

Nonlinear Channel Equalization with Gaussian Processes for Regression

Fernando Pérez-Cruz *Senior Member, IEEE* (*), Juan José Murillo-Fuentes *Member, IEEE* and Sebastián Caro

Abstract

We propose Gaussian processes for regression as a novel nonlinear equalizer for digital communications receivers. GPR's main advantage, compared to previous nonlinear estimation approaches, lies on their capability to optimize the kernel hyperparameters by maximum likelihood, which improves its performance significantly for short training sequences. Besides, GPR can be understood as a nonlinear minimum mean square error estimator, a standard criterion for training equalizers that trades-off the inversion of the channel and the amplification of the noise. In the experiment section, we show that the GPR-based equalizer clearly outperforms support vector machine and kernel adaline approaches, exhibiting outstanding results for short training sequences.

Index Terms

Gaussian Processes, Regression, Equalization, Nonlinear Equalization, Kernel Adaline, Support Vector Machines.

I. INTRODUCTION

Channel equalization is a major issue in digital communications, because the channel affects the transmitted sequence with both linear and nonlinear distortions. Channel equalization minimizes those

This work was partially funded by Spanish government (Ministerio de Educación y Ciencia TEC2006-13514-C02-01/TCM y TEC2006-13514-C02-02/TCM), the European Union (FEDER) and the Comunidad de Madrid (project 'PRO-MULTIDIS-CM', S0505/TIC/0223). Fernando Pérez-Cruz is supported by Marie Curie Fellowship 040883-AI-COM.

F. Pérez-Cruz is in the Electrical Engineering Department in Princeton University. Princeton, NJ 08544. He is also an associated Professor in Universidad Carlos III de Madrid. E-mail: fp@princeton.edu (*) **Corresponding Author**

J.J. Murillo-Fuentes and S. Caro are with the Dept. Teoría de la Señal y Comunicaciones, Escuela Técnica Superior de Ingenieros, Universidad de Sevilla, Paseo de los Descubrimientos s/n, 41092 Sevilla, Spain. E-mail: {murillo, scaro}@us.es

distortions to recover the transmitted sequence. In wireless communications, in which bandwidth is a scarce resource and we need to send a training sequence in every burst, short training sequences are a prerequisite. Traditionally, channel equalization has been considered equivalent to inverse filtering. The optimal solution, based on maximum likelihood sequence estimation (MLSE) [1], has a complexity that grows exponentially with the dimension of the channel impulsive response (Viterbi algorithm), and an unknown delay. Alternatively, machine learning can be used to approximate MLSE decisions at a lower computational cost. Several nonlinear detection procedures have been proposed to address this problem with varying degrees of success, such as multi-layered perceptrons (MLPs) [2], radial basis function networks (RBFNs) [3], recurrent RBFNs [4], self-organizing feature maps (SOFMs) [5], [6], wavelet neural networks [7], kernel Adeline (KA) [8] and support vector machines (SVMs) [9], [10]. Such structures usually outperform linear equalizers, especially when non-minimum phase channels are encountered. They can also compensate for nonlinearities in the channel. In [11], a comprehensive survey about nonlinear methods for digital communications, the author details the challenges and possible structures to solve the channel equalization problem.

In this paper we focus on Gaussian Processes for regression (GPR) [12]. This approach has been already successfully applied to multi-user detection in CDMA systems [13], [14] and we have presented some preliminary results for channel equalization in [15], in which we proposed GPR and carried out some basic experiments. In this paper, we extend the analysis of GPR for channel equalization. In addition, we show that GPR can be understood as a nonlinear MMSE (minimum mean squared error) estimator. Therefore GPR achieves optimal results from the MMSE viewpoint, a widely used criterion in digital communications.

Standard methods for optimally setting the hyperparameters (i.e. cross-validation [16]) require immense computational resources. For short training sequences an hyperparameter mismatch significantly affects the performance of digital communication receivers, while for longer training sequences this performance is not sensitive to variations in the hyperparameters. In the previously cited papers the authors propose fixed hyperparameters and sufficiently long training sequences, hence the exact setting of the hyperparameters is not critical. Our proposal introduces GPR as nonlinear equalizer with optimally trained hyperparameters for reducing the length of the training data sequence. We experimentally illustrate that previous fixed hyperparameters machine learning tools clearly underperforms compared to a GPR-based equalizer.

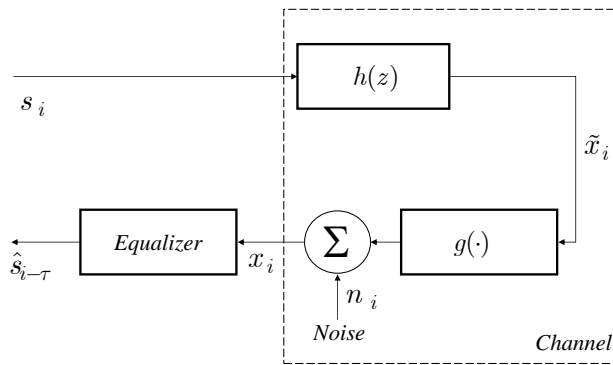


Fig. 1. Base-band discrete-time digital communication system for channel equalization proposed in this paper.

II. NONLINEAR CHANNEL MODEL AND MMSE EQUALIZATION

We have depicted in Figure 1 the based-band discrete-time digital communication system that we use throughout this paper. A dispersive channel model, which spreads the transmitted energy throughout several symbol intervals, can be modeled as a finite impulse response filter. The transmitted symbols are convolved with the channel model:

$$h(z) = \sum_{j=0}^{N_c-1} h_j z^{l-j}, \quad (1)$$

where l and N_c represents, respectively, the channel delay and its length. The nonlinearities in the channel, due to amplifiers and converters in the receiver can be modeled as [8]:

$$x_i = g(\tilde{x}_i) + n_i = \sum_{k=1}^{N_o} g_k \tilde{x}_i^k + n_i, \quad (2)$$

where $g(\cdot)$ is the nonlinear function the inputs face at the receiver (amplifiers and converters), N_o is the order of this nonlinearity and $\tilde{x}_i = \sum_{j=0}^{N_c-1} h_j s_{i-l-j}$. The transmitted signal value s_i comes from a zero-mean constellation and n_i is additive white Gaussian noise (AWGN). An equalizer collects several of the received symbols to predict each transmitted symbol,

$$\hat{s}_{i-\tau} = f(\mathbf{x}_i), \quad (3)$$

where $\mathbf{x}_i = [x_{i+\tau}, x_{i+\tau-1}, \dots, x_{i+\tau-m+1}]^T$, and m and τ are, respectively, the order and delay of the equalizer.

Although the linear transversal equalizer (LTE) is the typical structure to equalize the channel, it is not optimal from the MMSE point of view [17], as the unrestricted MMSE solution is given by

$$f_{mmse}(\mathbf{x}) = E[s|\mathbf{x}] = \underset{f(\cdot)}{\operatorname{argmin}} E[(s - f(\mathbf{x}))^2] \quad (4)$$

and it is usually a nonlinear function of \mathbf{x} , even if the channel is linear and has minimum phase. If the channel is nonlinear and/or non-minimum phase, the MMSE solution is highly nonlinear and a linear equalizer underperforms [2].

Given a labeled iid (independent and identically distributed) training dataset ($\mathcal{D} = \{\mathbf{x}_k, s_{k-\tau}\}_{k=1}^n$, where the input $\mathbf{x}_k \in \mathbb{R}^d$ and the output $s_k \in \mathbb{R}$), if we choose $f(\mathbf{x}) = \mathbf{w}^\top \phi(\mathbf{x})$, where $\phi(\cdot)$ is a nonlinear universal transformation to a higher dimensional feature space, the weights \mathbf{w} that minimize (4) can be computed analytically:

$$\mathbf{w}_{mmse} = \left(\Phi^\top \Phi \right)^{-1} \Phi^\top \mathbf{s}, \quad (5)$$

where $\Phi = [\phi(\mathbf{x}_1), \dots, \phi(\mathbf{x}_n)]^\top$ and $\mathbf{s} = [s_{1-\tau}, \dots, s_{n-\tau}]^\top$.

III. GAUSSIAN PROCESSES FOR REGRESSION

We present GPR from the (Bayesian) generalized linear regression viewpoint [12]. A generalized linear regressor for equalization expresses the input-output relation as

$$s_{i-\tau} = \mathbf{w}^\top \phi(\mathbf{x}_i) + \nu_i, \quad (6)$$

where ν_i is a random variable that models the deviation between $s_{i-\tau}$ and its estimate, $\mathbf{w}^\top \phi(\mathbf{x}_i)$. Given a statistical model for ν_i , $p(\nu_i) = p(s_{i-\tau} | \mathbf{w}^\top \phi(\mathbf{x}_i))$, a training dataset \mathcal{D} and a prior distribution for \mathbf{w} , $p(\mathbf{w})$, we can compute the posterior density for \mathbf{w} using Bayes rule:

$$p(\mathbf{w} | \mathbf{s}, \mathbf{X}) = \frac{p(\mathbf{w}) p(\mathbf{s} | \mathbf{w}, \mathbf{X})}{p(\mathbf{s} | \mathbf{X})} = \frac{p(\mathbf{w}) \prod_k p(s_{k-\tau} | \mathbf{w}^\top \phi(\mathbf{x}_k))}{p(\mathbf{s} | \mathbf{X})}, \quad (7)$$

where $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]^\top$.

GPR assumes a zero-mean Gaussian prior, $p(\mathbf{w}) = \mathcal{N}(\mathbf{w} | \mathbf{0}, \sigma_{\mathbf{w}}^2 \mathbf{I})$, which allocates probability mass to every possible \mathbf{w} , where \mathbf{I} is the identity matrix and $\sigma_{\mathbf{w}}$ is a parameter that controls the width of the prior. GPR also assumes a Gaussian likelihood model $p(s_{k-\tau} | \mathbf{w}^\top \phi(\mathbf{x}_k)) = \mathcal{N}(s_{k-\tau} | \mathbf{w}^\top \phi(\mathbf{x}_k), \sigma_{\nu}^2)$, thereby the posterior for \mathbf{w} , $p(\mathbf{w} | \mathbf{s}, \mathbf{X})$, is also Gaussian and its mean is given by the maximum a posteriori

(MAP) estimation criterion:

$$\begin{aligned}
\mu_{\mathbf{w}} &= \mathbf{w}_{MAP} = \underset{\mathbf{w}}{\operatorname{argmax}} \{ \log p(\mathbf{w} | \mathbf{s}, \mathbf{X}) \} = \\
&= \underset{\mathbf{w}}{\operatorname{argmax}} \{ \log p(\mathbf{s} | \mathbf{X}, \mathbf{w}) + \log p(\mathbf{w}) \} = \\
&= \underset{\mathbf{w}}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma_{\nu}^2} \sum_{k=1}^n (s_{k-\tau} - \mathbf{w}^{\top} \phi(\mathbf{x}_k))^2 + \frac{1}{2\sigma_{\mathbf{w}}^2} \|\mathbf{w}\|^2 \right\}. \tag{8}
\end{aligned}$$

The prediction for any new point \mathbf{x}_i , for $i > n$, is a Gaussian density function, $p(s_{i-\tau} | \mathbf{x}_i, \mathcal{D}) = \mathcal{N}(s_{i-\tau} | \mu_{s_{i-\tau}}, \sigma_{s_{i-\tau}})$, and its mean value in closed form yields

$$\mu_{s_{i-\tau}} = \mu_{\mathbf{w}}^{\top} \phi(\mathbf{x}_i) = \mathbf{k}^{\top} \mathbf{C}^{-1} \mathbf{s}, \tag{9}$$

where $\mathbf{k} = \phi^{\top}(\mathbf{x}_i) \Phi^{\top} = [k(\mathbf{x}_i, \mathbf{x}_1), \dots, k(\mathbf{x}_i, \mathbf{x}_n)]^{\top}$ and \mathbf{C} is the covariance matrix of the Gaussian process, also known as the kernel matrix,

$$(\mathbf{C})_{kj} = k(\mathbf{x}_k, \mathbf{x}_j) + \sigma_{\nu}^2 \delta_{kj} \quad \forall k, j = 1, \dots, n \tag{10}$$

where, in turn, δ_{kj} is the Kronecker's delta function and $k(\mathbf{x}_k, \mathbf{x}_j) = \sigma_{\mathbf{w}}^2 \phi^{\top}(\mathbf{x}_k) \phi(\mathbf{x}_j)$ is the kernel, or inner product, of the nonlinear transformation $\phi(\cdot)$.

The mean solution for GPR is obtained optimizing (8), which is the stochastic version of (4) plus a regularization term to avoid overfitting. This term vanishes as the number of training samples grows. Also, the mean of the weights used in equation (9) is of the form given in (5). Therefore the GPR mean prediction can be regarded as a nonlinear MMSE estimation for the nonlinear mapping $\phi(\cdot)$.

A. Hyperparameter optimization

As discussed in the introduction, the optimal setting of the hyperparameters could be obtained by cross-validation, as we could for any other nonlinear machine learning method. In this case the GPR would be as good as any of the other methods, as it would require either to try different settings or to rely on a pre-specified one. However, we are interested in optimally setting the hyperparameters. A two-step simple approach, as proposed in [12], suitable for digital communications problems works as follows: we first compute the likelihood of the hyperparameters of the kernel given the training data set; and then we estimate its maximum to obtain its optimal setting.

B. Covariance matrix

To optimize the kernel hyperparameters we need to describe a kernel in a parametric form, which is flexible to accurately solve our estimation problem. Kernel design is one of the most challenging open

problems in machine learning [18]. We need to incorporate our prior knowledge into the kernel, but, at the same time, we want the kernel to be flexible to explain previously unknown trends in the data. In this paper we propose the following kernel:

$$(\mathbf{C})_{ij} = \alpha_1 \exp \left(- \sum_{\ell=1}^d \gamma_{\ell} (x_{i\ell} - x_{j\ell})^2 \right) + \alpha_2 \mathbf{x}_i^{\top} \mathbf{x}_j + \alpha_0 \delta_{ij}. \quad (11)$$

The covariance function in (11) is a good kernel for solving the channel equalization using GPR, because it contains a linear and a universal nonlinear part. In minimum-phase channels a linear solution can provide a good approximation to the optimal solution, but it is still suboptimal and the nonlinear part is needed to improve the performance of the equalizer. In this sense the proposed covariance function is ideal for the problem. The linear part can mimic the best linear decision boundary and the nonlinear part modifies it, where the linear explanation is not optimal to obtain the expectation of s given \mathbf{x} . If the channel is nonlinear and/or non-minimum phase, the ML solution sets $\alpha_2 = 0$ and there is no-interference of the linear term with the nonlinear one. Also, using a radial basis kernel for the nonlinear part is a good choice to achieve nonlinear decisions for channel equalization, because the received symbols form a constellation of clouds of points with Gaussian spread around its centers.

IV. EXPERIMENTAL RESULTS

In this section we face the experimental study of the GPR-based equalizer, hereafter GPR-EQ. We compare the GPR-EQ to the linear MMSE, KA and SVM approaches, respectively, denoted as MMSE-EQ, KA-EQ and SVM-EQ. The BER figures included are computed as the average of 100 experiments using 10^5 test samples. The training sets used in each experiment are different and random values. In all experiments, the GPR-EQ uses the covariance matrix in (11).

For the GPR-EQ, we first learn the hyperparameters of the linear and noise terms of the kernel. Then, with these terms fixed, we train the nonlinear part. The SVM-EQ is trained using a Gaussian kernel with its width equal to the noise standard deviation and the soft margin parameter $C = 0.5$. We found these hyperparameters to provide the best mean equalizer. For the KA-EQ, following the guidelines in [8], we assign 76% of the training sequence to the training subset and 24% for regularization. We set the σ parameter to the noise standard deviation and $\eta = 1/n_{tr}$, where n_{tr} is the size of the training subset¹.

¹ η is the learning rate for KA as defined in [8].

A. Experiment 1: 16 QAM over nonlinear memoryless channel.

Through this text we just paid attention to polar digital modulation, as extension to multilevel complex symbols is immediate. The input vector is a vector containing the real and complex part of the symbols and we focus on detecting the real part or the complex part of the transmitted sequence. In this experiment the transmitted symbols are 16-QAM modulated, i.e., they are complex symbols with real and imaginary parts values $\{\pm 1, \pm 3\}$ and we face the equalization after a nonlinear channel such as an amplifier. We focus on 16-QAM as this modulation is quite sensitive to nonlinear distortion. The amplifiers usually rotate the symbols and compress their amplitude. The larger the amplitude is, the larger it is compressed. As a result, symbols with larger amplitude suffer greater distortion. In this experiment we focus on detecting one of the 4 bits transmitted every symbol. This bit is the one telling the real part of the symbol being $\{\pm 1\}$ or $\{\pm 3\}$. We do not include here the complete study of the linearization of the receiver; we just focus on the final BER. We modeled the amplifier by using the following Saleh model [19]:

$$g(x) = \frac{1.3 \cdot |x|}{1 + 0.4 \cdot |x|^2} \exp \left(j \left(\frac{0.3 \cdot |x|^2}{1 - 1 \cdot |x|^3} + \angle x \right) \right) \quad (12)$$

We set the input values to have a 0 dB back-off, i.e., we use the whole input range from 0 to saturation. In Fig. 2 we include the BER as a function of the length of the training sequence for an SNR of 20dB. The GPR-EQ provides the best overall equalizer. For short training sequence it shuts down the nonlinear part of the kernel, while SVM-EQ and KA-EQ report chance level performance. GPR-EQ is able to decide automatically if it has a long enough training sequence to train its nonlinear part and if it does not it reports the best linear equalizer. As the number of training samples increases, the GPR-EQ incorporates the nonlinear part to its kernel and it further improves its solution. The SVM-EQ has a better response than the KA-EQ but quite poor compared to the GPR-EQ. For very short training sequences ($n < 20$) the linear MMSE-EQ converges fastest, but it saturates to a very high BER. This result suggests that the GPR-EQ can be adopted as a good linearizer in predistortion schemes [20].

B. Experiment 2: BPSK over nonlinear multipath channel.

The transmitted symbols are BPSK modulated. In this experiment we face the equalization when a nonlinear function follows a multipath channel. We use the channel model proposed in [8]:

$$h(z) = 0.4 + 0.9z^{-1} + 0.4z^{-2}, \quad (13)$$

followed by a nonlinearity as proposed in [20], a GSM amplifier by ANDREW Corp.

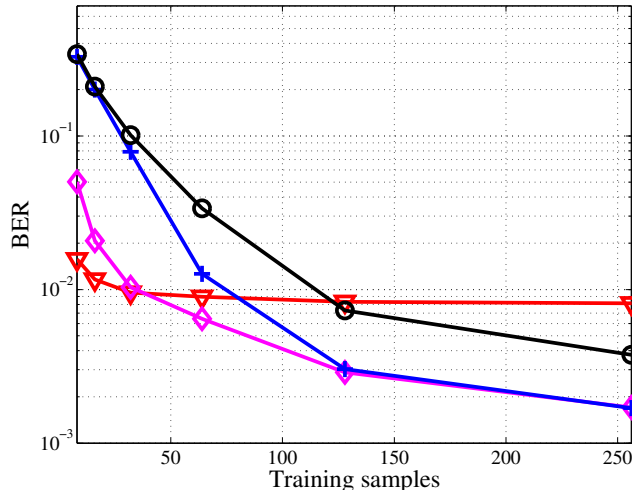


Fig. 2. BER of a 16QAM with nonlinear memoryless channel along the training sample length for $SNR = 20\text{dB}$, equalizer lengths and delays ($m = 4, \tau = 1$) for MMSE-EQ (∇), SVM-EQ ($+$), KA-EQ (\circ) and GPR-EQ (\diamond).

In Fig. 3 we include the averaged BER for a short training sequence of $n = 128$ samples. The GPR-EQ clearly outperforms the other approaches. The nonlinear SVM-EQ and GPR-EQ improve the linear response by 3 dB at a SNR of 12 dB. This means great improvements in encoded (e.g. with LDPC) transmissions. The KA-EQ cannot follow these approaches and it does not provide a significant reduction of the BER. Furthermore, the KA-EQ -and the SVM- has a irreducible BER and the linear solution outperforms them for SNR larger than 16dB. The poor performance of the KA-EQ and SVM-EQ can be improved by increasing the number of training samples. In Fig. 4 we include the BER as a function of the length of the training sequence for an SNR of 20dB. The behavior is even more striking than that of Fig. 2. As the number of training samples increases, the GPR-EQ greatly reduces its BER, while the KA-EQ is not able to improve the BER of the linear MMSE-EQ with training sequences of 512 symbols. The SVM-EQ has a better performance than the KA-EQ, but still quite poor compared to the GPR-EQ.

V. DISCUSSION AND CONCLUSIONS

Gaussian Processes for regression is a novel Bayesian machine learning tool that presents several advantages for designing nonlinear digital communications receivers. Given the covariance function, GPR provides an analytical solution to any regression estimation problem. Moreover, it does not only give point estimates, but it also assigns confidence intervals for them. In GPR, we can use flexible kernels with several hyperparameters that are individually tuned for each particular problem. In previous machine

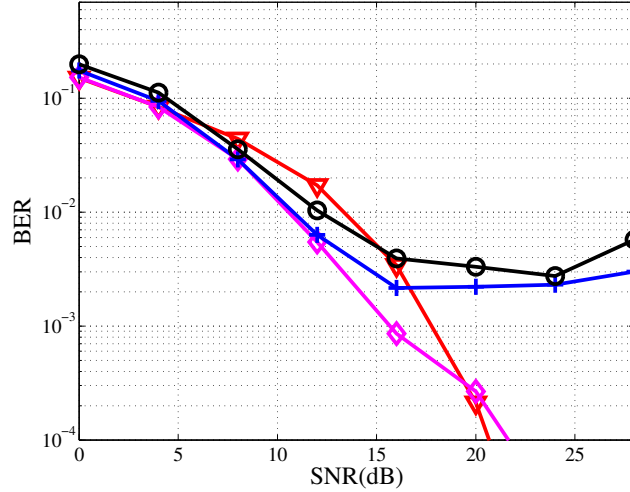


Fig. 3. BER with nonlinear channel $m = 4$, $\tau = 1$ and 128 training samples for MMSE-EQ (∇), SVM-EQ (+), KA-EQ (\circ) and GPR-EQ (\diamond).

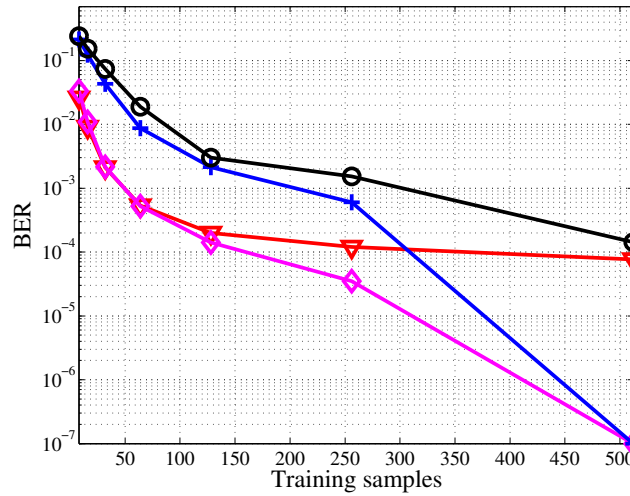


Fig. 4. BER of a BPSK with nonlinear channel along the training sample length for $SNR = 20\text{dB}$, equalizer lengths and delays ($m = 4, \tau = 1$) for MMSE-EQ (∇), SVM-EQ (+), KA-EQ (\circ) and GPR-EQ (\diamond).

learning based equalizers, such as SVM or KA, the hyperparameters have to be either pre-specified or estimated by cross-validation. SVM and KA can only tune one or two hyperparameters by cross-validation, due to its computational complexity, and they need longer training sequence. For long training datasets, the hyperparameters slightly influence the solution and, consequently, approaches as SVMs, KA or GPR provide similar solutions close to optimum BER performance. But for short training sequences, hyperparameters matter and an optimal setting significantly improves their performance. These remarkable

drawbacks limit the application of nonlinear tools, such as SVM or KA, in digital communications receivers, since we face complex nonlinear problems with reduced computational resources and short training sequences. By exploiting the GPs framework, as stated in this paper, we can avoid them.

In addition, GPR can be understood as a nonlinear minimum mean squared error estimator. MMSE is the standard method for training linear transversal equalizers, because it trades-off the inversion of the channel and the amplification of the noise in the zeros of the channel response. But the optimal MMSE estimator is the expectation of the transmitted symbols given the received information, which is a nonlinear function of the received symbols. Therefore GPR with optimally chosen hyperparameters can provide optimal solution in the MMSE sense.

REFERENCES

- [1] E. A. Lee and D. G. Messerschmitt, *Digital Communication*, 2nd ed. Boston, MA: Kluwer, 1994.
- [2] G. J. Gibson, S. Siu, and C. F. N. Cowan, "The application of nonlinear structures to the reconstruction of binary signals," *IEEE Transactions on Signal Processing*, vol. 39, pp. 1877–1884, Aug. 1991.
- [3] S. Chen, G. J. Gibson, C. F. N. Cowan, and P. M. Grant, "Reconstruction of binary signals using an adaptive radial-basis-function equalizer," *Signal Processing*, vol. 22, pp. 77–83, 1991.
- [4] J. Cid-Sueiro, A. Artés-Rodríguez, and A. R. Figueiras-Vidal, "Recurrent radial basis function networks for optimal symbol-by-symbol equalization," *Signal Processing*, vol. 40, pp. 53–63, 1994.
- [5] T. Kohonen, K. Raivio, O. Simula, O. Venta, and J. Henriksson, "Combining linear equalization and self-organizing adaptation in dynamic discrete-signal detection," in *Proc. International Joint Conf. on Neural Networks*, vol. 1, San Diego, CA, pp. 223–228, June 1990.
- [6] G. A. Barreto and L. G. M. Souza, "Adaptive filtering with the self-organizing map: A performance comparison," *Neural Networks*, vol. 19, no. 6, pp. 785–798, 2006.
- [7] P. R. Chang and B. C. Wang, "Adaptive decision feedback equalization for digital satellites channels using multilayer neural networks," *IEEE Journal on Selected areas of communication*, vol. 13, no. 2, pp. 316–324, Feb. 1995.
- [8] B. Mitchinson and R. F. Harrison, "Digital communications channel equalization using the kernel adaline," *IEEE Transactions on Communications*, vol. 50, no. 4, pp. 571–576, 2002.
- [9] F. Pérez-Cruz, A. Navia-Vázquez, P. L. Alarcón-Diana, and A. Artés-Rodríguez, "SVC-based equalizer for burst TDMA transmissions," *Signal Processing*, vol. 81, no. 8, pp. 1681–1693, Aug. 2001.
- [10] I. Santamaría, C. Pantaleón, L. Vielva, and J. Ibañez, "Blind equalization of constant modulus signals using support vector machines," *IEEE Transactions on Signal Processing*, vol. 52, no. 6, pp. 1773–1782, 2004.
- [11] M. Ibnkahla, "Applications of neural networks to digital communications - a survey," *Signal Processing*, vol. 80, no. 7, pp. 1185–1215, 2000.
- [12] C. K. I. Williams, "Prediction with gaussian process: From linear regression to linear prediction and beyond," in *Learning in Graphical Models*, M. I. Jordan, Ed. Cambridge, (MA): M.I.T. Press, pp. 599–621, 1999.
- [13] J. J. Murillo-Fuentes, S. Caro, and F. Pérez-Cruz, "Gaussian processes for multiuser detection in CDMA receivers," in *Advances in Neural Information Processing Systems*, Y. Weiss, B. Schölkopf, and J. Platt, Eds. Cambridge, MA: MIT Press, vol. 18, 2006.

- [14] F. Pérez-Cruz and J. Murillo-Fuentes, "Gaussian processes for digital communications," in *ICASSP*, vol. V, Toulouse, Fr, pp. 781–784, May 2006.
- [15] S. Caro, F. Pérez-Cruz, and J. Murillo-Fuentes, "Gaussian processes for regression in channel equalization," in *EUSIPCO*, Florence, Italy, Sept. 2006.
- [16] C. M. Bishop, *Neural Networks for Pattern Recognition*. Clarendon Press, 1995.
- [17] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261–1283, Apr. 2005.
- [18] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press, 2006.
- [19] A. Saleh, "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers," *IEEE Transactions on Communications*, vol. 29, no. 11, pp. 1715–1720, Nov. 1981.
- [20] F. J. Gonzalez-Serrano, J. J. Murillo-Fuentes, and A. Artes-Rodriguez, "GCMAC-based predistortion for digital modulations," *IEEE Trans. on Communications*, vol. 49, no. 9, pp. 1679–1689, Sept. 2001.