I, Kevin J Berndsen, hereby submit this original work as part of the requirements for the degree of Master of Science in Electrical Engineering.

It is entitled:
Signal Optimization for Efficient High-Power Amplifier Operation

Student's name: Kevin J Berndsen

This work and its defense approved by:

Committee chair: H. Howard Fan, PhD
Committee member: Emmanuel Fernandez, PhD
Committee member: Ali Minai, PhD
Signal Optimization for Efficient High-Power Amplifier Operation

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF CINCINNATI
BY

Kevin J. Berndsen
B.S. Electrical and Computer Engineering

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Howard Fan, PhD
Emmanuel Fernandez, PhD
Ali Minai, PhD

May 2011
Abstract

A passband RF or microwave signal is amplified for transmission by a high-power amplifier system. The quality of the transmitted signal degrades as the limits of the amplification range are reached. Signal pre-distortion can be optimized for operation with maximum power efficiency while achieving minimum performance requirements. A basic model and formulations are developed, leading to algorithmic solutions by advanced optimization techniques. Each formulation is verified for multiple amplifier systems and performance constraints.
© Kevin J. Berndsen 2011
Acknowledgements

I thank all of my colleagues at the Harris Corporation which supported this research. I would like to acknowledge contributions and support from Carlos Abascal, Dmitri Borodulin, George Cabrera, Ty Chizovsky, Dave Danielsons, David Geier, Scott Halozan, Ted Korte, Jeff Malec, Anders Mattsson, Geoff Mendenhall, Paul Mizwicki, Joe Seccia, and Brandon Sorn. Thank you Carlos for being a mentor and for believing in me. Thank you Ty for teaching me almost everything I needed to know, for all your excellent design work, and for knowing how to test all of it. George and Dmitri, thanks for your patience and for always challenging me with new problems to solve. Also thank you to John Meltzer for the discussions on MER.

I also wish to thank Dr. Fernandez for encouraging me to pursue this topic. Also a great thanks to my advisor, Dr. Howard Fan, for your expertise and for the many long discussions.

I am grateful for my family, who were all extremely supportive in their own ways. And most importantly, my wife and children, for your love, prayers, understanding, and encouragement – I love you.
# Table of Contents

1 INTRODUCTION .......................................................................................................................... 1

1.1 SUMMARY .................................................................................................................................. 2
1.2 ORGANIZATION OF PAPER .................................................................................................. 3
1.3 INTELLIGENT SYSTEMS AND OPTIMIZATION ..................................................................... 4

2 SYSTEM DESCRIPTION .............................................................................................................. 6

2.1 POWER AMPLIFIER .............................................................................................................. 6
2.2 SYSTEM BLOCK DIAGRAM .................................................................................................... 7
2.3 IN CONTEXT OF PREVIOUS WORK ...................................................................................... 9
2.4 DEFINITIONS ......................................................................................................................... 11

3 DISTORTION AND SIGNAL QUALITY ..................................................................................... 14

3.1 NONLINEAR DISTORTION .................................................................................................... 14
3.2 SIGNAL QUALITY .................................................................................................................. 14
3.3 MEASUREMENT AND ESTIMATION ................................................................................... 17

4 MODELS ....................................................................................................................................... 20

4.1 DISCRETE-TIME EQUIVALENT SYSTEM ............................................................................ 20
4.2 EFFICIENCY AND BACK-OFF .............................................................................................. 21
4.3 PARAMETRIC PA MODELS .................................................................................................... 23
4.4 PREDISTORTION MODEL ....................................................................................................... 27
4.5 SIMULATION MODEL .............................................................................................................. 29

5 PROBLEM FORMULATION ......................................................................................................... 31

5.1 PROBLEM STATEMENT .......................................................................................................... 32
5.2 GRAPHICAL DESCRIPTION .................................................................................................... 35
5.3 INVERSE FUNCTIONS ............................................................................................................ 38
5.4 LINEARIZATION ..................................................................................................................... 38
5.5 ANALYTIC SOLUTIONS .......................................................................................................... 40
5.6 TRAINING DESIGN .................................................................................................................. 48
5.7 PROBABILITY DENSITY TRANSFORMATION ...................................................................... 48
5.8 INDIRECT LEARNING ARCHITECTURE .................................................................................. 50
5.9 SOLUTION BASIS ................................................................................................................... 52
5.10 MULTI-PHASE VERIFICATION ............................................................................................. 54
5.11 OPTIMIZATION ..................................................................................................................... 56
5.12 OPTIMIZATION FORMULATION ........................................................................................... 61

6 SOLUTION METHODS ............................................................................................................... 67

6.1 ANALYTIC SOLUTION FOR PEAK LIMITED PA ................................................................... 67
6.2 LEAST SQUARES SEARCH ALGORITHM (LSS) .................................................................... 68
6.3 OPTIMIZATION SOLVER ....................................................................................................... 72
6.4 CONSTRAINED SOLVER ........................................................................................................ 73
6.5 PROJECTION AND GRADIENT DESCENT .......................................................................... 74

7 RESULTS ..................................................................................................................................... 88

7.1 PA MODEL VALIDATION ......................................................................................................... 88
7.2 PEAK REDUCTION .................................................................................................................... 90
7.3 LINEARIZATION ..................................................................................................................... 91
7.4 FULL SIMULATION..................................................................................................................93
7.5 SUMMARY OF RESULTS ......................................................................................................97
8 CONCLUSIONS .......................................................................................................................104
  8.1 DISCUSSION OF RESULTS ...............................................................................................104
  8.2 FUTURE RESEARCH .........................................................................................................106
9 REFERENCES ..........................................................................................................................107
10 APPENDIX A ..........................................................................................................................111
List of Figures

Figure 1: Signal Path Diagram ................................................................. 7
Figure 2: System Diagram with Feedback .................................................. 8
Figure 3: Performance Example ............................................................... 9
Figure 4: Variance of sample variance vs. sample size (n) ......................... 19
Figure 5: Spectrum of WF11 BW=0.25 [Matlab\NonLinear\WF11.mat] ...... 21
Figure 6: WF11 CCDF estimate. Pr[W > y] [Matlab\NonLinear\WF11.m] ... 21
Figure 7: Comparison of PA models [MATLAB/Nonlinear/Haggman/show_g7g8g9.m] ................................................................. 26
Figure 8: General AM/AM, PA (red), PD (blue) [MATLAB\Nonlinear\NLSystemSolve.m] ................................................................. 37
Figure 9: Solution of integral (20) and ideal clipping of WF11 [MATLAB\integral_def1.m] #114 ................................................................. 44
Figure 10: Estimation of \( \theta \) params for g8, IBO=8dB [Matlab\NonLinear\test_NL_PeakSlope_2.m] ................................................................. 46
Figure 11: Estimation of \( \theta \) params for g7, IBO=8dB [Matlab\NonLinear\test_NL_PeakSlope_2.m] ................................................................. 46
Figure 12: Estimate of \( \theta \) params for g11, IBO=9.0 [Matlab\NonLinear\test_NL_PeakSlope_2.m] ................................................................. 47
Figure 13: Indirect Learning Architecture .................................................. 50
Figure 14: Example 3-1 1/NMSE surface [Matlab\NonLinear\topo_1.m] #109 63
Figure 15: Example 3-2 n5 1/NMSE contours PD=\( u(1 + a_3u_2) \) [Matlab\NonLinear\topo_2.m] ................................................................. 65
Figure 16: Example 3-2 n5 1/NMSE contours PD=\( u(1 + a_5u_4) \) [Matlab\NonLinear\topo_2.m] ................................................................. 65
Figure 17: Example 5 performance surface 1/NMSE [dB] over two dimensions [Matlab\NonLinear\param_surf_a3_IBO_01.m] #108 ......................... 75
Figure 18: Example 6 performance surface 1/NMSE [dB] over two coeff dimensions at a fixed IBO=11dB [Matlab\NonLinear\param_surf_a3_a5_01.m] #103 ......................... 76
Figure 19: Example 6 performance surface 1/NMSE [dB] at IBO=10.2dB [Matlab\NonLinear\param_surf_a3_a5_01.m] #104 ......................... 76
Figure 20: Example 6 expansion, IBO=11.0dB [Matlab\NonLinear\param_surf_a3_a5_expansion_01.m] #107 ......................... 76
Figure 21: Example 6 expansion IBO=10.2dB [Matlab\NonLinear\param_surf_a3_a5_expansion_01.m] #106 ......................... 76
Figure 22: Expansion IBO=10.0dB [Matlab\NonLinear\param_surf_a3_a5_expansion_01.m] #107 ......................... 77
Figure 23: SNDR surface for Proof 1A/1B. Blue line with circle markers is LS .... 79
Figure 24: [Matlab\Nonlinear\POCS_4.m] ................................................ 85
Figure 25: \( f(m|a(i)) \) vs m. \( a_0= \)Blue, circles, \( a_25= \)Red, asterisk [Matlab\Nonlinear\POCS_4.m] #113 ......................... 85
Figure 26: g8 validation, PR y dB = 7.5, 8.0, 8.5, 9.0 [Matlab\NonLinear\Haggman\test_g8_PARS] ................................................ 89
Figure 27: g11 validation, PR $\gamma dB = 7.5, 8.0, 8.5, 9.0$

[Matlab\NonLinear\Haggman\test_g6_PARS.m] ..................................................... 89

Figure 28: Magnitude limiting function, $T=7.7$dB, slight deviations introduced by interpolation filtering. ................................................................. 91

Figure 29: G9 linearization [MATLAB/NonLinear\Haggman\run_g5.m] .......................................................... 92

Figure 30: IBO / OBO over search path ................................................................. 93

Figure 31: Comparison of Peak Estimate and known Backoff, g7

[Matlab\fullsim_IBO_02.m Figure 601] ................................................................... 94

Figure 32: PD-1 and PD-2 $Bu=10$, G8 [MATLAB\NonLinear\run_g7_IBO_03.m] .... 95

Figure 33: PD-1 and PD-2 $Bu=10$, G7 [MATLAB\NonLinear\run_g7_IBO_03.m] .... 96

Figure 34: g7 final CCDF ...................................................................................... 96

Figure 35: Results vs. theoretical bound, eqn. (20), of three PD algorithms for each PA model ......................................................................................... 97

Figure 36: PD-2 g7 linearization. Red: PA AM-AM ($y$ vs. $x$); Blue: PD, Black: PA,PD output ($z$ vs. $x$) ........................................................................ 98

Figure 37: Compare PD-1, PD-2 w/ Restrictive upper bound $Bu=0.79$ (9dB) .......... 99

Figure 38: Compare PD-1, PD-2 w/ Reasonable upper bound $Bu=1.2589$ (13dB) .... 99

Figure 39: g11 w/ $\Omega = 35dB, 30dB$ .................................................................. 100

Figure 40: Performance vs. Iterations, g11

[Matlab\NonLinear\fullsim_IBO_PD1vsPD3_g6] ..................................................... 100

Figure 41: PD-3 Parameters for g11 [Matlab\NonLinear\fullsim_IBO_PD1vsPD3_g6] 100

Figure 42: Convergence of PD-1, PD-2 for G11, $\Omega = 35$ .................................. 101

Figure 43: Coefficient magnitudes for PD-2, g11, $\Omega = 35dB$ ............................ 101

Figure 44: Coefficient magnitudes for PD-2, g11, $\Omega = 30dB$ ........................... 101

Figure 45: Gain of PD-2, G11, $\Omega = 30dB$, $B=1.259$ (13dB) .............................. 102

Figure 46: CCDF of g11 input and output for PD-1,PD-2, $\Omega = 30dB$, $B=1.259$ (13dB) ................................................................. 103
1 Introduction

The efficiency of a power amplifying system is a characteristic of its material composition and form. The system’s operating efficiency depends on its use. Consider a car with a gasoline engine. While much design effort is made to maximize the use of fuel energy in producing power, the car is seldom operated to yield its maximum efficiency. For a given engine with maximum efficiency at an engine speed around 5000 RPM, the transmission ratio to the wheels is deliberately set so that the engine runs at only 3500 RPM on a highway at full speed. This tradeoff is fundamental to automotive system design, and of course race cars and luxury cars and off-road vehicles are tuned for different purposes. You might say that the design is optimized for usefulness.

Similarly an amplifier is often not useful operating at its characteristic maximum efficiency. When amplifying an information-bearing signal, the signal itself is a dominant factor determining the power efficiency. In the application to wireless communication other desired benefits in the broader system weigh against power efficiency. The competition in the trade-off space includes spectral efficiency, latency, and processing resources, each with a corresponding financial cost for users. Today, the cost of licensed spectrum and pressure on energy consumption is increasing while the cost of digital and analog signal processing continues decreasing. Demand for communication is also increasing in self-reinforcing cycles, making efficiency optimization a valuable problem.
This work applies primarily to signals with high spectral efficiency (bits/Hz). Such signals have high peak-to-average power ratios and it is generally accepted that poor power efficiency in the amplifying system is the price that must be paid.

1.1 Summary

The thesis of this work is that it is possible to verifiably achieve the highest possible power efficiency within the limits of a realized system while satisfying all measures of quality and conformance.

When optimizing a complex system it is necessary to revisit some of the elementary principles of the system, which involve functional analysis, estimation theory, and adaptive modeling. In the domain of amplifier linearization, although the theory and practice have been well accepted and applied, holistic optimization is however still lacking a sufficient model. Such a model and problem description are developed in this work. The minimum requirements of the desired solution are expressed in the following statement.

**Minimum Problem Statement**

Given an amplifying system and a signal with known characteristics and specified performance limits, determine the operating power level and the predistortion linearization that maximizes operating efficiency.

Achieving maximum efficiency often translates to finding and executing the simultaneous minimum margin and operational limits. For reasons to be shown, this goal is possible given sufficient understanding of the relationship among the limits.
On the theoretical side, this work includes some basic functional analysis of the proposed cost and constraint functions, which is omitted in almost all contemporary literature on this subject. Also, the ideal linearization is defined relative to the operating point of the PA. On the practical side, the implementation constraints are expressed within the problem formulation. Other work in this domain has not defined potential constraints that are commonly encountered in practice. In optimization solutions, it is valuable to quantify the impact of constraints regardless of their mutability.

Experimentation with a real system or model shows qualitatively that the advantages of pre-distortion (PD) consistently diminish as an amplifier is operated beyond a certain power level. This necessitates an optimization of the operating point of the amplifier system. The algorithms developed in this work assume the efficiency is monotonically increasing with an increase in power, i.e. a decrease in power back-off. It is shown how to determine when this assumption is true and why it is commonly satisfied.

1.2 Organization of paper

Background information is given briefly considering the extensive coverage of this topic in literature. A contribution of this work is the comparison between algorithms based on a least-squares solution and other algorithms employing more general theory and additional computation. The problem is formulated using these techniques and the results are compared.
Specifically, three pre-distortion function models and two architecture paradigms are presented. Three specific combinations are constructed from these and tested against four PA models.

<table>
<thead>
<tr>
<th>PA MODELS</th>
<th>ARCHITECTURES</th>
<th>ALGORITHMS</th>
<th>RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>G7</td>
<td>Indirect</td>
<td>Least Squares</td>
<td>PD-1</td>
</tr>
<tr>
<td></td>
<td>Learning with</td>
<td>Search (LSS)</td>
<td></td>
</tr>
<tr>
<td>G8</td>
<td>Explicit Back-</td>
<td>Convex</td>
<td>PD-2</td>
</tr>
<tr>
<td></td>
<td>off</td>
<td>Optimization</td>
<td></td>
</tr>
<tr>
<td>G9</td>
<td>Direct</td>
<td>Solver</td>
<td>PD-3</td>
</tr>
<tr>
<td></td>
<td>Optimization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G11</td>
<td>with Implicit</td>
<td>Projections on</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Back-off</td>
<td>Constraint Sets</td>
<td></td>
</tr>
</tbody>
</table>

1.3 Intelligent Systems and Optimization

There is a crossover between the concepts of intelligent systems and optimization as well as a common set of mathematical theory and tools.

The initial question in an optimization is simply whether a set of optimal solutions is expected to exist. There is a class of problem specification where each function is described by a closed form expression, and a problem instance is described by parameter
values. Usually the structure of the functions pre-determines the type of solutions that exist but each problem still requires an optimization to compute the exact solution.

Alternatively, the oracle model perspective is similar to that of system identification. “In an oracle model, we do not know \( f \) explicitly, but can evaluate \( f(x) \) at any \( x \in \text{dom} \ f \) [1].” The true oracle approach to this problem is appealing because real system (PA) behavior is a function of the stimulus applied. If the behavior is a function of the average power, then a model which is constructed at a different average power is not accurate. And if the average power is an explicit parameter to be optimized, then the optimization is forced to query the oracle at any proposed power level. By executing the optimization algorithm on the functioning system, this eliminates the step of creating a system model for off-line optimization.

In order to predict or estimate performance over an unknown model, techniques of identification must be used, involving selection of an appropriate model of the unknown system, collecting stimulus and response data, and exploiting the data to estimate the model parameters.

In this work all of these perspectives are applicable. Regardless of the approach, the cost of collecting more data is considered minimal and the cost of risk is high when operating an unknown physical system near uncertain bounds.
2 System Description

A complete description of the subject system is partitioned between a general system block diagram and specific model details. This chapter includes background information about the problem context in §2.1-2.2, and motivation for the unique thesis problem in §2.3. The chapter concludes with brief mathematical definitions in §2.4. All models and signal vectors (unless specified) are referenced to complex baseband (envelope) representation [2].

2.1 Power Amplifier

A transistor used as a signal amplifier exhibits the characteristics of cutoff, linear amplification, and saturation, depending on the mode of operation. The pre-distortion (PD) technique of linearization most often applies to an amplifier which is tuned for Class AB operation. While this is a “linear” mode of operation, the power amplifier (PA) imposes various forms of non-linear distortion on the amplified signal.

A figure of merit for power amplifier operation is power-added efficiency (PAE). The higher the efficiency, the more supply power is delivered in the amplified signal. For most systems, and as assumed in this paper, the supply power, $P_{in,DC}$, is constant. Therefore, from (1), the efficiency increases with increased output power, $P_{out,RF}$.

$$\eta_{PAE} = \frac{P_{out,RF} - P_{in,RF}}{P_{in,DC}} \quad (1)$$
2.2 **System Block Diagram**

The transmitting signal path diagram is shown in Figure 1. The signal \( w[n] \) is the ideal discrete complex baseband waveform produced by the modulation process. The pre-distortion operation (PD) is applied in a manner similar to a digital filter but without the restriction of linear operations. The PD output signal \( x[n] \) is converted to an analog RF passband waveform by one of several typical means involving interpolation, mixing, and filtering. Finally, the signal is amplified by an amplifying system consisting of one or multiple stages both in series and in parallel.

![System Block Diagram](image)

**Figure 1: Signal Path Diagram**

The amplifier has a desired primary function – amplification – and an undesired function – nonlinear distortion. The purpose of the PD is to cancel the undesired function while preserving the amplifier’s linear signal gain. The cancellation is a function on the signal, in much the same way that a linear equalizer cancels the frequency response applied to a signal. The PD must correspond to the current operating mode characteristics of the amplifier. It is possible to pre-determine a fixed PD for each PA...
operating mode and apply the appropriate PD during operation. Alternatively, an automatic system can dynamically determine the PD while the PA is operating, at the cost of additional sensing and algorithm subsystems, as shown in Figure 2 below.

Again the signal $w$, generated by the waveform source, is the reference for any distortion introduced by the system. The PD operates its input $u$ which is equivalent to $w$ or a modified version of $w$. The PD output signal has sufficient sample rate to represent the excess bandwidth of the inter-modulation distortion (IMD) introduced by the non-linear transform. The signal $x$ is scaled and converted to the analog RF input signal to the amplifier. The output of the amplifier is coupled and down-converted to the digital baseband signal $y$. The PD algorithm may use samples of $w$, $x$, and $y$ to linearize the system. The equations governing these subsystems are described in the Models section.

![Figure 2: System Diagram with Feedback](image-url)
2.3 In Context of Previous Work

Significant work has explored the problem and solutions for linearization by pre-distortion methods. The research is often presented as depicted by the vertical arrow in Figure 3, where a system is to be operated at a desired power level, i.e. back-off level, and the proposed solution improves signal performance over previous methods. In [3], for example, the author performs the analysis with the signal level at 1.5dB peak back-off, which is typically the maximum level at which pre-distortion is highly effective. Alternatively, an author may show that a proposed model performs as well as a more complicated model or a less computationally efficient method, such as in [4] and [5].

Figure 3: Performance Example

Solution algorithms for computing the pre-distortion (PD) use least squares (LS) [4], [6], and recursive least-squares (RLS) [5], [7], [3] for a variety of linear basis PD models. Others use training approaches with artificial neural networks (ANN) [8] or a genetic algorithm (GA) [9]. Interestingly, the GA approach is shown to work when the
bandwidth of the feedback signal is too narrow to capture the inter-modulation products for analysis.

A related problem is to determine the optimal signal transformation to a constrained peak-to-average that results in the best signal performance. The SNDR is optimized for general signals in [10] and for OFDM in [11]. Another work [12] minimizes OFDM PAR subject to constraints on modification of the ideal signal while admittedly ignoring the effects of PA nonlinearity. All of these have similarities to this work in satisfying signal constraints. Constraining peaks with limited distortion is synergistic as pre-processing for the problems stated herein.

This work addresses the problem depicted by the horizontal arrow in Figure 3 - determine the optimal efficiency at which the power amplifier system can be operated for a given set of constraints on signal performance. In most cases, where operating efficiency increases with average output power, this is achieved by minimizing back-off. The performance metric shown in Figure 3 is the modulation error ratio (MER) and the curves A, B, and C represent the performance three abstract algorithms. It is also assumed at this point, and will be proven later, that there exists some ideal bound depicted by the dashed curve.

This work serves to advance the problem with a formulation in an optimization framework. From this framework, the primary contribution of this study is the determination of valid constraints and the effects of those constraints on overall performance. And despite the extensive coverage over a considerable time period of predistortion linearization, I did not find any work directly addressing this stated
problem. Furthermore, I did not discover a sufficient model of the entire scope of the specific problem domain and so I am presenting a full system model.

\section*{2.4 Definitions}

A discrete-time sampled signal is a complex valued sequence or n-tuple. The set of all such signals comprises a metric space and more specifically a Hilbert space which induces an inner product on the vector space over the complex numbers. The Euclidian distance metric is:

$$d_E(x, y) = \sqrt{\sum_{k=1}^{N} |x_k - y_k|^2} \quad x, y \in \mathbb{C}^N$$

The associated Euclidian norm is the distance between a vector and the origin:

$$\|x\|_2 = \sqrt{\langle x, x \rangle} = \sqrt{\sum_{k=1}^{N} |x_k|^2} \quad x \in \mathbb{C}^N$$

The Cauchy-Schwartz inequality is:

$$|\langle x, y \rangle| \leq \|x\|_2 \|y\|_2 \quad x, y \in \mathbb{C}^N$$

The magnitude limiting function is:

$$\Lambda_T(x) = \begin{cases} x, & |x| < T \\ T e^{i\phi_x}, & |x| \geq T \end{cases} \quad \forall \ x \in \mathbb{C}, T \in \mathbb{R}_+$$
The magnitude of $\Lambda_T(x)$ is always less than or equal to $T$ while the phase $\varphi_x$ is preserved for all $x$. The magnitude limiting function is only partially invertible, when $|x| < T$.

A general adaptive polynomial function is defined as:

$$f(x[n]|a^{[j]}(i)) = \sum_{k=0}^{K} a_k^{[j]} x[n]^k \quad \forall n \in p(i) \subseteq \mathbb{Z}_+$$

(5)

where the coefficient vector $a^{[j]}$ is applied to the signal $x$ during a corresponding period of time $p(i)$. More specialized versions of this function are used as models.

Following the convention presented in [13], let property sets $S_j$ be specified by

$$S_j = \{ x \in X: x \text{ possesses property } P_j \} \text{ for } 1 \leq j \leq M$$

Such property sets are used to define the feasible set under constraints. Given each constraint on a signal defined as a property set, the feasible set can be determined as:

$$S_{F_x} = S_1 \cap S_2 \cap \ldots \cap S_M$$

A random variable has a probability distribution function over a range. The range is a set of values for which the function is non-zero. When a function operates on a random variable, the co-domain of the function may differ from the range of the input r.v. For example, the magnitude limiting function is a mapping from a potentially unbounded set to a bounded set. For convenience, the co-domain set for this type of function is
given explicit notation. For example, when a normally-distributed r.v. $X$ is processed by a limiting function (4), the co-domain of the resulting r.v. $Y$ is denoted $S_T$.

$$X \sim \mathcal{N}(0,1) \in \{\mathbb{R}: (-\infty, \infty)\}$$

$$Y = \Lambda_T(X) \quad y \in S_T = \{\mathbb{R}: |y| \leq T\}$$

The magnitude of a noise-like signal has approximately a Rayleigh distribution from $|u| = \sqrt{I^2 + Q^2}$ where $I, Q \sim \mathcal{N}(0, \sigma^2)$.

The pdf of this distribution is

$$f_X(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} \quad x \in \mathbb{R}_+ \quad (6)$$

And the cdf is

$$F_X(x) = 1 - e^{-\frac{x^2}{2\sigma^2}}$$
3  Distortion and Signal Quality

3.1  Nonlinear Distortion

Any signal transformation which does not qualify as a linear transform is some variety of nonlinear transform. The criteria of a linear system are superposition and frequency preservation. For a linear power amplifier, the relevant type of nonlinear distortion is the gain and phase distortion over the input signal amplitude.

3.2  Signal Quality

A signal quality measurement is an estimator or a function of the observation of the signal. For example, a meter which measures RMS power of a signal observes the signal through a sensor and produces a value, either continuously or at discrete intervals. This measurement has statistical properties which depend on the noise and distortion in the observation from all sources and also on the function producing the value.

As a figure of merit, a signal quality measure, as with any statistical value, must be specified with the statistics of the measurement technique. In the absence of this information, a sufficient number of trials may be performed to estimate the statistics of the measurement such as the variance and bias.

3.2.1  SNR

Signal-to-noise ratio is the deviation of a signal from a reference. The deviation is caused by noise, distortion, or both. Unless specified, SNR refers to the “signal-to-noise plus distortion ratio” (SNDR) or “signal-to-noise and distortion ratio” (SINAD).
The SNR is related to the mean square error (MSE) of the time-domain sequence, a commonly used metric.

3.2.2 EVM / MER

The error vector magnitude (EVM) and modulation error ratio (MER) are two related measures of the deviation from the ideal modulated reference. As a metric the expected value over a sufficient number of symbols is implied. The EVM is typically expressed as a percentage, while the MER is expressed as a ratio in dB, each relative to a modulation-specific nominal signal level. Since these measurements are specific to the representation of symbols, they are only affected by noise and distortion within the bandwidth of the modulation.

3.2.3 ACPR

The adjacent channel power ratio (ACPR) is a measure of the power within the designated channel to the power within a specified bandwidth outside the channel. This ratio can be measured for both the baseband representation and by RF measurement equipment. The ratio is normalized by bandwidth, if the two are different.

\[
ACPR(X) = \frac{\sum_{\text{CHANNEL}} X^2(\omega)}{\sum_{\text{ADJACENT}} X^2(\omega)} \left( \frac{\text{BW}_{\text{ADJACENT}}}{\text{BW}_{\text{CHANNEL}}} \right) \quad X = F(x) \tag{7}
\]

Measurement specifications are provided with any communication standard or by the appropriate regulatory body. This metric is appealing because it does not require comparison to a reference signal.
3.2.4 Approximations

An approximation of one metric can be derived from another. Under certain conditions and assumptions, an approximation is used to substitute a simpler calculation for a more complicated calculation or immeasurable quantity. A reasonable approximation of MER includes only the portion of the noise and distortion which resides in the bandwidth of the modulated signal, as represented by the summation over the bandwidth (BW) in (8) below. The relationship to the total SNDR depends on the power spectral density of the noise and distortion.

\[
MER(x, \hat{x}) \equiv \frac{\sum_{BW} X^2(\omega)}{\sum_{BW} N^2(\omega)} \quad X = \mathcal{F}(x), N = \mathcal{F}(|x - \hat{x}|) \quad (8)
\]

This approximation could also be viewed as the MER after the signal plus noise has been filtered by an ideal matched filter, which would have the same effect of removing the contribution of noise and distortion outside the bandwidth of the modulated signal. The accuracy of this MER approximation is only important when the actual value matters, such as in a constraint of some optimization. When used for minimization, the gradient of SNR and MER are interchangeable. In this work, unless specified, MER refers to the approximation (8) above. Furthermore, given Parseval’s identity, and assuming that the signal bandwidth is strictly less than the sampling bandwidth, the MER is greater than the SNR because the noise plus distortion energy in a fixed bandwidth is less than or equal to the total noise plus distortion

\[
\sum_{BW} N^2(\omega) \leq \sum_{\omega} N^2(\omega)
\]

\[
\frac{\sum_{BW} X^2}{\sum_{BW} N^2(\omega)} \geq \frac{\sum X^2}{\sum N^2} = \text{SNR}
\]
3.2.5 Mean square error (MSE)

The difference between two signals or between a signal and its estimate is often measured by the square of the residuals.

\[ MSE(x, y) := \min_{k_n} E[|k_n * y - x|^2] = \min_{k_n} \frac{1}{L} \sum_{i=1}^{L} (k_n y_i - x_i)^2 \quad k_n \in \mathbb{C} \]  

\hspace{1cm} (9)

The minimization over \( k_n \) is easily found when \( y_i = \alpha e^{j\phi} x_i + n_i, \quad \bar{n} = 0 \).

Since \( x \) and \( y \) are complex baseband signals, they must be aligned in magnitude and phase to yield the minimum deviation. However, if \( y \) is a non-linear function of \( x \), Bussgang’s theorem [14] must be applied to decompose \( y \) into a scaled linear and an additive non-linear component.

Note that the MSE is most often normalized relative to the squared signal level \( \sigma_X^2 \) and expressed in logarithmic scale.

3.3 Measurement and Estimation

All estimated parameters of the system are computed using an estimator, which is a function of a r.v. thus having an inherent distribution. The statistical properties of the estimate, such as bias and variance must be quantified. In an engineering context, the errors in these estimates must be factored into the error budget of the system. For example, in normalizing the magnitude of a set of sample data, any variance in the RMS level translates to a variance in gain of any function that is computed from the normalized data. This variance in normalization occurs because a unique data set (sample) is used in
each estimate of signal level, i.e. sample variance. For the signals encountered in this
problem there is variance in the signal level estimate even with large sample sizes
because the magnitude limiting operation has significant impact but low-probability of
occurrence.

In precise statistical language, the variance in the sample variance of a normally
distributed r.v. is inversely proportional to the sample size, n. This approximates the
variance\(^1\) in the normalization of n-sized data sets for this problem and is also shown in
Figure 4 below.

\[
\text{var} \left[ \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right] = \frac{2\sigma_x^4}{n - 1}
\]

A sample size of n=200 real-valued samples results in a variance in the signal
level estimate and thus a variance in output power of 1% or -20dB. In the case of n=200
complex-valued samples, the variance in power is -23dB. This level would be significant
in the context of the broader problem because pre-distortion linearization, being non-
linear, is sensitive to signal level. Typically sample sizes greater than 20,000 are used,
practically eliminating any issues. Other possibilities are Bayesian tracking of signal
level or one-time (fixed) normalization with an independent regulation of power level.

\(^1\) http://en.wikipedia.org/wiki/Variance#Distribution_of_the_sample_variance
Figure 4: Variance of sample variance vs. sample size (n)

Finally, note that the variance of the expected value of the residual magnitude of a curve fit is inversely proportional to the probability density. This is of concern when estimating a curve in the region of the (low-probability) tail of a distribution, as with the noise-like signals used to estimate the PA and PD.
4 Models

In fulfillment of the system description begun in the previous chapter, this chapter describes the models that have been developed for this work. The reference waveform is introduced in §4.1. Next, the relationship (to be optimized) between efficiency and back-off is described in §4.2. The final three sections introduce the power amplifier (PA) and pre-distortion (PD) functions and give their context in the complete simulation model. The optimization problems are not given until chapter 5.

4.1 Discrete-time equivalent system

A waveform generated in the digital domain is a discrete-time representation of a band-limited continuous signal. This representation and the properties of the random process that generated the signal determine the probability distribution function (pdf), and the corresponding cumulative distribution function (CDF) of the signal samples.

The waveform WF11 is the reference for all simulations. The waveform is two adjacent channels of OFDM symbols, centered at complex baseband. Symbols are 8k 16-QAM carriers, oversampled by 4X. The SNR is set to approximately 52dB. The complementary cumulative distribution function (CCDF) of the signal amplitude (Figure 6) is the probability that the normalized amplitude \( w/\sigma_w \) is above the level \( \gamma \).
4.2 Efficiency and Back-off

The optimization criterion is the power efficiency of the system. This is known to be a function of the operating power level. The optimization result will provide an optimal power level. The efficiency is determined from a known function of signal level relative to output back-off (OBO). An equation for operating efficiency is based on the known or provided efficiency function and the signal’s probability density function.

The instantaneous efficiency of the system, $H_{PA}(x)$, can be provided from measurement in order to calculate absolute efficiencies of operation. This information is generally not needed as the relationship is affine and optimization is equivalent with or without the scaling factor. For example, if the efficiency is nominally 60% at the output saturation level (normalized to unity), assuming a linear relationship with input magnitude, the instantaneous efficiency at any level is

$$H_{PA}(x) = 0.6 \times x \quad x \in ([0,1])$$

Finally, the operating efficiency can be calculated from the magnitude distribution
To show the effect of back-off on overall efficiency, the integral above can be evaluated, assuming a specific probability density function, \( f_X \). For this example, a Rayleigh distribution (6) is used with a range of values of average signal level, \( \sigma_R \). The following table shows efficiencies calculated from (10) for this example with an ideal wideband noise signal for multiple values of back-off.

Table 1: Efficiency vs. OBO

| OBO dB | \( E[|x|^2]^{1/2} = \sqrt{\cdot \sigma_R} \) | \( \sigma_R \) | Efficiency \( \eta \) |
|--------|---------------------------------|-------------|----------------|
| 7.5    | 0.4217                          | 0.2982      | 22.4%          |
| 8.0    | 0.3981                          | 0.2815      | 21.2%          |
| 8.5    | 0.3758                          | 0.2658      | 20.0%          |
| 9.0    | 0.3548                          | 0.2509      | 18.9%          |

The method of deriving OBO from \( OBO_{dB} = 20 \log \left( \frac{\sqrt{\cdot \sigma_R}}{1} \right) \) is a sufficient approximation to show the strictly monotonic relationship with operating efficiency over this range. As the average power increases, the system operates at higher instantaneous efficiency with higher probability.

The proposal is that every system with amplifier function \( g \) has an optimal operating mode \( \xi^* \), given a reference signal with pdf, \( f_W \), and a set of constraints \( S_F \).

\[ \text{System}(g, f_W, S_F, \xi^*) \rightarrow J^* \]
For the standardized statement of a constrained optimization problem, the optimization variable vector $\xi$ contains signal parameters and the power level which is equivalent to the input back-off (IBO).

$$\xi = [a \quad \beta \quad \gamma]^T \in \mathbb{Z}$$

The vector $a$ is the coefficients of a PD function, $\beta$ is the back-off. Also $\gamma$ is a peak reduction parameter for pre-processing of the reference signal prior to the PD (to be described in §5.7). The absolute power output is ignored in lieu of a back-off.

### 4.3 Parametric PA models

Behavioral modeling of electronic circuits uses the physics of the elementary components and devices. Such relationships are used in circuit design and simulation tools such as Agilent ADS. Even with necessarily extensive models, often empirical methods determine behavior over the variations from component tolerances. There is one deficiency with static models - the behavior of a system changes with the average power applied. This is due to the effects of temperature on parasitic behavior and thermal noise.

Beyond the models based on the physics and circuitry of the amplifier devices, several canonical models have been developed, notably Rapp, Seleh, and Honkanen [15]. A canonical model is often an intermediate step in the context of linearization where the circuit model is not necessary as it would be in considering modifications of a design. Instead, the canonical models determine the impact of accounting for bulk features in the behavior, such as nonlinear AM-AM, AM-PM, and time-constants.

The simplest sufficient models for this problem are AM-AM models, such as Rapp, in which the output gain is not constant but rather a function of the input
magnitude. The output phase is not considered a function of the input magnitude (AM-PM) in this model. The feature of the Rapp model is the “smoothness parameter defining the transition smoothness from the linear region to the saturated region [15].” The Honkanen model enhances the Rapp model “to copy the low-voltage features of a real amplifier causing crossover distortion [15]” and also introduces an AM-PM conversion function. In order to have a higher degree of flexibility in the shaping and scaling of the AM-AM and AM-PM curves, additional models were developed based on polynomials of varying degrees as described below. The model g8 approximates the Rapp model.

For this work, the AM-PM distortion will be generally disregarded, although all PA and PD models have complex coefficients and thus readily represent AM-PM. Non-linear phase distortion, if present in the PA, must be absent from the output (by correction) when minimizing MSE, as shown in [10]. Since the PA models are AM-AM only, phase distortion is not present. All linearization calculations have been verified with AM-PM models, but results are not included. Future work (§8.2) may include independent analysis of AM-PM along with memory effects.

Described mathematically, the memory-less AM-AM is a member of the set of continuous non-decreasing monotonic functions.

\[ g: \mathbb{R}_+ \rightarrow \mathbb{R}_+, \text{dom } g = \{[0 b]\}, \exists A \in \text{dom } g \mid g'(x) \geq 0 \ \forall \ x \leq A \]

Furthermore, the function exhibits a saturation property. If the function has a critical point in the domain, the function is constant beyond the critical point. This property preserves monotonicity.

\[ \text{if } (\exists c \in \text{dom } g | g'(c) = 0) \text{ then } (g(x) = g(c) \ \forall x \in \text{dom } g | x > c) \]
These conditions are insufficient to declare the function concave. This is important because concavity or convexity is necessary for convex optimization. There are general models within the set that are concave. The Rapp model is a concave subclass of the set if the domain extends to the critical point, $dom \, g = \{[0..b] \mid b = c\}$.

Unfortunately the AM-AM functions cannot be restricted to the concave subset, as real systems exhibit multiple non-linearities. According to [1]§3.2.1, a weighted sum of convex functions is convex if and only if the weights are non-negative. Since positive integer powers over the positive real numbers are convex functions, then a general polynomial AM-AM function such as (5) is only concave (negative convex) if all coefficients (except the linear term) are non-positive. The models described below are not concave.

The plots in Figure 7 below show three of the PA models which will be studied. As described earlier, the horizontal axis is the input signal magnitude normalized to the RMS level and the vertical axis in dB is the gain, normalized to 0. The models were designed to have identical input $\theta_1$ levels and similar saturation levels. The $\theta_1$ parameter is the input level at which the output gain is suppressed by 1dB from the maximum (known as 1dB compression). The difference in the shape of the curves is accounted for by the difference in the model order. When comparing these models, the IBO is referenced from the input saturation level.
### Figure 7: Comparison of PA models

[MATLAB/Nonlinear/Haggman/show_g7g8g9.m]

<table>
<thead>
<tr>
<th>Name</th>
<th>Model Type</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>G8</td>
<td>3rd-order polynomial</td>
<td>[1.0048, 0.0857, -0.1905]</td>
</tr>
<tr>
<td>G7</td>
<td>5th-order polynomial</td>
<td>[0.9778, 0.0463, 0.2315, -0.5144, 0.1715]</td>
</tr>
<tr>
<td>G9</td>
<td>7th-order polynomial</td>
<td>[0.6928, 3.6562, -16.0852, 34.0345, -36.7929, 19.3501, -3.9354]</td>
</tr>
</tbody>
</table>

Additional models were created by curve fitting of data from real systems. These models are better predictors of actual performance because they include non-linearity over the entire signal range.

<table>
<thead>
<tr>
<th>Name</th>
<th>Model Type</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>G6</td>
<td>5th-order polynomial</td>
<td>[0.7662+j<em>0.0396, -1.6659-j</em>0.2743, 14.7374+j<em>3.1745, -27.5651-j</em> 9.4430, 14.7593+j*6.9268]</td>
</tr>
<tr>
<td>G11</td>
<td>5th-order polynomial</td>
<td>[0.7662, -1.6659, 14.7374, -27.5651, 14.7593]</td>
</tr>
</tbody>
</table>

26
4.4 Predistortion Model

The predistortion function is an unknown nonlinear function of the signal magnitude. A common method to define the function is to select a parameterized model and determine the parameters by an algorithm. This is similar to the approach of an adaptive linear filter in which the discrete-time model is typically an FIR transversal filter and the parameters are the tap weights. Other methods have been proposed such as by an artificial neural network (ANN) [8], [16] or a genetic algorithm (GA) [9].

Let us establish some properties of such a function, regardless of its form. Monotonicity in the amplitude is a necessary condition for invertibility as described in §5.3. The next practical restriction is unity gain, which applies only to the extrinsic back-off formulations such as the on-line indirect learning architecture described in §5.8. The magnitude of the expected value of the input and output must be equal for a given probability distribution.

Among parameterized models for adaptive pre-distorters, the primary division is between memory-less nonlinearity and a model with memory. The Volterra series [17] is the general nonlinear memory model. Other specific minimal-parameter models which can be mapped to the Volterra model include Wiener-Hammerstien and the memory polynomial [18][19]. The memory-less polynomial model operates independently at sample time \( n \) on sample \( u_n \) to produce an output sample \( x_n \).

\[
x_n = F_N(u_n) = u_n \ast \sum_{k=1}^{N} a_k \ast |u_n|^{k-1}
\]  

(11)
The properties of this model have both a benefit and a limitation. A polynomial of order \( N \) is \( N \)-times continuously differentiable and therefore smooth. Just as continuity of the function defines the relationship between two nearby points, the continuity of the derivative confines the nearby slope and the second derivative confines the change in slope. The slope of a normalized PD function is generally close to unity over the lower range and gradually increases over the higher range. Consider the case in which a polynomial, \( p(x) \), is used to approximate a piecewise continuous function \( q(x) \). While \( q'(x) \) is discontinuous at \( x_1 \), so \( q'(x_{1-}) \neq q'(x_{1+}) \), \( p'(x) \) must be continuous with \( p'(x_{1-}) = p'(x_{1+}) \). The higher the order of the polynomial, the more degrees of freedom allowing for closer approximation of functions with sudden changes is slope over a small range. So while smoothness and linear independence between terms are desireable properties, the model is not well-suited for piecewise continuous functions, especially when the model order is low. Multiple model orders for the PD are used, but the selection of the appropriate model order is not addressed here.

Piecewise functions have also been considered in other work. In [5], a piecewise linear approximation is used. And splines are another popular method for general curve fitting. The advantage of spline curve fitting is that parameters influence a confined range of the curve. But the disadvantage is the dependence among parameters to satisfy continuity requirements. In contrast, the parameters of a polynomial are independent, but they all influence the entire range of the function.
4.5 Simulation model

The model of the predistortion algorithm is described later in §5.8 “Indirect Learning Architecture.” Since back-off is an important parameter of the problem, its definition and implementation must be considered. A uniform mathematical description of back-off is now provided.

A small-signal amplifier model has properties such as gain which are nonlinear and therefore dependent on the input signal level. The input back-off (IBO) represents an input level relative to an absolute magnitude such as the input saturation level. Since the input saturation level is fixed for each model, the back-off is a fixed affine function of the input signal level. In simulation, the input saturation level is merely a fixed reference point. The parameter of interest is the estimated average output power relative to the estimated output saturated power.

There are two methods of varying the PA input level (and thus the back-off). The first method is to allow the signal level of the PD output to vary while applying a fixed scaling between the PD output and the PA input. This will be referred to as implicit back-off. The second option, explicit back-off, fixes the PD output level and introduces an additional parameter, $k_{in}$, to set the back-off. The linear scaling $k_{in}$ of the PD output signal $x_{PD, out}$ scales the RMS signal level to set the desired RMS input level to the PA, $\sigma_{PA, in}$.

$$x_{PA, in} = k_{in} \cdot x_{PD, out} \Rightarrow \sigma_{PA, in} = k_{in} \cdot \sigma_{PD, out}$$

Thus the input back-off, $\beta$, is the ratio of the PA input level to the saturation level. The RMS PD output level in terms of the PD function will be given in §5.11.2, eqn. (32).
\[ \beta = \frac{k_{in} \cdot \sigma_{PD,\text{out}}}{x_{sat,in}} \cdot \frac{\sigma_{PA,\text{in}}}{\theta_{sat,in}} \quad \forall_{sat,in} \in \mathbb{R}_+ \] (12)

For an optimization problem, implicit back-off is preferable. The back-off does not have to be controlled by an independent variable and so the number of parameters to optimize is reduced. The second method, explicit back-off, only becomes necessary in simulation and implementation. This will be discussed further in the problem formulation.

Input and output back-off is commonly expressed in \( dB_{\text{SAT}} \) with \( \beta_{dB} = 20 \log_{10} \beta \).

For an individual problem, the value of \( \beta \) has little significance, as the level could be expressed relative to any constant. Instead IBO and OBO are valuable as a uniform method of comparison between amplifiers and algorithms.
5 Problem Formulation

Thus far sufficient background for the problem is established, but it is the actual formulation that rigorously defines properties of the desired solution and determines appropriate algorithms. The formulation is adequate description of a solvable problem, based on the associated system models. The algorithm development process also involves design – making choices that affect the nature of the solution. Finally we arrive at optimization, which is a field of applied mathematics and as such it offers a set of theory and tools for solving problems in multiple ways.

A problem instance consists of an unknown but unchanging PA model, the properties of an available data set, and the set of constraints. The objective is a maximum within a constraint set and the available data should be used in any way that is helpful. One option is an exhaustive search within the known constraint set. This technique will almost always produce an optimal solution. This is undesirable for multiple reasons. In its most naïve implementation, the time required to complete the search excludes it from employment in all but the most critical cases. Also, the necessary exploration involved in mapping the solution space may drive systems into unstable operation or violate bounds which were not predicted and defined.

A formulation is classified broadly as requiring only a linear least-squares (LS) calculation or a robust optimization. Both the strength and weakness of LS methods is the resulting singular solution with a global minimum. This is desirable because the solution can be expressed analytically in terms of the problem data and because a global minimum is guaranteed to be optimal. But this is provided that one understands and
accepts the assumptions and limitations of such a formulation. Consequently an LS formulation is often adequate but not ideal. Once it is resolved for one reason or another to depart from LS, the realm of solution approaches explodes rapidly. A formulation involving LS is presented as a reference along with important non-LS candidates.

An appropriate solution method can be determined by classifying the formulation, understanding the intended use of the solution, and knowledge of the form of the data that is available for the problem. The data in this estimation problem is available in batch form and independent batches are available when needed, such as after a change in the system. For problems in a general form, selecting a solution technique is also a processing issue. If each sample were considered a constraint, it would be possible to optimize adaptively via relaxed projections as in [20], but this would not take advantage of the efficiency of the LS regression or other batch calculations.

In the general problem, a solution is drawn from a set. In the PD linearization domain, the functional space of the PD is constrained and solutions are further constrained by the expected performance. This is an important step in moving away from a minimization of the MSE as the objective. A solution may be projected onto the space or approached via a gradient method. Before describing these techniques, the functional space itself must be understood.

5.1 Problem Statement

The problem is first a constraint satisfaction problem. Among the constraint set, the solution with maximum efficiency is desired. As the problem is defined, the cost function has a single term in which system parameters determine efficiency. There are
no other competing terms against which the efficiency is traded. The constraint functions, $C$, define multiple constraint sets, each of which may be defined in a different space. The PD function is selected from a function space which is typically a linear parameterized complex Hilbert space. The selection of an appropriate function space is a design task.

**Problem 1**

$$\text{maximize } E[\eta_{PAE}] \Leftrightarrow \text{minimize } J$$

$$J(f_{PD}|f_{l}, \eta_{P}, g, C) = -E[\eta_{PAE}] \quad f_{PD} \in \mathcal{F} \quad (13)$$

$$E[\eta_{PAE}(X)] = \int_{-\infty}^{\infty} \frac{P_{out,RF}(g(x)) - P_{in,RF}(x)}{P_{in,DC}} f_{X}(x) \, dx \quad (14)$$

The power efficiency cost function (13) includes the unknown amplifier function, the parameters of the pre-distortion function, and the waveform. One use of this cost function is for comparison between PD functions for a given amplifier and waveform. Another use is for comparing amplifier configurations. In either of these cases, the absolute cost is used in the comparison. It is entirely possible that one amplifying system achieves a higher efficiency (lower $J$) than another, but it is necessary to confine the optimization to the proposed scope, as other factors such as financial cost, complexity, frequency range, and reliability also determine the best amplifier choice for an application.

Besides facilitating comparison, the function provides a unified space over which a minimum can be sought. The best way to find simple solution methods is to be familiar
with this space. The remaining sections provide that familiarity and the corresponding solution strategies.

Studying (14), the two basic strategies for maximizing efficiency are to maximize the output power or minimize the power supply input power. There are viable methods of varying (modulating) the supply voltage with the input signal voltage so that the efficiency curve is dynamic. These methods have been shown to approximately double the operating efficiency for a waveform with a high peak-to-average power ratio (PAPR). In this work, it is assumed that the power supply voltage is fixed and independent of the instantaneous input voltage. Another equally important opportunity for optimization, which is not explicitly addressed in this work, is when the output power is directly constrained, so the goal is to operate at maximum efficiency for a fixed power.

The optimization problem, Problem 1, can be solved by an equivalent problem, provided the necessary assumptions are met.

**Assumption 1**

The operating efficiency of (14) is monotonically increasing with the average output power.

When Assumption 1 is met, maximizing the average output power also maximizes operating efficiency, which leads to Problem 2. Since $P_{in,dc}$ is declared to be fixed and the probability distribution of the signal amplitude is non-increasing above the mean (true for noise-like signals), then Assumption 1 is sound. See §4.2 for an explanation and proof. As the output signal becomes increasingly saturated, the efficiency (and distortion) increases.
**Problem 2**

\[ \text{maximize } E[P_{out,RF}] \iff \text{minimize } \beta \]

The objective is simplified (by Assumption 1) from Problem 1, but the problem parameters are unchanged. The back-off $\beta$ is defined in (12). Output power is only bounded indirectly by the constraints. Problem 2 is not yet a problem formulation, so it cannot be solved. A formulation includes the details of a model and constraints that are necessary to apply a solution method.

### 5.2 Graphical Description

The AM-AM curve shown in Figure 8 is calculated from a data set from an operating amplifier system (1kW L-Band transmitter). The aspect ratio of the first plot is scaled so that the linear amplification reference (diagonal dotted segment) is approximately a 45 degree angle. This line has slope of unity and it is also the line of reflection for the inverse function. Any points where the function crosses the reference line are fixed points of the respective mapping.

First compare two points, A and B, on the PA curve. At point A and over a significant range in the locality of A, the curve is approximately linear. At the point B, the curve has deviated from the linear reference and the slope is significantly below unity.

The other two points, C and D, lie on the curve of an estimated inverse function. The inverse function is approximately a reflection over the reference line of the entire PA AM-AM curve and the point C is a reflection of the point B. Amplitude pre-distortion
can be described through the relationship of B and C (ignoring for now the small change in average power introduced by the PD). An ideal signal magnitude of 0.62 is mapped through PD to the value of 0.70 on the y-axis. This value now becomes the input (x-axis) value of the PA function, where that value, 0.70, maps to the original value of 0.62. The range of over which this processes can occur without limit is considered linearizable.

The points B and D, while they appear in the gain plot (bottom of Figure 8) to be the inverse of one another, actually have no direct inverse relationship (as B and C do). The point D corresponds to a point which is more severely compressed than B and therefore has a higher gain than C. The values of the points in Figure 8 are summarized in Table 2.

**Table 2: Points in AM-AM Example**

<table>
<thead>
<tr>
<th></th>
<th>Input</th>
<th>Output</th>
<th>Gain [dB]</th>
<th>Slope (\Delta\text{Output}/\Delta\text{Input})</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.30</td>
<td>0.3037</td>
<td>0.11</td>
<td>1.0</td>
</tr>
<tr>
<td>B</td>
<td>0.70</td>
<td>0.6204</td>
<td>-1.05</td>
<td>0.45</td>
</tr>
<tr>
<td>C</td>
<td>0.62</td>
<td>0.70</td>
<td>1.05</td>
<td>2.07</td>
</tr>
<tr>
<td>D</td>
<td>0.70</td>
<td>0.9081</td>
<td>2.26</td>
<td>3.30</td>
</tr>
</tbody>
</table>
Figure 8: General AM/AM, PA (red), PD (blue)
[MATLAB/Nonlinear/NLSysSolve.m]
5.3 Inverse Functions

A function is a mapping from a domain to a “co-domain” or range space. Any function which is a bijection is invertible. An example of a bijection is any continuous function over a domain which is both “one-to-one” and “on-to”. This ensures that the intuition about inversion is true – any point in either domain can be freely and unambiguously mapped back and forth using the function and its inverse.

The relationship is defined in terms of the function derivatives [21] (p.317).

\[
\frac{dy}{dx} = \frac{1}{dx/dy}
\]

\[
(G^{-1})'(G(x)) = \frac{1}{G'(x)} \tag{15}
\]

Since the function and its inverse are interchangeable in (15), neither derivative is permitted to equal zero at any point in either domain. The PA (red) and PD (blue) curves in Figure 8(a) are approximately inverses. The consequences of this relationship are discussed further in the following sections.

5.4 Linearization

Linearization involves companding in that the non-linear operations expand and compress the signal amplitude (along with continuous phase-shifting). When two functions are cascaded, as with a pre-distorter and an amplifier, the companding serves to alter the signal set. The image of the reference signal set through the first function becomes the pre-image of the output set of the second function. This notation becomes
necessary when attempting to confine any of the aforementioned sets. Every point in each of the sets has a mapping to a corresponding point in the other sets. This concept is used in this and other chapters in explicitly tracking sets through the transformations.

The required solution is not a model of the PA itself, but rather an estimate of the best linearizing PD function, \( f \). This solution is obtained from the desired linear function, according to the equation

\[
y_n = g(x_n) = \begin{cases} Ku_n, & |Ku_n| < T \\ \varphi_u, & |Ku_n| \geq T \end{cases}
\]  

(16)

where \( y \) is the PA output, \( g \) is the PA function of the PA input signal, \( x \), and \( u \) is the input of the PD function, \( f \). The PD output is given by

\[
x_n = f_N(u_n) = \sum_{k=1}^{N} a_k * u_n |u_n|^{k-1} \quad u_n, a_n \in \mathbb{C}
\]  

(17)

when the predistortion model is an \( N \)-order memory-less polynomial.

According to the criteria for inverse functions, the inverse exists only over the domain of the function which is invertible. The desired linear response is a piecewise function equivalent to a magnitude limiting function with limit \( T \). This function determines the interval over which the signal is linearized. With respect to the normalized signal \( u \), the input limit or the maximum magnitude \( u_p \) is a function of the output limit, \( T \), and the linear gain, \( K \), both of which may be time-varying.

\[
u_p = \frac{T(t)}{K(t)}
\]
From (16) the linearizable range of interest has the relationship
\[ g(f(u_n)) = Ku_n \]

The signal \( u \) is operated on by the PD function, \( f \) and then by the PA function, \( g \). Ideal linearization implies that the result is a scaled version of \( u \). From this, the PD function is defined in terms of the PA function.

\[ f(u_n) = K \cdot g^{-1}(u_n) \]

Finally, the inverse function is defined from (5)

\[ (g^{-1})'(g(u_n)) = \frac{1}{g'(u_n)} \quad \forall \, u_n \in S_T \]

In contrast, in the region of the domain of \( g \) beyond \( T \), this inverse criteria has no solution because

\[ g(x_n) = Te^{j\varphi u} \]
\[ g'(x_n) = 0 \]

Thus (16) proposed as the minimum required model for practical linearization.

Proper selection of \( T \) with respect to the domain of \( g \) (to be described in §5.7) can guarantee a valid inverse function.

### 5.5 Analytic Solutions

In the AM-AM plot of Figure 8, the normalized output saturation level is 0.67. Assuming that the curve continues with an output of 0.67 for any input greater than 0.90, the expected deviation in this input range is known. The minimum magnitude of \( x \) that produces the output saturation value is defined as \( \theta_{sat, in} \in dom \, g \).
The expected error above $\theta_{sat,in}$ is the sum of the constant error at $\theta_{sat,in}$ and the linearly increasing error. This ignores any error in the linear and transition regions of the curve.

$$e(x) = (0.90 - 0.67) + (x - 0.90) = x - 0.67 \quad x \geq 0.9$$

Now consider the effect of predistortion. The input saturation level of the PA is 0.9. If the PD is an exact inverse of this PA function, then (at point E) $g^{-1}(0.67) = 0.9$. The input value of the ideal PD that maps to the PA input saturation level is 0.67.

Any value at the input of the PD that is greater than 0.67 maps (under ideal PD) to the saturated output level. This relationship establishes a fundamental property of predistortion: The maximum linearizable range of the signal is determined by the normalized output saturation level of the PA.

Any PA input greater than $\theta_{sat,in}$ results in a saturated output. This output can be achieved either by limiting the input to the input saturation value or by allowing the PA to apply the saturation. In practical terms, the saturation level is not immediately or exactly known, so it is not possible to exactly limit the input.

Successful linearization implies that the total (mean squared) deviation from the reference signal at the output of the PA is less with PD than without PD:

$$\int_0^\infty |y_{PDPA} - u|^2 \, dt < \int_0^\infty |y_{PA} - u|^2 \, dt$$

This does not imply the error is less in any other sense, such as the maximum. The condition above can be expressed in terms of expected value. Consider the random
variable U to be the magnitude of the complex baseband signal u and the functions f and g are the PD and PA AM-AM functions.

\[ E \left[ \frac{1}{2} (g(f(U)) - U)^2 \right] < E[(g(U) - U)^2] \]

Expressed in terms of an explicit probability distribution (see (6) in §2.4):

Assuming \( r_n = |u_n| \), and \( r \sim f_R(x) \approx Rayleigh(\sigma) \)

\[
\int_0^\infty \frac{1}{2} \frac{(f(r) - r)^2}{f_R(r)} \, dr < \int_0^\infty \frac{1}{2} \frac{(g(r) - r)^2}{f_R(r)} \, dr \tag{18}
\]

The integrals above can be evaluated given the functions f and g. The weak condition of (18) is usually satisfied by reducing the expected error in higher probability regions while allowing higher expected error in the expanded lower probability tail.

Now consider the optimality condition, with \( f^* \) being the optimal PD function that satisfies:

\[ E \left[ \frac{1}{2} (g(f^*(U)) - U)^2 \right] \leq E \left[ \frac{1}{2} (g(f(U)) - U)^2 \right] \quad \forall f \in \Phi \tag{19} \]

If g were invertible over a range greater than or equal to the range of U, the optimal error would be zero.

5.5.1 Amplitude Limited MSE Bound

The range of the output of g (the PA function) is limited to the saturation level, \( \theta_{sat.out} \). Assuming, \( f(r) = r \), and \( g(t) = t \), giving ideal linearization with zero error (and unity gain) over a range of r, the range of the integral on the left side of (18) reduces to:
\[
\int_A \left( g^*(f^*(r)) - r \right)^2 f_R(r) \, dr, \quad A = g(\theta_{\text{sat, in}}) = \theta_{\text{sat, out}}
\]

Also with ideal saturation, all values of \( r > \theta_{\text{sat, out}} \) are mapped through the PD and PA to \( \theta_{\text{sat, out}} \), and the integral becomes

\[
e_{\text{sat}}(\theta_{\text{sat, out}}) = \int_A \left( \theta_{\text{sat, out}} - x \right)^2 \frac{1}{2\sigma_R^2} x e^{-\frac{x^2}{2\sigma_R^2}} \, dx \quad A = \theta_{\text{sat, out}} \tag{20}
\]

This integral (20) is the squared error between the expected value and the saturated value over the tail of the distribution. The integral can be evaluated by the method in Appendix A. The table below summarizes some results.

<table>
<thead>
<tr>
<th>std(y)</th>
<th>RMS</th>
<th>( \sigma_R )</th>
<th>( \theta_{\text{sat, out}} )</th>
<th>( \beta_{\text{sat}} ) [dB]</th>
<th>( e_{\text{sat}}(\theta_{\text{sat, out}}) )</th>
<th>Normalized MSE [dB]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.707</td>
<td>2.0</td>
<td>6.02</td>
<td>0.001734</td>
<td>-27.61</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.707</td>
<td>2.33</td>
<td>7.35</td>
<td>0.000325</td>
<td>-34.88</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.707</td>
<td>2.5</td>
<td>7.96</td>
<td>0.001270</td>
<td>-38.96</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.707</td>
<td>2.6</td>
<td>8.30</td>
<td>0.000071</td>
<td>-41.46</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.707</td>
<td>3.0</td>
<td>9.54</td>
<td>0.000006</td>
<td>-52.26</td>
</tr>
</tbody>
</table>

This result is important because it is the left side of the inequality (19). More importantly it establishes the performance bound for linearization under the assumptions in (20). This bound occurs when the maximum linearizable range of the PA is linearized by PD, where the linearizable range is determined exclusively by the saturation level. Uniquely, for this bound generating function, the output saturation level is equal to the normalized input saturation level. In the general case, \( \theta_{\text{sat, in}} > \theta_{\text{sat, out}} \). The same curve
is derived in [10], also for a Rayleigh distributed reference, by a different method. Being
an ideal model to estimate a bound, this does not include phase distortion which would be
present in the saturation region of the PA.

For a sufficiently high number of carriers in an OFDM signal, the Rayleigh pdf is
a good approximation for the magnitude. The upper limit of the integral could be reduced
from infinity to any number to approximate OFDM with reduced number of carriers or
bandwidth. Alternatively, the work of Ochiai in [22] produced a more accurate
approximation dependent on the number of carriers.

These results also align with experimental results when applying an ideal limiting
function to Gaussian noise. The figure below shows that results of WF11, our example
waveform, ideally saturated over a range of levels in comparison to a plot of (20) similar
to Table 3.

![Figure 9: Solution of integral (20) and ideal clipping of WF11](image)

[MATLAB\integral_def1.m] #114
The error between the true signal distribution, such as WF11, and the Rayleigh pdf is usually large enough to affect the problem outcome. The true amplitude distribution of an OFDM signal depends on the number of active carriers and the correlation between samples. For single-carrier modulation the distribution depends on the modulation order. For this reason, Monte Carlo evaluation using representative signals replaces probability density integrals for approximating integration over a signal with an exact probability density.

5.5.2 Parameters of Interest

Next consider the estimate of the analytic parameters directly from data, which can be achieved by a few simple methods. The parameters are the 1dB compression level (P1dB) \( \theta_1 \) and the input and output saturation level, \( \theta_{sat,in} \), and \( \theta_{sat,out} \). The value of the \( \theta_1 \) parameter itself is a point of interest in determining back-off when the saturation level is not known. Additionally the slope of the AM-AM curve at \( \theta_1 \) can generate a reliable upper bound of the saturation region. Figure 10 and Figure 11 show only the upper range of the AM-AM curves for two PA models and in each figure this bound is the blue line tangent to the PA AM-AM curve.

When \( \theta_{sat,out} \) cannot be measured safely, a quadratic fit over the data greater than \( \theta_1 \) could extrapolate the data to \( \theta_{sat,in} \) and give an estimate of \( \theta_{sat,out} \). This is shown in the black curves of Figure 10 and Figure 11, based on the “Sat Est data”.

45
The results for the ideal AM-AM models g8 and g7 are similar. Only the upper portion of the normalized range is shown. The data around $\theta_1$ and the corresponding linear fit is shown in blue. The data and fit at the peak are shown in red. The $\theta_1$ amplitude and the slope are similar between these models. The slope at the peak of the higher order model, g7, is lower, indicating a sharper transition to full saturation, as expected. Compare these two curves to PA model g11 shown in Figure 12.
These parameter estimates are used in the optimization.

Consider fitting a line to data of length \( n \) with additive noise, with the variance of the reference data being \( \sigma_x^2 \), and a noise variance \( \sigma_N^2 \). Then the variance of the estimated intercept was found to be

\[
\text{var}[\theta_b] = \frac{\sigma_N^2}{n}
\]  

(21)

The variance of the estimated slope was also found to be

\[
\text{var}[\theta_m] = \frac{\sigma_N^2}{\sigma_x^2 n}
\]  

(22)

If a fit over the peak range is used to estimate the peak magnitude or slope, then assuming the noise level is bounded, a desired confidence in the estimate is achieved by ensuring a sufficient sample size, \( n \).
5.6 Training design

An excitation signal must satisfy both limits of the amplifier while maximizing the exploration of the relevant space, in both amplitude and frequency. The pdf of all signals should be the same during training and operating phases. In [23] the author proposes that peak limiting should not be applied during a “single iteration” digital predistortion (DPD) in order to replicate the anticipated peak expansion of the PA input for characterization. For this optimization which is not adaptive, the only factor affecting the selection of the training data is the accuracy of the error estimate. The ideal estimator is identically the signal targeted for optimization.

5.7 Probability Density transformation

The probability density of the ideal transmitted signal is determined by the generating random process. In general the probability density of a sum of a large number of sinusoids of differing frequencies can be approximated as band-limited Gaussian noise. An important parameter of such a signal is the magnitude and frequency of occurrence of the signal peaks. This property is central to the analysis presented in this research.

Any processing of the signal is a modification of the probability density. For our purposes, the shape of the pdf in the upper range is most important. The nonlinearities introduced by the amplifier are most significant in this range. Additionally, the predistortion function has the greatest deviation at the highest magnitudes.
If a constraint takes the form of a magnitude limit, this limit defines a sub-set of the domain of the signal in the signal processing path. This sub-set may be mapped through the appropriate functions to determine the corresponding image or pre-image. For a magnitude bounded set, $S_T$, containing all values of an output signal, $y[n]$, the set of all values for which the magnitude is equal to $T_y$ (the boundary of $S_T$) has a pre-image through the PA function.

$$|g(S_b)| = T_y \quad (23)$$

The only necessary task is to refer a bound from one domain to another. If $g$ is a function of the magnitude of a single sample (memoryless) and $T_y$ is within the co-domain (output) of $g$, then the pre-image of $T_y$ can be calculated from

$$T_x = \arg\min_{x \in \mathbb{R}^+} [g(x) = T_y] \quad (24)$$

The argmin is only necessary because $g$ is only non-decreasing, not strictly one-to-one. If $g$ is a saturating function and $T_y$ is not in the co-domain of $g$ then $T_x$ does not exist and the set $S_y$ containing the output signal is

$$S_y = S_g \cap S_T = S_g \quad (25)$$

This simply asserts that any value, $y_n$, or set, $S_y$, at the output $g$ is in the co-domain of $g$.

For a given level of distortion, the method which results the lowest amplitude range is preferred for this problem. Methods of strictly limiting the pdf of the magnitude with minimum distortion are discussed in Section 7.2 on crest factor reduction.
5.8 Indirect Learning Architecture

The indirect learning architecture (ILA) for PD is first described in [24]. The method is a well accepted architecture for PD, cited in recent literature on the topic [25], [4]. The indirect learning architecture is derived from the definition of the inverse as a reflection over the identity relation, $x = y$. The graph of the inverse function can be produced by swapping the x and y axes so that y becomes the independent variable of the inverse function. A solution for determining the predistortion function using data is to fit a curve to the data set with the x and y vectors swapped.

The ILA diagram below (adapted from Figure 2) is followed by a mathematical description.

![Diagram of Indirect Learning Architecture](image-url)
Define a new variable, \( z = f(y|a^{[k]}) \), representing the predistortion function operating on the output signal. As the signal \( z \) approaches \( x \), so does the PA output \( y \) approach the reference signal \( w \) [24].

\[
x = f(u), \quad z = (f \circ g \circ f)(u)
\]

If the difference between \( z \) and \( x \) is defined as \( e \), the least-squares (LS) cost function to be minimized is

\[
J_{LS}(x^{[k]}, z^{[k]}) = \sum_{n=1}^{N} |e_n|^2 \approx E[|e|^2] \tag{26}
\]

This assumes that the functions PD and PA can commute (be applied in either order), which is valid only if the functions are true inverses (see §5.3). This is not the case for non-linear functions with memory nor when the domains of the functions are not identical. Consequently the error \( e \) is generally not equivalent to the error \( d \) between the output and the reference signal

\[
MSE(x, z) \neq MSE(w, y)
\]

There are two theoretical cases for any PD architecture, the distinguishing factor being the range of the set of values of the signal \( y \). In the first case, considered ideal, the domain of \( u \) is a subset of the relative co-domain of the PA function \( g \). In this case, if \( g \) is a one-to-one mapping, there is a potential for a complete linear mapping of the domain of \( u \) onto the co-domain of \( g \). With the appropriate PD, the minimum error between \( y \) and \( u \) approaches zero. In the second case, with saturation, the relative co-domain of \( g \) is a subset of the domain of \( u \) and consequently there is no PD for which the error between \( y \)
and \( u \) approaches zero. The second case can be translated to the first case by increasing the size of the relative co-domain of \( g \). This means scaling \( x \), which is equivalent to decreasing the input signal level to the PA. The same is true of any component in the signal chain with a finite range, such as a digital-to-analog converter.

The signal \( y \) has a measurable range. If \( y \) spans the co-domain of \( g \), this is the range to which the linearized signal must be mapped. Since the predistortion in the signal path is copied from an identical model operating on \( y \). It is necessary for \( u \) and \( y \) to have the same probability density so that the difference between \( x \) and \( z \) can be minimized. This can be achieved by estimating the range of \( y \) and actively limiting the signal \( w \) to this range at the input of the PD.

#### 5.9 Solution Basis

The unconstrained least-squares formulation of the parameter estimation problem is one of several possible foundations for a solution. The objective function is quadratic. The minimization estimates the globally optimal estimate of the parameter vector, \( a^* \), from the data.

The vector \( a \) is the coefficient vector of the memory polynomial predistortion model described in [4].

\[
x[n] = \sum_{k=1}^{K} \sum_{q=0}^{Q} a_k \ast u[n - q] \ast |u[n - q]|^{k-1} \quad a_k, u[n], x[n] \in \mathbb{C}
\]

For the memory-less model, \( Q = 0 \), and the model reduces to:

\[
x_n = u_n \ast \sum_{k=1}^{K} a_k \ast |u_n|^{k-1} \quad a_k, u_n, x_n \in \mathbb{C}
\]
The polynomial defined above is a linear combination (by complex numbers) of basis functions. The matrix form of the equation above is \( \mathbf{x} = \mathbf{V} \ast \mathbf{a} \) with \( \mathbf{V} \in \mathbb{C}^{L \times K} \) constructed from L samples of \( u \).

The maximum-likelihood estimation of \( \mathbf{a} \) for the indirect learning architecture is given in [4]:

\[
\text{minimize } E[\|\mathbf{e}\|^2_2]
\]

\[ e = \mathbf{B}a - \mathbf{x} \]  

The estimated parameterized function is not the PA but rather the inverse of PA, so the observation vector is constructed from the PA input signal \( \mathbf{x} = [x(0) \ x(1) \ldots x(n-1)]^T \). The data matrix \( \mathbf{B} \) is the basis vectors constructed from the sample vector \( \mathbf{y} \).

\[
b_{kq}(n) = y(n - q)|y(n - q)|^{k-1}
\]

\[
\mathbf{b}_{kq} = [b_{kq}(0) \ b_{kq}(1) \ldots b_{kq}(N - 1)]^T
\]

\[
\mathbf{B} = [\mathbf{b}_{10} \ \mathbf{b}_{20} \ldots \ b_{kq}]
\]

The LS solution of the normal equations, \( \mathbf{B}^T \mathbf{B}a = \mathbf{B}x \), is:

\[
\hat{\mathbf{a}} = (\mathbf{B}^T \mathbf{B})^{-1}\mathbf{B}x
\]  

By constructing \( \mathbf{B} \) from samples of \( y \), the matrix is corrupted by the additive noise contained in \( y \), but the advantage is that the probability distribution of \( y \) is represented in the data. An alternative, not explored here, is to construct \( \mathbf{B}^T \mathbf{B} \) from explicit basis functions over the known autocorrelation of the expected output signal. This approach is described and used successfully in [6].
5.10 Multi-phase Verification

Verification of the quality of a PD solution involves measurement of system output while the solution is applied. It is not sufficient to measure the performance from only the original data set that was used for the calculation.

The typical procedure is to calculate a predistortion function to be applied to the signal during the subsequent period. Consequently, data collected during that period is from the signal modified by the current PD. This assumes that the signal \( y \) is close to \( u \).

There are multiple classes of data sets. The PA input may or may not have PD applied. Additionally, as described in [26], the reference may be ideal or transformed, such as by crest factor reduction. The predistortion function changes the pdf of the function, so the only valid measurement of the output is when the signal is stable in its linearized transformation.

A naïve verification is as follows. Calculate a PD from a dataset with a normal or ideal reference. Apply the PD to the dataset’s output signal and measure the resulting signal. This is equivalent to exchanging the order of application of the PD and PA functions. Even if these operations were commutative (and they usually are not), there is a second problem with this approach. The solution is dependent on the probability density of the input signal. In order to excite the relevant range of the PA, the input must have PD applied. The naïve verification does not have this property. Since the PD usually expands the peaks of the signal, an extended range of the amplifier is explored when driven with a PD signal. This is especially important when operating in the saturation region, because the extended range introduces higher-order non-invertible distortion.
Notice that the minimization of the error in the indirect learning architecture uses the naïve approach as well. The error is minimized over the image of $y$ through the function $f(y|a^{[k+1]})$

As $f$ converges, so does the output of the PA converge to the image of $g(f(u|a^{[l]}))$.

$$f(y|a^{[l+1]}) = G(f(u|a^{[l]}))$$

$$y = G(f(u|a^{[l]}))$$

So when does

$$f(G(f(u|a^{[l]}))|a^{[l+1]}) = G(f(u|a^{[l]}))$$

The general minimization is

$$f(y|a^{[l+1]}) - f(u|a^{[l]}) = e$$

But there is no restriction on the domain of $f$ or on the pdf of $u$ and $y$. Here, a minimization of $e$ does not result in $y$ being equal to $u$.

A condition necessary for the output to equal the reference is when

$$f(y|a^{[l+1]}) = f(u|a^{[l]}) \Leftrightarrow a^{[l+1]} = a^{[l]} = a^*$$

(29)

which is the limit for global optimality given that $u$ is stationary. Restrictions on the domain of any of these functions may cause this condition to be unachievable. Optimality can still be defined in those cases.

Prior to exploring and comparing solution algorithms using models of the system, the models themselves must be verified to be correct or at least to exhibit the correct
behavior. This is achieved by comparing the behavior of the model to the behavior of a physical system under similar stimulus.

5.11 Optimization

Optimality is the property of a control vector which minimizes a function, commonly known as a cost function, while satisfying all declared constraints. Some constraint sets are naturally defined in the vector space of the signal while other constraint sets are defined in the function space.

5.11.1 Performance Constraint

There are several ways to arrive at the feasible set. Consider initially only the MER requirement on the final output signal. The relevant constraint in a communication system is usually expressed as a limit on the expected value of the MER, but occasionally as a limit on the minimum (worst case) over a duration or number of symbols. Since the estimated MER is always required to exceed some minimum performance level, this would become an inequality constraint in the problem formulation.

The expected value must be approximated over a data set. The calculation of MER (§3.2.2) is a mapping from a time-domain sequence to a single figure of merit, \( MER: \mathbb{C}^N \rightarrow \mathbb{R} \). Depending on the modulation type, there may be restrictions on \( N \), the required length of the sequence.

There exists a set of length-\( N \) vectors that map to a given MER value, \( M \), within a tolerance, \( \epsilon \).

\[
S_M \subseteq \mathbb{C}^N \quad MER: x \rightarrow M \pm \epsilon \quad \forall x \in S_M
\]
The ideal symbols are discrete but \( \epsilon \) is not, so theoretically \( S_M \) is a closed continuous set (not countable) and there is a closed pre-image of the amplifier function \( g \) for this set.

\[
g(P) = S_M
\]

The PD output should then be contained within the set \( P \).

Two simplifying approximations make the task of satisfying the constraint tractable.

1.) This metric concerns the expected deviation rather than the entire population of \( S_M \). So equivalently we may consider the set of all error distributions referenced from an explicit representative vector.

2.) Approximate MER by SNDR (“SNR”) and evaluate SNDR from normalized mean square error (NMSE).

The set \( S \) is a set of vectors produced by a random process. Of interest is the deviation of a vector in \( S \) from its reference vector. A reference vector, \( w \), is defined as a vector with zero distortion. The distortion of a signal \( y \) relative to its reference \( w \) is \( NMSE = 1/SNDR \). Therefore \( w \) can be defined by:

\[
\lim_{y \to w} SNDR(y) = \infty
\]

From the MSE definition (9), the normalized MSE estimate is

\[
NMSE(y|w) = \frac{1}{N} \sum_{n=1}^{N} \left\| \frac{y_n}{\sigma_y} - \frac{w_n}{\sigma_w} \right\|^2
\]

(30)

Since the signal energy is unity, i.e. \( \frac{1}{N} \sum_{n=1}^{N} w_n / \sigma_w^2 \) = 1, the estimator of SNDR is

\[
SNDR(y|w) = \frac{1}{\frac{1}{N} \sum_{n=1}^{N} \left\| \frac{y_n}{\sigma_y} - \frac{w_n}{\sigma_w} \right\|^2} = \frac{1}{NMSE}
\]

(31)
Performance Constraint #1

\[ SNDR(y|w) \geq \Omega \]

From the constraint mentioned above, only the sum of all error is bounded, equivalent to a norm ball in N-dimensional signal space. There is no further restriction on the direction of the error vector and thus on the distribution of the error. This degree of freedom is important in this problem, as the distribution of the error is determined in part by the PD function and the allocation of this error budget is to be optimized. The constraint value \( \Omega \) may be determined from a known approximation.

A secondary performance constraint is the amount of distortion allowed outside of the channel bandwidth. This is often defined as a mask on the transmitted power spectral density or as a limit on the spectral leakage or the amount of power generated in an adjacent bandwidth (ACPR). Depending on the exact specification, an ACPR limit is generally slightly more flexible, in much the same way as the SNR limit - the ACPR limit allows a degree of freedom in the frequency domain in allocating the distortion over the band. Non-linear distortion is the dominant contributor to adjacent channel distortion when the in-band performance constraint is tight. The relationship between the in-band SNR and the out-of-band ACPR can be altered by means of frequency shaping. A simple linear filter can discriminate between the frequency bands and vary the distortion levels independently. Filtering the PD modifies the linearization. A nonlinear function followed by a linear system is a Hammerstein system, which is a valid form for a PD with memory, as long as the model is estimated as a whole, as in [27].
5.11.2 Other Constraints

Beyond performance constraints, some other constraints arise directly from known or estimated physical properties of the system, e.g. the PA saturation level. Other constraints may be negotiable in the problem for the purpose of considering their impact or reducing the dimensions of variability, i.e. dynamic range of coefficients. The study of constraints in complex systems is important in design. Often one seeks to answer a question about the impact of the variation of a parameter on the rest of the system. If a relation between the parameter and the complete state of the system is already known, the answer can be calculated. If instead the parameter to be varied is a constraint, there may not be an established relationship.

Constraints are a restriction on the set of feasible solutions. The term restriction has a negative connotation, but there are often advantages. Restricting the set of solutions reduces the “size” or changes the “shape” of the feasible set. Solution methods based on searching can exploit this change. Constraints representing bounds of stability or predictability must be explicit so as to avoid those adverse conditions.

Constraint #2

The other constraint studied in this work is on the PD function itself. Consider the contrived requirement: “For safety reasons, you must guarantee that regardless of the average operating level, the peak input level must never exceed X”. When the magnitude of the PA input must be bounded independently, then the PD co-domain must be bounded. One method is to apply the limiting function to the output of the PD. Another possibility is to constrain the PD function (having a bounded domain) so that the co-domain satisfies the bound.
An implementation constraint for simulations or systems with explicit back-off is unity gain of the PD function. With explicit back-off, the PD is not responsible for changing the back-off and must be constrained to a fixed gain. Furthermore, since $f$ is a nonlinear function, the relationship between the parameters is dependant on the scaling of the input and output. The level of the PD output signal, defined as the root mean square or standard deviation, $\sigma_x$, must equal the PD input level. The level of the (zero-mean) PD output $x$ is also given by

$$\sigma_x = E \left[ \left( f(U|a) \right)^2 \right]^{1/2}$$

(32)

where $U$ is a r.v. representing the PD input signal.

Since the level of the PD input is constant, i.e. $\sigma_u = E[U^2]^1$, then $\sigma_x$ should be a constant to meet the constraint. This equality constraint is dependent on the probability distribution of the input signal $u$ and can only be expressed as an integral. Approximating the constraint with a predetermined fixed point of the function converts the constraint to a linear equality.

$$f(x_{RMS}|a) = (a, \kappa(x_{RMS})) = x_{RMS}$$

(33)

For the LS solution in the ILA, the constraint is met “implicitly” by scaling the data sets to the desired level. An alternate method of meeting this constraint is given in §6.5.

Other potential constraints not studied here are:

1.) The power variation from one symbol to the next.

2.) The derivative of the predistortion AM/AM is bounded above. (Less than $1/m$)

3.) The derivative of the PA function is bounded below. (Greater than $m$)
5.12 Optimization Formulation

The relationship between operating efficiency and output power has already been shown, leading to a goal of maximizing the output power. A further assumption is that the average output power of the amplifier is a non-decreasing function of the average input power. With this assumption, the task of maximizing the output power is equivalent to maximizing the input power. The first formulation, given in Problem 3 below, uses the implicit back-off model.

The real challenge in the optimization is estimating the MSE over the parameter space for the unknown non-linear PA function. The following analysis is performed with a known PA function. This analysis uses simulation to determine the MSE surface. In [28] the author was able to derive a closed form expression for SNDR for two cascaded low-order non-linearities

5.12.1 Problem 3: MSE Constraint

This section describes Problem 3, including a Formulation and several Examples.

The set of basis functions for a polynomial predistortion function operating on complex signals is

\[ f_N(x) = \{f_1, f_2, \ldots, f_N\} = \{x, x|x|, \ldots, x|x|^{N-1}\} \quad \forall x \in \mathbb{C} \]

The predistortion output is

\[ x_n = f_{PD}(u_n|a) = \sum_{k=1}^{N} a_k f_k(u_n) = \langle a, f_N(u_n) \rangle \quad a_k \in \mathbb{C} \]

An estimator of the signal level, \( \sigma_x \), is the sample standard deviation, given as a function of the PD function parameters
The variance $\sigma^2$ of a non-linear transform of a random variable is not a linear function of the variance of the random variable. Consequently $\sigma_{x_{PD,out}}$ can only be known from (32) and cannot be inferred from (33). However, because $(g \circ f^*)(u)$ is approximately a linear mapping under linearization, the scaling of a single value, $\sigma_u$, is proportional to the gain over the entire linearized range of the distribution. Using the approximation (33) of $\sigma_{x_{PD,out}}$, the signal level is an inner product of the parameter vector with a constant vector.

$$s_M(f_{PD}(u|a)) = \sqrt{\frac{1}{M} \sum_{j=1}^{M} \left( \sum_{k=1}^{N} a_k f_k (u_j) \right)^2}$$

Example 3-1

Assume the function $g$ (referenced as “n3”) is known to be $g_{n3}(x) = x \left( 1 - \frac{1}{3} |x|^2 \right)$, so the parameters of this model are $p = [1 \ 0 \ -\frac{1}{3}]$. The input and output saturation levels are determined from the critical point of the curve.

$$\frac{\partial g_{n3}(|x|)}{\partial |x|} = 1 - |x|^2 = 0 \Rightarrow \theta_{sat,in} = |x| = 1, \ \theta_{sat,out} = g_{n3}(x_{sat,in}) = \frac{2}{3}$$

The resulting expression for the PA output, $y$, is

$$y_n = g(f(u_n)) = f(u_n) \left( 1 - \frac{1}{3} |f(u_n)|^2 \right)$$
Assume that the PD model is \( f(u_n) = u_n(a_1 + a_3|u_n|^2) \), and the signal \( u \) is the reference waveform WF11 (§4.1) generated by a random process. Since \( NMSE(u, y) \) is a deterministic function of the parameters \([a_1, a_3]\), the function can be plotted over the parameter space. The dimension of the space is \( \mathbb{R}^2 \times \mathbb{R} \subseteq \mathbb{R}^3 \) and so it can be visualized as in Figure 14. For all such visualizations in Chapter 5 and 6, the surface is SNDR in logarithmic (dB) scale, and the values are computed by \( 10 \times \log_{10}(1/NMSE) \).

For this example, the IBO parameter is not determined independently. Rather the parameter values determine the IBO and the OBO. The parameter \( a_1 \) is the linear scaling parameter. The diagonal blue lines in Figure 14 are lines of constant OBO in \([dB_{sat}]\). A vector normal to these lines is the direction of increased or decreased signal level in the linear parameter space. Further analysis of this and similar surfaces is reserved to subsequent sections.
In Example 3-1, the function $g$ is known exactly. We can proceed to formulate an optimization for this example. However a practical algorithm would not complete without a method for estimating an unknown function $g(\cdot)$.

The objective is to minimize the back-off, and by the approximation of signal level given in (34), this is equivalent to a linear function of the parameters.

$$\min \beta \iff \max \sigma_{PA,in} \iff \max b^T a,$$

$$b = [f_1(\sigma_U)\ f_2(\sigma_U)\ ...\ f_N(\sigma_U)]^T$$

**Formulation 3**

$$\max b^T a$$

subject to

$$\frac{1}{L} \sum_{i=1}^{L} \left( \frac{\sigma_U}{\sigma_Y} y_i - u_i \right)^2 = \frac{1}{L} \sum_{i=1}^{L} \left( \frac{\sigma_U}{\sigma_Y} g \left( \sum_{k=1}^{N} a_k f_k (u_j) \right) - u_i \right)^2 \leq \frac{1}{\sigma_U \Omega}$$

Formulation 3 has a linear objective with a non-linear constraint. This is a direct optimization because the PA function must be known (estimated prior to the optimization). Conceptually, an algorithm to solve this formulation would trace out the SNDR surface at the constraint level over the parameter space, and find its intersection with the half-space defined by $b^T a$. The optimal point would be the tangent of the $b^T a$ hyperplane to the SNDR set.

**Example 3-2**
Identical to Example 3-1, with the modification: Assume the function \(g\) (referenced as “n5”) is known to be \(g_{n5}(x) = x \left( 1 - \frac{1}{5 \times 3} |x|^2 - \frac{4}{5 \times 5} |x|^4 \right)\), so \(\mathbf{p} = \left[ 1 \ 0 \ -\frac{1}{15} \ 0 \ -\frac{4}{25} \right]^T\).

Figure 15 and Figure 16 are the surfaces with 3\(^{rd}\) order and 5\(^{th}\) order PD. The minimum back-off with \(\Omega=37\)dB is slightly below 8dB for both PD models.

![Figure 15: Example 3-2 n5 1/NMSE contours](Matlab\NonLinear\topo_2.m)

![Figure 16: Example 3-2 n5 1/NMSE contours](Matlab\NonLinear\topo_2.m)

5.12.2 Problem 4: Peak-limited PD Output

This problem is identical to Problem 3 with the addition of constraints on the peak magnitude of the PA input. This constraint is new and separate from any constraints applied so far. It is used to satisfy Constraint #2, discussed in §5.11.2.

Consider an implementation of digital predistortion (DPD) and a representation of discrete signals by complex values with maximum magnitude \(T_d\). Also consider an
absolute maximum peak level $T_p$ to satisfy Constraint #2. Since these two thresholds apply to the same signal (the PD output), only the smaller of the two is active. In the formulation below, the peak of the PD output is represented as the p-infinity norm of the PD output treated as a vector. This vector is the product the PD parameters $\mathbf{a}$ and the PD basis matrix $\mathbf{V}$ formed from the PD input.

**Formulation 4**

\[
\text{maximize } b^T \mathbf{a} \\
\text{subject to} \\
\frac{1}{L} \sum_{i=1}^{L} \left( \frac{\sigma_u}{\sigma_y} y_i - u_i \right)^2 \leq \frac{1}{L} \sum_{i=1}^{L} \left( \frac{\sigma_u}{\sigma_y} \left( \sum_{k=1}^{N} a_k f_k(u_i) \right) - u_i \right)^2 \leq \frac{1}{\sigma_u \Omega} \\
\|\mathbf{V} \mathbf{a}\|_{\infty} \leq B = \min(T_d, T_p)
\]

where $T_d$ is the digital implementation threshold and $T_p$ is the peak-to-average threshold.
6 Solution Methods

The class of problems presented in the previous chapter and specifically Problem 3 and Problem 4 may be solved by various means. This chapter describes several solution methods that are appropriate for solving one or more of the Problems. The first analytic solution method is a theoretical reference. The Least-Squares Search (LSS) and the Constrained Solver are fully developed algorithms, and their results are given in Chapter 7. Additionally there are discussions on using a general Optimization Solver and on a partially developed method - Projection and Gradient Descent.

6.1 Analytic Solution for Peak Limited PA

This method is presented for theoretical reference. Assume Problem 3 is to be solved for an unknown memory-less PA, and the feedback, \( y \), is noiseless. The bound on performance under ideal linearization is given by (20) in §5.5.1. This bound is a mapping between a magnitude-limited linear range and expected distortion. In Problem 3, a limit on distortion is provided as an inequality constraint. Converting this to an equality constraint and applying the inverse of the mapping of (16) gives the minimum linear range. This range is the target OBO with ideal linearization. Assuming the amplifier is ideally linearized by predistortion, the system could be operated at the prescribed OBO with the specified distortion level.
A drawback of this approach is that the bound is waveform-specific, so it must be pre-computed for each waveform. Also, the inverse of the bound function, required in the first step, is not a direct analytic equation.

### 6.2 Least Squares Search Algorithm (LSS)

This algorithm is suitable for solving Problem 3 with an unknown amplifier in the ILA. The only constraint in this problem is the MSE constraint. The back-off parameter is explicit while the PD gain is normalized to unity (§4.5). The algorithm follows the
justification in §5.10. The search algorithm is an iterative process consisting of these stages:

1.) Create a data set from the model using the current parameters
2.) Measure the estimated performance of the current solution from the data
3.) Calculate a solution for the PD parameters from the data set.
4.) Update the back-off parameter for the next iteration.

The normal equations of the ILA and the analytic solution are given in §5.9. The block LS estimation of \( \mathbf{a} \) involves the inversion of the Hermitian positive-semidefinite covariance matrix \( \mathbf{B}^T \mathbf{B} \). The matrix has a high condition number for the memory polynomial. Also the signal samples used to construct \( \mathbf{B} \) are corrupted by additive noise. Following the justification given by Cadzow in [13] Theorem 3, the nearest positive semi-definite matrix of a given rank can be found by spectral decomposition. Singular value decomposition (SVD) is used to solve (28), so it is convenient to retain only the largest singular values within a pre-determined ratio of the maximum singular value \( \sigma_{max} \). In the simulations, this ratio threshold is \( 10^{-8} \).

The back-off parameter is updated as a step from the current value. The direction and size of the step are determined from the estimated performance. The constant \( \Omega \) is the specified absolute performance limit for the problem. This performance limit must be determined as an input to the algorithm. If the estimated performance is higher (better) than the limit, the step direction is negative and the amount of back-off decreases. If the performance is lower than the limit, the step direction is positive and the back-off increases. The size of the step is proportional to the difference between the current
performance and the limit. As the estimated performance approaches the limit, the step size also decreases.

$$\beta^{(k+1)} = \beta^{(k)} + \tau_{\text{step}} \ast (\Omega - \hat{\theta}_{\text{SNDR}})$$  \hspace{1cm} (36)

The stopping criterion is determined by a threshold on the magnitude of the step.

$$\beta^{(\text{final})} = \beta^{(k)} \quad |\beta^{(k+1)} - \beta^{(k)}| < \epsilon_{\text{stop}}$$  \hspace{1cm} (37)

Equation (36) is not a normal gradient descent equation, based on the gradient of the objective function with respect to the parameters. Rather this algorithm uses a fixed direction and is only valid where the SNDR decreases with decreased back-off.

A diagram of the LSS algorithm process follows.
Algorithm 2 (LSS)

Set IBO, $\beta^{[0]}$, to $\sim 20\text{dB}$
Set $\gamma = 20\text{dB}$
Set $f_{PD}^{[0]}(u) = u$

Set $\sigma_x$ to achieve $\beta^{[l]}$

Collect data set $\{w, x, y\}^{[l]}$

Calculate $f_{PD}^{[l]}$ from data set $\{x, y\}$ by LS (28)

Apply $f_{PD}^{[l]}$

Collect data set $\{w, x, y\}^{[l]}$

Estimate $SNDR(w, y)$

Estimate $\tilde{\theta}_{sat,out}$ from $y$ by method of §5.5.2

Set $\gamma = \tilde{\theta}_{sat,out}$

Update $\beta^{[l+1]}$ according to (36)

$\varepsilon$-stop criteria (37)

$[\beta^*, f_{PD}^*, \gamma^*]$
6.3 Optimization Solver

In order to use a general optimization solver, a problem should be converted to the nearest standard form. The standard form provides a classification of the problem for the purpose of selecting an appropriate solver. Formulation 3 has a linear objective and therefore both convex and concave. This objective is one of the simplest forms in $\mathbb{R}^n$, specifying a uniform direction. If the constraints were linear, the solver would be a linear program (LP). Unfortunately, the constraint of Formulation 3 is not simple because it has none of the desirable properties for a well-defined solution method.

$$\frac{1}{L} \sum_{i=1}^{L} \left( \frac{\sigma_u}{\sigma_y} y_i - u_i \right)^2 \leq \frac{1}{\sigma_u \Omega}$$

Even with the simplest PA model, $g(x) = x \ast (p_1 + p_3 |x|^2)$, the constraint function is a higher-order non-linear function of the parameter vector, $\mathbf{a}$.

$$f_1(a) = \frac{1}{\sigma_u \Omega} - \frac{1}{L} \sum_{i=1}^{L} \left[ k_n \ast f_i^{T} \mathbf{a} \ast \left( p_1 + p_3 |f_i^{T} \mathbf{a}|^2 \right) - u_i \right]^2 \leq 0$$

Also, by the non-convexity of realistic PA functions, as explained in §4.3, and by inspection of the super-level sets of 1/NMSE as in Figure 16, the constraint set is not convex over all of the parameter space, $\mathbf{a} \in \mathcal{A} \subset \mathbb{C}^N$. Formulation 3 cannot be solved as-is by the “CVX” optimization suite [29], because it cannot be classified within the rules of Grant’s disciplined convex programming [30]. There are methods for solving such non-linear, non-convex, constrained problems, as cataloged and theorized by Bertsekas [31] [32] and others. These methods are considered to be among the most
computationally demanding in the field. No results are presented herein, but future work with such methods is discussed in Chapter 8.

### 6.4 Constrained Solver

To prove that Problem 4 could be solved, Algorithm 3 was developed as a modification of the LSS algorithm (Algorithm 2). Unlike §6.3, the constrained optimization solver is only used to calculate the constrained LS parameter estimate. The problem is conveniently specified and executed using “CVX” [29], a state-of-the-art optimization suite and the optimization solver is Sedumi which runs in Matlab. The methods behind the solver are not presented here.

```
01 cvx_begin
02 variable p(n) complex;
03 minimize( norm(Y*p-x, 2) );
04 subject to
05 norm(Theta2*p,inf) < MAX_Theta_p; %1.1;
06 cvx_end
```

**Constrained LS Solver**

The statements above describe the LS approximation of $\mathbf{a}$ over an ILA data set. The peak magnitude constraint of Problem 4 is specified on line 5.

**Algorithm 3**

The iterative process is identical to Algorithm 2 with a change in the calculation of the PD function.
6.5 Projection and Gradient Descent

This section explores an alternative method of solving Problem 4, using projection onto convex sets (POCS) and gradient descent. This section does not include a working algorithm, but it does offer some analysis necessary for developing an algorithm.

While it is difficult to graphically represent a high-dimensional parameter space [33], the following plots show a surface for two-dimensional parameter space with other parameters fixed. The examples are analyzed with explicit back-off.

Example 5

\[ f_{PD}(x|a_3) = x \times (1 + a_3|x|^2) \]

\[ g_{PA}(x) = G8 \]

minimize \( \beta \)

subject to

\[ NMSE(w, y) \leq \frac{1}{\sigma_w} \]

The 1/NMSE surface for Example 5 is shown in the figure below.
Example 6

\[ f_{PD}(x|a) = x \cdot (1 + a_3|x|^2 + a_5|x|^4) \]

\[ g_{PA}(x) = GB \]

\textit{minimize} \( \beta \)

\textit{subject to}

\[ NMSE(w,y) \leq \frac{1}{\sigma_w} \]

Consider the coefficient parameter space of Example 6. Observe the change in the surfaces of Figure 18 and Figure 19 as the back-off is decreased. The maximum achievable performance decreases and the region representing 36dB decreases in size and shifts in location.
Figure 18: Example 6 performance surface 1/NMSE [dB] over two coeff dimensions at a fixed IBO=11dB
[Matlab\NonLinear\param_surf_a3_a5_01.m] #103

Figure 19: Example 6 performance surface 1/NMSE [dB] at IBO=10.2dB
[Matlab\NonLinear\param_surf_a3_a5_01.m] #104

Figure 20: Example 6 expansion, IBO=11.0dB
[Matlab\NonLinear\param_surf_a3_a5_expansion_01.m] #107

Figure 21: Example 6 expansion IBO=10.2dB
[Matlab\NonLinear\param_surf_a3_a5_expansion_01.m] #106
The surfaces are plotted over a transformed space in Figure 20, Figure 21, and Figure 22. The vertical dimension is the expansion (or gain) of the PD function at the reference limit level, $\gamma$. The constraint on the PA input which was introduced in Problem 4 is equivalent to a limit on the expansion. If the expansion were limited to 2dB, then the best possible SNDR performance is decreasing as the IBO (and thus OBO) is decreasing. Also, the area of the set in parameter space for the highest SNDRs is decreasing as the IBO decreases.

### 6.5.1 Proposition 1

The following proposition motivates the development of the optimization framework. The statement is followed by two methods of proof.
Proposition 1

An algorithm to estimate a parameterized PD function achieves the optimal solution when all constraints are satisfied and the cost is minimized. When the LS estimate of the PD function parameters lies outside the feasible set, the LS search algorithm (LSS) may not achieve the optimum.

Proposition 1 implies that the LSS is potentially useful, but one must first calculate the LS solution and determine if it is feasible. This does not have the appeal of an algorithm that works for much broader conditions, but it has the advantage of simplicity when applicable.

Since Proposition 1 contains the weak assertion that LSS may not achieve the optimum, a minimal proof consists of a single example for which the assertion holds, assuming the true optimum is provably so. Now consider a non-rigorous graphical Proof 1A of Proposition 1.
Proof 1A

This is non-rigorous explanation for Proposition 1, using Example 5 - the PA and simple PD model used for Figure 17. Consider the performance constraint $SNDR \geq 30dB$. In Figure 23 the LS estimate of the PD parameters is shown along the blue line. Using the LS solution, the optimal backoff, $\beta = 6.7dB$, occurs at the intersection of the line of LS solutions and the level set in parameter space where $SNDR = 30dB$. Next consider an additional bound, $|a_3| \leq 0.5$, represented by the vertical dashed line in Figure 23. The LS optimal solution violates this constraint and is therefore infeasible. If solutions are restricted to those calculated by LS, the optimal solution lies at the intersection of the LS solutions and the hyperplane $a_3 = 0.5$, where $\beta = 7.3dB$. Clearly in this simple example there exist feasible solutions with lower cost, the lowest
being $\beta^* = 6.75dB$. This is an example of Proposition 1. The important realization is that these solutions cannot be found using only the analytic LS estimate. This example only supports Proposition 1 because the constraint on $\alpha_3$ is active, meaning that the constrained solution is different than if this constraint did not exist. This concludes Proof 1A.

Proof 1A does not consider other types of constraints, either in the parameter space or on functions of the parameters. First, consider that the problem in Proof 1A may be partially solved by projecting the parameter vector onto the feasible set. The bound on $\alpha_3$ describes a closed halfspace and the projection $P_B(\xi_0)$ shifts the point $\xi_0$ so $\alpha_3 = 0.5$. The next explanation of Proposition 1 is carried out analytically using optimization theory.

Proof 1B

This is a non-rigorous explanation for the weak assertion of Proposition 1. For Example 5, consider the point $\xi_C$ to be the intersection of the LS solutions and the hyperplane $\alpha_3 = 0.5$. Since $SNDR(\xi_C) > 30dB$, the point satisfies the SNDR constraint and is located in the interior of the closed set. Therefore, at $\xi_C$ there is a gradient and a set of directions of descent of the SNDR that approaches the boundary. If the set contains any direction with a component decreasing in the direction of the optimization variable, then $\xi_C$ is not optimal and a better solution exists that satisfies the constraint.

$$\frac{\partial SNDR(\xi_C)}{\partial \beta} = \frac{\partial SNDR(w, g(f(w|0.5)))}{\partial \beta}$$

(38)
As $\beta$ decreases, the bounded support of ideal $g'(x)$ decreases (with $\sigma_x$ normalized), and the range of the bounded co-domain of $g$ decreases. With fixed $x = f(u|0.5)$ and $\Pr[X > x_{sat}(\beta - \Delta \beta)] > 0$, the error between $u$ and $y$ will increase. This condition on the distribution of $X$ is met whenever peak amplitudes extend to saturation, providing sufficient examples supporting Proposition 1. This concludes Proof 1B.

Due to the low dimension of the parameter space of Example 5 and the simplicity of the constraint, a solution method based on the Lagrange dual problem is straightforward to construct. However, without an expression for the function $g$ with respect to $\beta$, the dual optimization problem does not allow for a solution.

### 6.5.2 Unity Gain Constraint

Now consider a method of satisfying the PD unity gain constraint described in §5.11.2. The solution set $S_a$ of the equality constraint (35) is a hyperplane in the parameter space. Example 7 below demonstrates the appropriate projection technique.

**Example 7**

Using the PD function of Problem 2 and $x_{RMS} = 0.25$, $S_a := \{a \mid a_1 \ast 0.25 + a_3 \ast 0.0156 + a_5 \ast 0.0010 = 0.25\}$. The normal vector is $n = [0.25, 0.0156, 0.0010]^T$ with $\|n\| = 0.25049$.

The projection onto the hyperplane $S_a$ is given by (39) [20]

$$P_H(a_0|n, 0.25) = a_0 - \frac{(a_0, n) - 0.25}{\|n\|^2}n \tag{39}$$

With $a_0 = [1.0 \ 0.1 \ 1.0]^T$, the projected result is
If the constraint as expressed in (33) is part of the problem formulation, then other methods of satisfying the constraint are possible.

### 6.5.3 Peak Magnitude Constraint

One way to ensure that the co-domain satisfies a magnitude bound is to project the entire co-domain onto the bounded set. It is not sufficient to bind a sub-set of the co-domain, such as a single amplitude, because other regions of the set may violate the bound. This can be accomplished in practice by projecting a set of samples representative of the co-domain onto the bounded co-domain. The peak magnitude of the PD input may be constrained independently by another method. This method only controls the co-domain of the PD over its domain.

The constraint is expressed in Problem 4 as

\[
\| V \mathbf{a} \|_\infty \leq B = \min(T_d, T_p \sigma_x)
\]  

(40)

To satisfy (40), every sample through the linear PD mapping must lie within the bound.

\[
\langle \mathbf{a}[i], \mathbf{n}_i \rangle \leq B \quad \forall \ i \in \{1 \ldots L\}
\]

where \( \mathbf{n}_i \) are the row vectors of \( V \).

This is accomplished by the series of projections onto the closed half spaces

\[
\mathbf{a}[i+1] = P_H(\mathbf{a}[i], \mathbf{n}_i, B) = \begin{cases} 
    \mathbf{a}[i] & \langle \mathbf{a}[i], \mathbf{n}_i \rangle \leq B \\
    P_H(\mathbf{a}[i]) & \text{otherwise}
\end{cases}
\]
This series of projections will meet the constraint of (40) but may alter the corresponding curve to an invalid shape. For a PA operating into saturation, the corresponding PD function is monotonically increasing over the domain of the function. However, for a polynomial to mimic a magnitude limiting function, the function should be monotonically increasing up to a limit and then begin decreasing beyond the critical point. Can a polynomial of the original monotonic form be transformed to the limiting form? To answer this, we must analyze the two forms. Consider the PD function of Example 6, with $a_1 > 0$.

$$\lim_{u \to \infty} f(u|a) = \begin{cases} \infty & a_5 > 0 \\ -\infty & a_5 < 0 \end{cases}$$

If the sign of $a_1$ and $a_5$ are opposite, the derivative $f'(u)$ will have at least one positive and one negative real root. If the sign of $a_1$ and $a_5$ are the same, the derivative may have only imaginary roots. If $f'(u)$ does have a positive real root, then there is a critical point in the positive domain of the function. The conclusion is that the transform performed by the projection sequence must force the parameters $a_1$ and $a_5$ to have opposite signs. Equivalently, solutions are constrained to the second and fourth quadrant in the partial parameter space $\mathcal{P} := \{a_1, a_5\} \subset \mathbb{R}^2$.

Finally, note that because of the presence of the absolute value in the PD function, there are actually two separate constraints.

$$\|Ya\|_\infty \leq B$$
\[ \|Ya\|_\infty \geq -B \iff \|-Ya\|_\infty \leq B \]

Example 8 demonstrates a projection to satisfy a peak magnitude constraint.

**Example 8**

The signal level of the PD input is \( u_{RMS} = 0.30 \). The PD model is

\[ f_{PD}(x|a) = x \ast (a_1 + a_3|x|^2 + a_5|x|^4 + a_7|x|^6 + a_9|x|^8) \]

The initial parameter point is

\[ a^{[0]} = [1.0 - 1.5 \ 6.0 \ 0.0 \ 0.0]^T \]

The bound on peak magnitude of the PD output is as expressed in (40) with \( B = 4.8\sigma_x = 13.6 \) dB\( \text{RMS} \). The set of all samples in \( u \) or \( y \) is replaced by two representative peak samples, \( u_{peak} \).

\[ u_{peak} = \begin{bmatrix} 2.6\sigma_u \\ 2.8\sigma_u \end{bmatrix} \]

In addition to the peak bound, the projection should not change the average signal level of the PA input and thus the PD output, in the implicit back-off model. This constraint is similar to the unity gain constraint. The (hyperplane) equation (33) becomes

\[ f\left(u_{RMS}|a^{[i]}\right) = \langle a^{[i]}, \kappa(u_{RMS}) \rangle = x_{RMS} = f\left(u_{RMS}|a^{[0]}\right) \quad (41) \]

The projection sequence is generated by iteration of three projections – two to satisfy the peak inequalities and the third to satisfy the average gain constraint.

\[ a^{[i+1]} = P_H\left(P_{HS}\left(P_{HS}(a^{[i]}|\kappa(u_1), B)|\kappa(u_2), B\right)|\kappa(u_{RMS}), x_{RMS}\right) \quad (42) \]

The plots below show twenty-five iterations of (42) for Example 4.
The plots of $f(m|a^{[i]})$ $m \in (0,1)$ in Figure 25 show that the projection sequence successfully generated a solution that satisfied all three constraints. The rectangle in the upper left corner is the peak constraint $x_{peak} = 1.31$, which extends to the maximum
value of $u_{\text{peak}}$, 0.84. The original function violates the peak constraint, while the projected function does not. The difference between the two curves is significant for $m > 0.5$. If $a^{[0]}$ is an ideal PD for a corresponding PA, then replacing $f_{PD}(u|a^{[0]})$ with $f_{PD}(u|a^{[25]})$ would result in increased distortion at the output of the PA. This concludes Example 8.

This brings us to one of the basic questions of an algorithm based on projection. The goal of the method above is to project a point in parameter space by the minimum distance onto the intersection of all constraint sets. But what is needed is the point within this feasible set that maximizes the objective. An algorithm based on this projection alone is not sufficient to solve a problem such as Problem 4.

**Algorithm 4**

This algorithm is intended for direct optimization of Problem 4 with explicit back-off which includes constraints on peak amplitude. This algorithm is not complete, as it requires more development and a sufficient proof.

The intention is to generate a converging sequence of the parameters by iteration of POCS and a gradient descent method.

A. Initial condition

Choose an initial condition for the parameter vector which satisfies all constraints except the MSE constraint. PD parameters which satisfy the MSE constraint may not be known. This initial condition should be a safe stable point.

B. Unconstrained global minimum
An important piece of information to extract from an ILA data set is the global optimum. The MSE at this estimated point is the estimate of the best possible MSE performance for the power level represented by the data set. If this MSE does not satisfy the $\Omega_{MSE}$ constraint, there are two possible conclusions. The first is that it is simply not possible to meet the minimum performance at this power level. The other possibility is that the PD that was applied to the input signal caused degradation in the MSE due to the deviation in the probability distribution between the output signal and the reference signal.

While it is possible to project onto the feasible bounded set in parameter space, there is no guarantee that the closest projection to the optimum is the closest to satisfying the MSE. This occurs because of the shape of the bound in relation to the shape of the MSE set.

The two possible approaches to this problem are:

1. Estimate the PA function directly via a model. Then use a simulation to create data sets from this model. The only need for a data set is for calculating the MSE integration (Monte Carlo). If it was practical to evaluate the error integral, the data set would not be necessary.

2. Estimate the PD function directly via the ILA so that it satisfies all constraints.

Each data set used in the ILA was generated at a specific average IBO.

As stated, this algorithm is not complete and does not converge as Algorithm 3 does. No results are provided beyond the analysis contained in this section (§6.5) above.
7 Results

Simulations of Algorithm 2 and Algorithm 3 described in Chapter 6 solve the appropriate problems (Problem 3 and Problem 4) given in Chapter 5 for all of the PA models given in §4.3. Prior to comparing solution algorithms using models of the system, the models themselves must be verified to exhibit the correct behavior. Section 7.1 compares the behavior of the PA models to the expected behavior of a physical system. Section 7.2 verifies and demonstrates the crest factor reduction function, a part of both algorithms. Section 7.3 shows the result of linearizing two PA models according to the method described in current literature [19]. Finally §7.4 presents some important intermediate results from the simulations and §7.5 summarizes the final targeted results of the algorithms developed in this thesis.

7.1 PA Model Validation

The PA models developed represent a class or subset of real systems and should emulate many of their important characteristics. The pertinent characteristic is the gradual saturation of the output current as exhibited by the transconductance of a MOSFET [Sedra, 390]. For a fixed bias level, an amplifier scales the output current with the input voltage and as the input voltage increases beyond the linear region, the scaling of the output by the input becomes non-constant.

Based only on this known characteristic it can be concluded that as the input voltage signal extends into this region, the output will exhibit a degree of non-linear distortion in a form that is dependent on the signal’s frequency content. To verify this,
each of the models is driven with the representative signal, WF11, resembling a low-pass noise complex baseband signal. The output back-off (OBO) is defined as the ratio of the output envelope RMS level to the maximum (saturation) output envelope level. A lower OBO corresponds to a higher RMS PA output. Ignoring the other traces for now in Figure 26, the “PA Output” curve shows that as the OBO decreases, the adjacent channel power ratio (ACPR) decreases. The adjacent channel power is the inter-modulation distortion (IMD) generated outside the allotted channel. While ACPR only includes the distortion in a frequency segment outside of the channel bandwidth, this ratio is proportional to the signal-to-noise plus distortion ratio (SNDR) which includes the distortion at all frequencies. A decrease in the ACPR with an increase in output power (decrease in OBO) indicates that the signal is reaching into the PA model’s saturation region. In contrast, the g11 PA model exhibits non-linearity over the entire input magnitude range and consequently the ACPR is nearly constant over a wide back-off range from 6dB to 10dB (see Figure 27).

![Figure 26: g8 validation, PR](Matlab\NonLinear\Haggman\test_g8_PARS)  
![Figure 27: g11 validation, PR](Matlab\NonLinear\Haggman\test_g6_PARS.m)

\[ \gamma_{dB} = [7.5, 8.0, 8.5, 9.0] \]  
\[ \gamma_{dB} = [7.5, 8.0, 8.5, 9.0] \]
The remaining curves in Figure 26 and Figure 27 are associated with peak reduction (PR) and digital pre-distortion (DPD) linearization, described in the following two sections.

7.2 Peak Reduction

The domain of the PD input is limited to a designated peak-to-average ratio (PAR), as described in §5.7 and as specified in Algorithm 3 and Algorithm 4. Methods generally named crest factor reduction (CFR) have been developed for reducing the peaks of signals with high PAR. A preferred method for reducing peaks would cause no distortion to the modulated waveform, but for larger amounts of reduction this is not possible. It is assumed that the reference signal $w$ is distortion-free with a known probability density and the PAR bound on $u$, the input of the PD function, is met by a magnitude limiting function. The signals $w$ and $u$ and the function $\Lambda_T$ are shown in the ILA diagram (Figure 13).

The next observation is the effect of clipping the signal at the designated threshold, removing all energy in the signal peaks with no matched filtering. This reduces the amplitude range of the PD input signal. In Figure 28, for example, the signal level of $w$ is $\sigma_w = -11dBFS = 0.2818$ and the PAR limit is $7.7dB_{RMS}$, so the limit level is at $-3.3dBFS = 0.6819$. 
We now demonstrate the ability to linearize the PA model by a common digital pre-distortion (DPD) method presented in literature [19]. We also describe the connection and comparison to Algorithm 3.

In the indirect learning architecture (ILA), the cost function to be minimized is the mean square error (MSE) described in §5.8. While the MSE criteria is inherent in the least squares solution, the only figure of merit verified in [19] is the qualitative decrease in inter-modulation distortion (IMD) as shown by the PSD, and no numerical results are given. The typical figure of merit associated with IMD outside the channel bandwidth is the adjacent channel power ratio (§3.2.3 ACPR) (recall that a higher value corresponds to a higher SNDR). A validation was developed to verify the solution given in [19] and to produce numerical results for ACPR over a range of back-off levels.

Figure 28: Magnitude limiting function, T=7.7dB, slight deviations introduced by interpolation filtering.

7.3 Linearization

We now demonstrate the ability to linearize the PA model by a common digital pre-distortion (DPD) method presented in literature [19]. We also describe the connection and comparison to Algorithm 3.

In the indirect learning architecture (ILA), the cost function to be minimized is the mean square error (MSE) described in §5.8. While the MSE criteria is inherent in the least squares solution, the only figure of merit verified in [19] is the qualitative decrease in inter-modulation distortion (IMD) as shown by the PSD, and no numerical results are given. The typical figure of merit associated with IMD outside the channel bandwidth is the adjacent channel power ratio (§3.2.3 ACPR) (recall that a higher value corresponds to a higher SNDR). A validation was developed to verify the solution given in [19] and to produce numerical results for ACPR over a range of back-off levels.
The PD model is a 5th-order memory-less polynomial predistortion. The PD function parameters are determined by a block least-squares solution (as described in [19]) using a dataset from the indirect learning (ILA) configuration. The results “w/DPD” are plotted in Figure 26 and Figure 27. The fundamental observation is the comparison between “PA Output” (black) and “w/DPD” (blue). Over the entire range of IBO shown, the ACPR improves after pre-distortion is applied. Beyond this affirmation, there are two other notable observations. The first is that for a sufficient back-off, the level of PAR reduction determines the ACPR because PA saturation distortion is not significant. The second observation is at lower back-off the peak reduction neither improves nor degrades the performance of the PD linearization based on IBO. However, relevant to the optimization in this work, the linearized OBO for a given IBO increases with decreased PAR. This conclusion has been described in [26] where the author states that “the highest output power level and the best linearity performance were obtained for the signal with the lowest PAPR value.”

![Figure 29: G9 linearization](MATLAB/NonLinear\Haggman\run_g5.m)
7.4 Full Simulation

This section describes some details of the simulations including algorithm parameters. We also establish the exact algorithm-problem pairs for which results will be given and show some intermediate data from the simulations. These results are selected to support the thesis and to demonstrate the algorithms.

The relationship between OBO and IBO is described as monotonic in §4.5. This is a necessary assumption for Algorithm 2 and Algorithm 3 which holds for all models herein. Recall that both Algorithm 2 and Algorithm 3 are expected to start from a conservative back-off level and approach the performance bound with decreasing back-off. The trace of IBO vs. OBO for a simulation of G8 is shown in Figure 30.

![Figure 30: IBO / OBO over search path](image)

The next plot (Figure 31) shows that the method of estimating the peak using a quadratic fit (§5.5.2) tracks the back-off of a PA model with a known peak.
The following table shows the organization and designation of the results presented in the conclusion of this chapter. Results are given for the following three algorithm combinations:

<table>
<thead>
<tr>
<th>Problem</th>
<th>Algorithm</th>
<th>PD Model</th>
</tr>
</thead>
</table>
| PD-1    | Formulation 3 Algorithm 2 (LSS) | Polynomial  
          |          | K=5,Q=0 |
| PD-2    | Formulation 4 Algorithm 3 | Polynomial |
|         | B={9dB, 13dB} | K=7,Q=0 |
| PD-3    | Formulation 3 Algorithm 2 (LSS) | Polynomial |
|         |          | K=9,Q=0 |
The results are generated by a full simulation with multiple iterations, satisfying the condition on performance measurement discussed in §5.10. The PD calculated is applied and then the signal is run through the system in the forward direction (see Figure 13). The PA model is a fixed simulation, but the model is treated as an oracle or “black box” with the underlying nonlinear function “hidden.” The simulation follows the iteration sequence of Algorithm 2.

For PA models G8 and G7, the simulation is run for 25 iterations with an SNR limit of 35dB. In Figure 32 and Figure 33, the OBO steadily decreases while the ACPR and SNR are decreasing.

![Figure 32: PD-1 and PD-2 Bu=10, G8 [MATLAB\NonLinear\run_g7_IBO_03.m]](image)
At termination, each PD has linearized the PA output signal at the estimated highest feasible power. The CCDF in Figure 34 shows the probability distribution of the PA input and output signal relative to the average power. Both PD-1 and PD-2 expanded the peaks of the reference to $11.8\, dB_{RMS}$ (compare to Figure 6). The PA output, $y$, is limited by saturation to $7.6\, dB_{RMS}$.

Figure 33: PD-1 and PD-2 Bu=10, G7 [MATLAB\NonLinear\run_g7_IBO_03.m]

Figure 34: g7 final CCDF
7.5 Summary of Results

The results of the three PD algorithms are given below. Table 4 shows the estimated optimal back-off for an SNDR performance constraint of 35dB, i.e. -35dB NMSE. With this constraint, the back-off corresponding to ideal saturation is 7.25dB, which is the theoretical bound. Figure 35 also shows results at other NMSE levels.

Table 4: Optimal back-off estimates of three PD algorithms, $\Omega = 35\, dB$

<table>
<thead>
<tr>
<th></th>
<th>PD-1</th>
<th>PD-2</th>
<th>PD-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>G8</td>
<td>7.65</td>
<td>7.54</td>
<td>7.50</td>
</tr>
<tr>
<td>G7</td>
<td>7.66</td>
<td>7.48</td>
<td>7.48</td>
</tr>
<tr>
<td>G9</td>
<td>7.51</td>
<td>7.49</td>
<td>7.48</td>
</tr>
<tr>
<td>G11</td>
<td>7.70</td>
<td>7.50</td>
<td>7.69</td>
</tr>
</tbody>
</table>

Figure 35: Results vs. theoretical bound, eqn. (20), of three PD algorithms for each PA model
Example 9-1

For the g7 PA model with $\Omega = 35dB$, the difference in IBO and OBO between PD-1 and PD-2 is distinguishable. The values are summarized in Table 5.

<table>
<thead>
<tr>
<th></th>
<th>IBO [$dB_{sat}$]</th>
<th>OBO [$dB_{sat}$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>PD-1</td>
<td>10.73</td>
<td>7.686</td>
</tr>
<tr>
<td>PD-2</td>
<td>10.51</td>
<td>7.543</td>
</tr>
<tr>
<td>Difference</td>
<td>0.22</td>
<td>0.143</td>
</tr>
<tr>
<td>% Power Difference</td>
<td>5.20%</td>
<td>3.35%</td>
</tr>
</tbody>
</table>

The plots in Figure 36 show the results of linearization by PD-2. The PD functions are constrained by an additional upper bound on the magnitude of the PD output, as defined in Problem 4.

Figure 36: PD-2 g7 linearization. Red: PA AM-AM (y vs. x); Blue: PD, Black: PA,PD output (z vs. x)
The traces in Figure 37 and Figure 38 below show the convergence of the PD estimates toward the performance bound of 35dB. At initialization, no PD is applied and the back-off is set to within a few dB of the expected final back-off. At each iteration the back-off decreases while the measured SNDR of the output remains above the bound. At termination, PD-2 with a reasonable upper bound outperforms PD-1, while PD-2 with a restrictive upper bound does not match the performance of PD-1.

Figure 37: Compare PD-1, PD-2 w/ Restrictive upper bound Bu=0.79 (9dB)  
Figure 38: Compare PD-1, PD-2 w/ Reasonable upper bound Bu=1.2589 (13dB)

Example 9-2

Comparing PD-3 to PD-1 for the G11 PA model with $\Omega = 35dB$, and $\Omega = 30dB$, the resulting optimal back-off is nearly identical (Figure 39). At higher back-off, around 8.5dB, the PD-3 gives better SNDR performance, but the advantage diminishes to zero as the back-off decreases.
Figure 39: g11 w/ $\Omega = 35dB, 30dB$

The convergence over the iterations is also shown for SNR, ACPR and OBO in Figure 40 and for the odd coefficients of PD-3 in Figure 41.

Example 9-3

Comparing PD-1 and PD-2 for G11 with $\Omega = 35dB$ and the bound of PD-2 at 13dB, the optimal back-off obtained by PD-2 is 7.5dB, compared to 7.7dB for PD-1.
Figure 42: Convergence of PD-1, PD-2 for G11, $\Omega = 35$

Also the initial distortion level without PD is much higher than the initial distortion level of g7, g8, and g9 because g11 has non-linearity over the entire amplitude range while the other models only have significant non-linearity in the saturation range.

The coefficients of the 7th-order polynomial PD-2 are shown for two values of $\Omega$ in the figures below. The magnitude of the optimal coefficients is greater than 400 with $\Omega = 35 dB$ and greater than 50 for $\Omega = 30 dB$.

Figure 43: Coefficient magnitudes for PD-2, g11, $\Omega = 35 dB$

Figure 44: Coefficient magnitudes for PD-2, g11, $\Omega = 30 dB$
Finally, the gain of PD-2 is shown in Figure 45 for a 6.67dB OBO. The gain at an 8dB input level is 2.8dB. The curve also shows a critical point at a 9dB input level, above which the gain does not continue to increase. This is necessary to meet the 13dB peak-to-average bound which occurs at approximately 9.4dB input plus 3.6dB gain. The same limits are shown in the CCDF plots in Figure 46. The peak PA input level is much lower for PD-2 (8.6dB) compared to PD-1 (10.7dB).

Figure 45: Gain of PD-2, G11, Ω = 30dB, B=1.259 (13dB)
Figure 46: CCDF of g11 input and output for PD-1, PD-2, $\Omega = 30\, dB$, $B=1.259$ (13dB)
8 Conclusions

The motivation for this research was maximizing the operating efficiency of an RF amplifying system. The efficiency of an amplifying device is shown to be directly related to the average output power level when the supply voltage is constant. The significance of any improvement in efficiency compared to the effort required depends on the application. The author of [12] supports his closely-related work by arguing that “for certain applications, such as the base-station transmitter, reducing PAR by an extra 0.5dB may justify using a more complex algorithm.” Similarly, reducing the operating back-off by even a small amount is sometimes significant by other measures.

This work produced two novel optimization problem formulations for maximizing operating efficiency of a power amplifying system with a high peak-to-average ratio waveform. The most significant constraint, the fidelity of the PA output signal, was studied by analysis and simulation. Algorithms were developed and tested in solving the optimizations for several exemplary PA models. The results should be extended to variations on PA architectures, e.g. piecewise architectures such as Doherty and long memory duration architectures such as high-efficiency tube amplifiers and drain modulation. The results have answered some fundamental questions and the work has certainly created new questions and opportunities for research.

8.1 Discussion of Results

By inspection of the optimal back-off levels estimated by each algorithm (Table 4), both PD-2 and PD-3 resulted in lower back-off compared to PD-1 for a nominal
performance constraint. This shows that an increase in the model order of a memory-less polynomial pre-distortion function can improve linearization performance and allow for higher saturation distortion. Furthermore, reasonably constraining the PD output, represented by PD-2, does not degrade performance. From Example 9-1, PD-1 outperforms PD-2 if the PD output peak level is severely restricted (9dB) because ideal linearization must utilize the entire monotonic range of the PA. An active constraint on PA input signal peaks can only be remedied by increased back-off. However, when the PA input level is limited above the input saturation level, the conditions for linearization are met.

The results in Example 9-2 show that an increase in model order, from 5\textsuperscript{th}-order to 9\textsuperscript{th}-order in this case, does not guaranteed an improved result for the unconstrained PD problem of Formulation 3. This is most pronounced for the g11 PA model. A possible explanation is that the non-linearity of g11 over a large range is a penalty against polynomial models of all orders. The g9 PA model has a similar result, possibly because of the higher PA model order.

Example 9-3 demonstrates the advantage and disadvantage of PD-2. While the PD function has the desired unconventional shape for combined linearization and peak limiting, the penalty is in the magnitude of the coefficients. But recall that the goal of the algorithm is to solve the problem formulation, which is focused only on back-off and not on stability. If the PD model were implemented as an adaptive system, it would be beneficial to retreat slightly from optimized back-off and explicate constraints on the coefficient vector. In any case, the behavior of the constrained PD-2 algorithm in
Example 9-3 is significant in its own right and confirms that other potential algorithms may be able to satisfy Problem 4 in an improved way compared to Algorithm

As the signal performance constraint increases (less distortion), there is more differentiation among the choices of PD models and greater consequences of implementation constraints, such as maximum PA input level. This is also supported by the analysis of the SNDR surfaces and sets.

8.2 Future Research

First I restate the argument that Formulation 4 is a difficult problem to solve and is an excellent candidate for continued analysis and investigation into state-of-the-art optimization algorithms. In this work, I began a development of an algorithm using POCS, but the algorithm was not fully developed and could not be made to match the performance of Algorithm 3. The cutting-plane methods described in [32] rely on iterative approximation of a convex set by sub-gradients. Interior-point methods [1] replace bounds with a continuous approximation. These methods and others are promising because they are supported by theory and are the topics of continued research.

In addition to continued algorithm development, I would like to increase my understanding of the results by exercising a broader set of both synthesized and extracted PA models. Alternate PD models described in literature may also be applied to this novel problem.
9 References


10 Appendix A

This appendix gives the expansion and solution of the integral in §5.5.1.

\[
\theta_{s_{\text{at, out}}}^2 \int_A^\infty f_R(x) \, dx - 2 \int_A^\infty x \cdot f_R(x) \, dx + \int_A^\infty x^2 \cdot f_R(x) \, dx \quad A = \theta_{s_{\text{at, out}}}
\]

The integral in the first term is the complementary cumulative distribution function. The integral in the second term is the same as the expectation integral, and is proportional to the error function. The integral in the third term must be solved by integration by parts.

\[
\theta_{s_{\text{at, out}}}^2 \cdot (1 - F_R(A)) - 2 \int_A^\infty x \cdot f_R(x) \, dx + \int_A^\infty x^2 \cdot f_R(x) \, dx \quad A = \theta_{s_{\text{at, out}}}
\]

Second term:

\[
2 \int_A^\infty x \cdot f_R(x) \, dx = \frac{2}{\sigma^2} \int_A^\infty x^2 \cdot e^{-\frac{x^2}{2\sigma^2}} \, dx
\]

\[
f_R(x) = \frac{x}{\sigma^2} \cdot e^{-\frac{x^2}{2\sigma^2}}
\]

\[
\frac{2}{\sigma^2} \int_A^\infty x^2 \cdot e^{-\frac{x^2}{2\sigma^2}} \, dx = \frac{2}{\sigma^2} \cdot \sqrt{2\pi} \cdot \frac{\sqrt{\pi}}{2} \cdot \text{erfc} \left( \frac{A}{\sqrt{2\sigma^2}} \right) = \frac{\sqrt{2\pi}}{\sqrt{2\sigma^2}} \cdot \text{erfc} \left( \frac{A}{\sqrt{2\sigma^2}} \right)
\]

Third term:

\[
\int_A^\infty x^2 \cdot f_R(x) \, dx = \frac{1}{\sigma^2} \int_A^\infty x^3 \cdot e^{-\frac{x^2}{2\sigma^2}} \, dx = \frac{1}{\sigma^2} \cdot \sigma^2 \cdot (x^2 + 2\sigma^2) \cdot e^{-\frac{x^2}{2\sigma^2}} \int_A^\infty x \, dx
\]

\[
0 + (A^2 + 2\sigma^2) \cdot e^{-\frac{A^2}{2\sigma^2}}
\]