ADAPTIVE SIGNAL PROCESSING: A DISCUSSION OF TRADE-OFFS FROM THE PERSPECTIVE OF ARTIFICIAL LEARNING

A. R. Figueiras-Vidal, A. Artés-Rodríguez, J. Cid-Sueiro*, M. Martínez-Ramón
DSSR - ETSI Telecom, Universidad Politécnica de Madrid, Ciudad Universitaria, 28040 Madrid, Spain
Ph: +34 1 336 72 26; Fax: +34 1 336 73 50; E-Mail: anibal@gtts.ssr.upm.es
*DISA - ETSI Telecom, Universidad de Valladolid, C/Real de Burgos, s/n, 47011 Valladolid, Spain

ABSTRACT

Since many signal processing problems can be posed as sample-based decision and estimation tasks, we discuss how studies from other fields such as neural networks might suggest improved architectures (models) and algorithms for these types of problems. We then concentrate on PAM equalization, showing that a reordering of the equivalent classification problem generates a 'staircase' which, while retaining the simplicity of the classical equalizer, allows modifications to made in the outputs and in the training objectives which provide advantages even in the least complex cases.We go on to demonstrate that these advantages increase when one considers, for example, nonlinear channels with memory.

From our simulations we draw conclusions and suggest futher related research. We also present two new avenues of inquiry, offering significant practical advantages, which are motivated by the discussions.

1. INTRODUCTION

As is the case for many techniques in science and technology, signal processing deals with problems of decision and estimation, i.e., selecting an alternative among a countable or a continuous set. Examples arise in widely varying applications such as in radar detection or in time series prediction. Although for practical cases the operations can appear in more involved forms (vector quantization includes both deciding a class and defining a representative codeword), the elementary operations of decision and estimation are recognizeable.

From a statistical point of view, Bayesian formulations serve to solve these problems. A complete presentation of these theories can be found in [1]. Nevertheless, signal processing problems frequently appear in 'direct' forms, i.e., we can access only a limited number of examples of the problem to be solved and we must infer from these samples how to solve the problem. Thus, the system we design has to include a 'learning' strategy or algorithm for adapting itself to the specific case under consideration.

Learning is, in fact, a very complex task, and it can be considered at many levels and from many different perspectives. Since signal processing problems are solved by means of algorithms, an up-to-date general view of learning strategies, as presented in[2] is adequate. To be concise, algorithmic learning requires us to solve a compromise among three main issues:

-what is the structure to be used;

-what is the objective to be learned; and

-what search is to be done.

These issues offer trade-offs which cannot be separated from the specific application, such as obtainable generalization ability versus memorization capability, or static performance versus plasticity ('tracking' capabilities), or the global-local dilemma when selecting a search algorithm. From our point of view, the 'linear' hypothesis which is traditionally accepted in many signal processing tasks has precluded more extensive discussions addressing these aspects, but it seems clear that further refinements of many solutions cannot be reached without paying attention to these issues and trade-offs.

As an example, which we will discuss in more detail in Section 2, digital equalization is not a linearly separable classification problem (even considering minimum phase linear channels), nor is the square error criterion equivalent to minimizing the error probability. Furthermore, other objectives can be used for obtaining reasonable solutions for this minimization problem. Consequently, we can investigate whether other architectures, other objectives, and other algorithms provide advantages for equalization purposes. Similar comments can be made for many other signal processing problems, such as prediction or frequency estimation, which differ mainly according to their supervised or unsupervised character.

Fortunately, other fast developing areas have directed attention to these subjects; neural networks is the most prominent example since they constitute a reasonable tool to solve many signal processing problems (among others). We need not to detail here the huge number of papers which focus on the above points in this field: a simple look at general texts, such as [3], [4], [5], provides ample evidence. There is no possibility of synthesizing all signal processing problems in a tutorial paper. Consequently, we have decided to concentrate on digital equalization, and to illustrate how structures, objectives and algorithms, combined with a general concept of the equalization (decision, classification) tasks, can serve to open avenues providing practically usable methods which offer advantages with respect to the classical ones.

2. THE EQUALIZATION PROBLEM REVISITED

Channel equalization for digital transmission has been a very active research topic since the early proposals in the sixties. Classical approaches use a transversal filter at the receiver end, guided by the LMS (or its modifications such as NLMS) or the RLS (or its fast versions) algorithms, first in a supervised manner based on a known transmitted sequence, and followed by a decision-based supervision to track possible variations in the transmission path. There are many excellent overviews on this subject, such as [6], [7], [8]; therefore we will not discus the details.

On the other hand, there are many characteristics of the equalization problem which, although this is an established problem, have not until recently been considered in order to develop more efficient equalizer architectures.

First; equalization is a classification problem that, even in the case of minimum delay linear channels, has optimal separation surfaces different from the hyperplanes constructed by transversal filters [9]. In practical situations a low probability error is obtainable by means of transversal filters; furthermore, difficulties of sizing, convergence time, computational efforts and robustness appear when trying to use nonlinear structures such as neural networks [9], [10], [11], or even higher order/Volterra equalizers; these facts retarded advances with these architectures. Nevertheless, it must be noted that there are architectures which can be considered as combinations of linear filters, such as Jordan's Modular Neural Networks [12],[13]. These architectures alleviate many of the difficulties noted above(see [14], for example). Further, other neural network approaches to the optimal decision permit reasonably simplified alternatives [15].

Second; another infrequently addressed aspect of equalization is the selection of the adaption process objective; although minimizing the square error between the transmitted symbols (or decisions) is related to a low probability of error (which is the objective), it is not exactly equivalent, nor the only possible alternative. Again, satisfactory performance and simple implementations have closed this line of research. It is well known that probability of error is minimized by means of a 'Perceptron Rule' if the problem and structure allow a linear separation, which is not the case for digital communications. There are many appoximations to this, and a complete overview exceeds the bounds of this paper; we will mention only the use of L1-norm approaches [16], and the application of a Kullback-Leibler measure between the desired and the obtained outputs (since the optimal decision is equivalent to a comparison of *a posteriori* symbol probabilities) based on [17] [18]: this is frequently called entropy based training.

Following these ideas, attempts to improve the performance of conventional [19], neural net [20], or 'intermediate' [14] equalizers have been reported each showing slight but appreciable advantages (including sigmoidal soft decisions, which is mandatory in Kullback-Leiber based methods, and could help by itself, as seen in other studies [21]).

Perhaps the most interesting approach at the present time is to separate the use of elaborate structures from the 'objective' issue: this is the approach presented in [19]. However, the problem is that it is not obvious how to extend the approach to the more and more applied multilevel transmissions (accepting PAM as representative of complex constellations). That is, it is not easy to see how to modify the multilevel quantizer which serves as the hard decision step into an adequate soft version, because Potts activations (see [22] to understand their advantages) are not directly compatible with the classical transversal filter.

The rest of this paper discusses how one can combine an alternative view of equalization as a classification problem with the above ideas to realize a perspective with important advantages in digital equalization: this perspective allows us to both deal with the decision border separately and to directly include sigmoidal activations and entropy based training, thus allowing not only possibilities of controlling the complexity of the equalizer (in particular for practical nonlinear intersymbol interference cases), but also to take advantage of the characteristics of alternative algorithms with simple formulations. Section 3 presents the proposed structure and indicates its advantages; Section 4 details some simulation experiments which show that performance improvements are obtainable even in common situations. We close the paper with our conclusions and a series of suggestions for further research in this direction.

3. AN ALTERNATIVE TO PAM CLASSICAL EQUALIZERS AND ITS ADVANTAGES

In the practice, PAM transmissions consists of sending 2L equispaced signal values; say $\pm1,..., \pm2^{L-1}$. The classical equalizer for these signals is composed of a transversal filter, having the last R received values as inputs, followed by a uniform 2^L level quantizer, trained using an LMS algorithm to minimize the quadratic difference between the transmitted symbol and the output of the filter. It is obvious that this scheme will create a 'decision grid' of straight lines, as shown in Fig. 1 for L=4, R=2, and a channel having the nonlinearity o=th (i/10)/th(1/10) followed by $H(z)=1\text{-}02z^{-1}$.

The example demonstrates the (slight) flexibility which this scheme is allowed to provide: if there is a zero-memory, nonlinear distortion at the input of a linear channel, an (adaptive) nonuniform quantizer is sufficient to cope with the departure from the equispaced border of a linear channel.

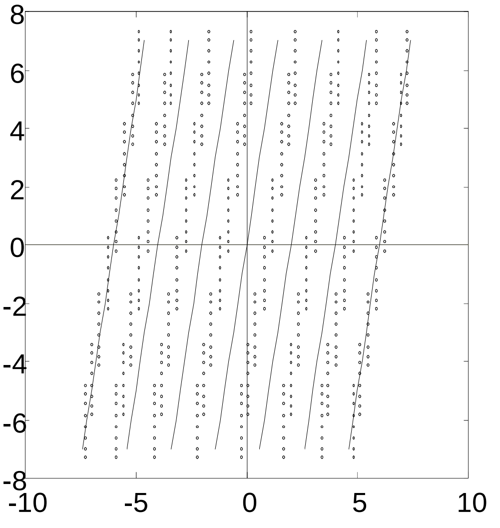

Fig 1: Classical equalization for a PAM transmission

Even though this scheme is not optimal, it is nearly optimal in many cases, provided that the intersymbol interference and the SNR are low or moderate. On the other hand, as previously mentioned, it is not clear how to apply a Kullback-Leibler measure as the objective, since we do not have (0, 1) (or (-1, 1)) separate outputs. Further, it is not clear how one might introduce 'soft' (sigmoidal) nonlinearities as might be needed to introduce an objective of the form

$$\sum_{j(outputs)} d_j \ln\left(\frac{d_j}{o_j}\right)$$

where $\{d_j\}$ are the desired and $\{o_j\}$ the obtained outputs.

To clarify, applying the above objective assumes that we decide on an output symbol via estimating the *a posteriori* symbol probabilities. While this is not strictly appropiate for general classification purposes, it is not unreasonable for equalization, since there are enough training samples to allow this estimation. In fact, LMS training uses this approach, see [23] for example. We argue that the Kullback-Leibler measure is a 'more clear' objective for this kind of estimation.

Furthermore, it is worth mentioning that there are many practical situations in which memory nonlinearities appear. For example, in radio transmission from a satellite or a base station, power amplifiers (TRWs) are present at the transmitter and they are memory nonlinear devices (see models in [24] [25]. Even if compensation schemes are applied, significent residual errors may subsist. In these cases, the borders shown in Fig. 1 must change to non-parallel lines, and even to nonlinear borders. Changes to this degree are not supported by conventional equalizers.

Now, let us consider the particular situation we are dealing with: we want to decide if a symbol has a value among $\pm 1, \ldots, \pm 2^{L-1}$. The decision can be presented as a 'cascade' of binary problems:

-first, what is the sign of the symbol?

-second, in which half of the positive/negative values is our symbol located?

etc. ('bit-by-bit'), until we locate the region of Fig. 1 in which we are operating.

Consequently, we can proceed using binary classifiers ('equalizers'). The first corresponds to the central border, the second to the central line of one of the semiplanes created by the previous border, etc.

All we need is to add to the conventional transversal 'equalizers' a bias term (except for the first one), to allow each border to depart from the origin. If, as usual, we are dealing with an (odd)-symmetric transmission, we must retain symmetry, by forcing corresponding coefficients to be equal and use biases with opposite signs for the cases corresponding to the same classification step. Additional degrees of freedom will introduce unnecessary convergence problems. Note that this follows the idea of 'hints' used in training other architectures [26] and has shown to provide advantages, and is routinely used in equalization.

In this way, i.e., constructing a 'staircase' of binary equalizers, we do not change any of the essential components of a conventional equalizer, but we open a series of interesting possibilities:

- nonlinear activations can be directly introduced;
- the entropic objective (and others) can be immediately used;
- in the case of general nonlinear channels, the structure allows the formation of a decision grid composed of nonparallel lines, which can improve error performance;
- we maintain reasonable convergence and moderate computational effort.

Additional comments could be made, but we will postpone them to the end of the paper because some options and conjectures require further analysis and experiments.

We will next show the results of simulations which demonstrate these advantages, even for the particular case corresponding to Fig. 1, a zero-memory nonlinearity plus a linear channel. This case is in the range of channels that can be managed by a 'conventional' equalizer, just to demonstrate that the approach is useful even in these cases. In other situations, the proposed approach can provide dramatic differences in performance.

4. A SIMULATION EXAMPLE

As noted, we consider a simple case which can be 'satisfactorily' solved, in principle, by a classical adaptive quantization transversal equalizer. The case is that corresponding to Fig 1: for N=4 (8-PAM), R=2, and a channel with a sigmoidal nonlinearity plus the linear filter $H(z)=1-0.2z^{-1}$.Initial values for the filter coefficients are (±0, ±2, ±4, ±6), 1, 0. We will apply the staircase equalizers with the constraint of parallel borders (both coefficients multiplying received samples are equal, and the bias terms are symmetric). We show simulations for:

- the usual LMS algorithm (thus, providing results nearly equivalent to the classical equalizer: only a discriminative training would be needed for equivalence);
- the LMS algorithm applied to outputs obtained through a hyperbolic tangent activation (reference level, 4);
- an instantaneous gradient minimization of the Kullback-Leibler measure, also using the nonlinear activation.

First of all, we choose a reasonable SNR, 30 dB, and we train the equalizers for different values of μ, changing from strictly supervised to decision based reference after convergence. Fig. 2 shown the performances (probability of error) versus selected μ (10^{-1} to 10^{-2}) for the three equalizers.

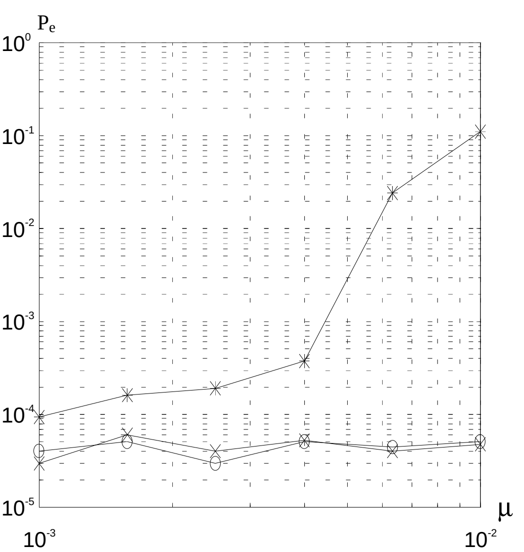

Fig. 2: Simulated performance of conventional (o), sigmoidal-LMS (*) and (sigmoidal) entropic equalizers (x) for the example.

It is obvious that, in the explorated margin, the second and the third equalizers offer clear advantages in (final) performance and robustness with respect to μ (this is partly due to the absence of discriminative training).

Nevertheless, note that the above is not a fair comparison since equal values of μ do not necessary result in the same rate of convergence for the different algorithms. Thus we need to be cautious before claiming 'overall' advantages.

To explore the convergence characteristics, we proceed as follows. We select two values of μ, μ_1=0.025 and μ_2=0.016, and apply the three algorithms in a supervised manner to equalize a sigmoidal plus linear channel, the linear part being $H(z)=1-0.05z^{-1}$, for 1000 samples (sufficient for approaching convergence); after this, H(z) is abruptly changed to $H(z)=1-0.2z^{-1}$, and training is done in the decision directed manner. We simulate 1000 realizations, and all of them that offer error frequences higher than $5\cdot10^{-2}$ 500 frequences after the change are eliminated, and the rest are averaged. P_e vs. training time is shown in Fig. 3 (μ_1=0.025) and Fig. 4 (μ_2=0.016).
The numbers of eliminated realizations are:

*μ_1=0.0025
- -LMS algorithm: 71
- -LMS with sigmoidal activation: 130
- -Entropic: 39

*μ_2=0.0016
- -LMS algorithm: 85
- -LMS with sigmoidal activation: 216
- -Entropic: 40

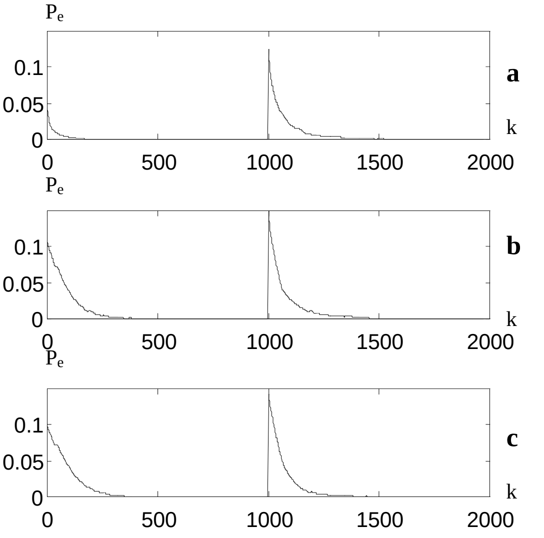

Fig. 3. Tracking capabilities of a: LMS; b: sigmoidal LMS; c: entropic equalizers for an abrupt change ($1-0.05z^{-1}$ to $1-0.2z^{-1}$); μ_1=0.0025 for all the cases, P_e vs. training steps (k) for selected realizations.

This demonstrates that convergence problems for the LMS and the entropic algorithms are similar while the (expected) final performance (P_e) for the second is clearly better, as predicted by Fig. 2 (although for converging cases there is a small advantage in speed for the first). With respect to a comparison between LMS with sigmoidal activation and entropic algorithms, when the (expected) performances are similar (same P_e), convergence problems are much more frequent for LMS.

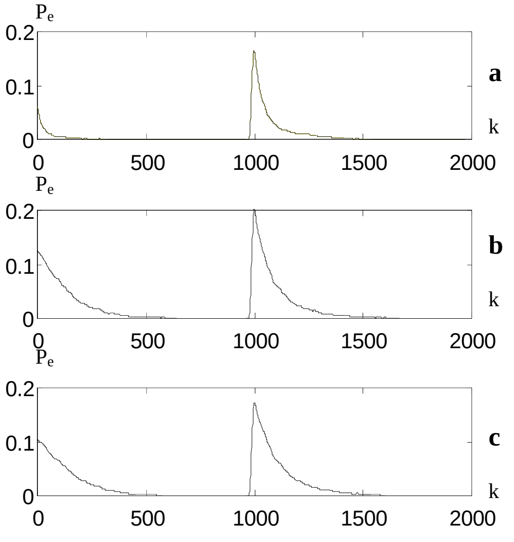

Fig. 4. Tracking capabilities of a: LMS; b: sigmoidal LMS; c: entropic equalizers for an abrupt change (1-$0.05z^{-1}$ to 1-02·z^{-1}); μ_1=0.0016 for all the cases, P_e vs. training steps (k) for selected realizations.

Thus, the potential advantage of using the staircase version appears clearly: using the Kullback-Leibler measure as the minimization objective provides the best balance between speed and final performance.These results are quantitatively similar to those of [20], but they utilize a more practical structure and include decision-based supervision.

The word 'robustness' which we have used above must not be misunderstood. If, for example, we repeat Fig. 2 for a SNR=25 dB, the main change is that the entropic equalizer does not converge for μ from 10^{-3} to 10^{-2}. Thus, it would seem that the equalizer is more sensitive to SNR changes. However, it is worth noting that the Pe values obtainable for this SNR and μ values by the other schemes are hardly of practical value.

5. CONCLUSIONS

Combining the concepts of equalization as a classification problem with the use of other objectives, we have introduced a 'staircase' linear-filter-based equalizer which keeps the basic computational effort of the classical scheme for PAM and shows advantage with respect to the speed vs. final performance compromise even in cases for which it can be said that the classical equalizer works 'properly'. It can be easily explained and shown that this advantage dramatically increases if the situation is not in the acceptable working range of a conventional equalizer, such as when memory nonlinearities are present at the transmission channel. In these situations, the 'parallelism' constraint imposed by the standard equalizer is more costly than the freedom obtained with the staircase approach (by training each symmetric pair of equalizers independently of the others).

6. FURTHER WORK

There are a number of immediate modifications of the proposed scheme, such as:

- using NLMS and other search algorithms;
- applying 'discriminative' training: i.e., correcting only those coefficients corresponding to the border nearer to the sample to be classified (and its symmetric border);
- establishing other objectives for training.

Furthermore, the staircase structure which we are proposing permits an analysis by near-immediate extension of known approaches.

We would also like to comment on two more involved lines which may prove especially useful:

1) In the case of nonlinear channels with memory, not only paralellism, but also linear borders for classification may become unacceptable.

Such a situation will force the introduction of complex structures in the place of the conventional equalizer. However, this is an 'excessive' approach. In practical cases difficulties arise mainly with the more 'extreme' borders, because nonlinearity effects are more pronounced for larger-valued (absolute value) received samples.

The staircase approach allows the introduction of (even simple) nonlinear equalizers which deal with these 'extreme' problems, while retaining the desired simplicity for the remainder of the elements.

2) Once one has opened the door to contemplate the equalization problem as a problem of classification, many other approaches are possible. A particularly interesting one (for its straighforward character) is to see the case as a Learning Vector Quantization task. Note that, when equalizing by means of a classical scheme, we are implicitly making decisions based on the proximity to the lines 'in the middle' of the borders. The idea of vector quantizing with a reference different from a point (which is needed for transmission coding) is not essentially different from traditional vector quantization (and, in particular, from supervised versions, both guided and decision-based). We can select different reference lines (in general,

hypersurfaces), even using regression over samples obtained by measuring the channel. There are a number of available tools to be used with this approach. In fact, we are getting some preliminary promising results.

Needless to say, to extend PAM results to complex constellation cases is not a difficult task.

REFERENCES

[1] H.L. Van Trees, *Detection, Estimation and Modulation Theory, Vols I, II and III*, New York: Wiley, 1968, 1970 and 1971.

[2] A. Hutchinson, *Algorithmic Learning*, Oxford: Clarendon Press, 1995.

[3] S. Haykin, *Neural Networks. A Comprehensive Foundation* Englewood Cliffs, NJ: Prentice-Hall, 1993.

[4] S. Y. Kung, *Digital Neural Networks* Englewood Cliffs, NJ: Prentice-Hall, 1993.

[5] S. I. Gallant, *Neural Network Learning and Expert Systems* Cambrigde, MA: The MIT Press, 1993.

[6] S. Benedetto, E. Biglieri, V. Castellani, *Digital Transmission Theory* Englewood Cliffs, NJ: Prentice-Hall, 1987, ch. 8.

[7] R.D. Gitlin, J. F. Hayes, S. B. Webstein, *Data Communication Principles* , New York, NY: Plenum Press, 1992, ch. 8.

[8] J. G. Proakis, *Digital Communications (2nd ed.)*, New York, NY: Mc Graw-Hill, 1989, ch. 6.

[9] G. J. Gibson, S. Siu, C. F. N. Cowan, “The Application of Nonlinear Structures to the Reconstruction of Binary Signals,” *IEEE Trans. Signal Processing*, vol. 30, pp. 1877-1889, 1991.

[10] S. Chen, G. J. Gibson, G. F. N. Cowan, P. M. Grant, “Reconstruction of Binary Signals Using an Adaptive Radial-Basis-Function Equalizer,” *Signal Processing*, vol 22, pp. 77-93, 1991.

[11] S. Chen, B. Mulgrew, P. M. Grant, “A Clustering Technique for Digital Communication Channel Equalization Using Radial Basis Function Networks,” *IEEE Trans. Neural Nets., vol 4, pp. 570-579, 1993.*

[12] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, G. E. Hinton, “Adaptive Mixtures of Local Experts,” *Neural Computation*, vol. 3, pp. 79-87, 1991.

[13] M. I. Jordan, R. A. Jacobs, “Hierarchical Mixtures of Experts and the EM Algorithm”, *Neural Computation*, vol. 6, pp. 181-214, 1991.

[14] J. Cid-Sueiro, A. R. Figueiras-Vidal, “Digital Equalization Using Modular Neural Networks, An Overwiew,” *Proc. 7th. Intl. Thyrrenian Workshop on Digital Comms.*, pp. 337-345, Viareggio, Italy, 1995.

[15] J. Cid-Sueiro, A. Artés-Rodríguez, A. R. Figueiras-Vidal, “Recurrent Radial Basis Function Networks for Optimal Symbol-by-Symbol Equalization,” *Signal Processing*, vol. 40, pp. 53-63, 1994.

[16] B.A. Telfer, H.H Szu, “Energy Functions for Minimizing Missclassification Error with Minimum-Complexity Networks,” *Neural Networks*, vol. 7, pp. 809-818, 1994.

[17] J. J. Hopfield, “Learning Algorithms and Probability Distributions in Feed-Forward and Feed-Back Networks,” *Proc. Nat. Academy Sci.* USA, vol 84, pp. 8429-8433, 1987.

[18] G. E. Hinton, “Connectionist Learning Procedures,” *Artificial Intelligence*, vol. 40, pp. 185-234, 1989.

[19] M. K. Sönmez, T. Adali, “Channel Equalization for Distribution Learning: The Least Relative Entropy Algorithm,” *Proc. First IEEE Intl Workshop on Appls. of Neural Networks to Telecomms.*, pp. 218-224, Princeton, NJ, 1993.

[20] T. Adali, X. Liu, M. K. Sönmez, “Conditional Distribution Learning with Neural Networks and Its Application to Channel Equalization,” submitted to *IEEE Trans. on Signal Processing*, 1996.

[21] S. J. Nowlan, G. E. Hinton, “A Soft-Directed LMS Algorithm for Blind Equalization,” *IEEE Trans. Comms.*, vol 41, pp. 275-279, 1993.

[22] S. I. Amari, “Backpropagation and the Stochastic Gradient Descent Method,” *Neurocomputing*, vol. 5, pp. 185-196, 1993.

[23] D. W. Ruck, S. K. Rogers, M. Kabrisky, M. E. Oxley, B. W. Suter, “The Multilayer Perceptron as an Approximation to a Bayes Optimal Discriminant Function,” *IEEE Trans. on Neural Nets.*, vol. 1, pp. 296-298, 1990.

[24] E. Biglieri, S. Barbieris, M. Catena, “Analysis and Compensation of Nonlinearities in Digital Transmission Systems,” *IEEE Trans. Selected Areas in Comms.*, vol. 6, pp 42-51, 1988.

[25] G. Lazzarin, S. Pupolin, A. Sarti, “Nonlinearity Compensation in Digital Radio Systems,” *IEEE Trans. on Comms.*, vol 42, pp. 988-999, 1994.

[26] Y. S. Abu-Mustafa, “Hints,” *Neural Computation*, vol. 7, pp. 639-671, 1995.

This work has been partly supported by E.U. Human Capital & Mobility contract number ERB CHRXCT 930405.