Chapter 6

Pre/Post-Distortion Algorithms

If the amplifier exhibit nonlinear characteristics invariable with time, which is a reasonable assumption in many low-power cases, a fixed pre-distorter is enough to achieve a good linear performance. However, power amplifiers operating under more stringent conditions may undergo slow but significant changes in their AM/AM and AM/PM characteristics basically due to factors like temperature, age of components, power level, biasing variations, frequency changes and so on. This means that, given the requirements of some applications, a pre-distorter must be an adaptive nonlinear processor, capable of following the changes in the HPA characteristics.

6.1 Adaptive scheme for PD, basic definitions

As seen in section 2.2.1, the response of RF High Power Amplifiers (HPA) can be characterized as a type of complex non-linear multiplicative distortion which is (whenever a frequency-independent approach has been justified) only dependent on the modulus $u_I(t)$ of the input to the HPA base-band signal $b_I(t) = u_I(t)e^{j\alpha_I(t)}$ in the form

$$b_O(t) = b_I(t)G_{HPA}(u_I(t)) \tag{6.1}$$

which is also expressed in terms of the modulus and phase non-linearities as

$$|b_O(t)| = |b_I(t)| \cdot |G_{HPA}\big(u_I(t)\big)| = A\big(u_I(t)\big) \tag{6.2}$$

$$\arg\{b_O(t)\} = \arg\{b_I(t)\} + \arg\{G_{HPA}\big(u_I(t)\big)\} = \alpha_I(t) + \Phi\big(u_I(t)\big) \tag{6.3}$$

with $b_O(t)$ the base-band signal at the HPA output, $A(\cdot)$ and $\Phi(\cdot)$ the AM/AM and AM/PM transfer characteristics, respectively.

When the presence of the memory effect of analog stages is considered, the HPA model should also include a base-band equivalent time delay parameter Δ, in order to represent the memory effect introduced along the up-conversion (U/C) plus HPA plus

down-conversion (D/C) chain. In such a case, the amplifier operation in (6.1) can be reexpressed as

$$b_O(t) = b_I(t-\Delta)G_{HPA}(u_I(t-\Delta)). \tag{6.4}$$

This approximation for representing the distributed electronic delays by a concentrated time shifting parameter is reasonable, provided that the response of the filters within U/C and D/C be sufficiently flat in the band of interest. This is a common and technologically feasible assumption which, as seen in previous chapters, leads to useful simplifications on the signal model. Thus, since the unknown time delay Δ constitutes a nuisance parameter for digital base-band pre-distortion, it must be compensated. On one hand, time delay has been shown to be somewhat dispensable information for estimating the AM/AM nonlinearity (in [37] it is shown that AM/AM characteristic can be obtained from the probability density function of the input and output modulus), but on the other hand it still proves to be particularly critical for the estimation and compensation of the AM/PM distortion, where the estimation of Δ will be required.

In this section we aim to propose and evaluate adaptive schemes focusing on the efficient estimation of the inverse HPA characteristics for their application in base-band. For this purpose, our choice is to operate a fast coarse time delay estimation (TDE) algorithm (like [54],[39] and [38]) previously to PD estimation, in order to reduce to a minimum (less than one sample period) the misalignment between the reference signals used in the adaptive structure and to preserve the major part of the available degrees of freedom for the adaptation of the PD coefficients regardless of major time corrections. Thus, the forthcoming algorithms will be devised to operate adaptive PD estimation independently from time alignment.

6.1.1 Basic schemes for PD estimation: Pre-distortion vs. Post-distortion

Let us consider the basic adaptive schemes shown in figure 6.1 where two alternative linearization structures are defined by positioning the nonlinear compensation block to operate either as a post-distorter or as a pre-distorter. These structures can be considered as counterparts in the sense that, as shown in section 2.2.1, whenever the coefficients defining the 'pre' or 'post' operation are such that the PD block optimally approaches the inverse HPA complex gain

$$P(\cdot) = G_{HPA}^{-1}(\cdot),$$

the compensation error signals $e_{post}(t)$ and $e_{pre}(t)$ are both minimized [1]. However, two different expressions for the compensation error signal are obtained from these structures.

[1] In the figures we refer to signal blocks using discrete-time vector notation. This will be useful in later algorithm derivations.

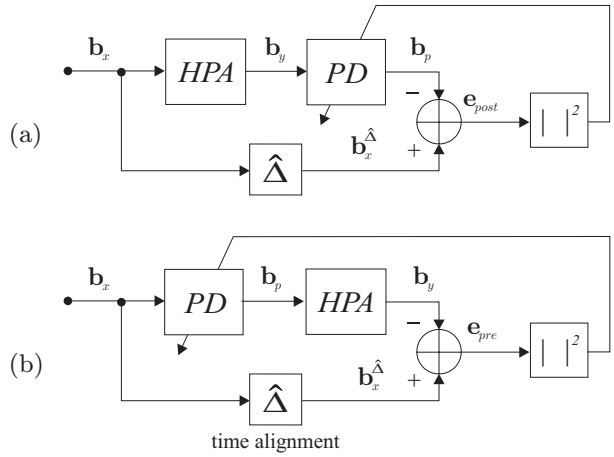


Figure 6.1: Basic adaptive PD schemes.

Assuming a perfect discrete-time alignment, or equivalently, a memoryless HPA operation in figures 6.1(a) and 6.1(b), we force $b_x^{\hat{\Delta}}[n] = b_x[n]$. Thence, the error signal we obtain for each case is

$$e_{post}[n] = b_x[n] - b_y[n]P\left(|b_y[n]|\right) \tag{6.5}$$

$$\begin{aligned} e_{pre}[n] &= b_x[n] - b_p[n]G\left(|b_p[n]|\right) \\ &= b_x[n] - b_x[n]P\left(|b_x[n]|\right)G\Big(\left|b_x[n]P\left(|b_x[n]|\right)\right|\Big) \end{aligned} \tag{6.6}$$

whence we can obtain different expressions for the squared error signal which will constitute the cost function needed for the optimization of the pre/post nonlinear gain $P(\cdot)$.

As it is well known, optimization algorithms based on the steepest descent method, perform a downhill gradient search for a minimum in a cost function (normally the squared error). The dependence of the cost function on the estimated coefficients should ideally be quadratic, otherwise, higher-order dependence may increase the risk of convergence to a sub-optimal minimum within a local neighbourhood. Hence, it becomes clear that the error signal in (6.6), associated to a pre-distortion structure, is not well conditioned for this kind of minimization, since the cost function based on $e_{pre}[n]$ will feature higher order dependence with respect to the coefficients of any descriptive model of the nonlinear function $P(\cdot)$. This is because the HPA gain function $G(\cdot)$ is itself a nonlinearity (it generates higher order products) and therefore the squared error signal $|e_{pre}\left(\boldsymbol{\alpha}, n\right)|^2$, where $\boldsymbol{\alpha}$ is a vector containing the coefficients that completely describe $P(\cdot)$, has many local minima. In these conditions, the treatment of the gradient could involve almost untractable algebraic expressions. In turn, the post-distortion structure yields (6.5) which is a better conditioned expression for the error, as will become more clear in the forthcoming

derivations. Therefore, our solution proposal will be to implement a previous training of the coefficients in $\boldsymbol{\alpha}$ using a *post-distortion* scheme and subsequently apply the optimum estimated value of this PD vector to set the *pre-distortion* gain that finally introduces the pre-correction of the input signal. Thence, two general remarks arise at this point:

1. Post-distortion is a good choice for defining a well conditioned structure for the adaptive training of the linearizer coefficients. However, its implementation at the transmitter implies dealing with high power analog signal and/or stages.

2. Pre-distortion allows several ways for its digital implementation in base-band but, in general, does not provide the most suitable structure for the adaptive estimation of the inverse characteristics by means of a steepest descent based technique.

For these reasons, we will concentrate our analysis on the adaptive structure of figure 6.1(a). Therein, the correction error signal $e_{post}[n]$ is measured during the training stage to adapt the set of complex coefficients $\boldsymbol{\alpha} = [\alpha_0 \cdots \alpha_{N_c-1}]$ that characterize the estimated inverse HPA gain function $P(\cdot)$. Note that $P(\cdot)$ is a *modulus-dependent multiplicative complex gain* that can be operated either in the form of a post-distorter or a pre-distorter using the same set of estimated coefficients $\boldsymbol{\alpha}$. Therefore, from now on, we will refer the pre/post distorter stage as PD, making explicit the difference whenever necessary. Finally, along with the comparison between the pre/post distortion structures, it is also worthy to focus our attention on the information used to construct the error signals in a general linearization problem. In appendix 6.D we review an important property regarding the error signal considered in our PD schemes.

6.1.2 Pre/post distortion block definition

Like the HPA transfer function in (6.1), the PD function in figure 6.1(a) will be modeled (now in discrete time) as a modulus-dependent multiplicative complex gain according to

$$\begin{aligned} b_p[n] &= b_y[n]P(|b_y[n]|) = b_y[n]P(u_y[n]) \\ &= b_y[n]\left(\sum_{k=1}^{N_c} \lambda_k\left(u_y[n]\right)\alpha_k\right) \\ &= b_y[n]\boldsymbol{\lambda}^T(u_y[n])\boldsymbol{\alpha} \end{aligned} \tag{6.7}$$

where the vector $\boldsymbol{\lambda}(\cdot) = [\lambda_0 \cdots \lambda_{N_c-1}]^T$ is called the *bin activation function* and contains, for each value of the input modulus $u_y[n] = |b_y[n]|$, the proportional contribution of each triangular activation (membership) function from the set shown in figure 6.2(a). These functions interpolate (using $\boldsymbol{\alpha}$) the complex PD factor that multiplies the input data $b_y[n]$ producing the PD output $b_p[n]$. For any given $c_k \leq u_y[n] < c_{k+1}$, with $k \in \{0, 1, \ldots, N_c - 2\}$, one or at most two consecutive triangular membership functions will be activated (with values ranging from 0 to 1) through their corresponding elements $\lambda_{(k-1)}$ and $\lambda_{(k)}$, while all the remaining terms will be set to zero, thus defining a null

contribution for their respective elements from $\boldsymbol{\alpha}$.

Once the error signal in the adaptive system of figure 6.1 has converged to a minimum, the PD coefficients in vector $\boldsymbol{\alpha}$ are expected to approximate the inverse complex gain characteristic $G_{HPA}^{-1}[\cdot]$ of the HPA. Thus, a perfect non-linear compensation would be expressed as a linear transference

$$b_p[n] = k \cdot b_x[n] = b_y[n]P(|b_y[n]|) \tag{6.8}$$

with a constant gain factor k.

In the same figure 6.2, the characteristic points c_k are called the *centroids* and are grouped within a vector $\mathbf{c} = [c_0 \dots c_{N_c-1}]$. Each centroid is associated to one triangular activation function λ_k and its respective PD coefficient α_k. The centroids also define the division of the whole dynamic range into $N_{\mathcal{I}} = N_c - 1$ activation intervals of width $\Delta_k = |c_k - c_{k-1}|$ for $k = \{1, 2, \dots, N_{\mathcal{I}}\}$. As a design criterion, since the dynamic response of a real HPA is up-bounded by the saturation output amplitude A_{sat}, we can assume that in the case shown in figure 6.1(a) there will be no input data beyond $c_{N_c-1} = A_{sat}$ at the post-distorter input. Hence, for simulations, we can restrict the maximum dynamic range for the PD input to be $[c_0, c_{N_c-1}] = [0, 1[$. Nevertheless, since the same assumption does not always hold for the operation of a pre-distorter, whose input may exceed A_{sat} according to the input signal variance, some definitions regarding the treatment of out-of-range inputs will be later included in the algorithm implementation section.

6.1.3 Definition of activation functions

From the membership functions shown in figure 6.2, we define three types of input activation with dependence on the input modulus u.

Activation of type 1:

This activation is associated to the set of N_c triangular membership functions shown in figure 6.2(a) which are respectively assigned to each centroid-gain coefficient pair (α_k, c_k). Therefore, this activation can be also referred as the *centroids* or *coefficients activation*.

Here, for a given input amplitude u_0 to the PD block, with $0 \leq u_0 < A_{sat}$, the elements of an activation (column) vector $\boldsymbol{\lambda}(u_0) = [\lambda_0(u_0) \cdots \lambda_{N_c-1}(u_0)]^T$ are defined according to

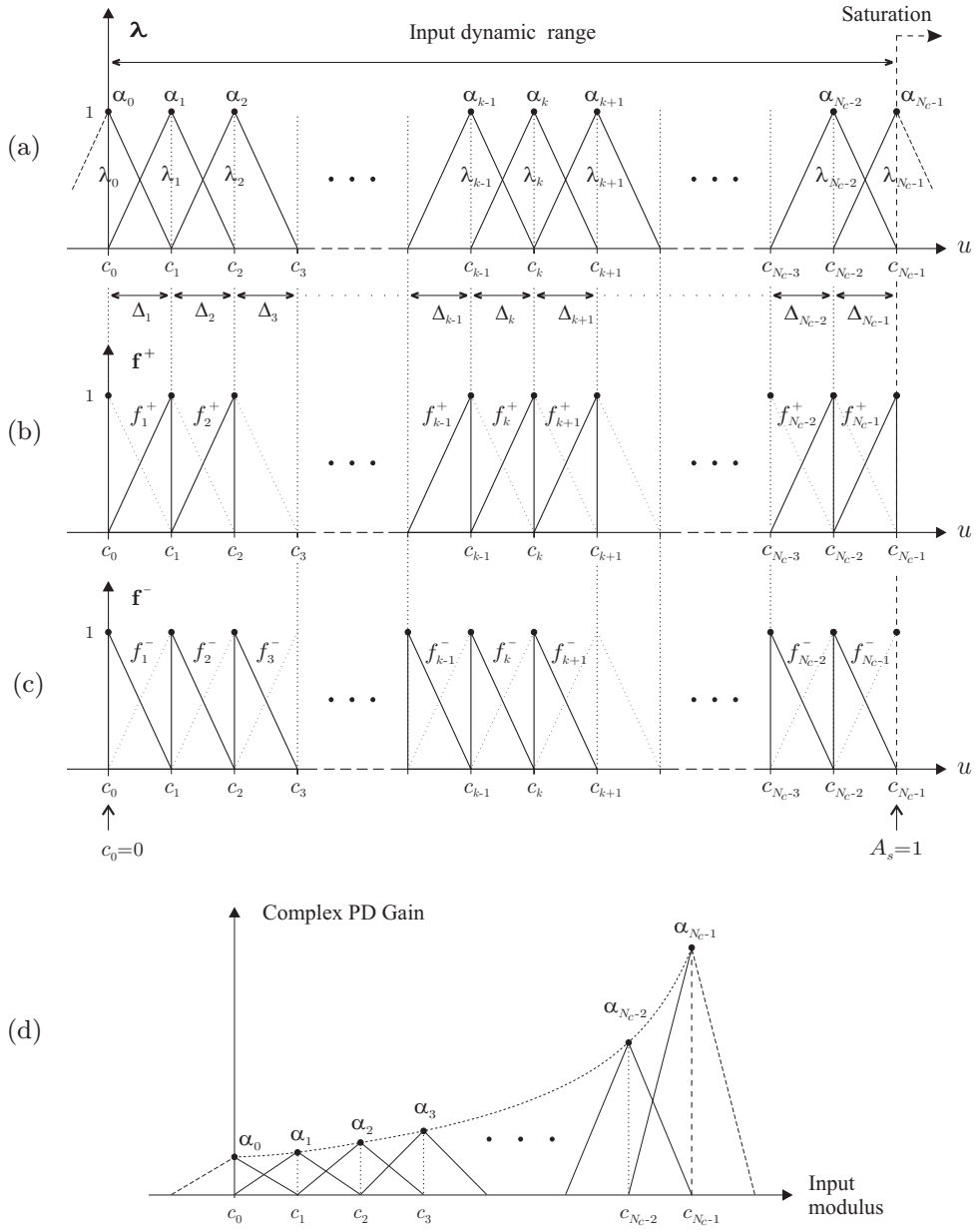


Figure 6.2: Activation functions used for PD gain interpolation. The input range to the pre/post distorter is divided into $N_{\mathcal{I}}$ intervals defined by a set of $N_c = N_{\mathcal{I}} + 1$ centroids. In (a),(b)and (c), the vertical axis is the membership degree for a given input modulus u to the PD block. In (a) we observe the set of N_c triangular fuzzy membership functions, associated to each complex PD gain coefficient from $\boldsymbol{\alpha}$. In (b) and (c) the membership of a given input is jointly defined through positive-negative slope activation functions for each interval. In (d) we show an example of the aspect of a PD gain curve with the trained coefficients that interpolate the PD gain factor for any given value of the input modulus.

$$\lambda_k(u_0) = \begin{cases} \dfrac{u_0 - c_{k-1}}{\Delta_k} & ; \quad c_{k-1} \leq u_0 < c_k \quad ,\forall\, k \geq 1 \\ 1 - \dfrac{u_0 - c_k}{\Delta_{k+1}} & ; \quad c_k \leq u_0 < c_{k+1} \quad ,\forall\, k \leq N_c - 2 \\ 0 & ; \quad \text{otherwise} \end{cases} \tag{6.9}$$

where a three piece expression has been necessary to consider the activation of any element in $\boldsymbol{\lambda}(u_0)$ accounting for the particular evaluation of the outermost activation functions λ_0 and λ_{N_c-1}. This definition corresponds to a first-order fuzzy distribution of the input data along the activation vector of length N_c, so that one or at most two consecutive elements in $\boldsymbol{\lambda}$ will be activated for a given u_0. In any case, the activation vector $\boldsymbol{\lambda}(u_0)$ should fulfil

$$\text{sum}\,\{\boldsymbol{\lambda}(u_0)\} = 1\ ;\ \text{for}\ 0 \leq u_0 < A_{sat}. \tag{6.10}$$

If all the intervals are defined to have the same width $\Delta_{fix} = A_{sat}/(N_c - 1)$ and a long stream of input samples $\mathbf{u}$ is considered, this type of activation leads to the construction of the *soft histogram* of the input modulus to the PD [61].

The activation functions of type 1 are suitable for applying PD as indicated in equation (6.7). However, when using this kind of activation definition we do not distinguish whether the gain coefficients α_k are being activated from the right or from the left, according to the position of the input data u_0 with respect to the center c_k. Besides, since the triangles associated to each λ_k may not be symmetric around their centers, using only this kind of activation definition may result in a lack of information for some specific calculations. Therefore, a more elaborate activation scheme is provided after the next example.

Example of activation type 1:

Let$A_{sat} = 1$ and consider the input range uniformly divided into $N_{\mathcal{I}} = 4$ intervals defining the vector of $N_c = 5$ centroids, $\mathbf{c} = [\ 0\ \ 0.25\ \ 0.5\ \ 0.75\ \ 1\ ]$. Given the input amplitude vector

$$\mathbf{u}_0 = \begin{bmatrix} 0 \\ 3/8 \\ 9/16 \\ 1/2 \\ 15/16 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.375 \\ 0.5625 \\ 0.5 \\ 0.9375 \end{bmatrix} \tag{6.11}$$

with length $N = 5$, the corresponding type 1 activation matrix $\boldsymbol{\Lambda}(\mathbf{u}_0)$, size $(N \times N_c)$, whose rows are the activation vectors $\boldsymbol{\lambda}^T(u_{0(i)})$ for each sample in $\mathbf{u}_0$, will result in

$$\mathbf{\Lambda}(\mathbf{u}_0) = \begin{bmatrix} \boldsymbol{\lambda}^T(u_{0(1)}) \\ \vdots \\ \boldsymbol{\lambda}^T(u_{0(N)}) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.75 & 0.25 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0.25 & 0.75 \end{bmatrix} \tag{6.12}$$

Activation of type 2:

In figures 6.2(b) and (c) we show two sets of positive and negative slope activation functions which now contain complementary information to describe the activation for a given input amplitude u_0 provided that $0 \leq u_0 < A_{sat}$. We define two simultaneous activation vectors $\mathbf{f}^+$ and $\mathbf{f}^-$ whose elements are given by

$$f_k^+(u_0) = \begin{cases} \dfrac{u_0 - c_{k-1}}{\Delta_k} & ; \quad c_{k-1} \leq u_0 < c_k \\ \\ 0 & ; \quad \text{otherwise} \end{cases} \tag{6.13}$$

and

$$f_k^-(u_0) = \begin{cases} 1 - \dfrac{u_0 - c_{k-1}}{\Delta_k} & ; \quad c_{k-1} \leq u_0 < c_k \\ \\ 0 & ; \quad \text{otherwise} \end{cases} \tag{6.14}$$

for $k = \{1, \cdots, N_c - 1\}$. Each piece-wise expression above determines that for a given input u_0 at most one element in each vector will be activated. The extreme case occurs when $u_0 = c_i$ which determines an all-zeros vector $\mathbf{f}^+ = \mathbf{0}$ while the total activation is associated to the element $f_{i+1}^- = 1$. By examining figure 6.2 we can easily observe the following relations regarding this type of activation distribution and the previous one:

$$f_k^- + f_k^+ = \lambda_{k-1} + \lambda_k = 1 \; ; \text{ for } 1 \leq k \leq N_c - 1 \tag{6.15}$$

and

$$\lambda_i(u_0) = \begin{cases} f_1^-(u_0) & ; \quad i = 0 \\ \\ f_{N_c-1}^+(u_0) & ; \quad i = N_c - 1 \\ \\ f_i^+(u_0) + f_{i+1}^-(u_0) & ; \quad \text{otherwise.} \end{cases} \tag{6.16}$$

According to this, the activation type 1 can be constructed as a specific combination of the positive/negative slope activation vectors $\mathbf{f}^+$ and $\mathbf{f}^-$ (type 2). These vectors now convey the information about the right or left hand activation of each coefficient α_k with respect to the centroid c_k. Such information will be useful later in defining the diferential

variations of the interpolated curve for an optimization algorithm.

Example of activation type 2:

Considering the same conditions of the previous example given for activation type 1, the corresponding type 2 activation matrix $\mathbf{A}(\mathbf{u}_0)$, size $(N \times 2N_c - 2)$, is defined by inserting the elements of both activation vectors $\mathbf{f}^-(\mathbf{u}_{0(i)})$ and $\mathbf{f}^+(\mathbf{u}_{0(i)})$ corresponding to the i-th sample from $\mathbf{u}_0$ into the i-th row as follows:

$$\mathbf{A}_{(i)}(\mathbf{u}_0) = [f_1^-(\mathbf{u}_{0(i)})\ f_1^+(\mathbf{u}_{0(i)})\ \cdots\ f_{N_\mathcal{I}}^-(\mathbf{u}_{0(i)})\ f_{N_\mathcal{I}}^+(\mathbf{u}_{0(i)})]. \tag{6.17}$$

Thence, the resulting type 2 activation matrix for the given example is

$$\mathbf{A}(\mathbf{u}_0) = \begin{bmatrix} 1.0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.75 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.25 & 0.75 \end{bmatrix} \tag{6.18}$$

Activation of type 3:

Since the input dynamic range of the PD has been divided into $N_\mathcal{I}$ intervals, using the previous definitions in (6.13) and (6.14) we can calculate the *interval activation* vector as

$$\mathbf{f}(u_0) = \mathbf{f}^+(u_0) + \mathbf{f}^-(u_0) \tag{6.19}$$

which is a vector of length $N_\mathcal{I}$ containing the absolute measure (1 for the activated interval and 0 for the rest) of the interval activation for a given input $0 \leq u_0 < A_{sat}$. Thus, if all the intervals are defined to have the same width $\Delta_{fix} = A_{sat}/(N_c - 1)$ and a long stream of input samples $\mathbf{u}$ is considered, this type of activation leads to the construction of the simple histogram of the input modulus to the PD.

Example of activation type 3:

With the same previous conditions, the corresponding type 3 activation matrix $\mathbf{F}(\mathbf{u}_0)$, size $(N \times N_\mathcal{I})$, is defined by combining the elements of both activation vectors $\mathbf{f}^-(\mathbf{u}_{0(i)})$ and $\mathbf{f}^+(\mathbf{u}_{0(i)})$ corresponding to the i-th sample from $\mathbf{u}_0$ into the i-th row of $\mathbf{F}(\mathbf{u}_0)$ as follows:

$$\mathbf{F}_{(i)}(\mathbf{u}_0) = \left[(f_1^-(\mathbf{u}_{0(i)}) + f_1^+(\mathbf{u}_{0(i)})) \cdots (f_{N_\mathcal{I}}^-(\mathbf{u}_{0(i)}) + f_{N_\mathcal{I}}^+(\mathbf{u}_{0(i)}))\right]. \tag{6.20}$$

Thence, the resulting type 3 activation matrix for the example under consideration is

$$\mathbf{F}(\mathbf{u}_0) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6.21}$$

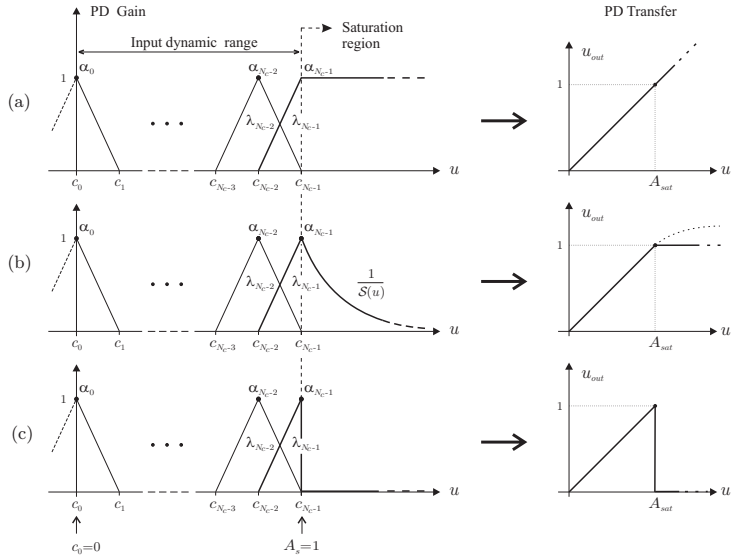


Figure 6.3: Alternative definitions for the extension of the set of activation functions to apply PD gain considering saturating input amplitudes: (a) to define linear unitary transference, (b) to define clipping response and (c) to define null output when the input exceeds saturation.

For inputs beyond the saturation amplitude, that is, for $u > A_{sat}$, we can consider three alternative PD responses by arbitrarily defining the extension of the upper activation function of type 1, $\lambda_{N_c-1}(u)$. This is equivalent to defining an additional type 2 activation function $f^-_{N_\mathcal{I}+1}$ for this region. These alternatives are shown in figure 6.3 and are defined as follows:

$$f^-_{N_\mathcal{I}+1}(u_0 \geq A_{sat}) = \begin{cases} 1 & ; \text{ linear transfer} \\ \dfrac{1}{\mathcal{S}(u_0)} & ; \quad \text{clipping} \\ 0 & ; \quad \text{null output} \end{cases} \tag{6.22}$$

In the expression defining a clipping response for the transfer function over saturation, if $\mathcal{S}(u_0) = |u_0|$, we obtain the hard clipping response shown in the figure 6.3(b). The soft clipping response, marked in dashed line, can be obtained, for instance, by adjusting the parameters γ and ν of a function $\mathcal{S}(u_0) = 1 + \gamma(u_0 - A_{sat})^\nu$. In any case, the resulting value for the activation of $f^-_{N_\mathcal{I}+1}$ will be associated to the last coefficient α_{N_c-1}. Therefore, the activation $f^+_{N_\mathcal{I}}$, which is zero for $u_0 > A_{sat}$, can be directly replaced by the resulting value of $f^-_{N_\mathcal{I}+1}$ at the corresponding position in the activation matrix. This allows us to perform the vectorial operations without requiring an additional column in the activation matrices already defined.

6.1.4 PD gain based on linear interpolation

In figure 6.2(d), an example for a trained set of estimated inverse gain coefficients has been represented considering only their absolute values $|\alpha_k|$ (AM/AM inverse). However, the PD operation over the complex base-band data $b_y[n]$ must be carried out considering both the modulus and phase of the complex PD coefficients α_k to compensate for the AM/AM as well as the AM/PM distortion.

In general, the modulus-dependent complex PD gain, as represented in figure 6.2(d), can be obtained in terms of the type 1 or type 2 activation functions and by means of linear interpolation through the complex PD coefficients in vector $\boldsymbol{\alpha}$ according to

$$P(u_y) = \frac{c_k - u_y}{\Delta_k}\alpha_{k-1} + \frac{u_y - c_{k-1}}{\Delta_k}\alpha_k \tag{6.23}$$

$$= \lambda_{k-1}(u_y)\ \alpha_{k-1} + \lambda_k(u_y)\ \alpha_k \tag{6.24}$$

$$= f^-_{k-1}(u_y)\ \alpha_{k-1} + f^+_k(u_y)\ \alpha_k \tag{6.25}$$

where $c_{k-1} \leq u_y < c_k$. Now, using (6.23) we can obtain the output of the PD block for the input modulus u_y. From (6.7), and skipping the discrete time dependence for simplicity, we can express the PD output as the input-to-output complex transference

$$\begin{aligned} b_p &= \mathcal{P}(b_y) = b_y P(u_y) = b_y\left(\frac{c_k - u_y}{\Delta_k}\alpha_{k-1} + \frac{u_y - c_{k-1}}{\Delta_k}\alpha_k\right) \\ &= u_y e^{j\alpha_y}\left(\frac{c_k - u_y}{\Delta_k}\alpha_{k-1} + \frac{u_y - c_{k-1}}{\Delta_k}\alpha_k\right) \\ &= e^{j\alpha_y}\left(u_y \underbrace{\left(\frac{c_k\alpha_{k-1} - c_{k-1}\alpha_k}{\Delta_k}\right)}_{(a)} + u_y^2 \underbrace{\left(\frac{\alpha_k - \alpha_{k-1}}{\Delta_k}\right)}_{(b)}\right). \end{aligned} \tag{6.26}$$

The inverse characteristics for the Saleh model were presented in section 2.2.1. Therein we obtained the inverse transfer characteristics (figure 2.3) as well as the corresponding inverse gain (figure 2.5) for this model. The use of linear interpolation with a reduced number of characteristic points will logically introduce a fitting error when approaching these continuous inverse curves. Nevertheless, from the expression obtained in (6.26), we observe that using simple linear interpolation to describe the PD gain leads to a 2^{nd} order interpolation in terms of the I/O transfer approximation. This results advantageous since it allows a better fitting to the shape of the inverse transfer curve, thus reducing the minimum adjusting error achievable with a given number of interpolation coefficients. In the expression above, the term denoted (a) controls the

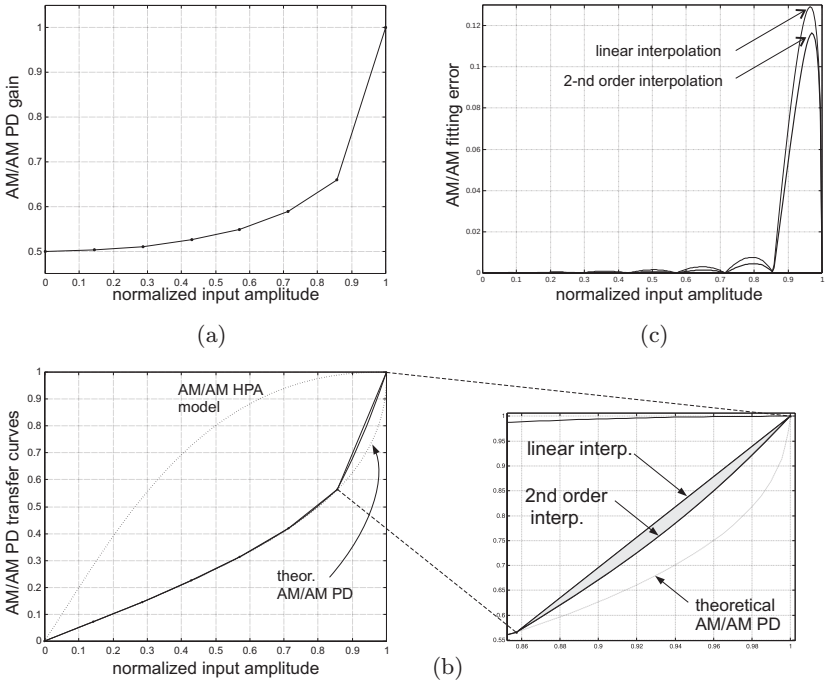


Figure 6.4: (a) Theoretical inverse AM/AM PD gain curve with $N_c = 8$ coefficients α for linear interpolation. (b) Linear and 2-nd order interpolated inverse AM/AM transfer curves. (c) Fitting error with respect to the ideal inverse AM/AM transfer characteristic.

linear part of the interpolation. Considering that $c_k > c_{k-1}$ the positive/negative sign of (a) will depend on the complex difference $\alpha_{k-1} - \alpha_k$. In turn, the term (b), which determines the contribution of the quadratic interpolation, is totally dependent on such difference: the greater the difference $\alpha_{k-1} - \alpha_k$ (in modulus or phase), the greater the influence of the quadratic interpolation term. As an example, in figure 6.4(a) we depict the theoretical inverse PD gain for the Saleh model using 8 coefficients obtained from the expression (2.39). Then, using these coefficients, we construct the resulting inverse AM/AM transfer characteristic according to (6.26). The inverse AM/AM transfer so obtained appears in figure 6.4(b) where the zooming box shows the 2^{nd} order interpolation effect with respect to the linear link of the estimated coefficients of the PD curve. Hence, figure 6.4(c) shows the absolute errors associated to the linear (used for reference) and 2^{nd} order AM/AM PD approaches with respect to the theoretical PD curve.

Since $c_k = c_{k-1} + \Delta_k$, rearranging terms in equation (6.26) we obtain

$$b_p = e^{j\alpha_y}\left(u_y\left(\alpha_{k-1} - \frac{c_{k-1}(\alpha_k - \alpha_{k-1})}{\Delta_k}\right) + u_y^2\left(\frac{\alpha_k - \alpha_{k-1}}{\Delta_k}\right)\right) \tag{6.27}$$

$$= e^{j\alpha_y}\left(u_y\left(\alpha_k - \frac{c_k(\alpha_k - \alpha_{k-1})}{\Delta_k}\right) + u_y^2\left(\frac{\alpha_k - \alpha_{k-1}}{\Delta_k}\right)\right) \tag{6.28}$$

where we can clearly observe that as well as $u_y \to c_{k-1}$ in (6.27) or $u_y \to c_k$ in (6.28), the resulting quadratic term will tend to a null value while the gain will be specified by the corresponding coefficient α_{k-1} or α_k.

6.1.5 Optimum PD set calculation

The former definitions, though generals, use notation restricted to the pre/post distortion of single scalar inputs. Hereafter we will consider that at any instant $[n]$ an input signal column vector

$$\mathbf{b}_x[n] = \left[\ b_x[n] b_x[n-1] \cdots b_x[n-N+1]\ \right]^T = [b_{x1} \cdots b_{xN}]^T$$

corresponding to a single sampled OFDM symbol in base-band, is fed into the adaptive system of figure 6.1(a). Then, the column vectors $\mathbf{b}_y[n]$ and $\mathbf{b}_p[n]$ will define the discrete output of the HPA and the PD block respectively. According to the notation used in figure 6.1, in the forthcoming derivations the dependence of signal vectors on discrete-time $[n]$ will be assumed implicit in order to simplify the expressions. Thus, equation (6.7) can be extended to vector notation using the same PD gain coefficients vector $\boldsymbol{\alpha}$ to pre/post distort the elements within the column vector $\mathbf{b}_y = [b_{y1} \cdots b_{yN}]^T$ as follows:

$$\mathbf{b}_p = \mathrm{diag}\{\boldsymbol{\Lambda}(\mathbf{u}_y)\boldsymbol{\alpha}\}\mathbf{b}_y \tag{6.29}$$

or equivalently

$$\mathbf{b}_p = \mathrm{diag}\{\mathbf{b}_y\}\boldsymbol{\Lambda}(\mathbf{u}_y)\boldsymbol{\alpha}. \tag{6.30}$$

where $\mathbf{u}_y = [|b_{y1}| \cdots |b_{yN}|]^T$. According with the previous definitions given in section 6.1.3, in the above expressions for PD, the input-modulus dependent matrix $\boldsymbol{\Lambda}(\mathbf{u}_y)$ is a type 1 activation matrix also called the *coefficients activation matrix* and defined by

$$\boldsymbol{\Lambda}(\mathbf{u}_y) = \begin{bmatrix} \boldsymbol{\lambda}_1^T \\ \vdots \\ \boldsymbol{\lambda}_N^T \end{bmatrix} \tag{6.31}$$

which is an $(N \times N_b)$ sparse array containing a fuzzy pattern of the activation of coefficients in $\boldsymbol{\alpha}$, which is dependent (for each instant $[n]$) on the N-length input modulus vector $\mathbf{u}_y$. The i-th row in $\boldsymbol{\Lambda}(\mathbf{u}_y)$ corresponds to the activation vector $\boldsymbol{\lambda}_i^T$ which is dependent on the modulus of the i-th element in $\mathbf{u}_y$ according to (6.9). Thence, $\boldsymbol{\Lambda}(\mathbf{u}_y)$ is an *information distribution matrix* (IDM) such that each row should fulfil the condition (6.10), while the sum along its columns gives a set of N_c values that compose the soft histogram [61] approximation to the bin activation probability which can be considered an estimation of the PDF of the input modulus u_y. Now, assuming that the activation matrix $\boldsymbol{\Lambda}$ is

implicitly dependent[2] on the vector $\mathbf{u}_y$ and by defining the diag$\{\cdot\}$ vectorial operation, the expressions (6.29) and (6.30) can be further developed as

$$\mathbf{b}_p = \text{diag}\{\mathbf{\Lambda}\boldsymbol{\alpha}\}\mathbf{b}_y = \left(\sum_{i=1}^{N} \boldsymbol{\delta}_i \boldsymbol{\delta}_i^T \mathbf{\Lambda}\boldsymbol{\alpha}\boldsymbol{\delta}_i^T\right)\mathbf{b}_y \tag{6.32}$$

and

$$\mathbf{b}_p = \text{diag}\{\mathbf{b}_y\}\mathbf{\Lambda}\boldsymbol{\alpha} = \left(\sum_{i=1}^{N} \boldsymbol{\delta}_i \boldsymbol{\delta}_i^T \mathbf{b}_y\boldsymbol{\delta}_i^T\right)\mathbf{\Lambda}\boldsymbol{\alpha} \tag{6.33}$$

where

$\boldsymbol{\delta}_i$: size $(N \times 1)$, is a pinning vector (i-th term set to one and the rest to zero) used to build $(N \times N)$ diagonal matrices from a column vector as shown in (6.32) and (6.33).

$\mathbf{\Lambda}$: size $(N \times N_c)$, is the coefficients activation matrix.

$\boldsymbol{\alpha}$: size $(N_c \times 1)$, is the PD complex gain coefficients vector.

Our purpose is now to formulate the optimization of $\boldsymbol{\alpha}$ and, therefore, we select the expression in (6.33) for convenience. Let us define the diagonal matrix

$$\mathbf{D}_y = \text{diag}\{\mathbf{b}_y\} = \sum_{i=1}^{N} \boldsymbol{\delta}_i \boldsymbol{\delta}_i^T \mathbf{b}_y \boldsymbol{\delta}_i^T. \tag{6.34}$$

Hence, the error signal for the simplified system in figure 6.1(a) can be expressed as the N-length column vector

$$\mathbf{e}_{post} = \mathbf{b}_x - \mathbf{D}_y\mathbf{\Lambda}\boldsymbol{\alpha}. \tag{6.35}$$

Now, the squared error $|\mathbf{e}_{post}|^2 = \mathbf{e}_{post}^H\mathbf{e}_{post}$ is given by

$$\begin{aligned} |\mathbf{e}_{post}|^2 &= (\mathbf{b}_x - \mathbf{D}_y\mathbf{\Lambda}\boldsymbol{\alpha})^H(\mathbf{b}_x - \mathbf{D}_y\mathbf{\Lambda}\boldsymbol{\alpha}) \\ &= \mathbf{b}_x^H\mathbf{b}_x - \mathbf{b}_x^H\mathbf{D}_y\mathbf{\Lambda}\boldsymbol{\alpha} - \boldsymbol{\alpha}^H\mathbf{\Lambda}^T\mathbf{D}_y^H\mathbf{b}_x \\ &+ \boldsymbol{\alpha}^H\mathbf{\Lambda}^T\mathbf{D}_y^H\mathbf{D}_y\mathbf{\Lambda}\boldsymbol{\alpha}. \end{aligned} \tag{6.36}$$

The minimization of $|\mathbf{e}_{post}|^2$ in terms of $\boldsymbol{\alpha}$ can be formulated by evaluating the gradient $\nabla_{\boldsymbol{\alpha}^*}|\mathbf{e}_{post}|^2$. Thus, from (6.36), we have

$$\nabla_{\boldsymbol{\alpha}^*}|\mathbf{e}_{post}|^2 = -\mathbf{\Lambda}^T\mathbf{D}_y^H\mathbf{b}_x + \mathbf{\Lambda}^T\mathbf{D}_y^H\mathbf{D}_y\mathbf{\Lambda}\boldsymbol{\alpha} \tag{6.37}$$

$$= -\mathbf{\Lambda}^T\text{diag}\{\mathbf{b}_y^*\}\mathbf{b}_x + \mathbf{\Lambda}^T\text{diag}\{\mathbf{b}_y\mathbf{b}_y^H\}\mathbf{\Lambda}\boldsymbol{\alpha} \tag{6.38}$$

$$= -\mathbf{\Lambda}^T(\mathbf{b}_y^* \odot \mathbf{b}_x) + \mathbf{\Lambda}^T\text{diag}\{\mathbf{R}_{yy}\}\mathbf{\Lambda}\boldsymbol{\alpha} \tag{6.39}$$

[2]In general $\mathbf{\Lambda}$ will be assumed dependent on the input modulus vector at the PD block.

where $\odot$ stands for the element by element Shur-Haddamard product. Note that in equation (6.39), only the main diagonal of the autocorrelation matrix $\mathbf{R}_{yy}$ is considered. Then, using the equivalence $\text{diag}\{\mathbf{b}_y\mathbf{b}_y^H\} = \text{diag}\{\mathbf{b}_y^*\}\text{diag}\{\mathbf{b}_y\} = \mathbf{D}_y^H\mathbf{D}_y$, we set $\nabla_{\boldsymbol{\alpha}^*}|\mathbf{e}_{post}|^2 = 0$ in (6.38) obtaining

$$\begin{aligned} -\boldsymbol{\Lambda}^T\text{diag}\{\mathbf{b}_y^*\}\mathbf{b}_x + \boldsymbol{\Lambda}^T\text{diag}\{\mathbf{b}_y\mathbf{b}_y^H\}\boldsymbol{\Lambda}\boldsymbol{\alpha} = 0 \\ -\boldsymbol{\Lambda}^T\mathbf{D}_y^H\mathbf{b}_x + \boldsymbol{\Lambda}^T\mathbf{D}_y^H\mathbf{D}_y\boldsymbol{\Lambda}\boldsymbol{\alpha} = 0 \\ -(\mathbf{D}_y\boldsymbol{\Lambda})^H\mathbf{b}_x + (\mathbf{D}_y\boldsymbol{\Lambda})^H(\mathbf{D}_y\boldsymbol{\Lambda})\boldsymbol{\alpha} = 0 \end{aligned}$$

thus resulting in the final optimum set of PD gain coefficients

$$\boxed{\boldsymbol{\alpha}_{opt} = \left[(\mathbf{D}_y\boldsymbol{\Lambda})^H(\mathbf{D}_y\boldsymbol{\Lambda})\right]^{-1}(\mathbf{D}_y\boldsymbol{\Lambda})^H\mathbf{b}_x} \tag{6.40}$$

Now, provided that the squared error given in (6.36) corresponds to a quadratic cost function with respect to the coefficients $\boldsymbol{\alpha}$, a typical LMS updating structure can be considered for the iterative estimation of the optimum set of PD coefficients. The general updating equation is

$$\boxed{\boldsymbol{\alpha}_{n+1} = \boldsymbol{\alpha}_n - \mu\nabla_{\boldsymbol{\alpha}^*}|\mathbf{e}_{post}|^2} \tag{6.41}$$

where the gradient of the squared error is required for each iteration. Thence, in order to provide to the PD algorithm a more direct measure of the gradient, involving less computational load than calculating (6.39) at each iteration, an alternative expression for $\nabla_{\boldsymbol{\alpha}^*}|\mathbf{e}_{post}|^2$ is needed. Therefore, from (6.37) and (6.35) the gradient can be expressed as

$$\begin{aligned} \nabla_{\boldsymbol{\alpha}^*}|\mathbf{e}_{post}|^2 &= -\boldsymbol{\Lambda}^T\mathbf{D}_y^H(\mathbf{b}_x - \mathbf{D}_y\boldsymbol{\Lambda}\boldsymbol{\alpha}) \\ &= -\boldsymbol{\Lambda}^T\mathbf{D}_y^H\mathbf{e}_{post} \\ &= -\boldsymbol{\Lambda}^T\text{diag}\{\mathbf{b}_y^*\}\mathbf{e}_{post} = -\boldsymbol{\Lambda}^T(\mathbf{b}_y^* \odot \mathbf{e}_{post}). \end{aligned} \tag{6.42}$$

This new expression can be easily calculated from the signal vectors acquired at each iteration and clearly implies less operations than (6.39). Hence, the updating equation for the estimation of $\boldsymbol{\alpha}_{opt}$ is finally

$$\boxed{\boldsymbol{\alpha}_{(n+1)} = \boldsymbol{\alpha}_{(n)} + \mu\boldsymbol{\Lambda}^T(\mathbf{b}_y^* \odot \mathbf{e}_{post})} \tag{6.43}$$

To perform this algorithm, the distribution of the centroids is first defined as uniform within the input range to the PD, $[0, A_{sat}]$. This implies that the saturation input

amplitude is a known data. The maximum input amplitude A_{sat} is sometimes provided as a nominal design parameter for each specific HPA. Nevertheless, some drifts in the HPA characteristics and specifically in the saturation point are normally expected due to several factors like working temperature, power supply level, average input power, output load matching, etc. Therefore, it is recommendable that the estimation of the input A_{sat} be frequently updated to suitably design the activation functions distribution along the input amplitude axis.

Results evaluating the influence of the adaptation step μ, the IBO at the input of the PD, the number of intervals and the distribution of the input modulus are shown later in chapter 7.

6.2 Advanced adaptive strategies for PD

In the previous section we operated a basic optimization algorithm to search for the best set of PD gain coefficients $\boldsymbol{\alpha}$ that minimizes the MSE. In the PD scheme therein presented, the centroids were uniformly distributed along the valid input range of the PD processor $[0, A_{sat}[$. However, the uniform distribution of symmetrical activation functions of type 1 does not account for the probability distribution of the input signal modulus employed by the adaptive algorithm. As a result of this, we observed a slow adaptation of those gain coefficients associated to centroids located at low probability regions. Moreover, the MSE converged (again slowly) in average to a floor also presenting large error peaks due to the presence of non-compensated low probability inputs.

The number of centroids can be seen as a limited resolution resource that must be optimally distributed. Using a minimum number of centroids to achieve a given level of nonlinear distortion compensation reduces the computational complexity in the PD estimation and execution. In this sense, another implementation goal will be to use signal blocks of reduced length to avoid huge increments in the computational load due to the calculation of the activation matrices.

In the results obtained for the basic PD scheme with fixed and uniformly distributed centroids, once the MSE has converged to a floor, we still observe a high peak residual error which cannot be further reduced since the coefficients $\boldsymbol{\alpha}$ have been optimized for the given fixed set of centroids. Comparatively, the probability of activation in some intervals is dramatically inferior than in others. Specifically, intervals which are closer to the saturation region present a very low activation probability. These bins require a long stream of input samples to assure the minimum activation that provides good conditions to update their respective gain coefficients through the adaptation algorithm. Low-probability input values are responsible for the peak error events which in turn are associated to important degradations in the data transmission performance (BER, MER). A first conclusion arises from considering these results:

The use of symmetrical activation functions of type 1 (triangular) with a uniform distribution of centroids to span the input dynamic range of the PD is not optimal for signals whose envelope has a non-uniform PDF.

A first strategy to confirm this hypothesis would be to adapt the centroid allocation according to the PDF of the input signal modulus. Maintaining the total number of centroids fixed, a higher density of centroids could be defined for those regions of the input range where the input probability density is larger. In order to completely cover the input range, including the lowest probability regions, the positions of the outermost centroids can be defined as constants at $c_0 = 0$ and $c_{N_\mathcal{I}} = A_{sat}$ while the remaining centroids are adapted. Simulations (see results in appendix 6.B) evaluating this first strategy reported worse results than the case where a uniform centroid distribution was used. The cost

function converged to a higher mean error floor even featuring higher peak error values due to the excessive concentration of resolution points at high probability regions in detriment of the lower probability regions of the input. The only improvement observed was a reduction in the peak to mean error ratio. In the aforementioned test, according to the Rayleigh distribution considered, very few centroids were assigned to interpolate the near-to-saturation region of the PD curve. At this region the AM/AM PD characteristics of HPAs normally feature significant and rapid variations with respect to input variations. Therefore, interpolation of such part of the curve using few coordinate points will logically involve a high fitting error when low probability inputs activate these wider intervals.

These considerations on the use of the PDF as the only criterion to place centroids lead us to the conclusion that the optimum allocation of centroids along the input amplitude axis defined in figure 6.2 depends mainly on two factors:

a) The probability density function of the input signal amplitude to the PD.

b) The nonlinear characteristic of the HPA.

When regarding the minimization of the MSE (or the error power), some trade-offs associated with the above factors can be pointed out:

- A denser grid of centroids in those regions along the amplitude range where the signal envelope has more probability density will result in more resolution and better approximation of the PD curve. This may, of course, reduce the error power for input events taking place in such regions.

- A denser grid is, however, only necessary in the high probability regions if the PD characteristic under estimation features rapid variations in amplitude or phase. A flat PD response within a given region of the input can be well approximated by linear interpolation using few coefficients, thence, although the input PDF is high within such intervals, squandering resolution resources by assigning an excessive number of centroids to approach the PD in that region is not justified.

Hence, minimization of the error power may require higher interpolation resolution in regions of smaller probability where the PD characteristic presents higher local variations. The question that we could formulate is then: What is the combination between the distribution of N_c centroids in $\mathbf{c}$ and an equal number of gain coefficients in $\boldsymbol{\alpha}$ that provides the minimum error power to interpolate the inverse nonlinear characteristics in our PD scheme? To solve this, in this section we propose an adaptive algorithm formulation for the joint estimation of these two vectors (centroids and gain coefficients) with a minimum MSE criterion. The reader will probably realize that the oncoming proposal can be related to some vector quantization criteria as those included in [63] where it is shown, in general terms, how an optimal nonlinear interpolation can be

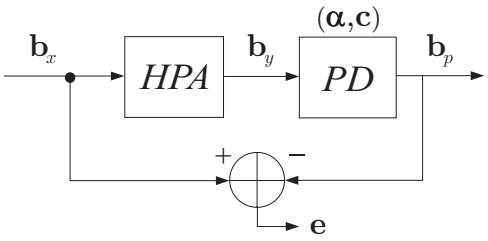


Figure 6.5: Block diagram of the PD system. The post-distortion block operation depends on: the estimated gain coefficients set $\boldsymbol{\alpha}$, and the centroids distribution $\mathbf{c}$.

formulated at a low complexity cost and disregarding ad-hoc interpolation techniques.

To begin, let us redefine in more detail the model on which our formulation will be based on. This is shown in the figure 6.5 and corresponds to the post-distortion operation described in (6.7). The digital post-distorter in the figure estimates the set of gain coefficients $\boldsymbol{\alpha}$ that, in association to a set of activation functions $\boldsymbol{\lambda}$ (type 1), define the instantaneous I/O error

$$e = b_x - b_y \, \boldsymbol{\lambda}^T(u_y) \, \boldsymbol{\alpha} \tag{6.44}$$

and minimize the following cost function

$$J = \mathrm{E}\left\{ \left| \, b_x - b_y \, \boldsymbol{\lambda}^T(u_y) \, \boldsymbol{\alpha} \, \right|^2 \right\} \tag{6.45}$$

which is the scalar expression for the expected I/O error power. Here, the discrete time dependence with $[n]$ has been assumed for the sake of brevity. Furthermore, the activation functions, as defined previously in section 6.1.3, feature a dependence on the input modulus to the PD block, $u_y = |b_y|$. This will be made explicit (as $\lambda(u_y)$ above) whenever needed for algorithm interpretation. Thus, this dependence will usually remain implicit along the forthcoming derivations in order to abbreviate the expressions.

After some algebra, the scalar cost function in (6.45) can be converted to

$$J = \mathrm{E}\left\{|b_x|^2\right\} - 2\mathrm{Re}\left\{\boldsymbol{\alpha}^T \mathrm{E}\left\{b_x^* b_y \boldsymbol{\lambda}\right\}\right\} + \boldsymbol{\alpha}^H \mathrm{E}\left\{|b_y|^2 \boldsymbol{\lambda}\boldsymbol{\lambda}^T\right\} \boldsymbol{\alpha}. \tag{6.46}$$

Now, denoting the expected input power as $\sigma_x^2 = \mathrm{E}\left\{|b_x|^2\right\}$ and defining

$$\mathbf{r} = \mathrm{E}\left\{b_x^* b_y \boldsymbol{\lambda}\right\} \tag{6.47}$$

and

$$\mathbf{R} = \mathrm{E}\left\{|b_y|^2 \boldsymbol{\lambda}\boldsymbol{\lambda}^T\right\} \tag{6.48}$$

the expression of the cost function becomes

$$J = \sigma_x^2 - 2\mathrm{Re}\left\{\boldsymbol{\alpha}^T \mathbf{r}\right\} + \boldsymbol{\alpha}^H \mathbf{R} \boldsymbol{\alpha} \tag{6.49}$$

$$= \sigma_x^2 - \boldsymbol{\alpha}^H \mathbf{r}^* - \boldsymbol{\alpha}^T \mathbf{r} + \boldsymbol{\alpha}^H \mathbf{R} \boldsymbol{\alpha}. \tag{6.50}$$

The gradient with respect to the set of gain coefficients can be expressed from this last expression as

$$\nabla_{\boldsymbol{\alpha}^H} J = R\boldsymbol{\alpha} - \mathbf{r}^*. \tag{6.51}$$

Setting this gradient to zero we obtain the optimum

$$\boldsymbol{\alpha}_o = \mathbf{R}^{-1}\mathbf{r}^* \tag{6.52}$$

for a fixed set of centroids $\mathbf{c}$. Nonetheless, here it is important to note that this optimum set of gain coefficients is expressed in terms of $\mathbf{R}$ and $\mathbf{r}$ which in turn are dependent on the centroid distribution $\mathbf{c}$ as it is clear from the definitions given in (6.47) and (6.48) since the expected activation $\boldsymbol{\lambda}(u_y)$ will be determined by the centroid positions with respect to the PDF of the input modulus.

Now, from the definitions in (6.48) and (6.9), the elements in $\mathbf{R} = \mathbf{R}^T$ are all real valued and $(\mathbf{R}^{-1})^H = (\mathbf{R}^{-1})^T = \mathbf{R}^{-1}$. Using these properties we replace (6.52) in (6.50) obtaining

$$J_{min}(\mathbf{c})\big|_{\boldsymbol{\alpha}_o} = \sigma_x^2 - \mathbf{r}^H\mathbf{R}^{-1}\mathbf{r} \tag{6.53}$$

which minimizes J for the $\boldsymbol{\alpha}_o$ determined for a specific centroid distribution $\mathbf{c}$. Therefore, we can aim to further reduce the value obtained through (6.53). This can be done by operating over the term $M(\mathbf{c}) = \mathbf{r}^H\mathbf{R}^{-1}\mathbf{r}$ where a new set of centroids $\mathbf{c}$ can be found in order to maximize

$$\mathbf{c} = \text{argmax}_{\mathbf{c}}(M(\mathbf{c})) = \text{argmax}_{\mathbf{c}}(\mathbf{r}^H\mathbf{R}^{-1}\mathbf{r}).$$

Along with this first criterion, we can introduce in the cost function (6.50) the relation found in (6.52), now rearranged as $\mathbf{r} = \mathbf{R}\boldsymbol{\alpha}_o^*$, thus obtaining an alternative expression for the error minimization

$$J_{min}(\mathbf{c}, \boldsymbol{\alpha}) = \sigma_x^2 - \boldsymbol{\alpha}^H\mathbf{R}\boldsymbol{\alpha} \tag{6.54}$$

where the optimum $\boldsymbol{\alpha}$ now depend on the centroid distribution in $\mathbf{c}$ through the matrix $\mathbf{R}(\mathbf{c})$. This new approach suggests an alternate maximization of

$$M(\mathbf{c}, \boldsymbol{\alpha}) = \boldsymbol{\alpha}^H\mathbf{R}\boldsymbol{\alpha}$$

where, if $M(\mathbf{c}, \boldsymbol{\alpha})$ is increased through iterations, $J(\mathbf{c}, \boldsymbol{\alpha})$ will converge since it is by definition positive. For this purpose, a general alternate iteration procedure for the minimization of the difference of non-constant terms in (6.49) is now proposed. First, for a given set $\mathbf{c}_{n-1}$ (where n is the iteration index) we have $\mathbf{R}_{n-1}$ and $\mathbf{r}_{n-1}$. At the next iteration we can maintain, for instance, the centroid set fixed, that is, $\mathbf{c}_n = \mathbf{c}_{n-1}$. Hence, we update the vector $\boldsymbol{\alpha}$ so that $\boldsymbol{\alpha}_n = \boldsymbol{\alpha}_{n-1} + \Delta_\alpha$. Thus, $\boldsymbol{\alpha}_n$ can be calculated to minimize

$$\boldsymbol{\alpha}_n = \text{argmin}_{\boldsymbol{\alpha}} \left(\boldsymbol{\alpha}^H\mathbf{R}_n\boldsymbol{\alpha} - 2\text{Re}\left\{\boldsymbol{\alpha}^H\mathbf{r}_n\right\}\right) \tag{6.55}$$

where $\mathbf{R}_n = \mathbf{R}_{n-1}$ and $\mathbf{r}_n = \mathbf{r}_{n-1}$. Then, maintaining the gain coefficients fixed at the next iteration, that is, $\boldsymbol{\alpha}_{n+1} = \boldsymbol{\alpha}_n$, the set of centroids is updated so that $\mathbf{c}_{n+1} = \mathbf{c}_n + \Delta_{\mathbf{c}}$. This is done according to the minimization

$$\mathbf{c}_{n+1} = \text{argmin}_{\mathbf{c}} \left(\boldsymbol{\alpha}_n^H\mathbf{R}\boldsymbol{\alpha}_n - 2\text{Re}\left\{\boldsymbol{\alpha}_n^H\mathbf{r}\right\}\right). \tag{6.56}$$

Note that the general procedure described above results similar to the Generalized Lloyd-Max algorithm for Vector Quantization [64][65] where alternate minimization is performed between the representation vectors, in this case $\boldsymbol{\alpha}$, and the assignment regions, which are determined here by the centroid vector $\mathbf{c}$. This scheme is specially suited to the PD estimation previously described in section 6.1.5, where the adaptation of $\boldsymbol{\alpha}$ according to (6.55) has been already defined through equations (6.40) to (6.43). Therefore, we will now lay the foundations for the required procedure for the general minimization expressed in (6.56).

6.2.1 Optimization of centroid distribution

In this section our interest will be centered in finding a model to estimate the gradient of the cost function J expressed in (6.45) with respect to the centroid distribution. Finding such an analytical model will allow us to implement the alternate optimization expressed in (6.56). For this purpose, let us first consider the gradient of the cost function with respect to the k-th element in the centroid vector. Thus, the partial derivative $\nabla_{c_k} J(\mathbf{c}) = \frac{\partial J(\mathbf{c})}{\partial c_k}$, derived from (6.46), gives

$$\nabla_{c_k} J(\mathbf{c}) = \mathrm{E}\left\{ |b_y|^2 \nabla_{c_k}(\boldsymbol{\alpha}^H \boldsymbol{\lambda}\boldsymbol{\lambda}^T \boldsymbol{\alpha}) - 2\mathrm{Re}\left\{ \boldsymbol{\alpha}^T b_x^* b_y \nabla_{c_k}(\boldsymbol{\lambda}) \right\} \right\}.$$

Then, developing the factor

$$\nabla_{c_k}(\boldsymbol{\alpha}^H \boldsymbol{\lambda}\boldsymbol{\lambda}^T \boldsymbol{\alpha}) = 2\mathrm{Re}\left\{ \boldsymbol{\alpha}^T \nabla_{c_k}(\boldsymbol{\lambda}) \boldsymbol{\lambda}^T \boldsymbol{\alpha}^* \right\}$$

this gradient becomes

$$\begin{aligned} \nabla_{c_k} J(\mathbf{c}) &= \mathrm{E}\left\{ 2\mathrm{Re}\left\{ |b_y|^2 \boldsymbol{\alpha}^T \nabla_{c_k}(\boldsymbol{\lambda}) \boldsymbol{\lambda}^T \boldsymbol{\alpha}^* \right\} - 2\mathrm{Re}\left\{ \boldsymbol{\alpha}^T b_x^* b_y \nabla_{c_k}(\boldsymbol{\lambda}) \right\} \right\} \\ &= 2\mathrm{Re}\left\{ \mathrm{E}\left\{ \boldsymbol{\alpha}^T \nabla_{c_k}(\boldsymbol{\lambda}) \left(b_y \boldsymbol{\lambda}^T \boldsymbol{\alpha} - b_x \right)^* b_y \right\} \right\} \\ &= -2\mathrm{Re}\left\{ \mathrm{E}\left\{ \boldsymbol{\alpha}^T \nabla_{c_k}(\boldsymbol{\lambda}) e^* b_y \right\} \right\} \end{aligned} \tag{6.57}$$

where including a straight measure of the conjugated error reduces the problem to finding a closed expression for the gradient $\nabla_{c_k}(\boldsymbol{\lambda}(u_y))$. This can be done by resorting to the definition of the activation functions of type 1 in terms of the functions of type 2. By observing the relations between type 1 and type 2 activation functions, it is clear that the gradient $\nabla_{c_k}\boldsymbol{\lambda}$ can be evaluated as

$$\nabla_{c_k}\boldsymbol{\lambda} = \nabla_{c_k} f_k^+ + \nabla_{c_k} f_k^- + \nabla_{c_k} f_{k+1}^+ + \nabla_{c_k} f_{k+1}^-.$$

The separate gradient terms showing the dependence on u, the input modulus to the PD, are given by

$$\begin{aligned}
\nabla_{c_k} f_k^+ &= \nabla_{c_k}\left(\frac{u-c_{k-1}}{c_k-c_{k-1}}\right) = -\frac{u-c_{k-1}}{(c_k-c_{k-1})^2} = -\frac{1}{\Delta_k} f_k^+ \\
\nabla_{c_k} f_k^- &= \nabla_{c_k}\left(\frac{c_k-u}{c_k-c_{k-1}}\right) = \frac{u-c_{k-1}}{(c_k-c_{k-1})^2} = \frac{1}{\Delta_k} f_k^+ \\
\nabla_{c_k} f_{k+1}^+ &= \nabla_{c_k}\left(\frac{u-c_k}{c_{k+1}-c_k}\right) = \frac{u-c_{k+1}}{(c_{k+1}-c_k)^2} = -\frac{1}{\Delta_{k+1}} f_{k+1}^- \\
\nabla_{c_k} f_{k+1}^- &= \nabla_{c_k}\left(\frac{c_{k+1}-u}{c_{k+1}-c_k}\right) = -\frac{u-c_{k+1}}{(c_{k+1}-c_k)^2} = \frac{1}{\Delta_{k+1}} f_{k+1}^-.
\end{aligned}$$

As seen in section 6.1.3, each activation function of type 2 is specifically associated to one gain coefficient from $\boldsymbol{\alpha}$ according to the index k and the sign of its slope. Therefore, the product $\boldsymbol{\alpha}^T \nabla_{c_k}(\boldsymbol{\lambda})$ included in (6.57) can be calculated using the expressions above as

$$\begin{aligned}
\nabla_{c_k}(\boldsymbol{\alpha}^T\boldsymbol{\lambda}) &= \alpha_k \nabla_{c_k} f_k^+ + \alpha_{k-1}\nabla_{c_k} f_k^- + \alpha_{k+1}\nabla_{c_k} f_{k+1}^+ + \alpha_k \nabla_{c_k} f_{k+1}^- \\
&= -\frac{\alpha_k}{\Delta_k} f_k^+ + \frac{\alpha_{k-1}}{\Delta_k} f_k^+ - \frac{\alpha_{k+1}}{\Delta_{k+1}} f_{k+1}^- + \frac{\alpha_k}{\Delta_{k+1}} f_{k+1}^- \\
&= \frac{(\alpha_{k-1}-\alpha_k)}{\Delta_k} f_k^+ + \frac{(\alpha_k-\alpha_{k+1})}{\Delta_{k+1}} f_{k+1}^-.
\end{aligned}$$

Inserting this result into the gradient expression in (6.57) we obtain

$$\nabla_{c_k} J(\mathbf{c}) = -2\mathrm{Re}\left\{\mathrm{E}\left\{b_y e^* \left(\frac{(\alpha_{k-1}-\alpha_k)}{\Delta_k} f_k^+ + \frac{(\alpha_k-\alpha_{k+1})}{\Delta_{k+1}} f_{k+1}^-\right)\right\}\right\} \tag{6.58}$$

where we observe that the gradient with respect to the k-th centroid considers two activation contributions: from its left and from its right. The centroid gradient is steered by the conjugate error and its sensitivity is proportional to the coefficient differences at each interval normalized to their respective bin widths. However, if we consider the calculation of this gradient at each learning step, we note that the activation of f_k^+ and f_{k+1}^- never takes place simultaneously. Therefore at each iteration only one of the terms in (6.58) will contribute to updating the centers through the steepest descent algorithm or to approximate the expectation in (6.58) through averaging.

We tested the applicability of this expression through simulations and found that this structure features some major disadvantages which are pointed out below.

- The variations of the gradient with respect to the centroid c_k have not been assumed to be influenced in any way by the variations with respect to other centroids. This

is equivalent to operating over each centroid position independently from the rest and assuming the risk of breaking the basic relations between centroids which, by definition, should be a monotonously increasing set of values within a well defined range $[0, A_{sat}]$.

- In the formulation, we restricted $c_1 = 0$ and $c_{N_\mathcal{I}} = 1$. Under adequate conditions, this fixes the positions of the outermost centroids and allow the remaining intermediate ones to be iteratively updated within the valid range. However, the rest of the centroids have not been in any way restricted to remaining within $[0, A_{sat}]$ through the minimization process. Therefore, some of them may be forced by the gradient to run beyond saturation or to become negative values when the gradient tends to a local minimum or when the error diverges.

- The inclusion of a minimum inter-centroidal distance Δ_{min} has not been considered. Therefore, two or more consecutive or non-consecutive centroids could merge into a single one making the cost function converge to a suboptimal local minima.

- Furthermore, even if a Δ_{min} is considered between consecutive centroids, the structure defined for adaptation does not necessarily avoid eventual centroid-crossings which occur when the condition $c_{k-1} < c_k$ is no longer valid although the consecutive centroids maintain the minimum distance defined between them. Thus, for the PD estimator considered, which is based on linear interpolation, centroid scrambling is logically unacceptable.

- Concluding, the use of an alternate adaptive structure based on $\nabla_{c_k} J(\mathbf{c})$ requires some heuristics in order to guarantee the convergence of the cost function to the absolute minimum.

A good solution to these drawbacks can be implemented if, instead of the gradient with respect to c_k, shown in (6.57), we calculate the gradient with respect to an indirect measure of the inter-centroid distance from which the centroid positions can be recovered while the basic conditions of monotonicity and range are respected. For instance, if we use directly Δ_k, some of these problems are still present, however, when the evaluation of the argument in (6.57) is based on finding $\nabla_{\Delta_k}(\boldsymbol{\alpha}^T \boldsymbol{\lambda}(u))$ we obtain

$$\nabla_{\Delta_k}(\boldsymbol{\alpha}^T \boldsymbol{\lambda}(u)) = \boldsymbol{\alpha}^T \frac{\partial \boldsymbol{\lambda}(u)}{\partial \Delta_k} = \alpha_{k-1} \frac{\partial f_k^-(u)}{\partial \Delta_k} + \alpha_k \frac{\partial f_k^+(u)}{\partial \Delta_k}$$

and, since $f_k^+ + f_k^- = 1$, this can be equivalently expressed as

$$\nabla_{\Delta_k}(\boldsymbol{\alpha}^T \boldsymbol{\lambda}(u)) = (\alpha_k - \alpha_{k-1}) \frac{\partial f_k^+(u)}{\partial \Delta_k}. \tag{6.59}$$

Hence, the gradient of the cost function in (6.58) takes the form

$$\nabla_{\Delta_k} J(\mathbf{c}) = -2\mathrm{Re}\left\{ \mathrm{E}\left\{ b_y e^*(\alpha_k - \alpha_{k-1}) \frac{\partial f_k^+(u)}{\partial \Delta_k} \right\} \right\}. \tag{6.60}$$

This last expression considers only the contribution of the effectively activated interval for the calculation of the gradient estimation. In the next section we normalize Δ_k and use it in this last gradient construction to formulate in more detail all the needed restrictions that provide a robust convergence to the optimal PD estimation.

6.2.2 ELASTIC* centroid allocation algorithm

Three related types of activation functions were previously defined in section 6.1.3 for input amplitudes to the PD within the interval $[c_{k-1}, c_k[$. Now, preserving the structure shown in figure 6.2, let us redefine the distance between any pair of consecutive centroids as a positive non-null number given by

$$\Delta_k = |c_k - c_{k-1}| = \delta_k^2 + \Delta_{min} \tag{6.61}$$

where δ_k^2 is a control parameter for adapting the distance Δ_k, while Δ_{min} is a small constant that sets the minimum distance between consecutive centroids. Let the first centroid be fixed at the origin: $c_0 = 0$. Then, the remaining ones, from $k = 1$ to $k = N_{\mathcal{I}}$, can be calculated as

$$c_k = \sum_{\ell=1}^{k} (\delta_\ell^2 + \Delta_{min}) = k\Delta_{min} + \sum_{\ell=1}^{k} \delta_\ell^2. \tag{6.62}$$

Hence, the maximum input range in which the PD will operate is defined by fixing the last centroid as

$$c_{N_{\mathcal{I}}} = A_{sat} = \sum_{\ell=1}^{N_{\mathcal{I}}} \Delta_\ell = N_{\mathcal{I}}\Delta_{min} + \sum_{\ell=1}^{N_{\mathcal{I}}} \delta_\ell^2. \tag{6.63}$$

Each interval width can also be expressed in proportional terms when normalized with respect to the total amplitude range, this is expressed as

$$P_k = \frac{\Delta_k}{A_{sat}} = \frac{\delta_k^2 + \Delta_{min}}{\sum\limits_{\ell=1}^{N_{\mathcal{I}}} (\delta_\ell^2 + \Delta_{min})}. \tag{6.64}$$

The adaptive centroid allocation can be formulated in terms of these interval proportions since any updating of the set $\mathbf{P} = [P_1 \dots P_{N_{\mathcal{I}}}]$ will lead to a corresponding update of the centroid positions (and vice versa) through the following relation:

$$c_k = A_{sat} \sum_{\ell=1}^{k} P_\ell. \tag{6.65}$$

The purpose of iteratively adapting the proportional value of each interval instead of the interval width itself, is to maintain the total range $[0, A_{sat}]$ unmodified provided that the condition in (6.63) is true for any centroid distribution fulfiling $c_k > c_{k-1}$. This last condition was guaranteed when we included the minimum distance Δ_{min} in (6.61).

From (6.13) and (6.65), the positive slope function of type 2 that accounts for the activation of the k-th interval, can be expressed in terms of the proportions in $\mathbf{P}$ as

$$f_k^+ = \frac{u - \sum\limits_{\ell=1}^{k-1} P_\ell A_{sat}}{P_k A_{sat}}. \tag{6.66}$$

*ELastic Allocation of a Set of Two-dimensional Interpolation Centroids.

Equations (6.64) and (6.66) are key in that they introduce the inter-dependence between centroids. On the one hand, the definition of any P_k considers the summation of all interval widths as the normalization factor. On the other hand, the activation function f_k^+ is now defined considering a cumulative summation of the proportions of inferior intervals and not only the centroids limiting the k-th interval. Additionally, any increment in P_k will require a distributed reduction over the rest of the intervals to account for the condition in (6.63). As a result of this, the next formulation will produce, for each given activation function f_k^+, a gradient vector that will determine proportional updates in the inter-centroid distances Δ_k along the whole centroid distribution. This will produce the reallocation of all other centroids within the range. We will refer henceforth to this property as the *elastic* adaptation of the centroids, hence the name of the proposed algorithm.

With the given definitions, particularly since the centroids ultimately feature dependence on the parameter δ according to (6.62), we can reformulate the gradient in (6.60) now taking into account that any activated $f_k^+(u)$ features a dependence on variations in any interval width along the valid range. Thence, the gradient can now be expressed as

$$\nabla_{\delta_i} J(\mathbf{c}) = -2\mathrm{Re}\left\{\mathrm{E}\left\{b_y e^*(\alpha_k - \alpha_{k-1})\frac{\partial f_k^+(u)}{\partial \delta_i}\right\}\right\} \tag{6.67}$$

wherein, as we similarly defined in (6.59), we must provide a closed expression for

$$\nabla_{\delta_i}(\boldsymbol{\alpha}^T\boldsymbol{\lambda}) = (\alpha_k - \alpha_{k-1})\frac{\partial f_k^+}{\partial \delta_i}. \tag{6.68}$$

Here, for a given k, the gradient will consider the contributions from all the intervals. This is done by applying the index i from 1 through $N_{\mathcal{I}}$ and solving the partial derivative according to the chain rule

$$\frac{\partial f_k^+}{\partial \delta_i} = \sum_{\ell=1}^{N_{\mathcal{I}}} \frac{\partial f_k^+}{\partial P_\ell}\,\frac{\partial P_\ell}{\partial \delta_i}. \tag{6.69}$$

A detailed formulation for this chain rule is presented in appendix 6.A where the following intermediate solutions are found:

$$\frac{\partial f_k^+}{\partial P_\ell} = \begin{cases} 0 & ; \quad \ell > k \\ -\dfrac{1}{P_k} f_k^+ & ; \quad \ell = k \\ -\dfrac{1}{P_k} & ; \quad \ell < k \end{cases} \tag{6.70}$$

and

$$\frac{\partial P_\ell}{\partial \delta_i} = \begin{cases} -\dfrac{2\delta_i}{A_{sat}} P_\ell & ; \quad \ell \neq i \\ \dfrac{2\delta_i}{A_{sat}}(1 - P_\ell) & ; \quad \ell = i. \end{cases} \tag{6.71}$$

Thus, the final expression for (6.69) is the following piece-wise function:

$$\boxed{\frac{\partial f_k^+}{\partial \delta_i} = \frac{2\delta_i}{P_k A_{sat}} \cdot \begin{cases} f_k^+ P_k + \sum_{\ell=1}^{k-1} P_\ell - 1 & ; \quad \text{for } i < k \\ f_k^+ P_k + \sum_{\ell=1}^{k-1} P_\ell - f_k^+ & ; \quad \text{for } i = k \\ f_k^+ P_k + \sum_{\ell=1}^{k-1} P_\ell & ; \quad \text{for } i > k \end{cases}} \tag{6.72}$$

The updating of the centroid vector $\mathbf{c} = [c_0 \cdots c_{N_c}]$ is then implemented using equations (6.67),(6.68) and (6.72). Recall that, according to the range constraint, the positions of the first and the last centroids in $\mathbf{c}$ should not be modified. Therefore, the iterative process described in (6.56) can be suitably implemented by updating the intermediate distance parameters contained in the vector $\boldsymbol{\delta} = [\delta_1 \cdots \delta_{N_\mathcal{I}}]$ according to

$$\boxed{\boldsymbol{\delta}_{(n+1)} = \boldsymbol{\delta}_{(n)} - \mu \widehat{\nabla}_{\delta_i} J(\mathbf{c})} \tag{6.73}$$

with $\widehat{\nabla}_{\delta_i} J(\mathbf{c})$ the raw gradient estimation obtained by ignoring the expectation operator in (6.67). The centroids will be updated at each iteration with the adapted values within $\boldsymbol{\delta}$ through the relation in (6.62).

The initial conditions for the algorithm are stated by defining first the minimum distance as

$$\Delta_{min} = \varepsilon \frac{A_{sat}}{N_\mathcal{I}}$$

where we take a small proportion ε of the interval's width considering the uniform distribution of the N_c centroids along the range $[0, A_{sat}]$. Then, the elements within the vector $\boldsymbol{\delta}_{ini}$ are initialized as

$$\delta_{ini} = \sqrt{\frac{A_{sat}}{N_\mathcal{I}}(1-\varepsilon)}$$

Vectorial extensions of the presented equations were used to implement the alternate optimization algorithm as a block-oriented process considering an N-length OFDM symbol as the input to the adaptive system at each iteration. Results for a preliminary test evaluating the behaviour of the ELASTIC algorithm in presence of two different nonlinearities are shown next in the appendix 6.C.

6.A Appendix: Gradient evaluation for elastic allocation of centroids

Using

$$P_k = \frac{\Delta_k}{A_{sat}} = \frac{\delta_k^2 + \Delta_{min}}{\sum_{r=1}^{N_{\mathcal{I}}} (\delta_r^2 + \Delta_{min})} \tag{6.74}$$

and

$$f_k^+ = \frac{u - \sum_{r=1}^{k-1} P_r A_{sat}}{P_k A_{sat}} \tag{6.75}$$

we must evaluate

$$\frac{\partial f_k^+}{\partial \delta_i} = \sum_{\ell=1}^{N_{\mathcal{I}}} \frac{\partial f_k^+}{\partial P_\ell} \frac{\partial P_\ell}{\partial \delta_i}. \tag{6.76}$$

For $i \neq \ell$ we have

$$\left.\frac{\partial P_\ell}{\partial \delta_i}\right|_{i \neq \ell} = -\frac{2\delta_i(\delta_\ell^2 + \Delta_{min})}{\left(\sum_{r=1}^{N_{\mathcal{I}}} (\delta_r^2 + \Delta_{min})\right)^2} = -\frac{2\delta_i}{A_{sat}} P_\ell \tag{6.77}$$

and if $i = \ell$

$$\left.\frac{\partial P_\ell}{\partial \delta_i}\right|_{i = \ell} = \frac{2\delta_\ell}{\sum_{r=1}^{N_{\mathcal{I}}} (\delta_r^2 + \Delta_{min})} - \frac{2\delta_\ell(\delta_\ell^2 + \Delta_{min})}{\left(\sum_{r=1}^{N_{\mathcal{I}}} (\delta_r^2 + \Delta_{min})\right)^2} = \frac{2\delta_\ell}{A_{sat}}(1 - P_\ell) \tag{6.78}$$

In turn, for the evaluation of $\frac{\partial f_k^+}{\partial P_\ell}$ three conditions must be considered. First, by observing the definition in (6.75), it is clear that for $\ell > k$

$$\left.\frac{\partial f_k^+}{\partial P_\ell}\right|_{\ell > k} = 0 \tag{6.79}$$

while for $\ell = k$

$$\left.\frac{\partial f_k^+}{\partial P_\ell}\right|_{\ell = k} = -\frac{1}{P_k^2} \frac{\left(u - \sum_{r=1}^{k-1} P_r A_{sat}\right)}{A_{sat}} = -\frac{1}{P_k} f_k^+ \tag{6.80}$$

and for $\ell < k$

$$\left.\frac{\partial f_k^+}{\partial P_\ell}\right|_{\ell < k} = -\frac{1}{P_k}. \tag{6.81}$$

Then, from (6.76) and applying the intermediate results above, we obtain

$$\begin{aligned}\frac{\partial f_k^+}{\partial \delta_i} &= \sum_{\ell=1}^{k} \frac{\partial f_k^+}{\partial P_\ell}\,\frac{\partial P_\ell}{\partial \delta_i} \\ &= \frac{\partial f_k^+}{\partial P_k}\,\frac{\partial P_k}{\partial \delta_i} + \sum_{\ell=1}^{k-1} \frac{\partial f_k^+}{\partial P_\ell}\,\frac{\partial P_\ell}{\partial \delta_i} \\ &= -\frac{1}{P_k} f_k^+ \frac{\partial P_k}{\partial \delta_i} - \frac{1}{P_k} \sum_{\ell=1}^{k-1} \frac{\partial P_\ell}{\partial \delta_i}. \end{aligned} \tag{6.82}$$

To evaluate (6.82) we use (6.77) and (6.78). Thus, for $i = k$ we obtain

$$\begin{aligned}\left.\frac{\partial f_k^+}{\partial \delta_i}\right|_{i=k} &= \frac{1}{P_k} f_k^+ \frac{2\delta_k}{A_{sat}}(P_k - 1) + \frac{2\delta_k}{P_k A_{sat}} \sum_{\ell=1}^{k-1} P_\ell \\ &= \frac{2\delta_k}{P_k A_{sat}} \left(f_k^+ P_k + \sum_{\ell=1}^{k-1} P_\ell - f_k^+ \right) \end{aligned} \tag{6.83}$$

while, for $i < k$

$$\begin{aligned}\left.\frac{\partial f_k^+}{\partial \delta_i}\right|_{i<k} &= \frac{1}{P_k} f_k^+ \frac{2\delta_i P_k}{A_{sat}} - \frac{1}{P_k} \left(\frac{\partial P_i}{\partial \delta_i} + \sum_{\ell \neq i} \frac{\partial P_\ell}{\partial \delta_i} \right) \\ &= f_k^+ \frac{2\delta_i}{A_{sat}} - \frac{1}{P_k} \left(\frac{2\delta_i}{A_{sat}}(1 - P_i) + \frac{2\delta_i P_i}{A_{sat}} - \sum_{\ell=1}^{k-1} \frac{2\delta_i P_\ell}{A_{sat}} \right) \\ &= \frac{2\delta_i}{P_k A_{sat}} \left(f_k^+ P_k + \sum_{\ell=1}^{k-1} P_\ell - 1 \right) \end{aligned} \tag{6.84}$$

and for $i > k$ we get

$$\begin{aligned}\left.\frac{\partial f_k^+}{\partial \delta_i}\right|_{i>k} &= \frac{1}{P_k} f_k^+ \frac{2\delta_i P_k}{A_{sat}} + \frac{1}{P_k} \sum_{\ell=1}^{k-1} \frac{2\delta_i P_\ell}{A_{sat}} \\ &= \frac{2\delta_i}{P_k A_{sat}} \left(f_k^+ P_k + \sum_{\ell=1}^{k-1} P_\ell \right). \end{aligned} \tag{6.85}$$

Therefore,we can summarize

$$\frac{\partial f_k^+}{\partial \delta_i} = \begin{cases} \dfrac{2\delta_i}{P_k A_{sat}} \left(f_k^+ P_k + \displaystyle\sum_{\ell=1}^{k-1} P_\ell - 1 \right) & ; \quad \text{for } i < k \\ \dfrac{2\delta_i}{P_k A_{sat}} \left(f_k^+ P_k + \displaystyle\sum_{\ell=1}^{k-1} P_\ell - f_k^+ \right) & ; \quad \text{for } i = k \\ \dfrac{2\delta_i}{P_k A_{sat}} \left(f_k^+ P_k + \displaystyle\sum_{\ell=1}^{k-1} P_\ell \right) & ; \quad \text{for } i > k. \end{cases} \tag{6.86}$$

6.B Appendix: Test for centroid allocation using the PDF of the input signal to the PD

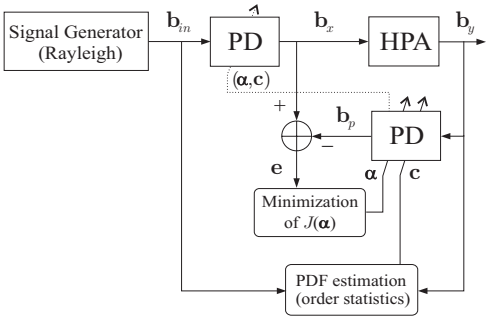
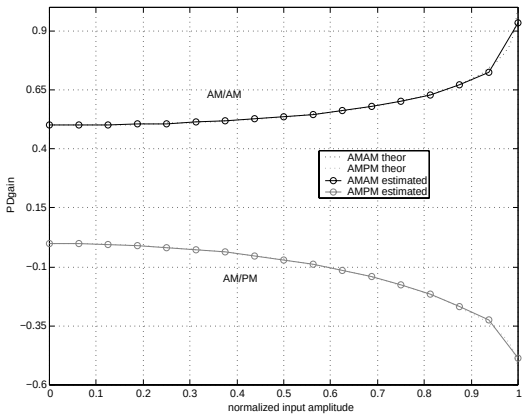
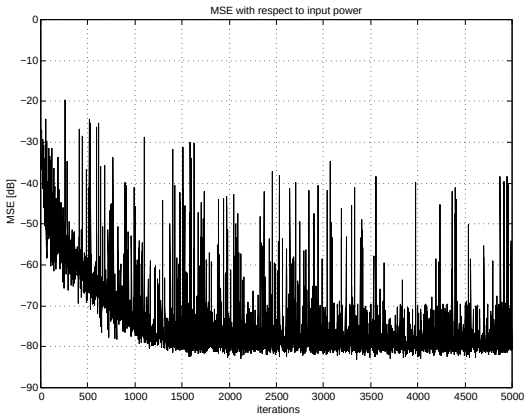


Figure 6.6: Scheme for adaptive PD where the optimization algorithm of section 6.1.5 is applied to estimate the PD gain coefficients $\boldsymbol{\alpha}$ while the centroid distribution is adapted according to a PDF estimation at the input of the PD block.

The test included in this appendix is solely intended to show that the use of the probability distribution of the input data as an absolute criterion for the centroid distribution adaptation is not optimal and can even lead to results worse than the use of a simple uniform centroid distribution to interpolate the PD curve. In figure 6.6 a signal generator is used to drive the input of a nonlinear HPA. For this block, we used the normalized Saleh model whose characteristics appear in section 2.2.1.

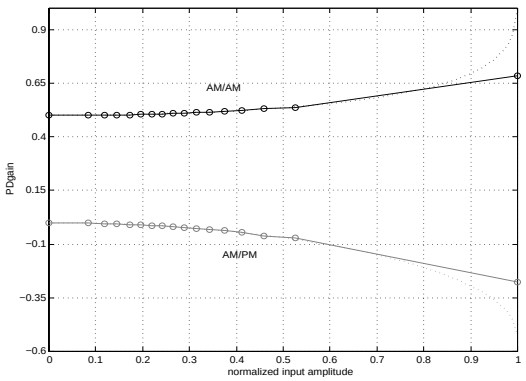


(a)

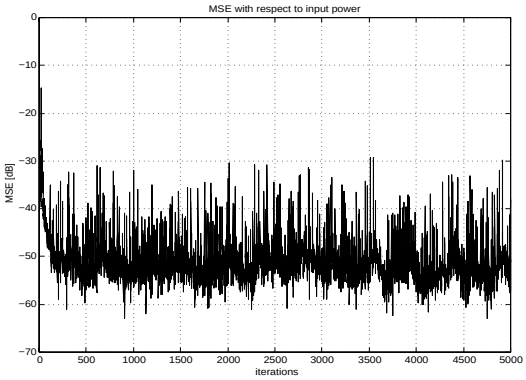


(b)

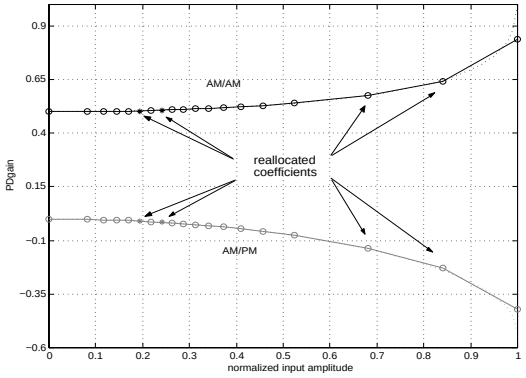
Figure 6.7: Test results for a uniform centroid allocation. In (a) we present the resulting set of estimated gain coefficients using a uniform distribution of $N_c = 17$ centroids. In (b) we observe the evolution of the mean square adaptation error signal $\mathbf{e}$ which is measured in dB with respect to the input power. The gain coefficients converge after 2500 iterations (each iteration considers a block with 128 data samples). Note that the MSE still registers high peak error values after convergence due to the large fitting errors near the saturation region.



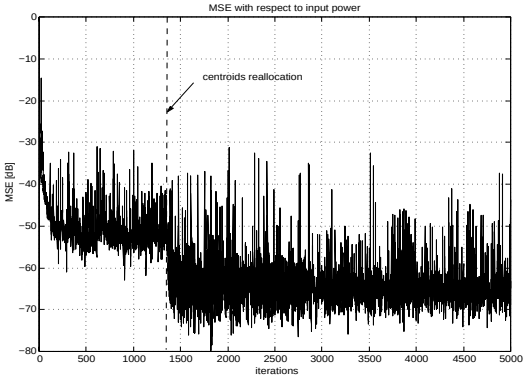
(a)



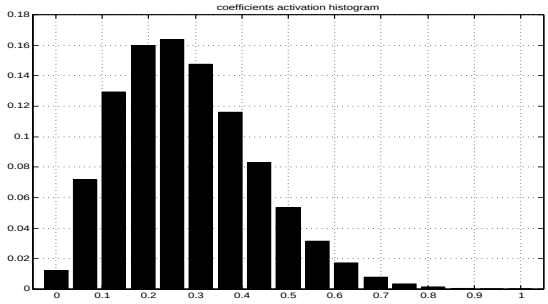
(b)



(c)



(d)



(e)

Figure 6.8: Test results for evaluating the PDF-based centroid allocation. In (b) we see the MSE evolution for the estimation of the PD gain coefficients of figure (a) where the adapted centroid distribution is denser in the higher probability regions. The MSE converges to a level 30 dB worse than the level achieved previously using the uniform centroid distribution of figure 6.7(a). Then, in (c) and (d) we tested the reallocation of two coefficients from the dense group to improve the resolution of the PD curve in the lower probability region. Two centroids were reallocated when the difference between the estimated input and output PDFs was under a given threshold at the iteration n=1360. This resulted in a significant improvement of the MSE, indicating that low probability signal values also introduce an important degradation of the adaptation error.

6.C Appendix: Test for ELASTIC using an input signal with uniform PDF and two different HPA models

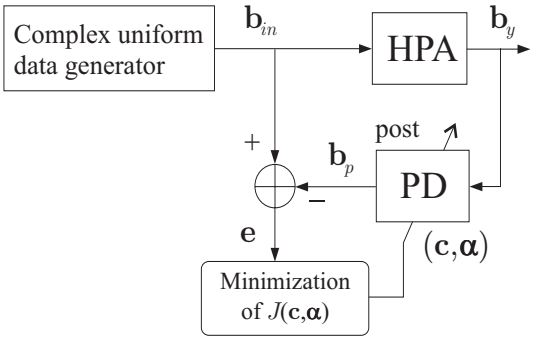


Figure 6.9: Basic adaptive architecture for the PD estimation of two different nonlinearities using the ELASTIC algorithm.

In this appendix, a preliminary test to observe the behaviour of the ELASTIC algorithm in presence of two different models of nonlinear HPAs is considered by applying, to the adaptive PD block in figure 6.9, a complex signal with uniform distribution in its modulus and phase components. The first HPA considered in this test is the Saleh model which has been extensively described in section 2.2.1. The multiplicative gain characteristics of such model, and the corresponding theoretical inverse gain curves for AM/AM and AM/PM PD, where shown in figures 2.5 and 2.3 and will be used here as a reference.

We also test an alternative to Saleh's model by replacing the AM/AM characteristic in (2.26) by a different nonlinearity while preserving the same AM/PM nonlinear distortion expressed in (2.27). The new resulting AM/AM nonlinearity is an arbitrary[3] nonlinear function given by

$$A_\kappa[u] = \frac{\arctan\left(\kappa\left(u-\frac{1}{2}\right)\right)}{2\arctan\left(\frac{\kappa}{2}\right)} + \frac{1}{2} \tag{6.87}$$

whose corresponding inverse transfer function is

$$A_\kappa^{-1}[u] = \frac{1}{\kappa}\left[\tan\left((2u-1)\arctan\left(\frac{\kappa}{2}\right)\right) + \frac{1}{2}\right]. \tag{6.88}$$

Then, according to the formulations in section 2.2.1, and specifically in equations (2.36) and (2.39), the direct and inverse gain associated to the above transfer expressions can

[3]It has not been designed to fit the characteristics of any specific nonlinear device.

be respectively obtained as

$$G_\kappa[u] = \frac{A_\kappa[u]}{u} \tag{6.89}$$

$$G_\kappa^{-1}[u] = \frac{A_\kappa^{-1}[u]}{u}. \tag{6.90}$$

Here it is worthy to mention that although the AM/PM characteristic for this alternative HPA model remains the same as Saleh's AM/PM presented in (2.27), the inverse AM/PM function is different since the AM/PM PD depends on the inverse AM/AM as expressed in (2.34). Thence, the inverse AM/PM for the new model is obtained as

$$\Psi_\kappa[u] = -\Phi\left[A_\kappa^{-1}[u]\right] \tag{6.91}$$

which does not correspond to the same inverse AM/PM for the Saleh model in (2.27) that can be obtained through (2.34).

Henceforth, we will refer to this new nonlinearity as the *Kappa* model. The AM/AM transfer characteristics for different values of κ are depicted in figure 6.10(a) while their corresponding direct and inverse gain curves appear in figures 6.10(b) and 6.10(d) respectively. For the curve identification test we choose $\kappa = 10$ whose associated theoretical transfer characteristics are depicted in figure 6.10(c). This setting will allow us to observe the behaviour of the coefficients facing a gain morphology significantly different in comparison to the PD gain shown in figure 2.5. Since the rapid variations of this new model coincide with the flat part of Saleh's PD gain, the coefficients and centroids may tend to move and concentrate in a distinct way according to the new nonlinearity.

In the figures included in this appendix, we observe the results obtained by driving the input of the PD with a signal $b_{in}[n] = u_{in}[n]e^{j\phi_{in}[n]}$ whose modulus and phase have uniform PDFs

$$P_{u_{in}} = \begin{cases} 1/M & ; \quad 0 \leq u \leq M \\ 0 & ; \quad \text{otherwise} \end{cases}$$

and

$$P_{\phi_{in}} = \begin{cases} 1/2\pi & ; \quad -\pi \leq \phi \leq \pi \\ 0 & ; \quad \text{otherwise} \end{cases}$$

respectively.

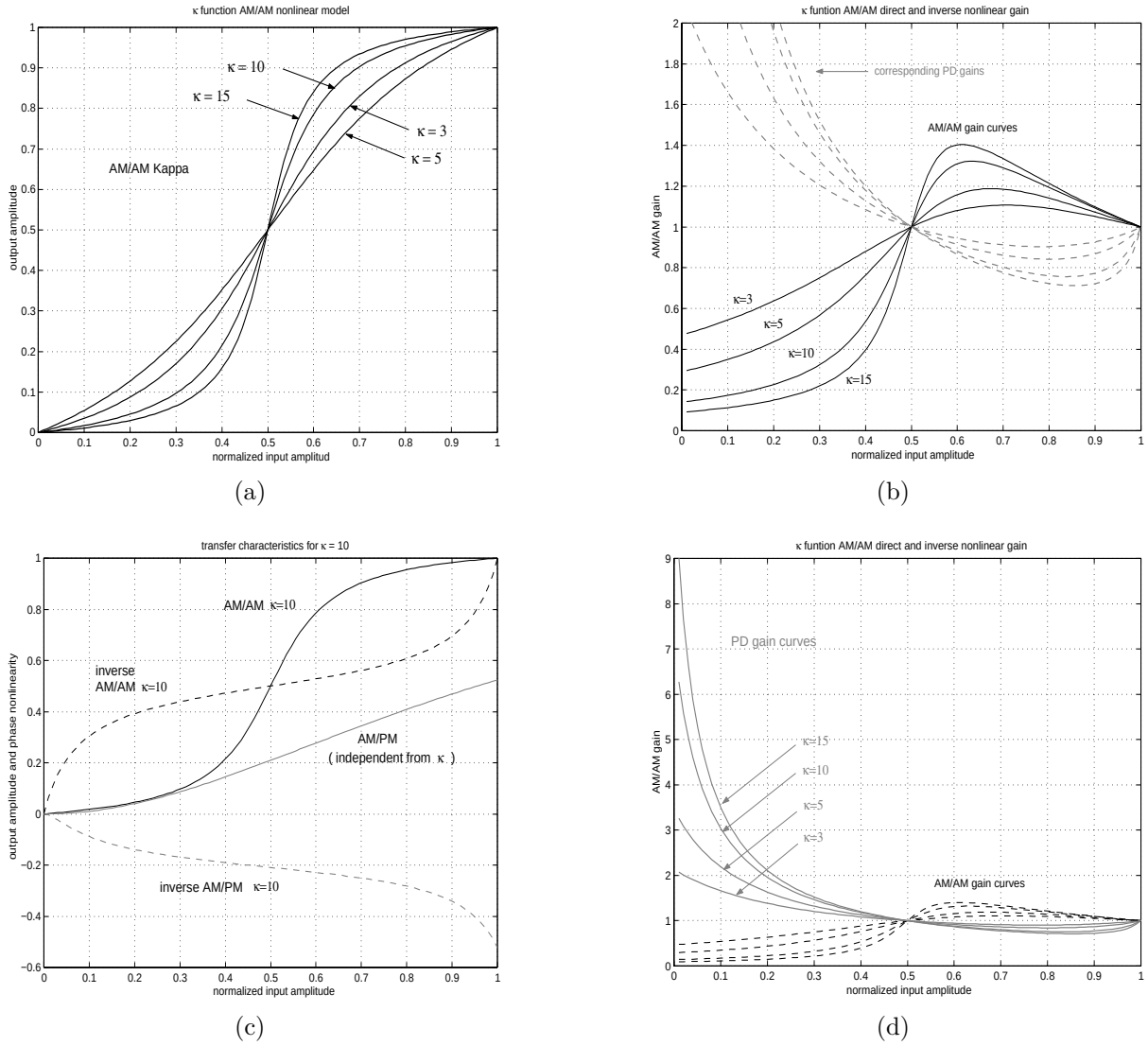


Figure 6.10: Alternative "κ" nonlinearity model to evaluate the ELASTIC PD algorithm. (a) AM/AM transfer curves for different values of the parameter κ. (b) the corresponding modulus-dependent gain curves. (c) direct and inverse transfer characteristics for $\kappa = 10$ (used in the present evaluation).(d) ideal PD gain curves associated to the set in figure (b).

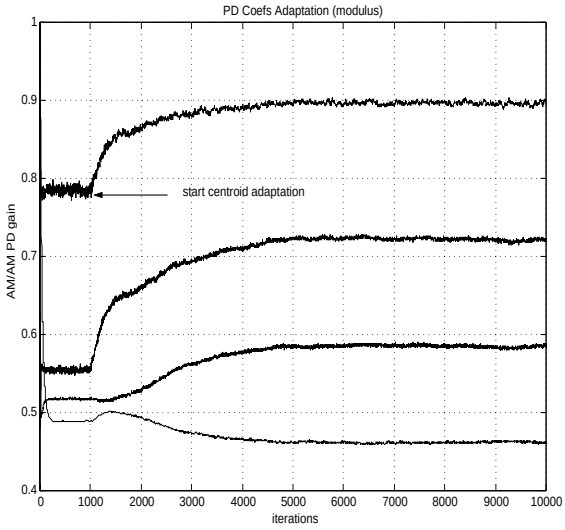


Figure 6.11: Evolution of AM/AM PD gain coefficients using a test input signal with uniform distribution in $[0, 1]$ and the Saleh model for the HPA. Note that the adaptation of these four centroids (which initially are uniformly spaced) starts at $n = 1000$.

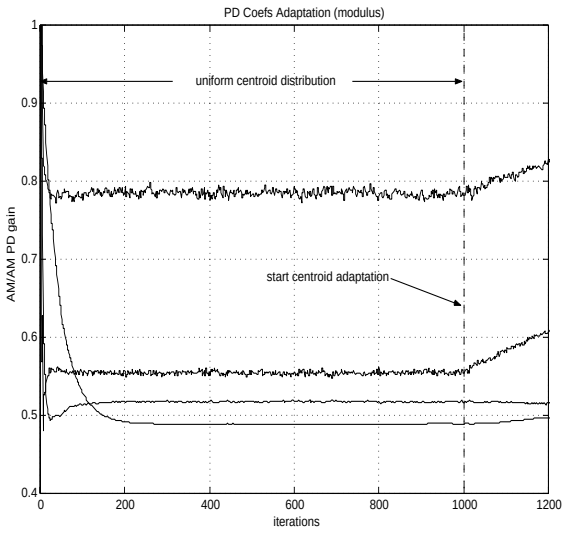


Figure 6.12: Adaptation of AM/AM PD coefficients before initiating centroid reallocation. The PD gain coefficients converge quite rapidly (500 iterations approximately) to their estimated optimum values.

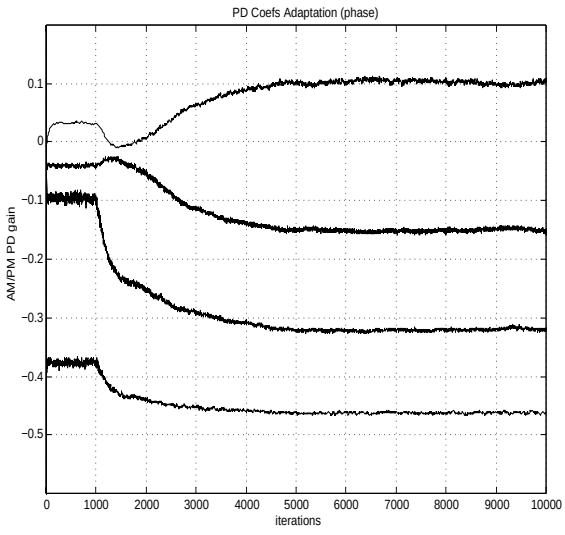


Figure 6.13: Evolution of the AM/PM PD gain coefficients associated to those in figure 6.11.

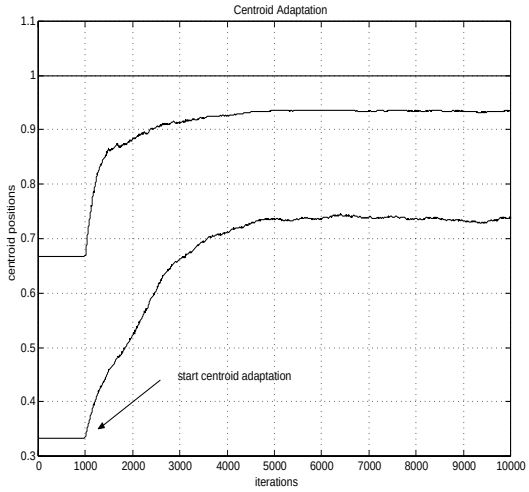
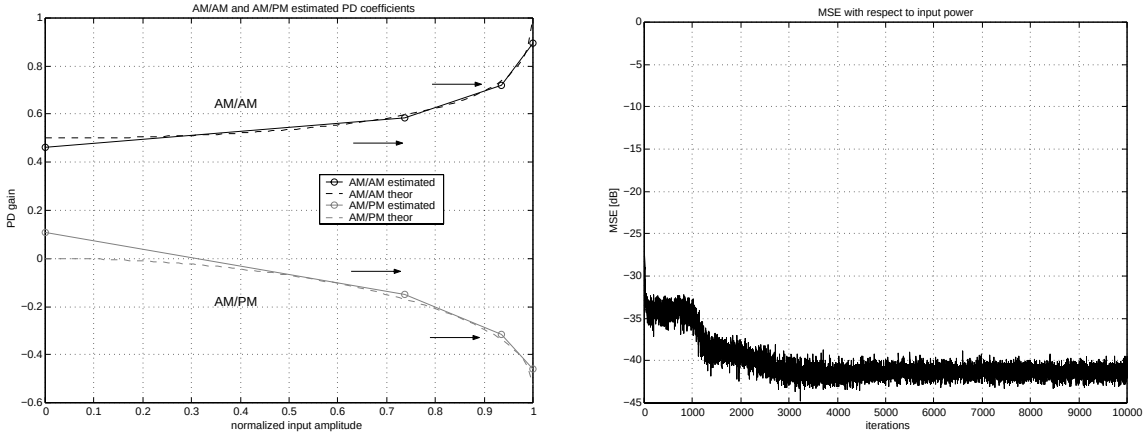


Figure 6.14: Adaptive reallocation of centroids using the uniform PDF and the Saleh model. Note that outermost centroids remain fixed and the inner two centroids adapt their position using ELASTIC.



(a) (b)

Figure 6.15: (a) Estimated PD gain coefficients for the uniform input test using the Saleh model for the HPA. Note that, since the input PDF in this case is uniform, the centroid density is determined by the variations of the HPA's characteristics, thus assigning low resolution to the flat part of the PD gain curve and higher resolution to the near-saturation zone. The theoretical PD gain characteristics, calculated from eqs.(2.39) and (2.40), are shown in dash line for reference. The circles mark the estimated PD coefficients related to the optimized centroid distribution. In (b), the adaptation error evolution illustrates the good convergence of the algorithm. During the first 1000 iterations, the centroids are fixed in a uniform initial distribution and only the gain coefficients are adapted. Then, the centroids become adaptive which results in a MSE 12dB better than the level achievable without centroid adaptation. Note that, the convergence times for the initial adaptation of the PD gain coefficients and centroids appearing in the figures, should be estimated considering only half of the iterations since the alternate adaptation has been performed from the beginning and, therefore, the coefficients $\boldsymbol{\alpha}$ where updated only at odd iterations $n = 1, 3, 5, \ldots$ for $n < 1000$.

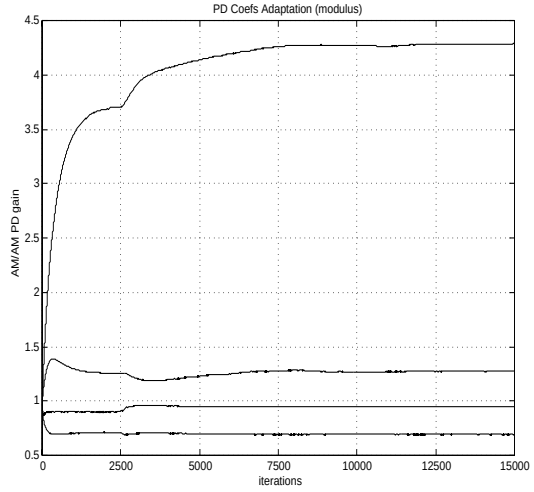


Figure 6.16: Evolution of the AM/AM PD gain coefficients using a test input signal with uniform distribution in $[0, 1]$ and the Kappa model for HPA. Note that the adaptation of centroids (which initially are uniformly spaced) starts at $n = 2500$.

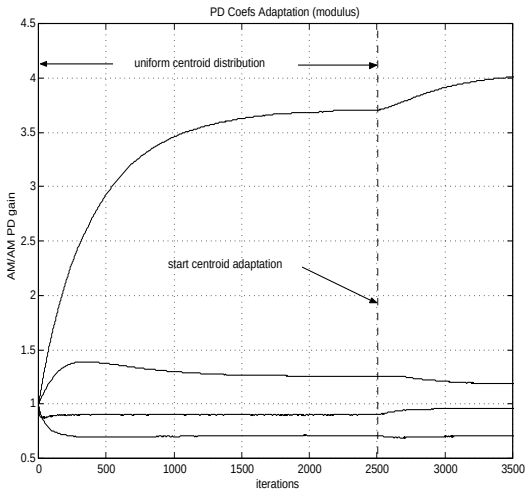


Figure 6.17: Adaptation of the AM/AM PD coefficients before initiating centroid reallocation. The PD gain coefficients approximately converged to the optimum values for the case of uniform centroid distribution.

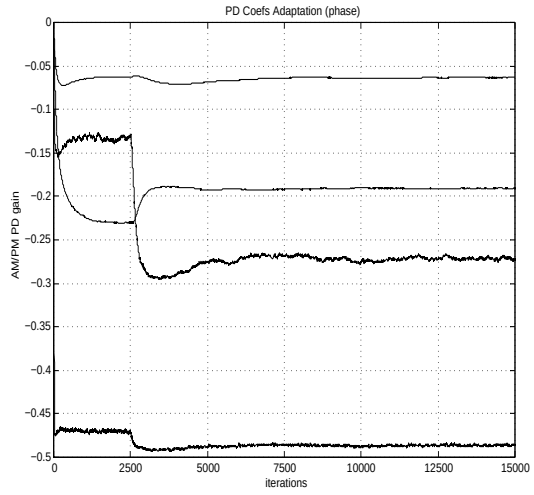


Figure 6.18: Evolution of the AM/AM PD gain coefficients associated to those in figure 6.16 for Kappa HPA model.

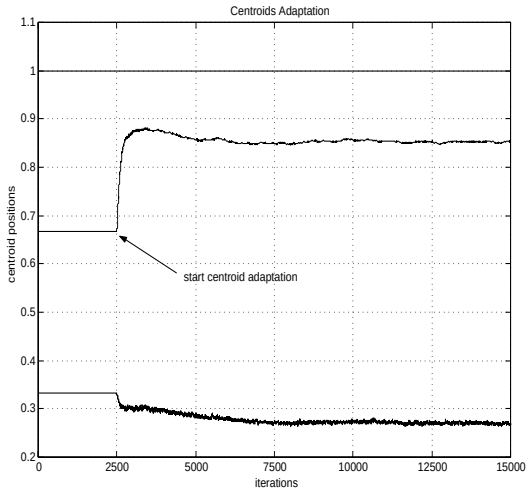
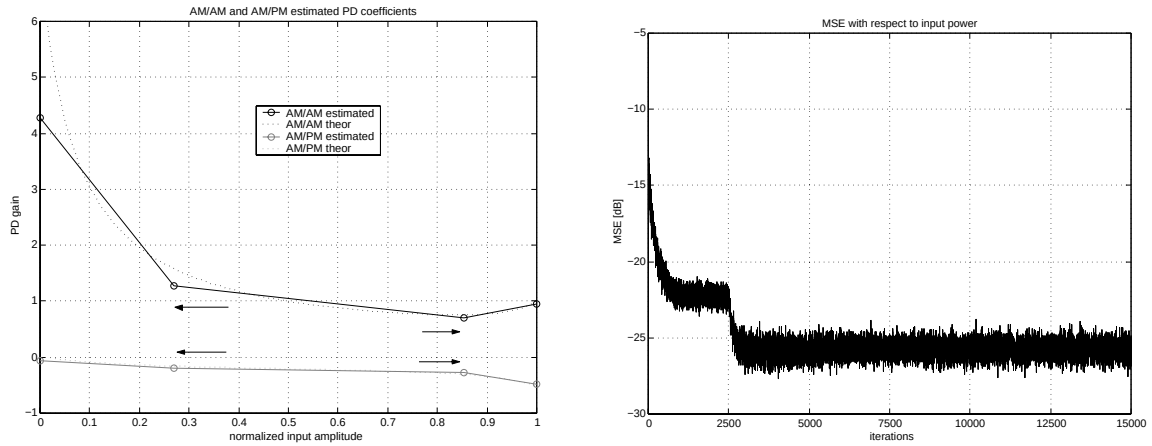


Figure 6.19: Centroid adaptation using the uniform PDF signal and Kappa nonlinear model for the HPA.



(a) (b)

Figure 6.20: (a) Estimated PD gain coefficients for the uniform input test with the Kappa nonlinearity. Note that, in this case, the centroid distribution is determined by the variations in the HPA's characteristics, thus assigning low resolution to the flat part of the PD gain curve, which is now in the middle, and higher resolution to both the low amplitude and near-saturation zones. Although the curve variation in this last region is much smaller, the error power associated to sub-compensated samples in this region is high and, therefore, the variations on the left part of the PD curve should be significantly more accentuated to force the second centroid to move leftwards. The arrows in the figure show the directions followed by the centroid adaptation to emphasize the difference with respect to the previous test using the Saleh model. The theoretical PD gain characteristics, calculated from eqs.(6.90) and (6.91), are shown in dash line for reference. The circles mark the estimated PD coefficients related to the optimized centroid distribution. In (b), the adaptation error evolution shows the convergence of the algorithm. During the first 2500 iterations, the centroids were fixed in a uniform initial distribution and only the gain coefficients were adapted. The MSE improvement in this case registered around 4dB with respect to the achievable level without adapting the centroids.

6.D Appendix: Equivalence between data and signal squared errors for PD estimation

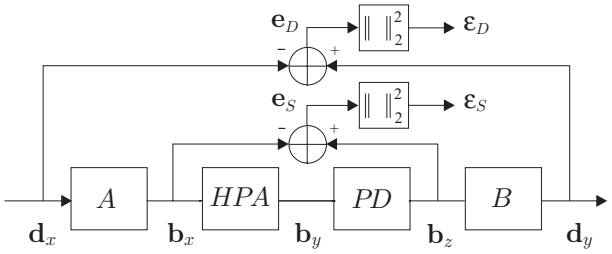


Figure 6.21: Block diagram of a general HPA linearization scheme. The adaptation of the post-distorter (PD) may depend on two different squared error measurements: M-QAM data (ε_D) or base-band signal (ε_S).

In figure 6.21, a basic linearization scheme is shown, where the input of a post distortion (PD) device is driven by the distorted base-band signal from the output of a non-linear HPA, thus expecting to obtain a global linear response. The base-band modulation and demodulation of M-QAM data is considered a general N-length block oriented discrete process (like the basic OFDM generation previously described), which is represented by the matrices $\mathbf{A}$ and $\mathbf{B}$ respectively. Then, from this structure, two different input-output error signal vectors with length N, can be acquired, namely: The M-QAM data error $\mathbf{e}_D = \mathbf{d}_y - \mathbf{d}_x$, and the base-band signal error, given by $\mathbf{e}_S = \mathbf{b}_z - \mathbf{b}_x$. Here, $\mathbf{d}_y = \mathbf{B}\mathbf{b}_z$ is the demodulated data at Rx, and $\mathbf{b}_x = \mathbf{A}^H\mathbf{d}_x$ is the base-band input vector to the HPA, which is expected to be identical to $\mathbf{b}_z$ after linearization with the PD. The square error can be then calculated from these data–signal error vectors as the squared norms: $\varepsilon_D = \mathbf{e}_D^H\mathbf{e}_D$ and $\varepsilon_S = \mathbf{e}_S^H\mathbf{e}_S$ respectively, providing two alternative scalar compensation parameters to adjust the post-distorter device through an adaptive algorithm.

In general, it is reasonable to say that the error signals used for PD coefficients adaptation should be defined in terms of the signals directly affected by the non-linear distortion under consideration. However, if in the case shown in figure 6.21, the modulation-demodulation process is characterized as a unitary-linear transformation, with the matrices $\mathbf{A}$ and $\mathbf{B}$ such that,

$$\mathbf{A}^H\mathbf{B} = \mathbf{B}\mathbf{A}^H = \mathbf{I}_{(N\times N)} \tag{6.92}$$

and

$$\mathbf{A}^H\mathbf{A} = \mathbf{B}^H\mathbf{B} = \mathbf{I}_{(N\times N)} \tag{6.93}$$

then, we can straightforwardly formulate the equivalence,

$$\begin{aligned} \varepsilon_D &= \mathbf{e}_D^H\mathbf{e}_D = \|\mathbf{d}_y - \mathbf{d}_x\|_2^2 = \|\mathbf{B}\,\mathbf{b}_z - \mathbf{d}_x\|_2^2 = \|\mathbf{B}\,[\,\underbrace{\mathbf{b}_z - \mathbf{A}^H\mathbf{d}_x}_{\mathbf{e}_S}\,]\|_2^2 \\ &= \|(\mathbf{B}\,\mathbf{e}_S)^H\mathbf{B}\,\mathbf{e}_S = \mathbf{e}_S^H\underbrace{\mathbf{B}^H\mathbf{B}}_{\mathbf{I}}\mathbf{e}_S = \mathbf{e}_S^H\mathbf{e}_S = \varepsilon_S. \end{aligned} \tag{6.94}$$

whence we may observe that, although the HPA and the PD operate over the base-band signal, the signal error ε_S can be used as a linearization performance parameter, equivalent to the data error ε_D. Thus, the use of base-band signal samples (time domain) to construct the adaptation error, leads us to a PD estimation that considers the correction of the non-linear distortion of the symbol constellation (frequency domain). This equivalence, based on the conditions (6.92) and (6.93), is in agreement with the OFDM signal model described in section 3.1.1 where the base-band modulation is implemented through the fast Fourier transform. When formulating such equivalence between error signals, we must be careful and recall that this property do not apply for the use of the data error ε_D to perform data PD. This is because data pre-distortion is a technique that, unlike the signal PD considered in this case, only compensates the distortion at the specific sampling times and, therefore, does not provide good compensation of the out-of band nonlinear distortion.