\frac{\theta_3 + \tilde{\theta}_3 + \pi}{2} - \theta_1\right)}{\frac{A_1}{2} + A_2 \sin\left(\frac{\theta_2 - \tilde{\theta}_2}{2}\right) \cos\left(\frac{\theta_2 + \tilde{\theta}_2 + \pi}{2} - \theta_1\right) + A_3 \sin\left(\frac{\theta_3 - \tilde{\theta}_3}{2}\right) \cos\left(\frac{\theta_3 + \tilde{\theta}_3 + \pi}{2} - \theta_1\right)} \tag{6.5}$$

and since the bandwidths of $\left(\frac{\theta_2-\tilde{\theta}_2-\pi}{2}-\theta_1\right)$ and $\left(\frac{\theta_3-\tilde{\theta}_3-\pi}{2}-\theta_1\right)$ are the same as the bandwidth of $\theta_1$and $\sin\left(\frac{\theta_2-\tilde{\theta}_2-\pi}{2}-\theta_1\right)$ has a bandwidth that is as many times wider than that of θ_1 as its modulation index. and arctan is a nonlinear continuous function. it can be assumed that the bandwidth of $\tilde{\theta}_1$ is greater than that of θ_1. This is a contradiction: therefore. the solution is unique. If $A_1 > 2(A_2 + A_3)$. then the series expansion. Equation 5.11. from the previous chapter can be used. implying that the bandwidth of $\tilde{\theta}_1$ is greater than that of θ_1. The same applies to the other phases as well. so the previous equation . 6.5. can be generalized to read:

$$\tilde{\theta}_p = \theta_p + \arctan \frac{\sum_{\substack{j=1 \\ j\neq p}}^{N} A_j \sin\left(\frac{\theta_j - \tilde{\theta}_j}{2}\right) \sin\left(\frac{\theta_j + \tilde{\theta}_j + \pi}{2} - \theta_p\right)}{A_p/2 + \sum_{\substack{j=1 \\ j\neq p}}^{N} A_j \sin\left(\frac{\theta_j - \tilde{\theta}_j}{2}\right) \cos\left(\frac{\theta_j + \tilde{\theta}_j + \pi}{2} - \theta_p\right)} \tag{6.6}$$

The argument that the bandwidth of $\tilde{\theta}_p$ must be wider than that of θ_p can now be extended to the case of more than three interfering signal. The reason it is assumed that the non-linear function arctan(x) results in a widening of the spectrum of x is simply that this is something that is hard to prove. It is well known that non-linearites does cause a widening of the spectrum: however. this does not mean that this always happens. One fairly strong argument for the assumption that the spectra of $\tilde{\theta}_p$ is wider than θ_p. is that Equation 6.6 can at some instances be approximated as:

$$\tilde{\theta}_p \cong \theta_p + \sum_{\substack{j=1 \\ j\neq p}}^{N} A_j \sin\left(\theta_j - \tilde{\theta}_j\right) \sin\left(\frac{\theta_j + \tilde{\theta}_j}{2} - \theta_p\right) : \left|\sum_{\substack{j=1 \\ j\neq p}}^{N} A_j \sin\left(\theta_j - \tilde{\theta}_j\right)\right| << A_p \tag{6.7}$$

In this expression it is clear that the spectrum of $\tilde{\theta}_p$ is wider than that of θ_p, but since the expression is not always valid it is not possible to conclude that the spectrum of $\tilde{\theta}_p$ is spread compared to that of θ_p.

Stability of the solution

It is paramount to know if the solution is stable, i.e.. if there is a small pertubation. will this lead to ever increasing errors? If the solution would be unstable, it would be futile to find it because any noise would cause the separation process to diverge. This part will show that if the three (or more) signal are approximately known then the following procedure will converge to the true signals: Create an estimate of the ith signal's phase by subtracting the other signals from the received signal and filter the phase of this residue to create an estimate of the ith signal's phase. Use this estimated phase and the amplitude of the ith signal to create an estimate, S_{iest}, of S_i. Repeat the procedure for the next signal, etc.

Assume that the solution has been found and that a small pertubation, $A_e e^{i\theta_e}$, is added to any of the signals. To make the arguments that follow a bit more mathematically stringent, it can be assumed that $A_e e^{i\theta_e} \in L^2$, is bounded and has compact support. It is not necessary to assume that $A_e e^{i\theta_e} \in C_c$, albeit a sufficient assumption, because one might want to allow the pertubation to initially have discontinuities. In the last case, the assumption that the function has compact support is essential because this implies boundedness. Assuming the error term to be small compared to the i'th signal, the introduced phase error is approximated by:

$$\theta_{iest} - \theta_i = \frac{A_e}{A_i} \sin(\theta_e - \theta_i) + \epsilon \qquad \|\epsilon\| \ll A_i \tag{6.8}$$

This will result in an error given by:

$$|S_i - S_{iest}| = 2A_i \left| \sin\left(\frac{\theta_{iest} - \theta_i}{2}\right) \right| \cong |A_e \sin(\theta_e - \theta_i)| \tag{6.9}$$

Note that this error is independent of i, which is very important, since this implies that not doing anything to the error will not cause any extra damage (at least in the immediate future). This is not the case when the error term is large because then the error in the estimate can be larger than the pertubation.

Because the term $\sin(\theta_e - \theta_i)$ has, by assumption, a greater bandwidth than $\theta_e - \theta_i$, this follows from the assumption that $\sin(\theta_i)$ has a wider bandwidth than θ_i, the filtering of the phase signal will suppress this error term. Assuming the filter's impulse response to be continuous, the filtered error terms will also be continuous. So, estimating the phase of θ_i will suppress the phase error, resulting in a new error term that is smaller than the previous one (in the L^2 norm, i.e., in the mean squared sense). The estimate of the phase of the next signal will further decrease the error. The conclusion is that any small pertubation will thus be iteratively suppressed until it is eliminated, i.e., convergence in the L^2 norm. Therefore, the solution is stable. The convergence must be pointwise, because of the fact that for continuous functions on compact sets the convergence in L^2 norm, assuming convergence to a continuous function, implies convergence in the L^∞ norm.

7 SUFFICIENT CONDITIONS FOR ITERATIVE SEPARATION

Necessary error suppression

This chapter will show which conditions are necessary for separation and also some sufficient conditions.

If S_1 has phase $\theta_1$and S_{1est} has phase θ_{1est}, both having amplitude A_1, then it follows from simple geometry that:

$$\begin{aligned} |S_1 - S_{1est}| &= 2A_1 \sin\left(\frac{|\theta_1 - \theta_{1est}|}{2}\right) \\ \angle S_1 - S_{1est} &= \frac{\pi}{2} + \frac{\theta_1 + \theta_{1est}}{2} \end{aligned} \tag{7.1}$$

A sufficient condition for separation is that S_2 dominates in the signal $S_R - S_{1est}$. This implies:

$$|S_1 - S_{1est} + S_3| < |S_2|. \tag{7.2}$$

More explicitly:

$$\begin{aligned} &[A_1 (1 - \cos(\theta_{1est} - \theta_1)) + A_3 \cos(\theta_3 - \theta_1)]^2 + \\ &[-A_1 \sin(\theta_{1est} - \theta_1) + A_3 \sin(\theta_3 - \theta_1)]^2 < A_2^2. \end{aligned} \tag{7.3}$$

It is tempting to say that this is a necessary condition, and it is if one chooses to use the method of cascaded estimators. In that case, the estimated signal is subtracted from the received signals and then it is assumed that the second (weaker) signal dominates. For that case, the above condition is necessary; however, there might be other ways of doing the separation. Geometrically the equation can be seen as a circle of radius A_1 (Figure 7.1) centered at $(A_1 + A_3\cos(\theta_3 - \theta_1), A_3 \sin(\theta_3 - \theta_1))$. Before solving this inequality, we will use the triangular inequality to get

$$|S_1 - S_{1est} + S_3| \leq |S_1 - S_{1est}| + |S_3|. \tag{7.4}$$

A sufficient condition is thus

$$|S_1 - S_{1est}| \quad \leq \quad |S_2| - |S_3| \Rightarrow \tag{7.5}$$

$$
\begin{aligned}
A_1\sqrt{2-2\cos(\theta_{1est}-\theta_1)} &= A_1 2\left|\sin\left(\frac{\theta_{1est}-\theta_1}{2}\right)\right| \le A_2 - A_3 \\
\left|\sin(\frac{\theta_{1est}-\theta_1}{2})\right| &\le \frac{A_2-A_3}{2A_1} \Leftrightarrow \\
\cos(\theta_{1est}-\theta_1) &\ge 1-\left(\frac{A_2-A_3}{2A_1}\right)^2 \Leftrightarrow \\
\theta_{1est}-\theta_1 &\le \arccos\left(1-\left(\frac{A_2-A_3}{2A_1}\right)^2\right) \Leftrightarrow \\
|\theta_{1est}-\theta_1| &\le 2\arcsin\left(\frac{A_2-A_3}{2A_1}\right).
\end{aligned}
$$

This simple criteria is, in most cases, very conservative, since it implies that there can be no error when $A_2 = A_3$. An alternative solution can be found by making use of the geometric properties. (see Figure 7.1). This figure shows the residue of $S_1 + S_3 - S_{1est}$ whose absolute value must be less than A_2. The critical angle when this happens is θ_t.

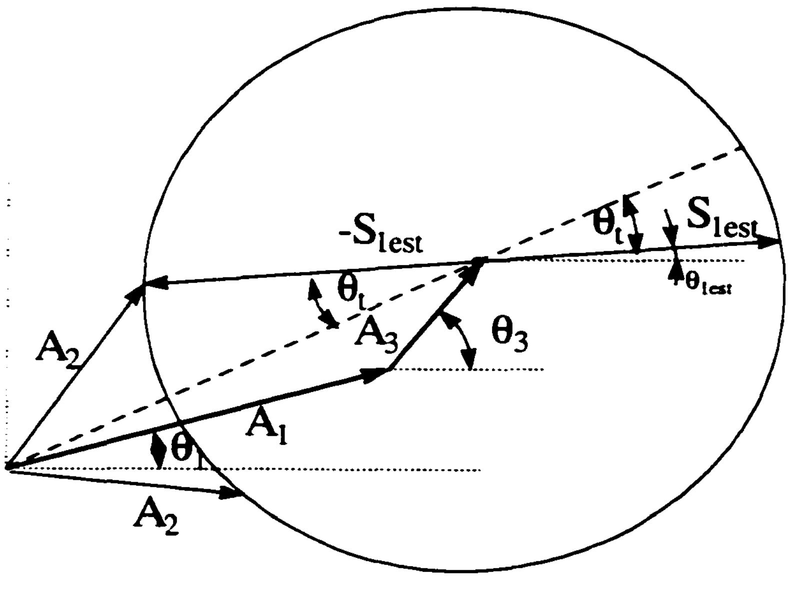


Figure 7.1 Geometric interpretation of the estimation problem.

$|S_1 + S_3 - S_{1est}| = A_2$ implies that θ_t is found from

$$A_2^2 = A_1^2 + A_u^2 - 2A_1A_u\cos(\theta_u), \tag{7.6}$$

where

$$A_u = \sqrt{A_1^2 + A_3^2 + 2A_1A_3\cos(\theta_3-\theta_1)}. \tag{7.7}$$

Finally, θ_{1est} is given by

$$\theta_{1est} = \theta_1 + \underbrace{\arctan\left(\frac{A_3\sin(\theta_3-\theta_1)}{A_1+A_3\cos(\theta_3-\theta_1)}\right)}_{\text{Odd function of } \theta_3-\theta_1} \pm \underbrace{\arccos\left(\frac{-A_2^2+2A_1^2+A_3^2+2A_1A_3\cos(\theta_3-\theta_1)}{2A_1\sqrt{A_1^2+A_3^2+2A_1A_3\cos(\theta_3-\theta_1)}}\right)}_{\text{Even function of } \theta_3-\theta_1} \tag{7.8}$$

In this expression, it is important to know which of the two solutions should be used. From Figure 7.1, it is seen that the first solution (the difference) should be used when $\theta_2 < \theta_u$ and the second solution (the sum) if $\theta_2 \geq \theta_u$.

What one wants to find is the minimum error suppression needed to ensure that the filtered received phase is sufficiently close to the phase of the stronger signal (S_1). Two entities that are of interest, namely, the ratio $(\theta_R-\theta_1)\,/\,(\theta_{1est}-\theta_1)$, where θ_{1est} is such that it guarantees that the residual signal is dominated by S_2. The second entity of interest is $\theta_R-\theta_{1est}$.

Assume the filtered phase θ_{1est} is $\theta_1+(\theta_R-\theta_1)/\beta$, where β can be seen as the necessary error suppression. To satisfy Equation 7.2, β must be sufficiently large. A critical value is when:

$$\begin{aligned}
\theta_{1est}-\theta_1 &= \arcsin\left(\frac{\dfrac{-A_1A_3\sin(\theta_3-\theta_1)\left(A_2^2-2A_1^2-A_3^2-2A_1A_3\cos(\theta_3-\theta_1)\right)}{2\left(A_1^4+2A_1^3A_3\cos(\theta_3-\theta_1)+A_1^2A_3^2\right)}\pm}{\dfrac{\left(A_1^2+A_1A_3\cos(\theta_3-\theta_1)\right)\sqrt{4A_1^2A_2^2-\left((A_2^2-A_3^2)-2A_1A_3\cos(\theta_3-\theta_1)\right)^2}}{2\left(A_1^4+2A_1^3A_3\cos(\theta_3-\theta_1)+A_1^2A_3^2\right)}}\right)\\
\Leftrightarrow \quad \frac{1}{\beta}\arcsin&\left(\frac{A_2\sin(\theta_2-\theta_1)+A_3\sin(\theta_3-\theta_1)}{\left[\begin{array}{c}(A_1+A_2\cos(\theta_2-\theta_1)+A_3\cos(\theta_3-\theta_1))^2+\\ (A_2\sin(\theta_2-\theta_1)+A_3\sin(\theta_3-\theta_1))^2\end{array}\right]}\right)\\
&= \arcsin\left(\frac{\dfrac{-A_1A_3\sin(\theta_3-\theta_1)\left(A_2^2-2A_1^2-A_3^2-2A_1A_3\cos(\theta_3-\theta_1)\right)}{2\left(A_1^4+2A_1^3A_3\cos(\theta_3-\theta_1)+A_1^2A_3^2\right)}\pm}{\dfrac{\left(A_1^2+A_1A_3\cos(\theta_3-\theta_1)\right)\sqrt{4A_1^2A_2^2-\left((A_2^2-A_3^2)-2A_1A_3\cos(\theta_3-\theta_1)\right)^2}}{2\left(A_1^4+2A_1^3A_3\cos(\theta_3-\theta_1)+A_1^2A_3^2\right)}}\right)
\end{aligned} \tag{7.9}$$

$$\Rightarrow \quad \beta\left(\begin{bmatrix}\theta_2-\theta_1\\ \theta_3-\theta_1\end{bmatrix}\right) \tag{7.10}$$

$$= \frac{\arcsin\left(\dfrac{A_2\sin(\theta_2-\theta_1)+A_3\sin(\theta_3-\theta_1)}{A_1^2+A_2^2+A_3^2+2A_1A_2\cos(\theta_2-\theta_1)+2A_1A_3\cos(\theta_3-\theta_1)+2A_2A_3\cos(\theta_2-\theta_3)}\right)}{\arcsin\left(\dfrac{\dfrac{-A_1A_3\sin(\theta_3-\theta_1)\left(A_2^2-2A_1^2-A_3^2-2A_1A_3\cos(\theta_3-\theta_1)\right)}{2\left(A_1^4+2A_1^3A_3\cos(\theta_3-\theta_1)+A_1^2A_3^2\right)}\pm}{\dfrac{\left(A_1^2+A_1A_3\cos(\theta_3-\theta_1)\right)\sqrt{4A_1^2A_2^2-\left((A_2^2-A_3^2)^2-2A_1A_3\cos(\theta_3-\theta_1)\right)^2}}{2\left(A_1^4+2A_1^3A_3\cos(\theta_3-\theta_1)+A_1^2A_3^2\right)}}\right)}$$

The next step is to find the maximum value of β, as a function of θ_2 and θ_3, which could be obtained by taking the derivative with respect to θ_2 and θ_3. Heaver, it is possible to avoid taking the first derivative. Note that in the expression, the numerator is the received phase and the denominator is independent of θ_2. By examining Figure 7.2, it is obvious that the maximum β as a function of θ_2 occurs when S_2 is orthogonal to the received signal (S_R). If $|S_1 + S_3| < |S_2|$, the maximum happens when $\theta_2 = -\theta_R$. In this case, the received signal cannot be orthogonal to S_2; however, if one assumes that $A_1 \geq A_2 + A_3$, then $A_1 - A_3 \geq A_2$ and $|S_1 + S_3| < |S_2|$ will never happen because of the triangular inequality.

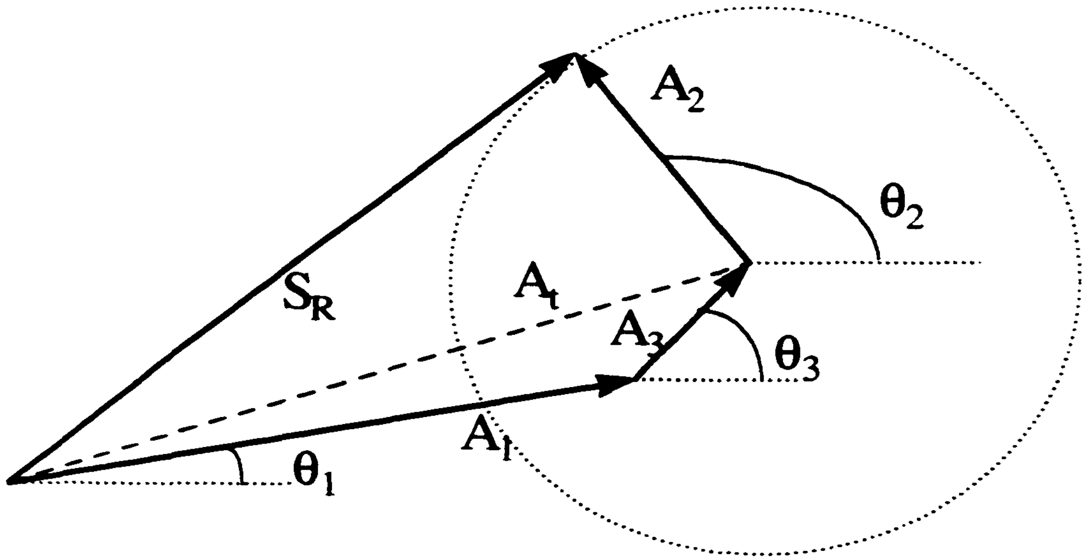


Figure 7.2 The received signal.

In the case when the signals S_2 and S_R are orthogonal,

$$\theta_2 - \theta_1 = \pi - \arccos\left(\frac{A_2}{A_t}\right) + \arccos\left(\frac{A_1 + A_3\cos(\theta_3 - \theta_1)}{A_t}\right) \Leftrightarrow \tag{7.11}$$

$$\theta_2 - \theta_1 = \arccos\left(\frac{-A_2}{\sqrt{(A_1 + A_3\cos(\theta_3 - \theta_1))^2 + (A_3\sin(\theta_3 - \theta_1))^2}}\right) + \tag{7.12}$$

$$\arccos\left(\frac{A_1 + A_3\cos(\theta_3 - \theta_1)}{\sqrt{(A_1 + A_3\cos(\theta_3 - \theta_1))^2 + (A_3\sin(\theta_3 - \theta_1))^2}}\right)$$

$$= \arccos\left(\frac{-A_2}{\sqrt{A_1^2 + A_3^2 + 2A_1A_3\cos(\theta_3 - \theta_1)}}\right) + \arccos\left(\frac{A_1 + A_3\cos(\theta_3 - \theta_1)}{\sqrt{A_1^2 + A_3^2 + 2A_1A_3\cos(\theta_3 - \theta_1)}}\right)$$

or

$$\theta_2 - \theta_1 =$$

$$-\arcsin\left(\sqrt{1-\frac{A_2^2}{(A_1+A_3\cos(\theta_3-\theta_1))^2+(A_3\sin(\theta_3-\theta_1))^2}}\right)+ \tag{7.13}$$

$$\arcsin\left(\sqrt{1-\frac{(A_1+A_3\cos(\theta_3-\theta_1))^2}{(A_1+A_3\cos(\theta_3-\theta_1))^2+(A_3\sin(\theta_3-\theta_1))^2}}\right)$$

$$=-\arcsin\left(\sqrt{\frac{A_1^2-A_2^2+A_3^2+2A_1A_3\cos(\theta_3-\theta_1)}{A_1^2+A_3^2+2A_1A_3\cos(\theta_3-\theta_1)}}\right)+\arcsin\left(\frac{A_3\sin(\theta_3-\theta_1)}{\sqrt{A_1^2+A_3^2+2A_1A_3\cos(\theta_3-\theta_1)}}\right). \tag{7.14}$$

This also implies

$$\max(\theta_R-\theta_1)=\arctan\left(\frac{A_2}{A_1+A_3\cos(\theta_3-\theta_1)}\right)+\arcsin\left(\frac{A_3\sin(\theta_3-\theta_1)}{\sqrt{A_1^2+A_3^2+2A_1A_3\cos(\theta_3-\theta_1)}}\right). \tag{7.15}$$

Then

$$\cos(\theta_2-\theta_1)=\frac{-A_2(A_1+A_3\cos(\theta_3-\theta_1))+A_3\sin(\theta_3-\theta_1)\sqrt{(A_1^2+A_3^2+2A_1A_3\cos(\theta_3-\theta_1)-A_2^2)}}{A_1^2+A_3^2+2A_1A_3\cos(\theta_3-\theta_1)} \tag{7.16}$$

and

$$\sin(\theta_2-\theta_1)=\frac{-A_2A_3\sin(\theta_3-\theta_1)+(A_1+A_3\cos(\theta_3-\theta_1))\sqrt{(A_1^2+A_3^2+2A_1A_3\cos(\theta_3-\theta_1)-A_2^2)}}{A_1^2+A_3^2+2A_1A_3\cos(\theta_3-\theta_1)}. \tag{7.17}$$

Assuming that the phase of S_2 is the worst possible, maximizing with respect to $(\theta_2-\theta_1)$, gives the needed error suppression as follows:

$$\underset{\max(\theta_2-\theta_1)}{\beta(\theta_3-\theta_1)}=\frac{\arctan\left(\frac{A_3\sin(\theta_3-\theta_1)}{A_1+A_3\cos(\theta_3-\theta_1)}\right)+\arctan\left(\frac{A_2}{\sqrt{A_1^2+A_3^2+2A_1A_3\cos(\theta_3-\theta_1)}}\right)}{\arctan\left(\frac{A_3\sin(\theta_3-\theta_1)}{A_1+A_3\cos(\theta_3-\theta_1)}\right)+\arccos\left(\frac{-A_2^2+2A_1^2+A_3^2+2A_1A_3\cos(\theta_3-\theta_1)}{2A_1\sqrt{A_1^2+A_3^2+2A_1A_3\cos(\theta_3-\theta_1)}}\right)}. \tag{7.18}$$

I spent many hours trying to find a clever way of actually finding the maximum value but in vain. It is easy to find the maximum with respect to $(\theta_2-\theta_1)$, but it is beyond my capability to find the maximum with respect to $(\theta_3-\theta_1)$. Taking the derivative of Equation 7.18 is easy, but setting it equal to zero results in a highly complicated equation. Fortunately, it is easy to find the maximum using a computer.

An alternative way of finding the error suppression (sufficient but overly pessimistic) is as follows: Max phase error is

$$\max(\theta_R-\theta_1)=\arcsin\left(\frac{A_2+A_3}{A_1}\right), \tag{7.19}$$

The sufficient condition for the acceptable error is

$$\theta_{1est} - \theta_1 \leq 2\arcsin\left(\frac{A_2 - A_3}{2A_1}\right). \tag{7.20}$$

Therefore,

$$\beta \leq \frac{\arcsin\left(\dfrac{A_2 + A_3}{A_1}\right)}{2\arcsin\left(\dfrac{A_2 - A_3}{2A_1}\right)} \tag{7.21}$$

is a sufficient condition for the error suppression. In the case when $A_1 \gg A_2 + A_3$

$$\beta \approx \frac{A_2 + A_3}{A_2 - A_3} \tag{7.22}$$

and when $A_1 = A_2 + A_3$,

$$\beta = \frac{\pi}{4\arcsin\left(\dfrac{A_2 - A_3}{2(A_2 + A_3)}\right)} \leq \frac{\pi}{2}\frac{A_2 + A_3}{A_2 - A_3}. \tag{7.23}$$

More generally, since $\frac{\pi}{2}x \geq \arcsin(x) \geq x \geq 0$,

$$\beta \leq \frac{\arcsin\left(\dfrac{A_2 + A_3}{A_1}\right)}{2\arcsin\left(\dfrac{A_2 - A_3}{2A_1}\right)} \leq \frac{\arcsin\left(\dfrac{A_2 + A_3}{A_1}\right)}{\dfrac{A_2 - A_3}{A_1}} \leq \frac{\frac{\pi}{2}\dfrac{A_2 + A_3}{A_1}}{\dfrac{A_2 - A_3}{A_1}} = \frac{\pi}{2}\frac{A_2 + A_3}{A_2 - A_3} \tag{7.24}$$

Still, when A_2 is close to A_3, the value gets very large.

The major problem not addressed here is that when filtering the received phase one can only guarantee suppression in the mean squared sense (L^2 norm) while the necessary and sufficient requirements are essentially in the L^∞ norm.

Using the amplitude information

The received amplitude does provide information about the phase of the strongest signal. Only looking at the received phase obviously discards this information. For example, when the received amplitude is either $A_1 + A_2 + A_3$ or $A_1 - A_2 - A_3$, all of the phases are known. Using the received amplitude, it is possible to give the range of the phase of S_1. Similar to the case of only two signals, there is still an ambiguity about the range that can be eliminated provided that it is possible to tell whether the phase error is positive. For a graphical illustration, see Figure 7.3. In this figure, it is assumed that the phase error is positive. The phase difference (error) between S_1 and S_R must be in the range $[\theta_{\min}, \theta_{\max}]$, otherwise $S_1 + S_2 + S_3$ cannot add up to S_R. Note that the difference could also be in the range $-[\theta_{\min}, \theta_{\max}]$. Also, when $A_3 \to 0$, there are only two possible values, i.e.,

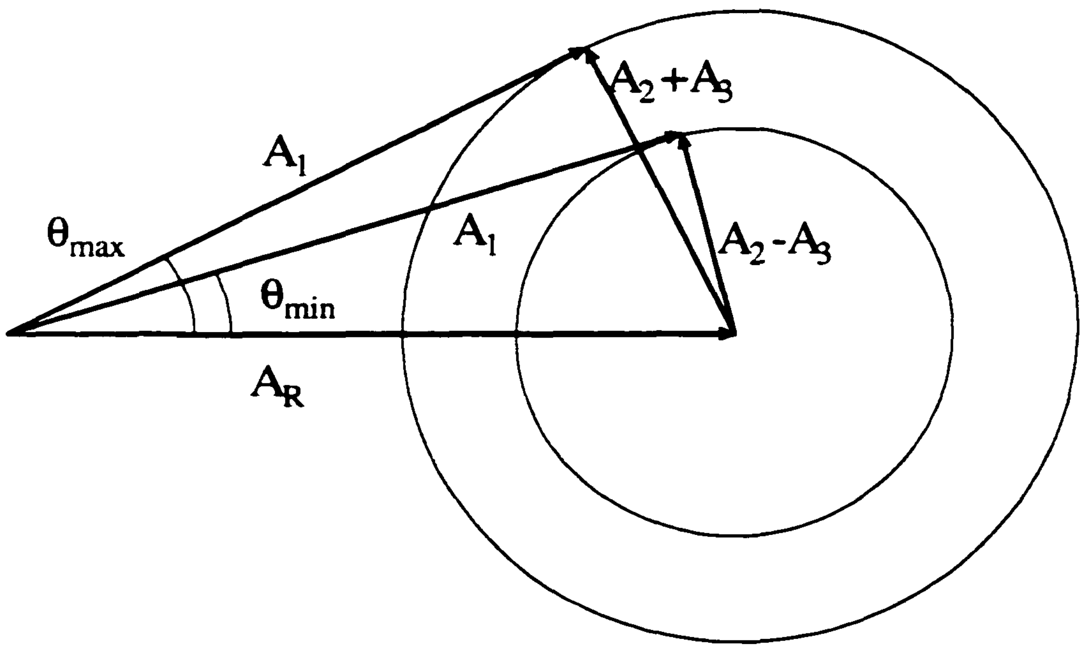


Figure 7.3 The maximum and minimum phase error as a fuction of the received amplitude.

$\theta_{\min} = \theta_{\max} \Rightarrow \theta_1 = \pm\theta_{\min}$. This is what one would expect as in the limit when $A_3 \rightarrow 0$, the problem becomes that of separating two signals. It has already been shown that in this case, one can separate the signals using the law of cosines. The method approaches that of separating two signals.

The idea is to use as the first estimate of S_1's phase, the "average" of $\theta_{\min}, \theta_{\max}$. The bounds on the phase errors are given by the law of cosines:

$$\begin{aligned} \theta_{\min} &= \arccos \frac{A_R^2 + A_1^2 - (A_2 - A_3)^2}{2A_R A_1} \\ \theta_{\max} &= \arccos \frac{A_R^2 + A_1^2 - (A_2 + A_3)^2}{2A_R A_1} \end{aligned} \tag{7.25}$$

The maximum difference between the two angles happens when $A_R = A_1 - A_2 + A_3$. This is seen by noting that the minimum occurs when the vectors of length $A_2 + A_3$ and $A_2 - A_3$ in Figure 7.3 are in phase. From the same figure, it is clear that the difference between the angles increases when moving away from this point. Hence, the maximum occurs when $\theta_{\min} = 0$, which corresponds to the two cases $A_R = A_1 - A_2 + A_3$ and $A_R = A_1 + A_2 - A_3$, for which:

$$\begin{aligned} \theta_{\max} &= \arccos \frac{(A_1 - A_2 + A_3)^2 + A_1^2 - (A_2 + A_3)^2}{2\,(A_1 - A_2 + A_3)\,A_1} \\ &= \arccos \frac{A_1^2 + (A_2 - A_3)^2 - 2A_1\,(A_2 - A_3) + A_1^2 - (A_2 + A_3)^2}{2\,(A_1 - A_2 + A_3)\,A_1} \end{aligned} \tag{7.26}$$

$$= \arccos \frac{2A_1^2 - 4A_2A_3 - 2A_1(A_2 - A_3)}{2(A_1 - A_2 + A_3)A_1}$$

or:

$$\begin{aligned}
\theta_{\max} &= \arccos \frac{(A_1 + A_2 - A_3)^2 + A_1^2 - (A_2 + A_3)^2}{2(A_1 + A_2 - A_3)A_1} \quad (7.27)\\
&= \arccos \frac{A_1^2 + (A_2 - A_3)^2 + 2A_1(A_2 - A_3) + A_1^2 - (A_2 + A_3)^2}{2(A_1 + A_2 - A_3)A_1}\\
&= \arccos \frac{2A_1^2 - 4A_2A_3 + 2A_1(A_2 - A_3)}{2(A_1 + A_2 - A_3)A_1}
\end{aligned}$$

Which of the two angles is the greatest? Consider:

$$\begin{aligned}
\arccos \frac{2A_1^2 - 4A_2A_3 - 2A_1(A_2 - A_3)}{2(A_1 - A_2 + A_3)A_1} &\lesseqgtr \arccos \frac{2A_1^2 - 4A_2A_3 + 2A_1(A_2 - A_3)}{2(A_1 + A_2 - A_3)A_1} \quad (7.28)\\
&\Leftrightarrow\\
\frac{2A_1^2 - 4A_2A_3 + 2A_1(A_2 - A_3)}{2(A_1 + A_2 - A_3)A_1} &\lesseqgtr \frac{2A_1^2 - 4A_2A_3 - 2A_1(A_2 - A_3)}{2(A_1 - A_2 + A_3)A_1}\\
\frac{A_1^2 - 2A_2A_3 + A_1(A_2 - A_3)}{(A_1 + A_2 - A_3)} &\lesseqgtr \frac{A_1^2 - 2A_2A_3 - A_1(A_2 - A_3)}{(A_1 - A_2 + A_3)}
\end{aligned}$$

$$\begin{aligned}
&A_1^2(-A_2 + A_3) - 2A_2A_3(-A_2 + A_3) + A_1(A_2 - A_3)(-A_2 + A_3) \quad (7.29)\\
\lesseqgtr\ &A_1^2(+A_2 - A_3) - 2A_2A_3(+A_2 - A_3) + A_1(A_2 - A_3)(-A_2 + A_3)
\end{aligned}$$

$$\begin{aligned}
2A_2A_3(A_2 - A_3) &\lesseqgtr A_1^2(A_2 - A_3) \quad (7.30)\\
2A_2A_3 &\lesseqgtr A_1^2
\end{aligned}$$

Once more. using of the assumption that $A_1 > A_2 + A_3$ gives

$$\begin{aligned}
A_1^2 &\geq (A_2 + A_3)^2 = A_2^2 + 2A_3A_2 + A_3^2 \Rightarrow \quad (7.31)\\
A_1^2 &\geq 2A_3A_2
\end{aligned}$$

Therefore. the maximum angle is as follows:

$$\begin{aligned}
\theta_{\max} &= \arccos \frac{2A_1^2 - 4A_2A_3 - 2A_1(A_2 - A_3)}{2(A_1 - A_2 + A_3)A_1} \quad (7.32)\\
&= \arccos \frac{A_1^2 - 2A_2A_3 - A_1(A_2 - A_3)}{(A_1 - A_2 + A_3)A_1}
\end{aligned}$$

The next question is what is the range of angles for θ_2? This turns out to be an easy problem. The maximum angle difference occurs when S_1 and S_3 are in phase.

$$\max(\theta_2 - \theta_R) \Rightarrow \quad (7.33)$$

$$
\begin{aligned}
(A_1 + A_3)^2 &= A_R^2 + A_2^2 + 2A_RA_2\cos(\theta_2 - \theta_R) \Rightarrow \\
\cos(\theta_2 - \theta_R) &= \frac{A_R^2 + A_2^2 - (A_1 + A_3)^2}{2A_RA_2}
\end{aligned}
$$

The minimum value happens when, provided that $A_R \le A_1 - A_3 + A_2$,

$$
\begin{aligned}
\min(\theta_2 - \theta_R) &\Rightarrow \\
(A_1 - A_3)^2 &= A_R^2 + A_2^2 + 2A_RA_2\cos(\theta_2 - \theta_R) \Rightarrow
\end{aligned}
\tag{7.34}
$$

$$
\begin{cases}
\min(\theta_2 - \theta_R) = \arccos\left(\frac{A_R^2 - A_2^2 - (A_1 - A_3)^2}{2A_RA_2}\right) : \left|\frac{A_R^2 - A_2^2 - (A_1 - A_3)^2}{2A_RA_2}\right| \le 1 \\
\min(\theta_2 - \theta_R) = 0 : \left|\frac{A_R^2 - A_2^2 - (A_1 - A_3)^2}{2A_RA_2}\right| > 1
\end{cases}
$$

For the separation process to work, it is required that:

$$
|S_1 - S_{1,est} + S_3| < A_2 \tag{7.35}
$$

Because of the fact that $\theta_3 \in [-\pi, \pi]$, the previous condition can only be guaranteed to hold if:

$$
|S_1 - S_{1,est}| < A_2 - A_3 \tag{7.36}
$$

Choosing the estimate of θ_1 to be $\theta_{1est} = \frac{\theta_{\max} - \theta_{\min}}{2}$, will minimize the maximum error. Using the previous results, the maximum of error, $\theta_1 - \theta_{1est}$, happens when

$$
\begin{aligned}
\theta_1 - \theta_{1est} &= \frac{\theta_{\max} - \theta_{\min}}{2} = \frac{\arccos\frac{A_1^2 - 2A_2A_3 - A_1(A_2 - A_3)}{(A_1 - A_2 + A_3)A_1}}{2} \\
|S_1 - S_{1,est}| &= 2A_1\sin\left(\frac{\arccos\frac{A_1^2 - 2A_2A_3 - A_1(A_2 - A_3)}{(A_1 - A_2 + A_3)A_1}}{4}\right)
\end{aligned}
\tag{7.37}
$$

The requirement that:

$$
|S_1 - S_{1,est}| < A_2 - A_3 \tag{7.38}
$$

implies:

$$
\begin{aligned}
\sin\left(\frac{\arccos\frac{A_1^2 - 2A_2A_3 - A_1(A_2 - A_3)}{(A_1 - A_2 + A_3)A_1}}{4}\right) &< \frac{A_2 - A_3}{2A_1} \Leftrightarrow \\
\sin\left(\frac{\arccos\frac{A_1^2 - A_1A_2 + A_3(A_1 - 2A_2)}{(A_1 - A_2 + A_3)A_1}}{4}\right) &< \frac{A_2 - A_3}{2A_1}
\end{aligned}
\tag{7.39}
$$

Using the identity,

$$\left|\sin\left(\frac{x}{4}\right)\right| = \sqrt{\frac{1-\cos(x/2)}{2}} = \sqrt{\frac{1}{2}-\frac{1}{2}\sqrt{\frac{1+\cos(x)}{2}}} \tag{7.40}$$

Equation 7.39 can be written as the algebraic equation:

$$\begin{aligned} \sqrt{\frac{1}{2}-\frac{1}{2}\sqrt{\frac{1+\dfrac{A_1^2-2A_2A_3-A_1(A_2-A_3)}{(A_1-A_2+A_3)A_1}}{2}}} &< \frac{A_2-A_3}{2A_1} \Rightarrow \\ \frac{1}{2}-\frac{1}{2}\sqrt{\frac{1}{2}+\frac{A_1^2-2A_2A_3-A_1(A_2-A_3)}{2(A_1-A_2+A_3)A_1}} &< \left(\frac{A_2-A_3}{2A_1}\right)^2 \end{aligned} \tag{7.41}$$

Since the order of the equation is more than three, it is not meaningful to try to solve it analytically. Instead, it is better to solve it numerically and plot the maximum allowable value of A_3 as a function of A_2. Because the equation can be rewritten in terms of A_2/A_1 and A_3/A_1, it is also in this case convenient to normalize with respect to A_1, so one can again assume that $A_1 = 1$ without loss of generality.

Plotting Equation 7.39 (Figure 7.4) shows that the third signal must be fairly small compared to the second signal to ensure that the second signal does indeed dominate, once the estimate of the strongest signal is subtracted from the received signal. The plot shows that there is a maximum occurring for $A_2/A_1 \approx 0.7$. An alternative way to estimate the initial phases is to use the criteria that if the received amplitude, A_R, is in the range $[A_1 - A_2, A_1 + A_2]$, chose S_1 and S_2 such that $S_R = S_{1est} + S_{2est}$. This will guarantee that $|S_1 + S_2 - S_{1est} - S_{2est}| = A_3$. This idea has not been pursued any further.

Maximum error

A slightly different way to look at the problem of separating the three signals is to see what range the signals can be within (see Figure 7.5). If the signals S_1 and S_2 are outside the regions marked by the thick line in Figure 7.5, the difference will be either greater or less than A_3. This can also be seen as the range for the cosine of the maximum difference between the estimated phase and the received phase for the three signals. For most cases, one cannot say anything about the phase of the weakest signal. The bound on $\cos(\theta_{1est} - \theta_R)$ is the one given by Equation 7.25. Similarly the bounds on $\cos(\theta_{2est} - \theta_R)$ are given by (see Figure 7.5)

$$(A_1 \pm A_3)^2 = A_R^2 + A_2^2 + 2A_RA_2\cos(\theta_{2est} - \theta_R) \tag{7.42}$$

Finally, if the received amplitude is in the range $A_R > A_1 + A_2 - A_3$ or $A_R < A_1 - A_2 + A_3$, then

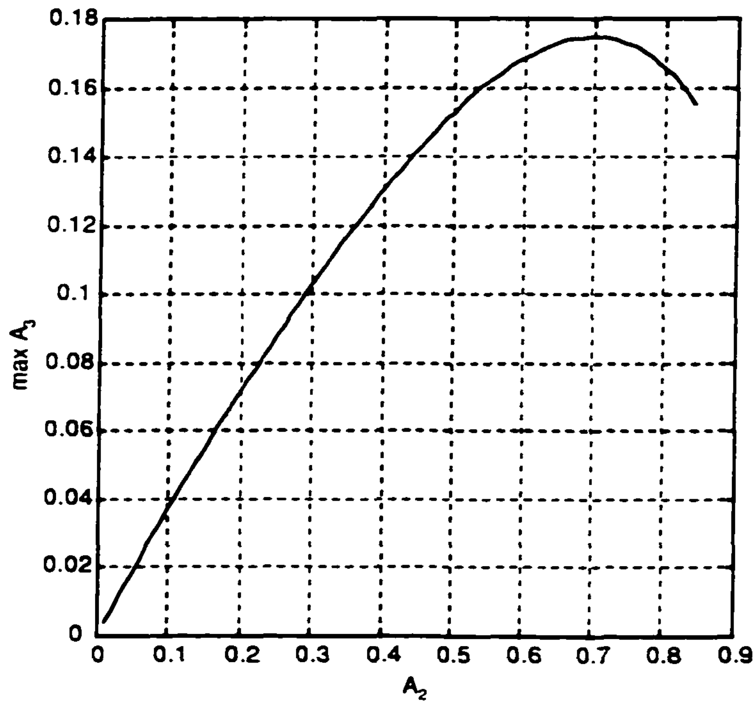


Figure 7.4 Maximum allowable value of A_3 as a function of A_2. A_1 assumed to be unity, to ensure that $|S_1 - S_{1,est} + S_3| < A_2$.

it is possible to say something about θ_3:

$$\cos(\theta_{3est} - \theta_3) > \frac{A_R^2 + A_2^2 - (A_1 \pm A_3)^2}{2A_R A_2} \tag{7.43}$$

Similar to the finding the maximum error for $\theta_{1,est}$, the maximum error of $\theta_{2,est}$ is found from

$$\begin{aligned}
(A_1 - A_3)^2 &= A_2^2 + (A_1 - A_2 + A_3)^2 + 2A_2(A_1 - A_2 + A_3)\cos(2[\theta_{2,est} - \theta_2]) && (7.44)\\
(A_1 - A_3)^2 &= A_2^2 + (A_1 + A_3)^2 + A_2^2 - 2A_2(A_1 + A_3) + 2A_2(A_1 - A_2 + A_3)\cos(2[\theta_{2,est} - \theta_2]) \\
-2A_1A_3 &= A_2^2 - A_2(A_1 + A_3) + A_2(A_1 - A_2 + A_3)\cos(2[\theta_{2,est} - \theta_2])
\end{aligned}$$

$$\cos(2[\theta_{2,est} - \theta_2]) = \frac{A_2(A_1 + A_3) - 2A_1A_3 - A_2^2}{A_2(A_1 - A_2 + A_3)} \tag{7.45}$$

The maximum error, $|S_2 - S_{2,est}|$ is then

$$\max|S_2 - S_{2,est}| = 2A_2 \sin\left(\frac{\arccos\left(\frac{A_2(A_1 + A_3) - 2A_1A_3 - A_2^2}{A_2(A_1 - A_2 + A_3)}\right)}{4}\right) \tag{7.46}$$

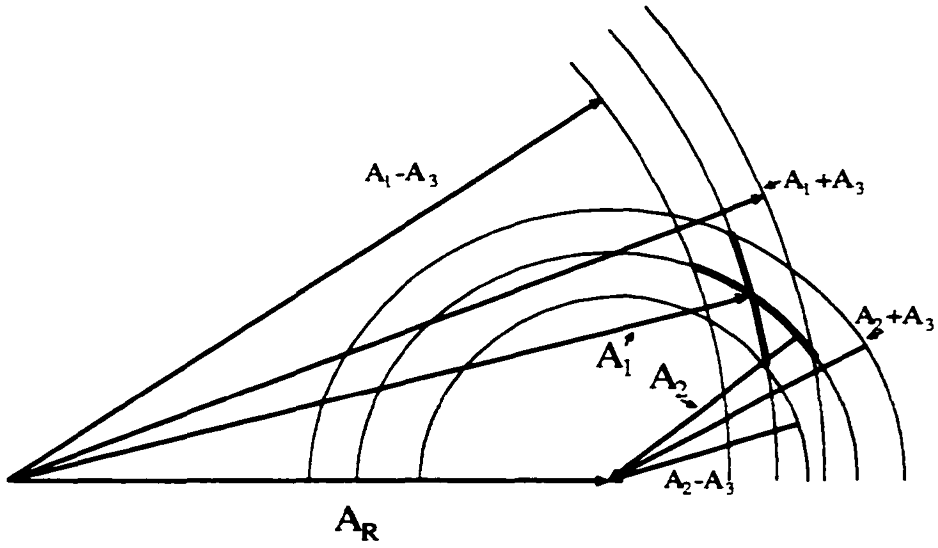


Figure 7.5 The allowable ranges.

Note that

$$\lim_{A_3 \to 0} \max |S_2 - S_{2,est}| = 0 \tag{7.47}$$

i.e., one has once more reached the case of two interferers. Compared to the previous section, the main point here is that it is possible to get more stringent restrictions on the phase of the three signals. This is very useful because it puts bounds on the maximum error on the phases, something that cannot be guaranteed when filtering the received phases.

8 THEORETICALLY GUARANTEED SEPARATION

This chapter presents a separation method that is guaranteed to separate any three FM signals. It is unfortunately not practical to implement unless one has access to almost, by today's standards, unlimited computer power. Since it has been shown that there is a unique, stable solution and that the amplitudes of the three signals can be found, the only things left to find are the phases of the three signals. This Chapter will show how this can be done. One could theoretically find the signals through an exhaustive search. Because the problem is posed as a problem containing continuous (nondiscrete) functions, there will be an uncountable number of functions to search through which is obviously not practical, because it cannot be done in finite time even with the fastest computer. Note that there is one degree of freedom in the problem of separating three FM signals, i.e., once one of the phases is known, the other two follow immediately from the law of cosines. This happens since subtracting the known signal transforms the problem into one of separating two FM signals, which can always be solved. In the case of N signals, there are $N-2$ unknowns, since the last two phases are given by the law of cosines. Since it is assumed that the phases are bandlimited, the Sampling Theorem ensures that they can be approximated as follows:

$$\begin{aligned} \theta_i(t) &\cong \sum_{i=1}^{n} a_i \chi_i(t) \\ \chi_i(x) &= \frac{\sin 2\pi f_c(t - nT_0)}{2\pi f_c t} \end{aligned} \tag{8.1}$$

where f_c is the maximum bandwidth of θ_i. Note that the function θ_i is approximated by a finite sequence. Because the functions are assumed to be bandlimited they cannot be in $C_c(C)$ because a bandlimited function cannot exist on a compact set. Since our functions (phases) are in the closure of $C_c(C)$, they can be approximated by functions given any L^P norm, including the L^∞ norm and, more importantly, the L^2 norm by a function in $C_c(C)$. For this reason, it is acceptable to see the phases as bandlimited, continuous, and living on a compact set, with the understanding that this introduces an arbitrarily small error. It is easily seen that $\theta_i \in \overline{C_c(C)}$ (with respect to the L^2 norm) since $\theta_i \in L^2$. For a thorough discussion about norms, see [32]. Since the phase signals are bounded, the range will be finite and can be discretized. i.e.. a_i in Equation 8.1 will only take on a finite set of values (a_i belongs to a finite set). The actual number of values will be determined by the acceptable error tolerance. assuming that θ_{iest} is a function as given by Equation 8.1 and a_i belongs to a finite set. In electrical engineering terms. assume that the phases are sampled at the appropriate sampling frequency (Nyquist frequency) and for some resolution of the A/D converter. Then the error

$$\begin{aligned} \varepsilon &= \|\theta_i - \theta_{iest}\|_P \\ \theta_{iest}(t) &= \sum_{i=1}^{n} a_i \chi_i(t) \end{aligned} \tag{8.2}$$

can be made arbitrarily small by increasing the resolution of the A/D converter. Conversely. given a certain error $\epsilon > 0$. it is possible to find a finite sampling frequency and resolution (a_i belonging to a finite set) such that the norm of the error is less than ε.

This indicates that it is possible to search a finite set of possible θ_{iest}. This can be completed in finite time. something that searching through an uncountable set of possible θ_{iest} is not. Even the slowest computer will find this solution. although it might take a very long time. In essence. this approximation transforms the problem from searching an uncountable set to searching a finite set. The separation procedure simply takes all the possible choices of θ_{iest} (a finite number) for a certain $i \in [1, 2, 3]$. finds a θ_{iest} that is within the bandwidth and gives the least error. defined as follows:

$$\varepsilon = \sum_{i \in [1,2,3]} \|\theta_i - \theta_{iest}\|_P \tag{8.3}$$

This error is simply the sum of the errors for each of the θ_{iest}. It is possible to come up with another type of error function based on the norms. but they will be equivalent due to the equivalence of norms. Note that although there might be more than one solution that minimizes the error. it is enough to pick one. In the section "Stability of the solution" beginning on page 6. it was shown that once the error is small enough. the iterative process will converge to the true solution. As described in the previously mentioned section. the next step is to iteratively find the right solution.

This separation process has one very serious disadvantage: the amount of computations increases exponentially. If a_i is discretized into 1024 levels (10 bits) and there are 1000 samples there will be $1024^{1000} > 10^{3000}$ possible combinations! This is obviously not practical to work with. On the other hand the advantage with the method is that it will work independently of the amplitudes of the three signal. There is no need to assume that $A_1 > A_2 + A_3$.

9 SIMULATIONS

The method discussed in the previous chapter is not practical since it is too computationally intensive. This chapter focuses on a different, but practical method of separating the three signals. Before discussing the actual simulations, it is beneficial to summarize the results so far. It has been shown that:

- The phase of the strongest signal will be enhanced by filtering the signal when $A_1 > A_2 + A_3$.
- The amplitudes of the three signals can be found.
- There are bounds on the possible phases given by the received amplitude (S_R).
- The suppression must be sufficiently large for iterative separation to work.
- The method that is guaranteed to work is too computationally intensive.

Because the guaranteed method is not practical, it is necessary to find a feasible method for separating the three signals. The obvious choice is to use an iterative method similar to the idea behind CCPLLs. The problem with this approach is that just any suppression of the error will not be enough. Filtering the received phase, knowing that it is bandlimited, will suppress the error in the mean-squared sense, but it does not ensure that the instantaneous error will decrease. One way of alleviating this problem is to make sure that θ_{1est} and θ_{2est} stay within the bounds given by Equations 7.25 and 7.42 and also ensure that they stay within their respective bandwidths. In other words, one can ensure a certain bound on the error. If the third signal, S_3, is sufficiently weak, the second signal, S_2, can be guaranteed to dominate once the estimate of S_1 is subtracted from the received signal.

The algorithm used in the simulations is as follows (see Figure 9.1): Take the received phase, filter it, and us this as the first phase estimate of the strongest signal (θ_{1est}). Check if the cosine of the phase estimate is within its bounds; if it is not; force it to be within the bound. The last step will result in a phase estimate that is not necessarily within the right bandwidth. Therefore, the whole process is repeated ten times. Ten times was chosen because experimentation showed that more iterations did

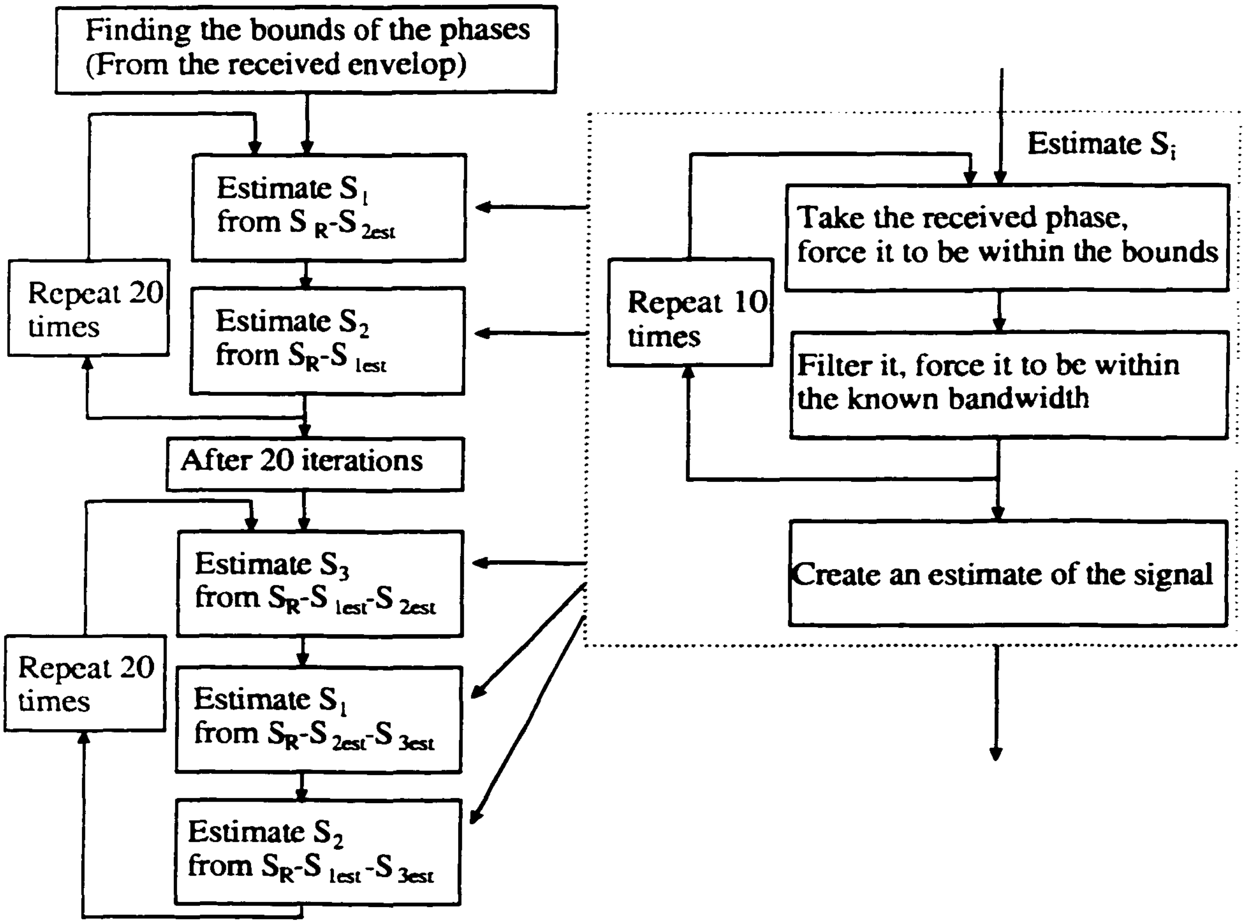


Figure 9.1 Separation that takes the allowable phase limits into account.

not seem to add any extra performance. This phase estimate is then used to form the estimate of the stronger signal, (S_{1est}), which is subsequently subtracted from the received signal.

This should leave the second signal dominating. In the previous chapter, it was shown that one can guarantee that the second signal is dominating when the third signal is sufficiently weak. The whole process is now repeated for the next signal (making sure that it is within the right bandwidth and that the cosine of the phase is within its bound).

Next the estimate of S_2 is subtracted and the phase of S_1 is again estimated. A natural question is "Why are S_{1est} and S_{2est} not subtracted and one tries to estimate the third signal?" The reason is because it is better to separate the two stronger signals as much as possible before estimating the weaker signal. Therefore, S_1 and S_2 are iteratively estimated 20 times. Once again, experimentation showed that one does not seem to gain anything by iterating more than 20 times. If S_3 is sufficiently weak, one can guarantee that S_2 dominates when S_{1est} is subtracted from the received signal and vice versa. In effect, this separates S_1 and S_2 as much as possible. The phase of S_3 is then estimated, providing an estimate of S_3. Subsequently S_{3est} and S_{2est} are both subtracted from the received signal and S_1 is estimated. Then S_{3est} and S_{1est} are both subtracted from the received signal, and S_2 is estimated, etc.

There is no guarantee that this will always separate the two signals; however, by checking the magnitude of the residual error, $|S_R - S_{1est} - S_{2est} - S_{3est}|$, one can tell if the separation was successful. If the residual is not small compared to the weakest signal, there is obviously no separation. If there is a strong capture effect, there might not be any need to support the algorithm by imposing the bound on the phase estimates. In this case, one can use the flowchart shown in Figure 9.2. This is essentially just an extension of the CCPLL idea to three interfering signals.

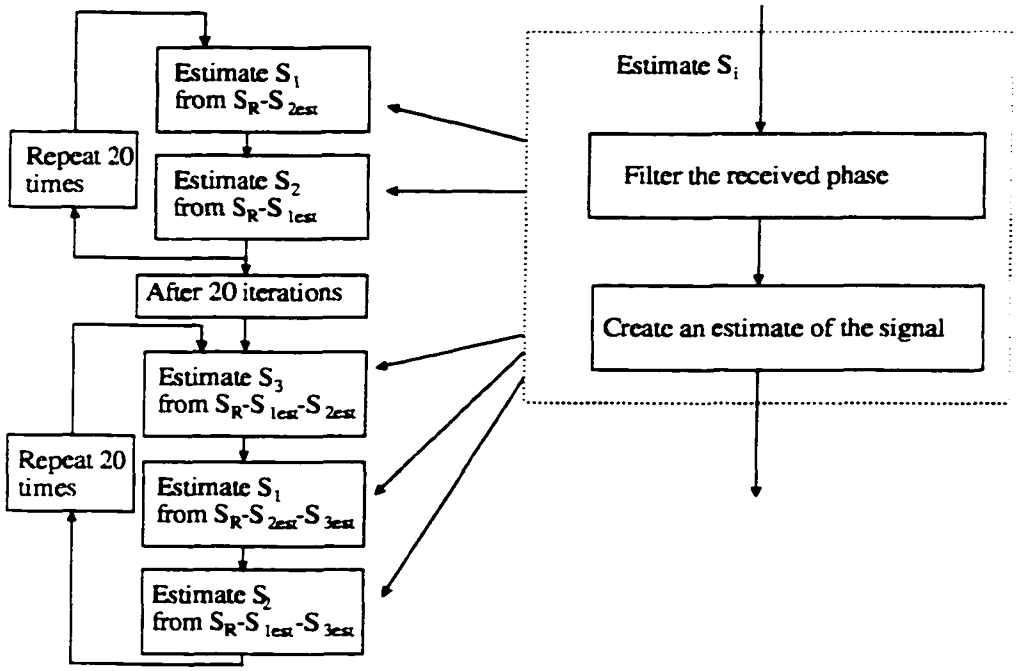


Figure 9.2 Separation relying on the capture effect only.

Practical considerations

One problem is the frequency distortion and dispersion caused by imperfect filters, i.e., sending a bandlimited signal through a filter that is not an ideal lowpass filter (constant amplitude and linear phase) will result in distortion. The distortion is of two kinds: phase distortion and amplitude distortion. This distortion will create a problem if it is of similar amplitude as the weakest signal, as it will "drown" the weakest signal. For this reason, it is important to use high-order filters with linear phase. Alternatively, one can perform the signal processing off-line and use either of two different methods for implementing the filters. The first method is to run the sequence through the filter and time reverse the output, which is sent through the filter a second time. This results in no phase distortion and a magnitude response that is the square of the original filter's magnitude response. The second method is simply to take the Fast Fourier Transform (FFT) (the normal Fourier transform will also work), simply set the high-frequency terms to zero, and then take the Inverse (Fast) Fourier Transform (IFFT). This will in effect implement an ideal lowpass filter.

Another practical problem is frequency folding, because the received phase has a theoretically infinite bandwidth. To minimize the effects of frequency folding, the sampling rate needs to be higher than twice the bandwidth of the signals (S_i). In the simulations, the sampling rate was 10 to 20 times the bandwidth of the phase signals. This introduces another problem. It is hard to implement narrowband digital filters in the time domain: for this reason, the filters were implemented using the Fourier transform method, as previously described.

Another reason for keeping the sampling rate high is to minimize phase slips. In a sampled system, the only way to determine if there is a phase jump is if the phase difference between subsequent samples is more than π radians. If the sampling rate is low, there will be many cases when the phase jumps are close to π radians. This will result in both erroneous phase jumps as well as true jumps that are missed.

Results

This section begins with a description of the signal spectra and how the distortion is measured, and then shows and discusses in detail some typical results. Following this, the results from more simulations are presented in a statistical form containing the number of successful separations and the mean and standard deviation for the demodulated signal.

Figure 9.3 shows the spectra of the phase and the transmitted signal. The modulation index is

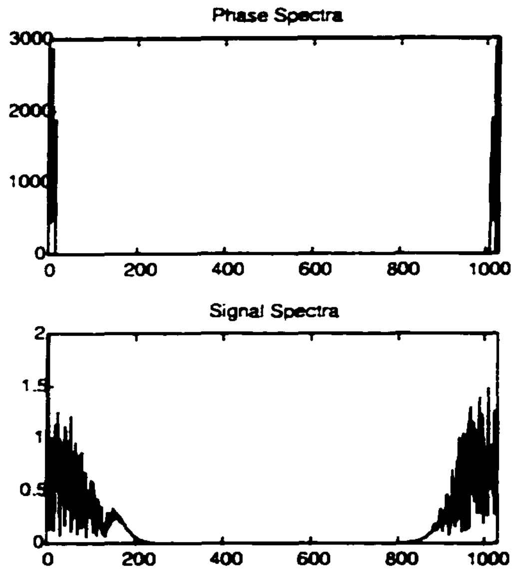


Figure 9.3 Phase spectrum and signal spectrum when the modulation index is approximately 5.

approximately four or five when defining it as the ratio of the 3 dB bandwidths. A modulation index of five is fairly common; it is used in FM broadcast systems as well as in analogue cellular radio. Running the simulations for different values of the amplitudes of the two weaker signals (A_2 and A_3), the amplitude of the strongest signal was unity ($A_1 = 1$). Note that the choice of A_1 is not important, but the ratios A_2/A_1 and A_3/A_1 are, so there is no loss in setting $A_1 = 1$. Figure 9.4 shows the spectra of the phase and the transmitted signal when the modulation index is increased to ten, also defined as the 3 dB bandwidths.

The output from the simulations is the received estimated phase of the weakest signal, and the distortion on its derivative is defined as follows:

$$SDR_{FM3} = 10 \cdot \log_{10} \frac{\int_0^{\infty} \left(\frac{\partial}{\partial t}\theta_{3est} - \frac{\partial}{\partial t}\theta_3\right)^2 dt}{\int_0^{\infty} \left(\frac{\partial}{\partial t}\theta_3\right)^2 dt}. \tag{9.1}$$

This measure was chosen because it represents the distortion in an FM receiver, and it shows the impact of phase slips without being totally "thrown off" by a phase slip. When measuring the distortion on the received phase, a phase slip will cause a very large error in the calculated distortion figure, when in effect the error is only an offset value, i.e., it could indicate poor performance when in reality performance is satisfactory. Since the challenge lies in finding the weakest signal, only the estimated phase of the weakest signal is presented. Not too surprising, there has not been a single case when the weakest signal

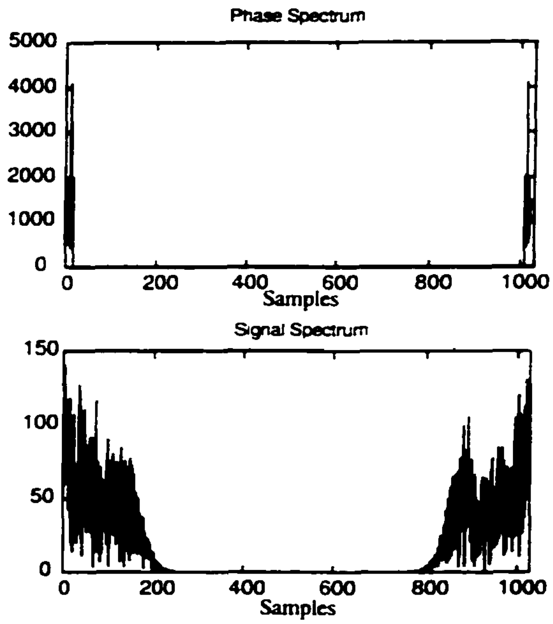


Figure 9.4 Signal and phase spectra when the modulation index is approximately 10.

was successfully demodulated and either or both of the two stronger signals were not.

In the first case (Figure 9.5), the third signal S_3 is so weak that Equation 7.41 guarantees that the second signal will always dominate when the estimate of the strongest signal is subtracted from the received signal. One would suspect that the separation process should work well and indeed it does. The distortion as measured by the SDR_{FM3} is 23.5 dB. After examining Figure 9.5, it is clear there is very little distortion except at the point where the estimated phase is slightly below the true phase.

The next figure, Figure 9.6, shows the effect of phase slips. The received signal follows the true phase very well except where there are phase slips, which temporarily cause a high error. The fact that the phase slips are the main culprit is evident in the next figure, Figure 9.7, where the SDR_{FM3} is 46.5 dB. As an extreme case, letting S_3 be 80 dB below S_1 (see Figure 9.8), shows it is still possible to successfully demodulate S_3. Increasing the value of A_2 to 0.8 (see Figure 9.9), shows the separation (demodulation) still works fine, with an SDR_{FM3} of 11.3 dB. The error is mainly caused by a few phase slips appearing at the end of the signal. Finally, with $A_2 = 0.8$ and $A_3 = 0.2$, the separation fails (see Figure 9.10). According to the theoretical results, one is not guaranteed separation in this case because the condition $A_1 > A_2 + A_3$ is violated. On the other hand, the theory does not say that separation will necessarily fail.

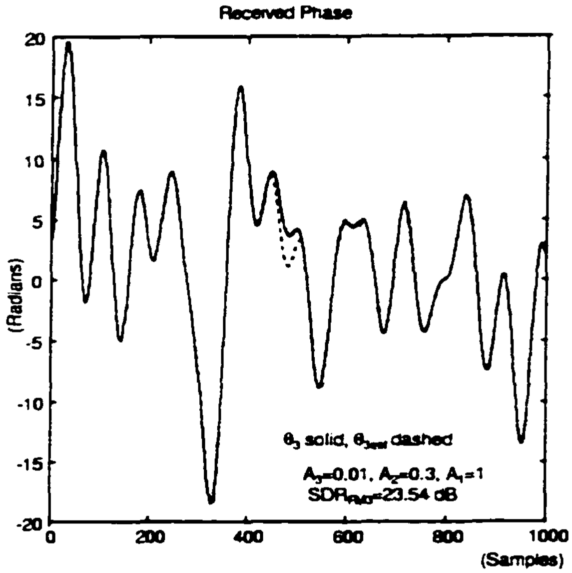


Figure 9.5 Good demodulation. S_3 is 40 dB below S_1.

It is of interest to examine, in more detail, how the separation process behaves. In Figure 9.11, the received phased and the residuals $|S_R - S_{1est} - S_{2est}|$, $|S_R - S_{1est} - S_{2est} - S_{3est}|$ are shown. The upper right plot is the most interesting because it shows that the residual is not always dominated by S_3. The phase slips only happen when this situation occurs. The opposite is not true. Although the residue is larger than A_3, this does not necessarily cause a phase slip.

Because the separation does not always work, it is important to obtain some statistical measure of how well it works. The simulations were repeated 50 times for the cases $A_2 = 0.1$, 0.3, 0.5, and 0.8. For each of these cases, A_3 was 0.1, 0.01, 0.001, and 0.0001 except for the case when $A_2 = 0.1$ for which the case of $A_3 = 0.1$ was excluded. Finally, the modulation index was 5 or 10, and in all cases, the three carrier frequencies were the same. This resulted in 30 different cases and a total of 1500 trials. Because of the high number of different combinations of possible amplitudes, modulation indices, filters, and carrier offsets, it is impractical to make simulations that cover all combination.

Figures 9.12 and 9.13 show the percentages of successful separations. The separation process was deemed successful if the standard deviation, i.e. power, of the residual signal $S_R - S_{1est} - S_{2est} - S_{3est}$ was less than A_3. This criteria was chosen because this is something that can be measured in practice, and it correlates very well with the separation process being successful or not. Choosing a more conservative limit would lower the success rate but would increase the average SDR_{FM3} when the separation was

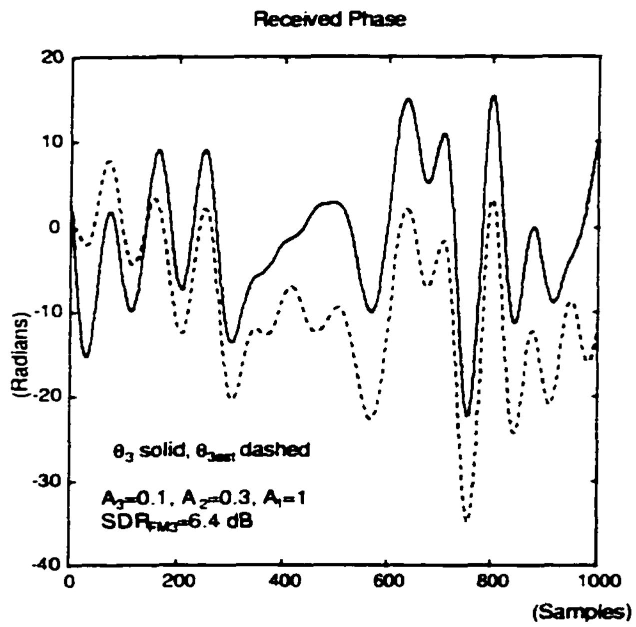


Figure 9.6 Fairly good separation, except for phase slips.

deemed successful. As one would suspect, the probability that the separation will be successful increases as the modulation index increases, rising from 50% on average to 70%. This is also predicted by the theory since the capture effect gets more pronounced as the modulation index increases.

For the cases when the separation was deemed successful, the average distortion on the demodulated FM signal for the weakest signal was calculated (Figures 9.14 and 9.15). The worst case is when the modulation index is approximately 5 and the third signal's (A_3) amplitude is 0.0001. For this case, the average SDR_{FM3} is approximately 15 dB, and on average, the SDR_{FM3} is approximately 20 dB for all cases. When the modulation index is increased to approximately 10, the SDR_{FM3} is on average 30 dB. In both cases, the demodulated signal degrades as the third signal gets weaker. It is seen that if the separation was successful, the distortion on the weakest signal is quite low. Furthermore, the distortion is mainly cased by phase slips. In between the phase slips, the distortion is negligible. As a quality measure, the spread of the average distortion is shown in Figures 9.16 and 9.17.

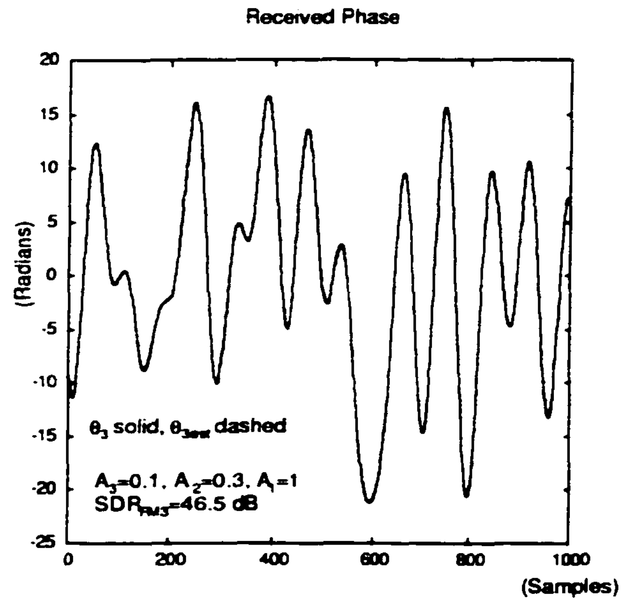


Figure 9.7 Hardly any distortion.

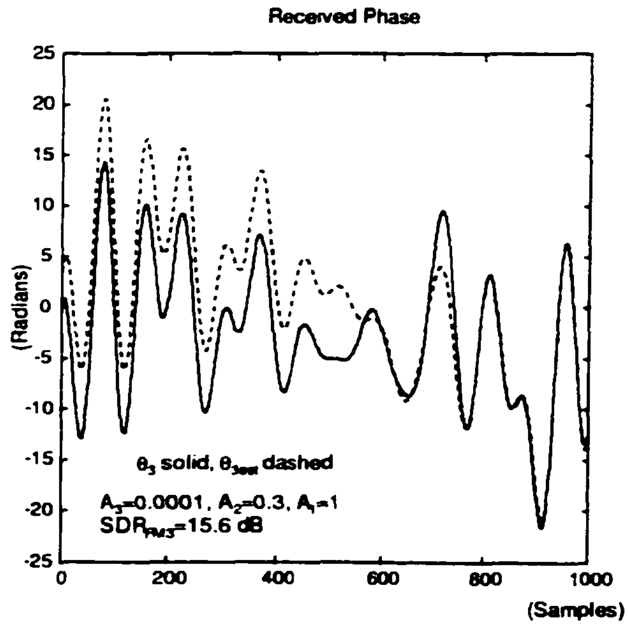


Figure 9.8 Sucessful demodulation of very weak signal.

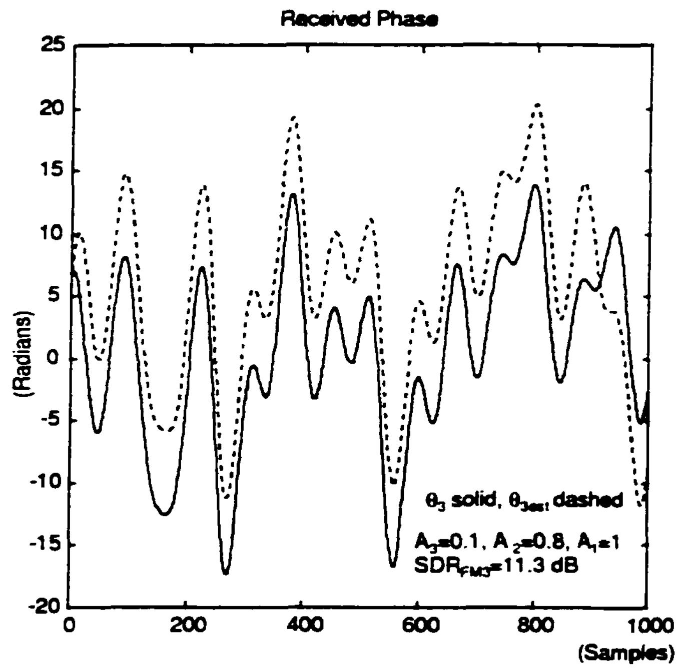


Figure 9.9 Acceptable demodulation of the weakest signal.

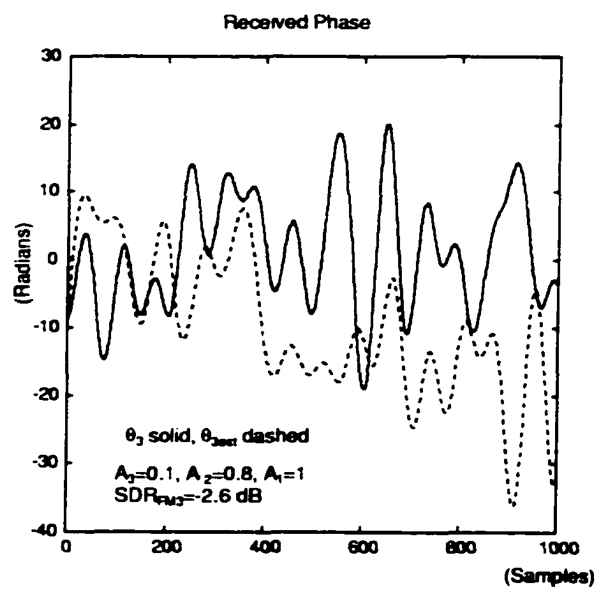


Figure 9.10 Unsucessful separation.

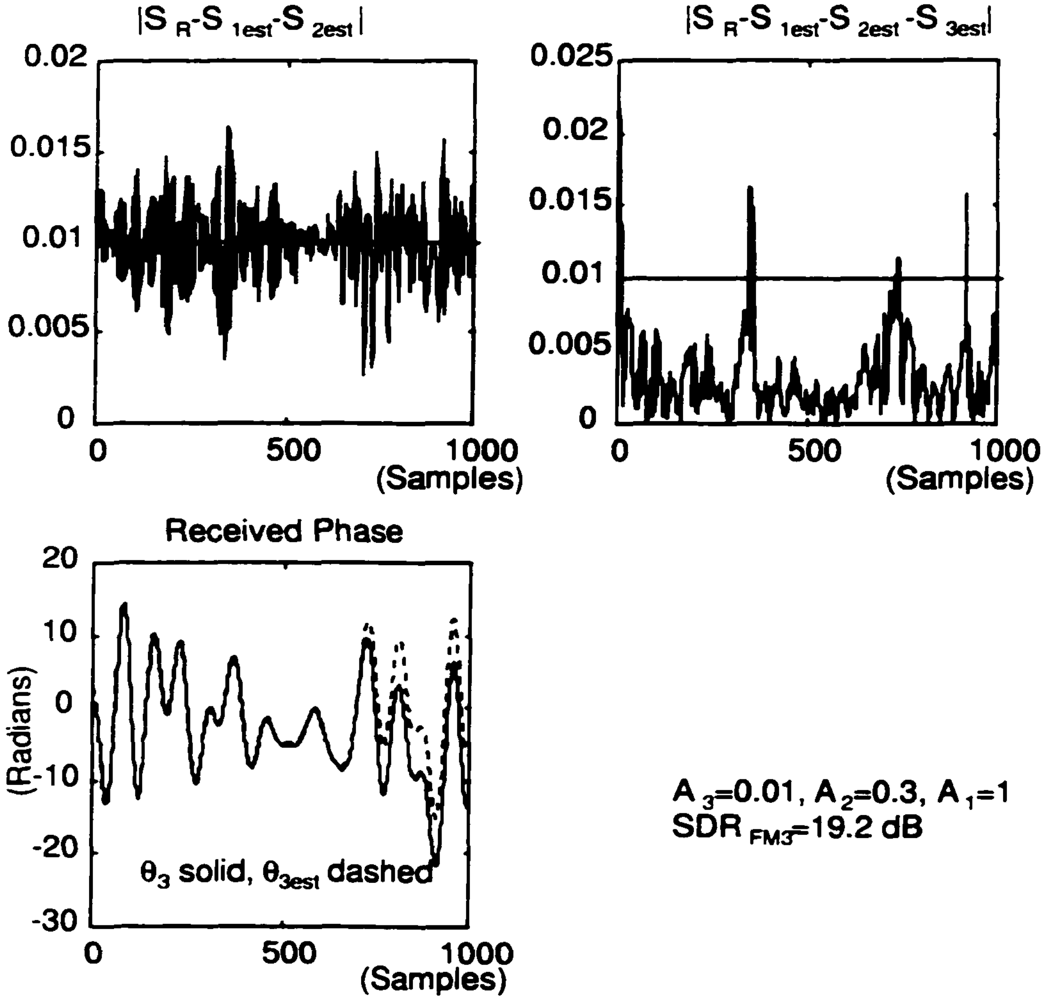


Figure 9.11 The residuals and demodulated signal.

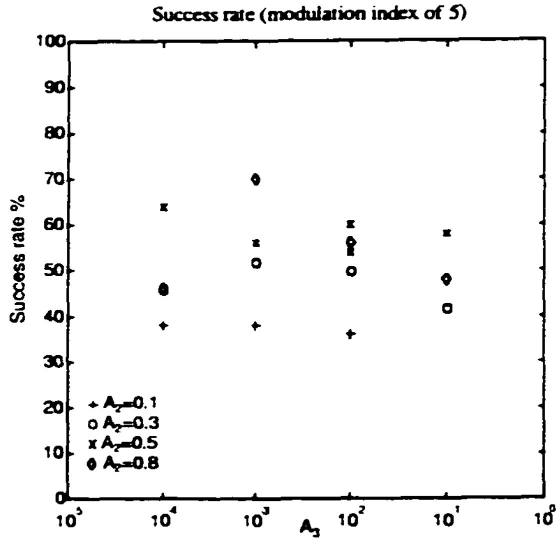


Figure 9.12 Percentage of successful separation.

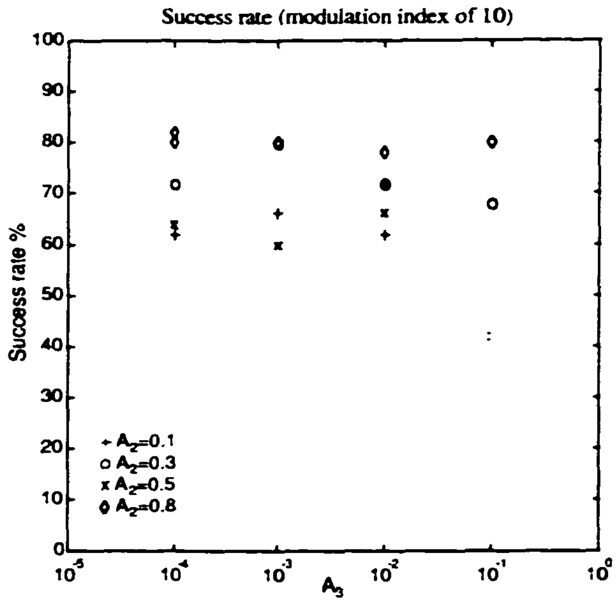


Figure 9.13 Percentage of successful separation.

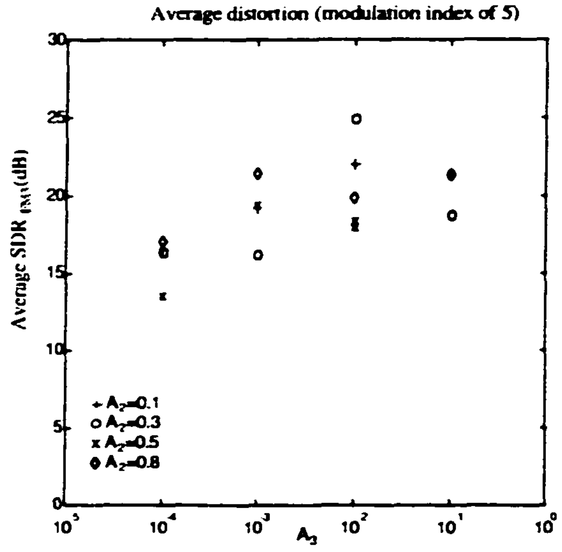


Figure 9.14 Average distortion when the modulation index is 5.

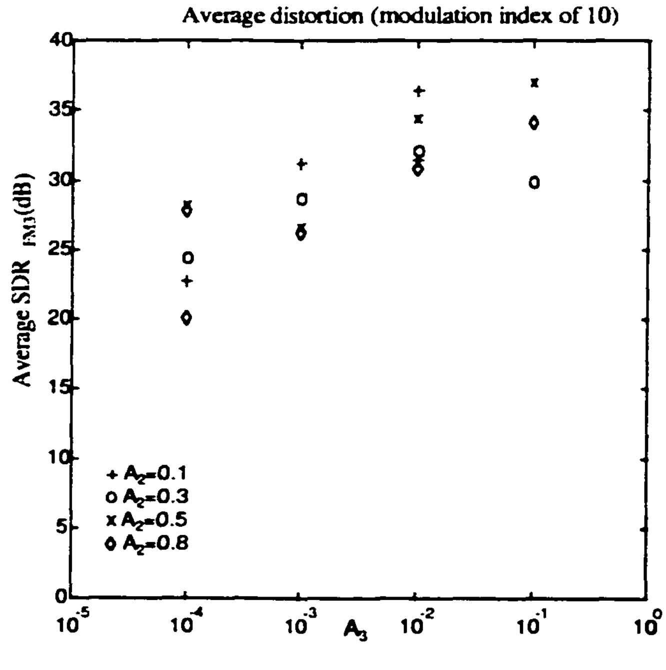


Figure 9.15 Average distortion when the modulation index is 10.

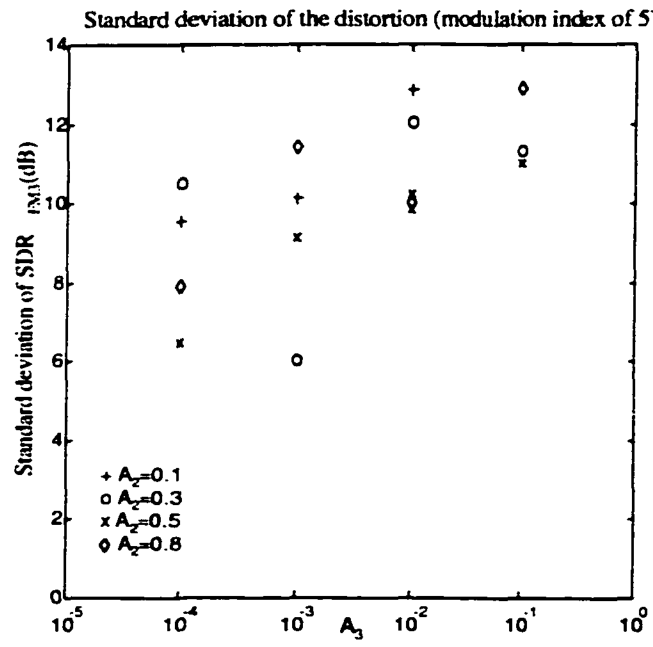


Figure 9.16 Standard deviation of the distortion. modulation index of 5.

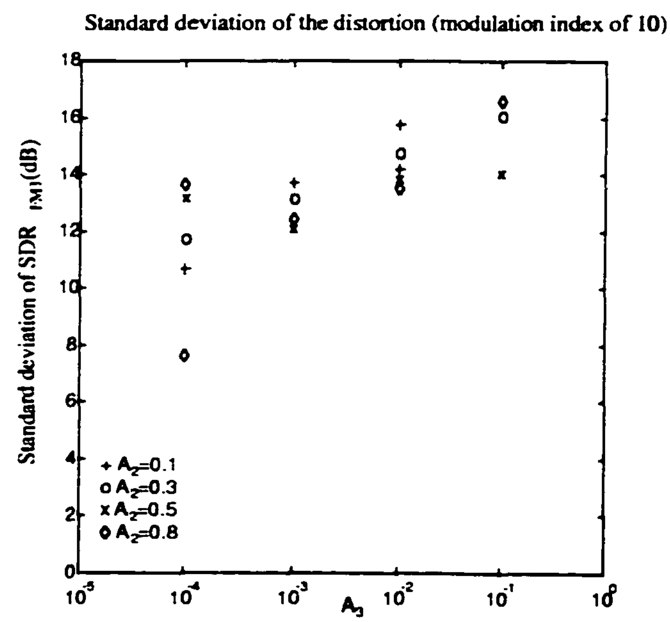


Figure 9.17 Standard deviation of the distortion, modulation index of 10.

10 HIGH-ORDER VOLTERRA SERIES BASED PRE-DISTORTION FOR HIGH-EFFICIENCY RF POWER AMPLIFIERS

A paper submitted to IEEE Transactions on Communications.

S. Anders Mattsson and Steve F. Russell

Abstract

In this paper, we derive high-order Volterra based pre-distorters for a broad class of nonlinear amplifiers. The traditional problems associated with high-order Volterra kernels are bypassed by using the structure of the amplifier. It is seen that the problem of pre-distortion is mainly that of estimating the amplifier. Once this is completed, finding the pre-distorter is straightforward. The derived pre-distorters are actually high-order Volterra operators. The proposed method results in easily implemented, but high-order pre-distorters. The method is applicable to both pre-distorting RF and baseband signals. If baseband signals are used, the in-band distortion is cancelled while the out-of-band distortion is essentially unchanged. The work was founded by Rockwell International, Cedar Rapids, Iowa, contract CFW0495-12.

Introduction

In the quest for highly linear amplifiers, the latest approach is to accept that the amplifier is nonlinear and to pre-distort the signal in such away as to compensate for the nonlinear behavior, [2] to [11]. All the references have, sometimes unknowingly, used Volterra series to accomplish this goal. Note that a series expansion can be seen as a Volterra series. Using the Volterra series theory to describe nonlinear amplifiers is very attractive because it makes it possible to immediately find a pre-distorter, provided the pre-distorter exists. This pre-distorter is, by itself, another Volterra series. Furthermore, this pre-distorter will also work as a post-distorter. The problem with this idea is that both the mathematical treatment and the implementation gets complicated due to the nonlinearity of the problem. As the order increases, the practical implementation of Volterra kernels becomes excessively complicated. Also, finding the Volterra kernels is in general awkward [1]. These problems are painfully evident as in references [1] to [7] where only up to third-order and simplified fifth-order Volterra pre-distorters have been implemented for nonlinear systems with memory. The standard implementation of Volterra series of higher order than three is not practical.

To achieve satisfactory performance, it is necessary to implement high-order pre-distorters. We will show that it is possible to do this in a highly efficient way for a broad class of amplifiers, including all those used in [2] to [11].

Preliminaries

This section contains a very brief introductory description of Volterra series since they are not commonly used. A good, very readable reference is [1].

Notation:

$$\text{The Volterra operator is defined by } \mathbf{H} = \sum_{n=1}^{\infty} h_n(\tau_1, \ldots, \tau_n) \tag{10.1}$$

where $h_n(\tau_1, \ldots, \tau_n)$ is the n'th-order Volterra kernel.

If $\mathbf{H}$ has only a finite number of kernels, then the notation $\mathbf{H}_n$ is used to indicate that the highest order kernel is n or it will be used to indicate a Volterra operator that has only an n'th order kernel. The context should make it clear which case is used. The Volterra operator maps an input $x(t)$ to an output $y(t)$, according to the rule:

$$\begin{aligned} y(t) &= \mathbf{H}[x(t)] = \\ &= \sum_{n=1}^{\infty} \underbrace{\int \cdots \int a_n h_n(\tau_1, \ldots, \tau_n) * x(t-\tau_1) \ldots * x(t-\tau_n) d\tau_1 \ldots \tau_n}_{n \text{ times convolution}} . \end{aligned} \tag{10.2}$$

Note that the kernel h_1 is just the impulse response of a normal linear filter. For our purpose, it is prudent to assume that $\sum \|h_n\|_2 < \infty$ and $x(t) \in L^{\infty}$ then $y(t) \in L^{\infty}$. Volterra operators can be used to describe systems both nonlinear and linear with memory and can be seen as a combination of impulse responses and series expansion or as an extension of the impulse response.

A good example for clarifying these concepts is to consider a filter with impulse response $h(t)$ and a nonlinearity described by $y(z) = z + c_3 z^3$. Consider the three cases: (1) sending a signal $x(t)$ through the nonlinearity, (2) sending it through the filter and then the nonlinearity, and finally (3), sending it through the nonlinearity and then the filter.

$$\text{Case 1 } y(t) = x(t) + c_3 x^3(t) \Leftrightarrow y(t) = \mathbf{H}[x(t)] \quad \text{where} \tag{10.3}$$

$$\mathbf{H} = \delta(t_1) + c_3\delta(\tau_1)\delta(\tau_2)\delta(\tau_3).$$

$$\text{Case 2 } y(t) = h * x(t) + c_3\left[h * x(t)\right]^3 \Leftrightarrow y(t) = \mathbf{H}\left[x(t)\right] \quad \text{where} \tag{10.4}$$
$$\mathbf{H} = h(\tau_1) + c_3 h(\tau_1)h(\tau_2)h(\tau_3).$$

$$\text{Case 3 } y(t) = \left[x(t) + c_3 x^3(t)\right] * h(t) \Leftrightarrow y(t) = \mathbf{H}\left[x(t)\right] \quad \text{where} \tag{10.5}$$
$$\mathbf{H} = h(\tau_1) + c_3\delta(\tau_1 - \tau_3)\delta(\tau_2 - \tau_3)h(\tau_3).$$

The first example shows how a series expansion of a function can be seen as Volterra operator and how it expands on the concept of impulse response. The Volterra operator in the last example is not unique. e.g.. $\mathbf{H} = h_1(\tau_1) + c_3\delta(\tau_1 - \tau_2)h(t_2)\delta(\tau_3 - \tau_2)$ gives the same output. To get a unique Volterra operator. one can impose the condition that it must be symmetric. which in this example is $h(t_1) + c_3\frac{1}{3}\left[\delta(\tau_1 - \tau_3)\delta(\tau_2 - \tau_3)h(\tau_3) + \delta(\tau_1 - \tau_2)h(\tau_2)\delta(\tau_3 - \tau_2) + h(\tau_1)\delta(\tau_2 - \tau_1)\delta(\tau_3 - \tau_1)\right]$.

Provided that h_1 is invertible. the theory guarantees a p'th order pre-distorter $\mathbf{K}_p$ exists such that

$$\mathbf{Q} = \mathbf{H}\left\{\mathbf{K}_p\left[x(t)\right]\right\} = \delta(t) + \sum_{n=p+1}^{\infty} q_n(\tau_1, \dots, \tau_n), \tag{10.6}$$

this pre-distorter ($\mathbf{K}_p$) will eliminate all 2nd- to p'th-order intermodulation. note that $q_2 \dots q_p$ are all zero. Finding this pre-distorter is all that is needed: unfortunately. this involves two difficult problems. One problem is to find the operators of the pre-distorter. but this is messy because they are increasingly hard to measure as the order increases [1]. The second problem is to implement the inverse; the rest of this section is devoted to this problem.

To find the kernels of the pre-distorter. a formula for series connection of Volterra operators is needed [1]. The easiest way to find the kernels q_n is to equate powers of c in the following equation (Eq. 10.7):

$$\mathbf{Q} = \sum_{n=1}^{\infty} c^n q_n(\tau_1, \dots, \tau_n) = \sum_{m=1}^{p}\sum_{n_1=1}^{\infty} \cdots \sum_{n_m=1}^{\infty} c^{n_1 + \cdots + n_m} k_m h_{n_1} h_{n_2} \cdots h_{n_m - 1} h_{n_m} \tag{10.7}$$

i.e..

$$q_n = \sum_{U_m} k_m h_{n_1} h_{n_2} \cdots h_{n_m - 1} h_{n_m} \tag{10.8}$$
$$U_m = \{n_1, \cdots, n_m : n_1 + \cdots + n_m = n\}$$

As an example. the operator for a third-order pre-distorter is given by

$$\mathbf{Q}_1 = \mathbf{K}_1\mathbf{H}_1 = 1 \tag{10.9}$$
$$\mathbf{Q}_2 = \mathbf{K}_1\mathbf{H}_2 + \mathbf{K}_2\mathbf{H}_1 = 0$$
$$\mathbf{Q}_3 = \mathbf{K}_1\mathbf{H}_3 + \mathbf{K}_2\left(\mathbf{H}_1 + \mathbf{H}_2\right) - \mathbf{K}_2\mathbf{H}_1 - \mathbf{K}_2\mathbf{H}_2 + \mathbf{K}_3\mathbf{H}_1 = 0$$

For a description of how this is actually derived, see [1]. Therefore

$$\begin{aligned} \mathbf{K}_1 &= \mathbf{H}_1^{-1} \\ \mathbf{K}_2 &= -\mathbf{K}_1\mathbf{H}_2\mathbf{K}_1 \\ \mathbf{K}_3 &= -[\mathbf{K}_1\mathbf{H}_3 + \mathbf{K}_2(\mathbf{H}_1 + \mathbf{H}_2) - \mathbf{K}_2\mathbf{H}_1 - \mathbf{K}_2\mathbf{H}_2]\,\mathbf{K}_1 \end{aligned} \tag{10.10}$$

The implementation of $\mathbf{K}_3$ is shown in Figure 10.1.

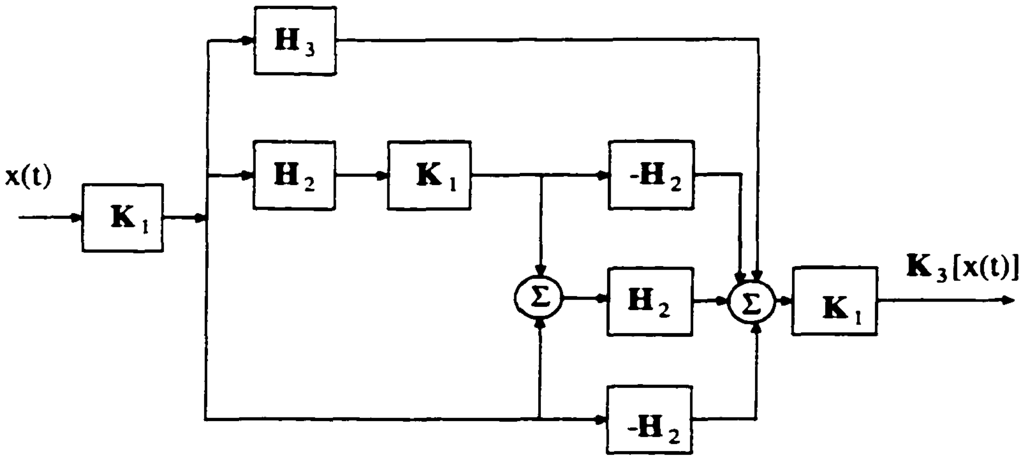


Figure 10.1 Implementation of the operator K_3 in terms of first- and second-order operators, using $\mathbf{K}_2 = -\mathbf{K}_1\mathbf{H}_2\mathbf{K}_1$.

One demanding problem is how to implement the kernels $\mathbf{H}_3$and $\mathbf{H}_2$. Generally, implementing a third-order kernel is a messy operation. It will consist of a possible infinite parallel connections of "basic" third-order building blocks (Figure 10.2). Obviously, the traditional implementation of even a third-order pre-distorter is quite complex. Higher order pre-distorters are even more complicated. For these reasons, only third-order pre-distorters or simplified fifth-order pre-distorters have been proposed [2] to [8]. However, the implementation of Volterra kernels is not unique. For example, by using a computer, it is possible to numerically perform the n'th-order convolutions using the kernel k_n. This might be an easier method, given the increasing availability of computing power, but we are not aware of anyone who has tried this.

The theory of Volterra series ensures that there exists a pre-distorter only if the first Volterra kernel is invertible. Recall that the first Volterra kernel is just the impulse response of a linear filter; theoretically, the inverse of this filter might not be stable. Implementing the inverse filter if it is unstable

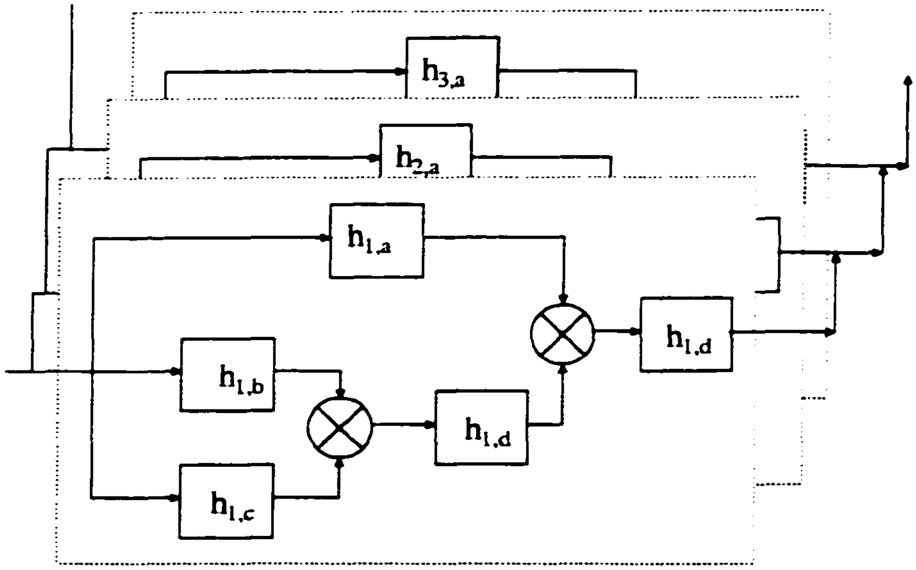


Figure 10.2 Implementation of a third-order Volterra kernel. using only linear filters and multipliers.

is not realistic since any noise will cause the inverse filter to become unstable. Therefore. the filter (first Volterra kernel) will be required to be minimum phase.

For a pre-distorter. the goal is to find a quasi inverse to the system described by Equation (10.11).

$$\mathbf{H} = \sum_{n=1}^{N} h_n(\tau_1, \ldots, \tau_n). \tag{10.11}$$

A true inverse would result in unity gain in the system, which is not desirable. To preserve the desired gain. we will instead find the true inverse of the modified system given by Equation (10.12).

$$\tilde{\mathbf{H}} = \frac{1}{g}\mathbf{H} = \frac{1}{g}\sum_{n=1}^{N} h_n(\tau_1, \ldots, \tau_n), \tag{10.12}$$

where g is the desired gain of our system. In a nonlinear system, the gain g can be chosen somewhat at will. but unreasonable choices will result in unreasonable outputs from the pre-distorter or even in a non-invertible system. It is obviously impossible to get an arbitrarily high gain out of an amplifier by pre-distorting the signal. For most practical amplifiers, the gain g will probably be chosen to be that of h_1 at the center frequency.

Pre-distorting of RF vs. baseband signals

With pre-distortion, a subtle difference exists between RF vs. baseband signals. Theoretically, a pre-distorter that compensates for all distortion, including that at harmonics of the carrier frequency, requires the creation of a pre-distorted signal with, in most cases, an extreme bandwidth. If one is restricted to pre-distort the baseband signal only, one can create an RF signal with no distortion at the carrier frequency (in-band distortion), but one cannot at the same time eliminate the harmonics at $2\omega_c, 3\omega_c$, etc. (out-of-band distortion).

At this point, it is useful to look at an example. Assume an amplifier is described by $y_1(x) = c_1 x + c_3 x^3$, $c_i \in \mathbb{C}$. The use of complex coefficients ($c_i \in \mathbb{C}$) allows the modeling of amplitude dependent phase shifts in the amplifier. If the input signal to the amplifier is $x(t) = a(t)\cos(\omega t + \theta(t))$, then the output is as follows:

$$\begin{aligned} y_1(a(t)\cos(\omega t + \theta(t))) &= c_1 a(t)\cos(\omega t + \theta(t)) + c_3 a^3(t)\cos^3(\omega t + \theta(t)) \\ &= \left(c_1 a(t) + c_3 \frac{3}{4} a^3(t)\right)\cos(\omega t + \theta(t)) + c_3 \frac{1}{3} a^3(t)\cos(3\omega t + 3\theta(t)). \end{aligned} \tag{10.13}$$

This equation reveals that the distortion of the component at the carrier frequency can be seen as the baseband signal passing through a nonlinearity looking like $y_2 = \left[c_1 x + c_3 \frac{3}{4} x^3\right]/c_1$.

Ideally, one would find the inverse y_1^{-1} of y_1 and use this inverse as the pre-distorter, which would result in an output signal with no in-band distortion and no harmonics. However, this would require the pre-distorter to work with the RF signal, resulting in very wideband signals, at least a number of times the carrier frequency. For example, in a 900 MHz cellular system, the pre-distorted signal could easily have a bandwidth of 5 GHz.

An alternative (and in our opinion [11], for many cases, the only practical way) is to compensate only for the distortion of the component at the carrier frequency. This choice also has the advantage of requiring pre-distortion only of the baseband signal, drastically reducing the required signal processing because it does not require an unreasonable sampling rate. Returning to the example, finding an inverse y_2^{-1} of $y_2 = \left(c_1 x + c_3 \frac{3}{4} x^3\right)/c_1$ and using this to pre-distort the baseband signal gives

$$y\left[y_2^{-1}(a(t))\cos(\omega t + \theta(t))\right] = c_1 a(t)\cos(\omega t + \theta(t)) + \frac{1}{3}\left[y_2^{-1}(a(t))\right]^3 \cos(3\omega t + 3\theta(t)). \tag{10.14}$$

This results in no distortion around the carrier frequency and a gain at the carrier frequency of c_1, but the downside is the presence of harmonics. Unless the amplifier is highly nonlinear, the out-of-band distortion will be of the same magnitude as without pre-distortion. To summarize, it is theoretically possible to pre-distort only the baseband signal and eliminate the distortion in the signal around the carrier frequency. This will not eliminate the out-of-band distortion, so reasonable linear amplifiers still need to be constructed to minimize out-of-band distortion.

Novel method for finding pre-distorters

Volterra series deals with an extremely wide class of amplifiers. By looking at a smaller class of amplifiers, the pre-distortion procedure should be simplified, which happens to be true. Many nonlinear systems can be modeled by cascading linear systems (filters) and nonlinear memoryless elements. The simplest amplifier (system) is shown in Figure 10.3 where f is a memoryless nonlinearity and h is a linear filter (i.e., linear with memory). This can be thought of as a power amplifier followed by a harmonic filter. The nonlinearity f is assumed to create amplitude dependent distortion, i.e., the distortion does not depend on the phase of $x(t)$, only its amplitude. The amplifier model shown in Figure 10.3 describes all of the amplifier models that were used in Refs.[2] to [7].

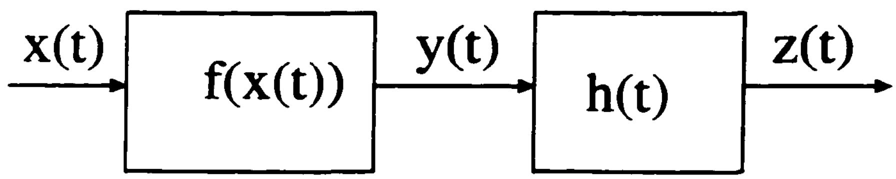


Figure 10.3 Amplifier modeled as a nonlinear memoryless operator followed by a filter (linear operator). Also basic building block for more advanced amplifier models.

It is possible to find the Volterra operator of this system as well as its inverse system and to build a pre-distorter based on this, but it is obvious that the pre-distorter of the system shown in Figure 10.3 is that shown in Figure 10.4, where it is assumed that the inverse functions exists. It is sufficient to assume that the function f is invertible for the range of $z(t)$ in Figure 10.4. This assumption is more realistic since no amplifier can produce an infinite output. This also relates to the previous discussion about choosing a suitable gain for the pre-distorted system Interestingly, the system shown in Figure 10.3 is also the inverse of the system shown in Figure 10.4.

If the nonlinearity causes only amplitude dependent distortion, which is generally assumed [2] to

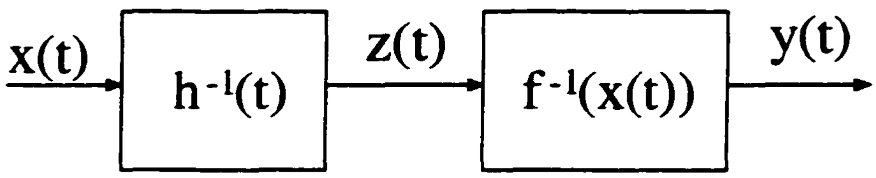


Figure 10.4 The inverse of the previous system.

[7], the condition that the filters must be minimum phase can be relaxed. Every non-minimum phase filter can be seen as a minimum phase filter followed by an allpass filter, i.e., a minimum phase filter exists with exactly the same *amplitude* response as the non-minimum phase filter. Using the inverse of this minimum phase filter produces a pre-distorter that will, in theory, give no amplitude distortion but does not affect phase distortion. As Figure 10.5 shows, this process of finding a pre-distorter is easily extended to deal with more complicated systems.

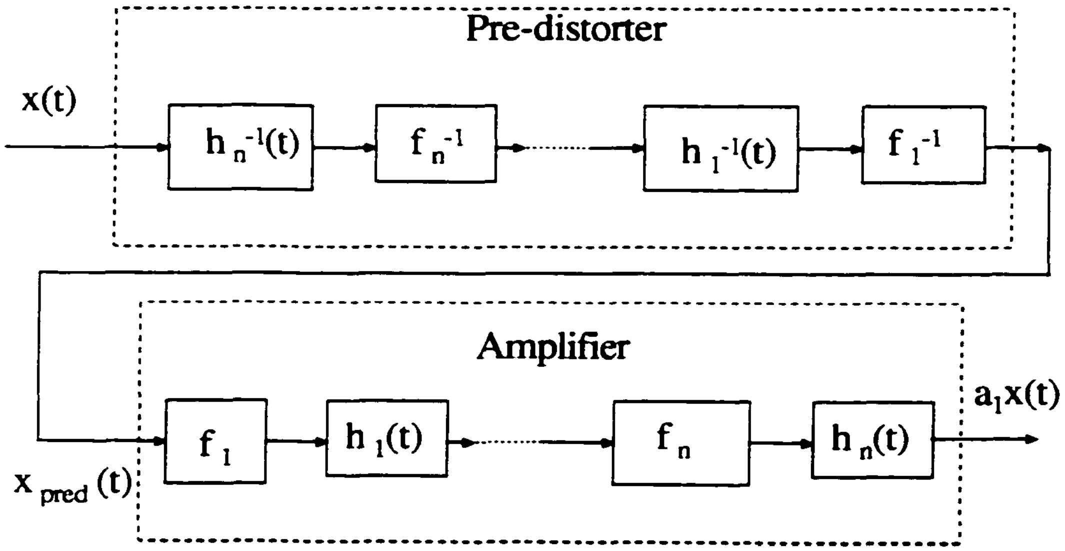


Figure 10.5 Generalized pre-distorter when cascading an arbitrary number of elements.

The process shown in Figure 10.4 can be described by

$$y(t) = f^{-1}\left[h_1^{-1}(t) * x(t)\right] = \sum_{n=1}^{\infty} a_n \left[h_1(t) * x(t)\right]^n \qquad h_1^{-1} \in L^2, f^{-1} \in L_{Loc}^{\infty} \tag{10.15}$$

assuming that $f^{-1}(x)$ is given by a series expansion $f^{-1}(x) = \sum a_n x^n$. The notation $f^{-1} \in L_{Loc}^{\infty}$ states that f^{-1} exists and is bounded for the input $h_1^{-1}(t) * x(t)$, and $h_1^{-1} \in L^2$ implies that h_1 is minimum phase. From this, we can derive the Volterra series expansion of the pre-distorter by noting that

$$a_n \left[h_1(t) * x(t)\right]^n = \int\int ... \int a_n h_1(\tau_1)h_1(\tau_2)....h_1(\tau_n)x(t-\tau_1)x(t-\tau_{n-1}).....x(t-\tau_n)d\tau_1....d\tau_{n-1}d\tau_n. \tag{10.16}$$

In this case, the n^{th} order Volterra operator is defined by

$$h_n(\tau_1, ..., \tau_n) = a_n h_1(\tau_1)h_1(\tau_2)....h_1(\tau_n) \tag{10.17}$$

which is the unique n'th-order Volterra operator describing Equation (10.15) since it is symmetric. In a similar way, it is possible to derive the Volterra operator for the pre-distorter shown in Figure 10.5, by a series connection of Volterra operators as described by Equations (10.7).

Clearly, there is no need to use the awkward traditional method of implementing the Volterra operator to get the pre-distorter. Instead, one will simply implement the inverse systems shown at the top of Figure 10.5. *The problem of pre-distorting the signal is thus reduced to that of estimating the filters and nonlinearities in the model of the amplifier.* Estimating the inverse filters is equivalent to estimating the filters used in the pre-distorter because one would only have to interchange the poles and zeroes to get the inverse filter. Finding the inverse function of the nonlinearities generally requires one more step, namely, to approximate the inverse, since there is usually no way of finding the inverse given the function.

It is theoretically possible to estimate the inverse of the nonlinearities, but this is not recommended because the functions might not be invertible for all inputs. To first estimate the nonlinearities makes it possible to investigate for what input they are invertible. From a practical point, trying to estimate an inverse function that does not exist will result in very poor estimates. The same argument is applicable to estimating the inverse filters: first finding the filters makes it possible to check if they are invertible or not.

One advantage with minimum phase filters is that they can be approximated by auto regressive (AR) filters [11], and using this, models of the non-linear amplifier can be found that are linear in their parameters.

For an amplifier modeled as a nonlinear memoryless element followed by an AR filter as in Figure 10.3, in the discrete case, the output is as follows:

$$\begin{aligned} y(t) &= h(t) * f(x(t)) = h(t) * \sum_{n=0}^{\infty} b_n x(t)^n \Leftrightarrow \\ \sum_{m=0}^{\infty} a_n y(t-mT) &= \sum_{n=0}^{\infty} b_n x^n(t) \Leftrightarrow \\ y(t) &= \sum_{n=0}^{\infty} b_n x^n(t) - \sum_{m=1}^{\infty} a_n y(t-mT) \approx \sum_{n=0}^{N} b_n x^n(t) - \sum_{m=1}^{M} a_n y(t-mT). \end{aligned} \tag{10.18}$$

where $h(t)$ is given by

$$H(z) = \frac{1}{1+\sum_{m=1}^{\infty} a_n z^{-m}} \Rightarrow H^{-1}(z) = 1 + \sum_{m=1}^{\infty} a_n z^{-m}. \tag{10.19}$$

Most importantly, this model is linear in its parameters, which can easily and efficiently be estimated using least square methods (LMS) or recursive least squares (RLS). So, estimating the nonlinear amplifier comes down to finding the vector $\begin{bmatrix} b_1 & \ldots & b_n & . & -a_1 & \ldots & -a_m \end{bmatrix}^T$. From this vector, the inverse filter is immediately found, while one is forced to do one more LMS estimation to find the inverse of the nonlinear memoryless device. Define

$$\mathbf{X} = \begin{bmatrix} x^0(t) & x^1(t) & \cdots & x^N(t) & y(t-T) & \cdots & y(t-MT) \\ x^0(t+T) & x^1(t+T) & \cdots & x^N(t+T) & y(t-T+T) & \cdots & y(t-MT+T) \\ x^0(t+2T) & x^1(t+2T) & \cdots & x^N(t+2T) & y(t-T+2T) & \cdots & y(t-MT+2T) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \end{bmatrix} \tag{10.20}$$

$$\mathbf{Y} = \begin{bmatrix} y(t) \\ y(t+T) \\ y(t+2T) \\ \vdots \end{bmatrix} \tag{10.21}$$

then the least squares solution to (let X^H be the Hermitian transpose of X)

$$\mathbf{Y} = \mathbf{X} \begin{bmatrix} b_1 \\ \vdots \\ b_n \\ -a_1 \\ \vdots \\ -a_M \end{bmatrix} \tag{10.22}$$

is

$$\begin{bmatrix} b_1 \\ \vdots \\ b_n \\ -a_1 \\ \vdots \\ -a_M \end{bmatrix} = \left(\mathbf{X}^H\mathbf{X}\right)^{-1}\mathbf{X}^H\mathbf{Y}. \tag{10.23}$$

It is also possible to use more advanced methods. e.g.. to perform a singular value decomposition (SVD) of $\mathbf{X}^H\mathbf{X}$. This method becomes necessary in numerically solving Equation (10.22) when $\mathbf{X}^H\mathbf{X}$ is close to singular. In practice. this will happen sooner or later as the order of the model increases. The easiest way to find the LMS estimation is probably to use LMS. which has the added benefit of avoiding the matrix inverses.

The next step is to find f^{-1}. To do this. $z(t)$. the signal between the nonlinear amplifier and the filter must be predicted. This is easily performed from the existing data by calculating

$$Z = \begin{bmatrix} z(t) \\ z(t+1) \\ \vdots \end{bmatrix} = \mathbf{X} \begin{bmatrix} b_1 \\ \vdots \\ b_N \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{10.24}$$

Define

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} f^{-1}(z_1) \\ \vdots \\ f^{-1}(z_n) \end{bmatrix} \tag{10.25}$$

and

$$Z = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}. \tag{10.26}$$

Form the matrix $\mathbf{Z}$ by calculating the powers of Z:

$$\mathbf{Z} = \begin{bmatrix} z_1^0 & z_1^1 & \dots & z_1^P \\ \vdots & \vdots & \ddots & \vdots \\ z_n^0 & z_n^1 & \dots & z_n^P \end{bmatrix}. \tag{10.27}$$

Now find the LMS solution to

$$X = \mathbf{Z}\begin{bmatrix} c_0 \\ \vdots \\ c_P \end{bmatrix}. \tag{10.28}$$

so

$$\begin{bmatrix} c_0 \\ \vdots \\ c_P \end{bmatrix} = (\mathbf{Z}^H\mathbf{Z})^{-1}\mathbf{Z}^H X \tag{10.29}$$

The inverse function is then approximated by $f^{-1}(x) = \sum_{j=0}^{P} c_j x^j$. The pre-distorter becomes

$$x_{pred}(t) = f^{-1}\left(\frac{h^{-1} * x(t)}{g}\right) \tag{10.30}$$

where g is the desired gain, as previously discussed.

The process can be simplified by noting that

$$f^{-1}(\frac{x}{g}) = \sum_{j=0}^{P} \frac{c_j}{g^j} x^j = \sum_{j=0}^{P} d_j x^j. \tag{10.31}$$

the coefficients d_j can be estimated directly by redefining the vector Z to be

$$Z' = \begin{bmatrix} f(x_1)/g \\ \vdots \\ f(x_n)/g \end{bmatrix} = \begin{bmatrix} z_1/g \\ \vdots \\ z_n/g \end{bmatrix} \Rightarrow \tag{10.32}$$

$$\begin{bmatrix} d_0 \\ \vdots \\ d_P \end{bmatrix} = (\mathbf{Z}'^H\mathbf{Z}')^{-1}\mathbf{Z}'^H X. \tag{10.33}$$

This procedure has the added advantage of offering less numerical sensitivity (less unbalanced matrixes) because $f(x)$ tends to be much greater than x, while $f(x)/g$ is almost always the same magnitude as x.

The pre-distorting process can now be written as the following two processes:

$$z(t) = [x(t-T), x(t-2T), \ldots x(t-MT)]\begin{bmatrix} -a_1 \\ \vdots \\ -a_M \end{bmatrix} \tag{10.34}$$

$$x_{pred}(t) = \left[z^0(t), z^1(t), \ldots z^P(t)\right]\begin{bmatrix} d_0 \\ \vdots \\ d_P \end{bmatrix}, \tag{10.35}$$

where $x_{pred}(t)$ is the pre-distorted signal that is sent to the actual amplifier. Note that the pre-distortion process only involves additions and multiplications, making for simple implementations.

This process is easily extended to systems of the form shown in Figure 10.6. In this case, it is only necessary to measure the signals $y_1 \cdots y_{n-1}$, (Figure 10.6), because the output from one building block is the input to the next. For the above case, it was assumed that the input and output signals were known but the signal between the nonlinearity and the filter was not. If this signal is known, the estimation process is even simpler, but this assumption is not always realistic.

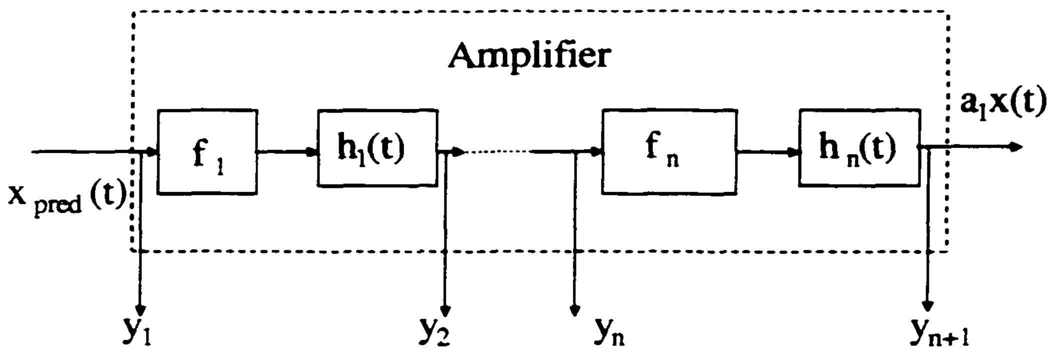


Figure 10.6 Measuring points on the nonlinear amplifier.

If only the input and output of the system shown in Figure 10.6 is known, it is still possible to estimate the filters and nonlinearities in the LMS sense but the model will not be linear in its parameters. This method forces one to use gradient descent based methods that suffer from the weakness of possible converging to a local minima or to use other methods like genetic algorithms.

Our method can also be applied to other types of amplifiers, e.g., the amplifier shown in Figure 10.7, Ref., [9]. The amplifier in Figure 10.7 works by separating the signal into an FM and AM part. The AM part is used to regulate the supply voltage to the power amplifier. The power supply will be bandlimited, which is modeled by the filter h_1.

The two causes of distortion are the bandlimiting effect of h_1 and the nonlinear behavior of the power amplifier (10.36). The former problem is accentuated by the fact that the envelope is generally much more bandwide than that of the input. In practice, the bandwidth of h_1 is generally at least ten times that of the input signal. The final filter h_2 mainly suppresses the harmonics and will not normally add any significant distortion at the carrier frequency. For this amplifier, the nonlinear elements are the envelope detector, the removal of AM (hard limiter), and the power amplifier (Amp in Figure 10.7).

The latter amplifier is described by

$$y_1 = x_1 f(x_3) \tag{10.36}$$

where $f(x_3)$ is a nonlinear but invertible function. Due to the removal of AM modulation (hard limiter), x_1 will have constant amplitude. It is easy to verify that the pre-distorter of this amplifier is represented by the signal processing block diagram in Figure 10.8. This system is a bit different from the previous systems since it has two branches, each made from cascading nonlinear elements and filters.

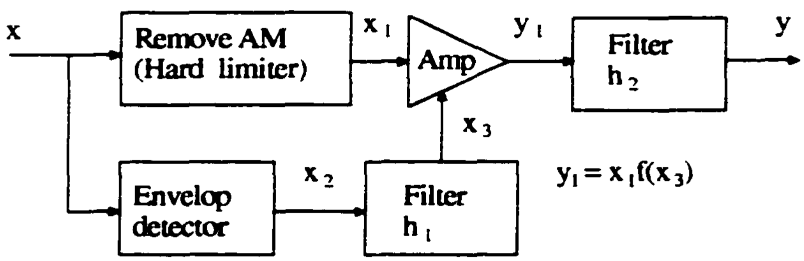


Figure 10.7 Amplifier based on envelope restoration.

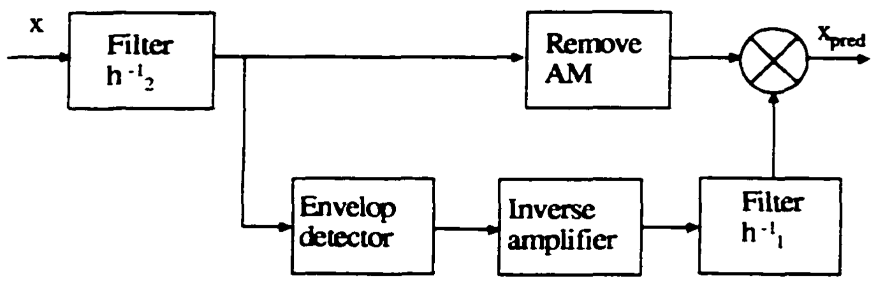


Figure 10.8 Pre-distorter to the previous system.

Simulations

We simulated ten different amplifiers [14] with five different input signals, to assess the robustness of the method. For all types of amplifiers that we simulated, we were always able to find a pre-distorter that worked. In the simulations, we focused on pre-distorting a baseband signal, compensating for the distortion at the carrier frequency. The results are, in fact, valid in the case of pre-distorting RF signals, although in these cases one would use a filter and amplifier with real-valued impulse responses.

In this section, we will present the results for three different amplifiers using two different signals. The first two amplifiers (#1 and #2 in Table 10.1) being described by Figure 10.3 each have a different nonlinear element. Amplifier #2, which models a traveling wave tube, has been used by several other authors [2]-[4]. This amplifier is complex valued, which models the AM to FM conversion that takes place in the amplifier. The last one (#3) is represented by Figure 10.7.

Table 10.1 The three amplifier elements.

Amplifier Model #	f(x)
1	$x + x^5/10$
2	$\frac{2x}{1-x^2}e^{\frac{i\pi x}{3(1+x^2)}}$
3	$5x - 0.2x^2 + 0.01x^3$

The first signal consisted of two independent bandlimited white Gaussian noise processes, one for the I-channel and one for the Q-channel. The envelope was normalized to have unity amplitude since amplifier #2 is only invertible in the range $[0-1]$. The second signal was a normalized two-tone single-sideband-signal. The fact that amplifier #2 is not invertible for all inputs does not cause any problems in our estimation process due to the way we make the estimations. It does make it impossible to come up with a pre-distorter unless the signal is in the range $[0-1]$; thus, the need to normalize.

For the first two amplifiers, the filter, $h(t)$, in Figure 10.3 was as follows:

$$H(z) = \frac{1}{1+(-1.7798+0.1176i)z^{-1}+(1.3885-0.1046i)z^{-2}+(-0.5340+0.0353i)z^{-3}+0.090z^{-4}}. \tag{10.37}$$

This filter has a complex impulse response and is used as the baseband equivalent of an RF filter. It adds significant distortion. The last amplifier (#3) used the following filters (Figure 10.7)

$$H_1(z) = \frac{0.3905}{1.0000-0.9428z^{-1}+0.3333z^{-2}} \tag{10.38}$$

$$H_2(z) = \frac{1}{1-0.1z^{-1}}. \tag{10.39}$$

The first filter $H_1(z)$ has a 3 dB point of $0.1 \cdot f_s$ where f_s is the sampling frequency.

The pre-distortion process for the first two amplifiers begins with the measurement of the I and Q components of the input and output, represented by the matrixes $\mathbf{X}$ and $\mathbf{Y}$ in Equation (10.20) and (10.21), respectively. From these measurements, the filter and amplifier inverse are estimated, using Equations (10.23) and (10.29). The next step is to pre-distort the signal $x(t)$ using the inverse filter and inverse amplifier as described by Equation (10.34) and (10.35). The pre-distorted signal $x_{pred}(t)$ is sent to the amplifier, producing an output $y(t)$. To measure how well this works, the figure of merit is the signal-to-distortion ratio (SDR) defined by

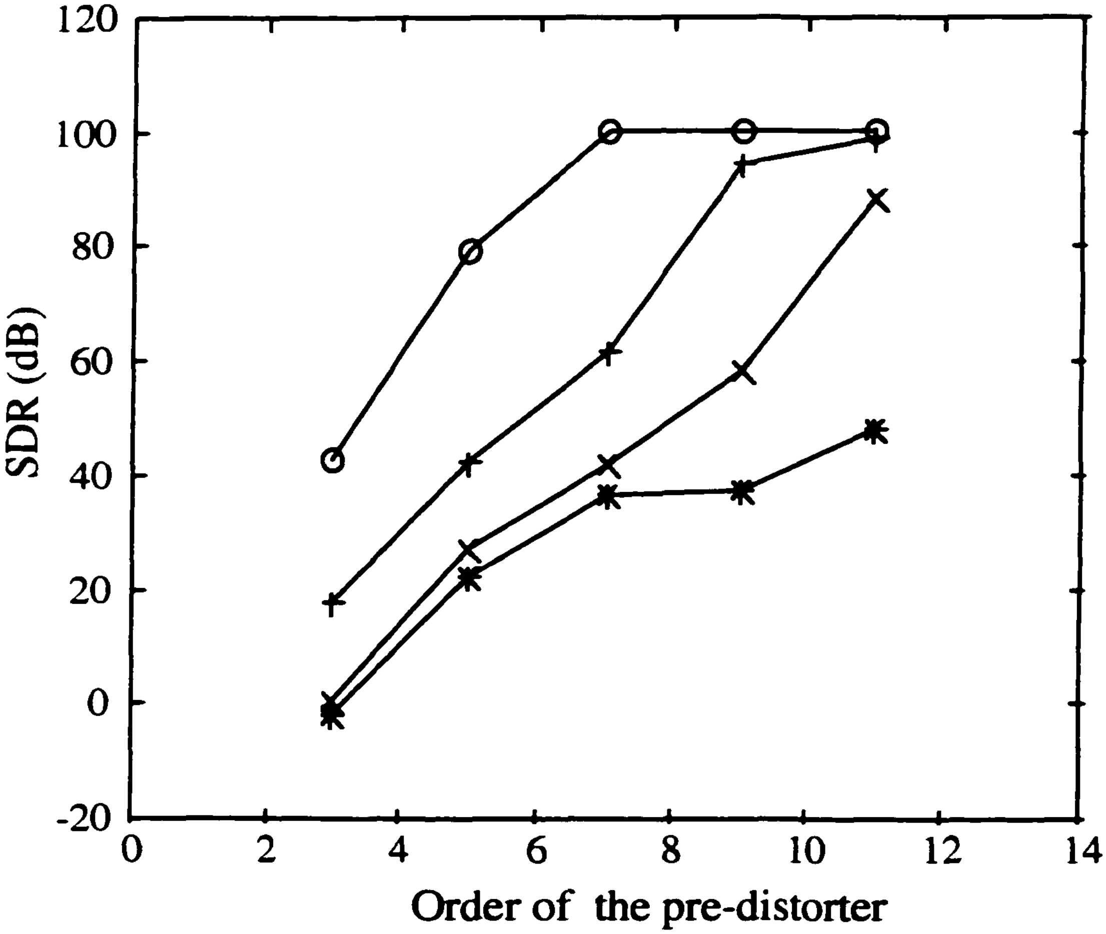
$$\begin{aligned} SDR &\triangleq -10\log_{10}(\frac{E(y-gx)^2}{E(gx)^2}) \\ E(x)^2 &\triangleq \int_0^{\infty} x^2(t)dt \end{aligned} \tag{10.40}$$

Note that $E(x)^2$ is simply the energy of $x(t)$; it is assumed that $x \in L^2$. Finally, we choose the parameter g, the desired gain. In the simulations, the choice $g = c_1$ was made, i.e., the approximate gain of the amplifier.

For the first two amplifiers, the simulations were done for pre-distorters of order 3, 5, 7, 9, and 11. These pre-distorters are equivalent to the same order Volterra based pre-distorter, which are significantly higher order than any previously published. For amplifier #1, a third-order Volterra based pre-distorter is not satisfactory while higher order pre-distorters are (Figure 10.9). Referring to Figure 10.9, an 11th order pre-distorter gives a SDR of between 50 dB or 90 dB, depending on the input signal. The fact that there are such large differences depending on the particular input signal should not be surprising since the systems are nonlinear. The difference is due to the distribution of the signal amplitude. As the pre-distorter order increases, the SDR increases regardless of the signal used. When the input signal is the random signal (denoted random signal in Figure 10.9), both pre-distorters perform slightly worse than when the signal is a two-tone SSB signal. When amplifier #2 is used, the performance of the pre-distorter also improves as the order increases. For an 11th-order pre-distorter, the SDR is 100 dB for both amplifier #2 input signals. At these SDR levels, the problem is no longer the pre-distorters but the computer's precision. For both amplifiers, the improvement in SDR when going from a third-order Volterra based pre-distorter to an 11th order is a minimum of 50 dB, showing that significantly improved pre-distorters can be achieved the order is increased beyond three.

Before presenting the result for the third amplifier, we will describe how it behaves. If the input signal is narrow band, the first filter will add very little distortion but the nonlinear device will. On the other hand, if the input is wideband, the distortion is dominated by the bandlimiting effect of h_1. The input signal is again a two-tone single-sideband signal, but to show the effects of the different types of distortion, the frequency spacing of the two tones was varied. The spacing varied from $0.01f_s$ to $0.15f_s$, implying that the bandwidth of the envelope was approximately $0.1f_s$ to $0.3f_s$. The signals x, x_3, y_1, and y in Figure 10.7 were measured and used to estimate the filters and the amplifier.

When nothing is done to the signal (Figure 10.10), the SDR starts at 30 dB and then decays as the frequency offset increases, mainly due to the effects of h_1, and to a lesser degree, aliasing. For a relative



x Amplifier #1, two-tone SSB

* Amplifier #1, random signal

o Amplifier #2, two-tone SSB

+ Amplifier #2, random signal

Figure 10.9 SDR vs. the order of the inverse.

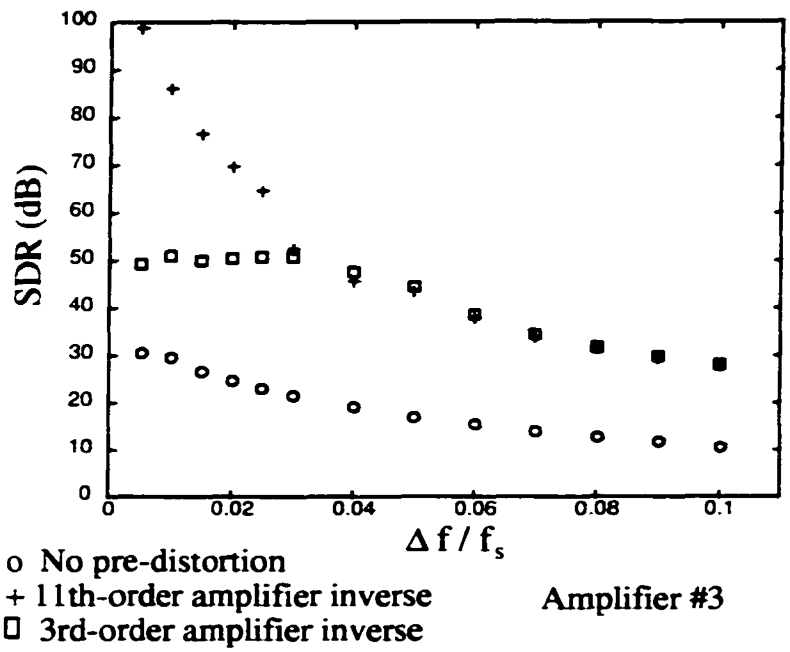


Figure 10.10 For amplifier #3, the performance of the pre-distorter as a function of the frequency offset and order of the inverse.

frequency spacing of less than 0.04, the 11th-order pre-distorter is superior to the third order, giving an SDR of 100 dB when the relative frequency spacing is 0.005. In practice, the original amplifier would not be designed for a frequency spacing (bandwidth) of more than 0.01.

When the spacing is more that 0.04, the different pre-distorters are equivalent. While this may seem strange, it has to do with frequency folding when computing the envelope. Essentially, the simulations show that frequency folding is the dominate problem in this case, and the only solution is to increase the sampling frequency. The order of the inverse in Figure 10.10 refers to the order of the inverse of the $f(x)$, i.e., n'th order implies that $f^{-1}(x) = \sum_{i=0}^{n} a_i x^i$, which implies that the corresponding Volterra operators will be of at least this order although there is no need to derive them. This shows once more that the method is powerful. When the high-order pre-distorter is used, the SDR improvement starts at 70 dB and then decreases to 20 dB due to frequency folding. The low-order pre-distorter gives a constant SDR of 50 dB until frequency folding starts to dominate; this corresponds to an SDR improvement of 20 to 30 dB. Even for the maximum frequency offset, the pre-distorted signal has an SDR of at least 30 dB while the non pre-distorted signal has a SDR of 10 dB, an improvement of 20 dB. This can be interpreted as a tenfold increase in the usable bandwidth.

Conclusion

The proposed method has the definite advantage of allowing for arbitrarily high-order pre-distorters for many, if not most, practical amplifiers. Because of their complexity, this is not possible with the traditional method of finding and implementing the Volterra operators. The new method also reduces the process of finding the pre-distorters to that of LMS estimations of parameters, a well-known method that is easily implemented, making it easy to adaptively measure the amplifier and update the pre-distorter. In contrast, the traditional process of measuring the Volterra kernels [1] rapidly becomes very awkward as the order increases.

The simulations show that the method works extremely well, easily giving SDR improvements of 50 to 80 dB. By increasing the order of the inverse, the only limitations on SDR improvement are the numerical precision of the computer and frequency folding. The estimation process is computationally effective due to the use of LMS. Performing the actual pre-distortion is even less computer intensive and should be easily implemented even with a very modest DSP. If the amplifier does not change over time, the parameter estimation can be made once and for all, further cutting down on the computer requirements.

Bibliography

[1] Martin Schetzen, *The Volterra and Wiener Theories Of Nonlinear Systems*, John Wiley & Sons, New York: 1980.

[2] Maria-Gabriella Di Benedetto and Paolo Mandarini, "A new analog predistortion criterion with application to high efficiency digital radio links," *IEEE Transactions on Communications*, vol. 43, no.12, pp. 2966-2974, December 1995.

[3] Giovanni Lazzarin, Silvano Pupolin, and Augusto Sarti, "Nonlinearity compensation in digital radio systems," *IEEE Transactions on Communications*, vol. 42, no.2/3/4, pp. 988-999, February/March/April 1994.

[4] Aldo N. D'Andrea, Vincenzo Lottici, and Ruggero Reggiannini, "RF power amplifier linearization through amplitude and phase predistortion," *IEEE Transactions on Communications*, vol. 44, no. 11, pp. 1477-1484, November 1996.

[5] Kurt Metzger and Rolf Valentin, "Intersymbol interference due to linear and nonlinear distortion," *IEEE Transactions on Communications*, vol. 44, no. 7, pp. 809-816, July 1996.

[6] Georges Karam and Hikmet Sari, "A data predistortion technique with memory for QAM radio systems," *IEEE Transactions on Communications*, vol. 39, no. 2, pp. 336-344, February 1991.

[7] Georges Karam and Hikmet Sari, "Analysis of predistortion, equalization, and ISI cancellation techniques in digital radio systems with nonlinear transmit amplifiers," *IEEE Transactions on Communications*, vol. 37, no. 12, pp. 1245-1253, December 1989.

[8] Lars Sundström, Michael Faulkner, and Mats Johansson, "Quantization analysis and design of a digital predistortion lineraizer for RF power amplifiers," *IEEE Transactions on Vehicular technology*, vol. 45, no. 4, pp. 707-719, November 1996.

[9] F.H. Raab and D. J. Rupp, "High efficiency single sideband HF/VHF transmitter based upon envelope elimination and restoration," *IEE, HF Radio Systems and Techniques*, 4-7 July 1994, Conference Publication no. 392, pp. 21-25, 1994.

[10] M. Ghaderi, S. Kumar, and D. E. Dodds, "Adaptive predistortion lineariser using polynomial functions," *IEE Proceedings on Communications*, vol. 141, no. 2, pp. 49-55, April 1994.

[11] Steve F. Russell and Anders Mattsson, "Signal distortion for power amplifiers," Project report contract no. CFW0495-12, Rockwell International, Cedar Rapids, Iowa State University, Ames Ia, 1996.

11 FUTURE WORK

Separation of three FM signals

It has been argued that in the case of three or more FM signals a unique solution most likely does exist, and a method to find the signals that is guaranteed to work has been derived. The first thing would be to rigorously prove that the solution is unique. It has not been possible to find a practical method that will always find this solution. The obvious future problem is developing a practical method that will always work. One possible approach might be to start with the theoretically guaranteed method and create a clever search algorithm that will do the job with a reasonable amount of calculations. In the separation algorithm used for the simulations, it was assumed that the stronger signal dominates, i.e., $A_1 > A_2 + A_3$, but the separation algorithm presented on page 8 does not depend on this fact. Because it is always possible to separate three FM signals irrespectively of their amplitudes, an obvious problem is then to improve the simplified algorithm so that it will always separate the three signals and will work for any arbitrary choice of amplitudes. Another problem that needs analyzing is how error propagates in the iterative separation process. Because the system is nonlinear, there is a possibility that an error will cause a bigger error in the next step, which causes an even bigger error, etc. This is exactly what happens when the separation process fails. The error gets so large that the next signal does not dominate in the residue. When the received amplitude is close to its maximum or minimum values, the phases are well known, suggesting that one should start the separation process in these regions and work away from there. A related problem comes from the observation that phase slips occur only when the residual signal $|S_R - S_{1est} - S_{2est} - S_{3est}|$ is greater than A_3. It seems possible that one should use this residue to find phase slips and eliminate them.

Another topic would be to compare how well extended Kalman filtering would work when trying to separate three FM signals. One published article [29] uses the extended Kalman filter for separating two signals. It does not seem to work too well, but it would still be interesting to run some simulations and see how well it would separate three signals.

In the beginning of this dissertation, the idea of noise tracking in an FM receiver was introduced.

The idea is based on the fact that there is a capture effect in the FM receiver, making it possible to estimate the noise. It is an interesting, counterintuitive observation that the stronger the signal is compared to the noise, the better the estimate of the noise will be. Is it possible to partially track the noise in a way similar to the how it is done in cross-coupled phase-locked loops when separating two FM signals? Theoretically analyzing the aspects of noise tracking/suppression in an FM receiver seems like a very challenging as well as intriguing problem, which I encourage someone to tackle. I have some preliminary results, but due to the lack of time, I have not been able to pursue this idea. This problem can, and probably should, be treated analytically by using stochastically differential equations. References [27] and [28] are standard texts within this field.

Pre-distortion

Using Volterra series is a theoretically very impressive method for pre-distortion but it does not work in practice except for low-order pre-distorters. It has been shown in this dissertation that for many amplifiers the pre-distorter can be found in a form much simpler than the traditional Volterra series representation. A natural question is "What are the limits on this class of amplifiers? I suspect that nonlinear systems with feedback might not fall into this category. A related question is "Is there any way of designing a pre-distorter that is simpler than the Volterra series approach?" This might not be as far-fetched as one might think, provided that the system can be approximated by a system without feedback. This would be similar to approximating an Infinite Impulse Response (IIR) filter with an Finite Impulse Response (FIR) filter. On the other hand, some systems probably cannot be described in any other way.

A different problem associated with the use of Volterra series is the complicated measurements needed to find the Volterra kernels. Is there any other way to find the Volterra kernels than by using Wiener's method of measuring them? Finally, with the constantly increasing availability and speed of computers, how practical would it be to implement a true fifth- or seventh-order Volterra pre-distorter?

12 CONCLUSIONS

Separation of three or more FM signals

The study on the separation of three mutually interfering FM signals culminated in five main results. It was shown that for a received signal, which is the sum of any three FM signals with bandlimited phases, there are strong reasons to believe that only one set of FM signals can form this received signal. In other words, there is most likely a unique solution. Furthermore, it was shown that this solution is stable in the sense that a small introduced error will be suppressed, forcing the estimates to approach the true solution. This was accomplished by showing that the amplitudes of the three signals can always be found (the third result). The fourth result is a series expansion of the received phase in the case of three FM signals, which describes a capture effect. The fifth result demonstrates that there is a way to always separate any three, or more, FM signals in finite time. However, in some situations, this time might be too long to be practical.

The third result, that the amplitudes of the three FM signals can be found, is given under very general and reasonable assumptions. For example, if the phases are evenly distributed over the interval $[0..2\pi)$, the amplitudes can be found. The fifth result is given in the more restricted case: when the amplitude of the strongest signal is greater than the sum of the two other signals, a series expansion of the received phase is derived showing that the received phase is composed of the dominating signal and error terms. These error terms are, in effect, FM signals, showing that there is a capture effect also for three FM signals, assuming the amplitude requirement is met. This series expansion was used to show that the solution is stable. This result applies to all cases; i.e., there is no restriction on the amplitudes of the three FM signals.

Starting from simple geometrical considerations, an expression showing the instantaneously needed error suppression was derived. This unfortunately also implies that some level of suppression may not be good enough, suggesting that iteratively filtering the signals, actually filtering the received phase, and subtracting the estimate from the received signal is not guaranteed to result in a successful separation of the three signals. The simulations show that this is indeed the case. When the modulation index was small, there were many cases where it was impossible to separate the signals using this method. The series expansion of the received phase does, however, imply that separation through this method will work, provided the modulation index is high enough. A somewhat more sophisticated method is introduced that uses the received amplitude to find limits on the phases of the three signals. This method shows that in some cases one can ensure that the estimate of the stronger signal will always be within limits, which will guarantee that the second signal will dominate once the estimate of the stronger signal is subtracted. This happens when the third signal is weak compared to the second signal. Unfortunately, this does not ensure that one will obtain good enough estimates of the two strongest signals such that the third signal will dominate once the estimates of the two strongest signals are subtracted from the received signal. The simulations support these findings. It is easier to separate the three signals when the third signal is weak compared to the second signal.

A foolproof way that will always separate the three signals was presented on page 60 in the Chapter 10. Unfortunately, this algorithm is too computationally intense to be of practical use. Instead, a practical algorithm was described in the Chapter 9. Using this algorithm, the simulations show that the weakest signal can be successfully demodulated in 50 to 70% of all cases, and when it is successfully demodulated, the distortion on the demodulated signal (FM) is reasonable low, SDR_{fm3} is above 15 dB, in most cases above 20 dB. It was also shown that it is possible to determine if the separation (demodulation) process was successful by simply looking at the residue $|S_R - S_{1est} - S_{2est} - S_{3est}|$.

Pre-distortion

When it comes to pre-distortion, the main result is that one can find simple pre-distorters for any amplifier that can be modeled as a cascade of nonlinear memoryless elements and linear elements with memory, i.e., non-linear elements and filters. It was shown that a basic building block consisting of one nonlinear element and one filter can be efficiently estimated, using regular LMS since the model is linear in its parameters. It was also shown that these concepts can be extended to another type of amplifier. The implementation of these proposed pre-distorters is extremely simple, which is one of their strengths. The other strength is that one can implement arbitrarily high-order pre-distorters, something that is impractical with the traditional approach. They can also be seen as an implementation of the Volterra series based pre-distorter, note that the implementation of a Volterra series is not unique. The simulations validate these claims, as they show the pre-distorter to essentially linearize the amplifier.

BIBLIOGRAPHY

[1] Zorka D. Stojanovic, Miroslav I. Dukic, and Ilija S. Stojanovich, "A new demodulation method improving FM system interference immunity," IEEE transactions on communications, Vol. Com-29, no. 7, pp. 1001-1011, July 1981.

[2] H.W. Hawkes, "Study of adjacent and co-channel FM interference reduction," IEE Proceedings-I, Vol. 138, no. 4, pp. 319-326, August 1991.

[3] Krijn Leentvaar and Jan H. Flint, "The capture effect in FM receivers," IEEE transactions on communications, Vol. Com-24, no. 5, pp. 531-539, May 1976.

[4] Zorka D. Stojanovic, Miroslav I. Dukic, and Ilija S. Stojanovich, "A new robust FM demodulator reducing adjacent radio-channel interference noise," IEEE transactions on communications, Vol. Com-32, no. 11, pp. 1001-1011, November 1984.

[5] H. W. Hakes, "Study of adjacent channel AM interference suppression," IEE Proceedings, Vol. 136, PT. I, no. 4, pp. 241-254, August 1989.

[6] Hongya Ge and Donald W. Tufts, "Estimating the frequencies of two sinusoids using only the phase angles of complex-valued data," IEEE transactions on signal processing, Vol. 41, no. 5, pp. 1982-1986, May 1993.

[7] C. Craig Ferris, "Spectral characteristics of FDM-FM signals," IEEE Transactions on communication technology, Vol. Com-16, no. 2, pp. 233-238, April 1968.

[8] Benjamin A. Pontano, Jorge C. Fuenzalida, and Nand Kishmore M. Chitre, "Interference in angle-modulated systems carrying multichannel telephony signals," IEEE transactions on communications," Vol. Com-21, no. 6, pp. 714-727, June 1973.

[9] B. C. Sarkar "Phase error dynamics of a first order phase locked loop in the presence of cochannel tone interference and additive noise". IEEE transactions on communications. Vol. 38. no. 7. pp. 926-968. July 1990.

[10] Frank A. Cassara, Harry Schachter, and Gerald H. Simowitz. "Acquisition behavior of the cross-coupled phase locked loop FM demodulator," IEEE transaction on communications. Vol. Com-28. no. 6. pp. 897-904. June 1980.

[11] Jonathan N. Bradely and R. Lynn Kirlin. "Phase locked loop cancellation of interfering tones." IEEE transaction on signal processing. Vol. 41. no. 1. pp. 391-395. January 1993.

[12] Tippure S. Sundresh, Frank A. Cassara, and Harry Schachter. "Maximum a posteriori estimator for suppression of interchannel interference in FM receivers," IEEE transactions on communications. Vol. Com-25. no. 12, pp. 1480-1485. December 1978.

[13] Y Bar-Ness, F. A. Cassara, H. Schachter, and R. Difazio. "Cross-coupled phase locked loop with closed loop amplitude control." IEEE transactions on communications. Vol. Com-32. no. 2. pp. 195-199. February 1984

[14] Brad A. Hedstrom and R. Lynn Kirlin. "Co-Channel Signal separation Using Coupled Digital Phse-Locked Loops," IEEE transactions on communications. Vol. 44. no. 10. pp. 1373-1384. October 1996

[15] H. K. Khalil. Nonlinear Systems. New York. Macmillian Publishing Company. 1992

[16] L.W. Couch. Digital and analog communication systems. Ed. 4. New York. Macmillian Publishing Company. 1993

[17] F. Gardner. Phaselock techniques. Ed. 2. New York. John Wiley & Sons, 1979

[18] H. Blanchard. Phased locked loops. New York. John Wiley & Sons, 1976.

[19] I.M. Rysik. I.S. Gradstein. Summen Product und Integral Tafeln. (in German) (East) Berlin, VEB Deutsche verlaug der wissenschaft. 1963.

[20] D.L. Abbey, "Composite signal separator and FM demodulator," Internal Rockwell document. Rockwell International, Cedar Rapids Iowa 1993.

[21] D.L Abbey, "Composite signal separator and PSK/AM/FM demodulator," Internal Rockwell document. Rockwell International, Cedar Rapids Iowa 1993.

[22] A. Mattsson, Steve F. Russell, "Separation of mutually interfering FM/AM signals," accepted by IEEE transactions on communications.

[23] A. Mattsson, "Separation of mutually interfering FM/AM signals," MS Thesis, Iowa State University, Ames, Iowa, 1995.

[24] Bengt Schmidtbauer,Reglerteori (in Swedish), Chalmers University of Technology, Gothenburg, 1987

[25] Frank L. Lewis and Vassilis L. Symoros, Optimal Control, Ed. 2, New York, John, Wiley & Sons, 1995

[26] H. Vincent Poor, An itroduction to signal detection and estimation, Ed. 2, New York, Springer-Verlag, 1994

[27] Ioannis Karatzas and Steven E. Shreve, Brownian motion and stochastic calculus, Ed. 2, New York, Springer-Verlag, 1991

[28] Bernt Øksendal, Stochastic differential equations, Ed. 4, Berlin-Hiedelberg, 1991

[29] Andrew P. Sage and James L. Melsa, Estimation theory with applications to communications and control, New York, Robert E. Krieger publishing company, 1979

[30] Lennart Råde and Bertil Westergren, BETA-Mathematics Handbook, Lund, Studentlitteratur, 1988.

[31] Anders Mattsson and Steve F. Russell, "High-order Volterra series based pre-distortion for high efficiency RF power amplifiers," submitted to IEEE transactions on communication.

[32] Walter Rudin, Real and Complex Analysis, 3rd Ed., McGraw-Hill series in Higher Mathematics, New York, 1987.