filters," *IEEE Trans. Automat. Cont.*, vol. AC-31, no. 3, pp. 283–287, Mar. 1986.
[9] E. Eweda, "Analysis and design of a signed regressor LMS algorithm for stationary and nonstationary adaptive filtering with correlated Gaussian data," *IEEE Trans. Circuits Syst.*, vol. 37, no. 11, pp. 1367–1374, Nov. 1990.
[10] H. Cramer and M. R. Leadbetter, *Stationary and Related Stochastic Processes*. New York: Wiley, 1967.

On the Optimal Weight Vector of a Perceptron with Gaussian Data and Arbitrary Nonlinearity

Arie Feuer and Roberto Cristi

Abstract—**In this correspondence we investigate the solution to the following problem: Find the optimal weighted sum of given signals when the optimality criteria is the expected value of a function of this sum and a given "training" signal. The optimality criteria can be a nonlinear function from a very large family of possible functions. A number of interesting cases fall under this general framework, such as a single layer perceptron with any of the commonly used nonlinearities, the LMS, the LMF or higher moments, or the various sign algorithms.**

Assuming the signals to be jointly Gaussian we show that the optimal solution, when it exits, is always collinear with the well-known Wiener solution, and only its scaling factor depends on the particular functions chosen. We also present necessary constructive conditions for the existance of the optimal solution.

I. Introduction

We consider the configuration of a single perceptron, which consists of a linear combination of N input sequences passing together with a training sequence through a nonlinearity. This configuration is depicted in Fig. 1. Typically, the perceptron output try to track the training sequence. The nonlinearity block here contains the nonlinear function of the perceptron, any nonlinear effects in measuring the training sequence (e.g., quantization, hard limiter), and the tracking criteria chosen. Similar configurations with specific nonlinearities have been discussed in [1]–[3]. Specifically, [3] deals with an adaptive algorithm which updates the weight vector in order to achieve optimal tracking.

It is well known from the adaptive filtering literature that the ability to determine the existence and structure of the optimal solution is at the basis of any adaptive algorithm. Since the perceptron weights are updated adaptively, the existence of an optimal solution and its nature must be addressed. That is exactly the purpose of this correspondence. We show here that for a *very large family of nonlinearities*, with Gaussian data, the optimal solution, if it exists, is *collinear with the well-known Wiener solution*. We also provide necessary conditions, testable for any particular nonlinearity, for the existance of the optimal solution.

II. The Main Result

Let us consider the configuration in Fig. 1 with the processes

$$X(n) = [x_1(n), x_2(n), \cdots, x_N(n)]^T, d(n)$$

Manuscript received June 9, 1991; revised November 13, 1992. The associate editor coordinating the review of this correspondence and approving it for publication was Dr. J. J. Shynk.
A. Feuer is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel.
R. Cristi is with the Department of Electrical and Computer Engineering, Naval Postgraduate School, Monterey, CA 93943.
IEEE Log Number 9208211.

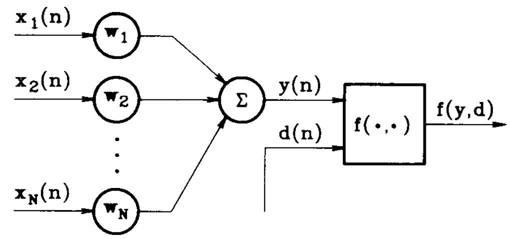

Fig. 1. Problem configuration.

being stationary and jointly Gaussian. We denote

$$R = E\{X(n)X(n)^T\} \tag{1}$$

$$P = E\{d(n)X(n)\} \tag{2}$$

$$R_d = E\{d(n)d(n)\}. \tag{3}$$

Clearly, $y(n) = W^T X(n)$ and $d(n)$ are also jointly Gaussian, with a density distribution function given by

$$p(y, d) = \frac{1}{2\pi|r|^{1/2}} \exp\left\{-\frac{1}{2|r|}(R_d y^2 - 2P^T W y d + W^T R W d^2)\right\} \tag{4}$$

where $|r|$ is the determinant of the covariance matrix r

$$r = \begin{bmatrix} W^T R W & P^T W \\ P^T W & R_d \end{bmatrix}. \tag{5}$$

Define the general tracking criterion as

$$J(W) = E\{f(y, d)\} \tag{6}$$

where $f(y, d)$ can be any nonlinearity which is bounded by an exponentially increasing function. Clearly, to be a valid tracking criterion, it must satisfy certain conditions but at this point this is not our concern. The problem we want to solve is to find a vector (set of gains) W_{opt} which minimizes $J(W)$, namely,

$$W_{\text{opt}} = \arg\min_{W \in R^N} J(W). \tag{7}$$

Noting that $R^{-1}P$ is what is commonly referred to as the Wiener solution we have Theorem 1, as follows.

Theorem 1: W_{opt} in the sense of (7), if it exists, is of the form

$$W_{\text{opt}} = \alpha P^{-1} P \tag{8}$$

where α is a scalar.

Proof: The proof makes use of some of the ideas in Price's theorem [4]. It is well known that the Gaussian p.d.f can be written in the form

$$p(y, d) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-j(w_1 y + w_2 d)} \Phi(w_1, w_2)\, dw_1\, dw_2 \tag{9}$$

where $\Phi(w_1, w_2)$ is the joint characteristic function of y and d. So, by (6) and (9)

$$J(W) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y, d) p(y, d)\, dy\, dd$$
$$= \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y, d) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdot e^{-j(w_1 y + w_2 d)} \Phi(w_1, w_2)\, dw_1\, dw_2\, dy\, dd. \tag{10}$$

1053-587X/93$03.00 © 1993 IEEE

Since y and d jointly Gaussian, using (4) and (5) we have

$$\Phi(w_1, w_2) = \exp\{-\tfrac{1}{2}(W^TRWw_1^2 + 2P^TWw_1w_2 + R_dw_2^2)\} \tag{11}$$

so that

$$\frac{\partial J(W)}{\partial W} = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y, d)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[-RWw_1^2 - Pw_1w_2] \cdot e^{-j(w_1y + w_2d)}\Phi(w_1, w_2)\,dw_1\,dw_2\,dy\,dd. \tag{12}$$

But, it can be readily shown (using (9)) that

$$\frac{\partial^2 p(y, d)}{\partial y^2} = -\frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w_1^2\Phi(w_1, w_2)e^{-j(w_1y + w_2d)}\,dw_1\,dw_2$$

and

$$\frac{\partial^2 p(y, d)}{\partial y\partial d} = -\frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w_1w_2\Phi(w_1, w_2) \cdot e^{-j(w_1y + w_2d)}\,dw_1\,dw_2.$$

So, substituting in (12) we observe that

$$\frac{\partial J(W)}{\partial W} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y, d) \cdot \left[RW\frac{\partial^2 p(y, d)}{\partial y^2} + P\frac{\partial^2 p(y, d)}{\partial y\partial d}\right] dy\,dd. \tag{13}$$

Since $f(y, d)$ is bounded exponentially the following integrals are well defined:

$$\alpha_1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y, d)\frac{\partial^2 p(y, d)}{\partial y^2}\,dy\,dd \tag{14}$$

$$\alpha_2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y, d)\frac{\partial^2 p(y, d)}{\partial y\partial d}\,dy\,dd. \tag{15}$$

(Note that α_1 and α_2 depend on R, P, R_d, and W itself; however, the important fact is that both α_1 and α_2 are scalars.)

Hence, W_{opt} must satisfy $\partial J(W)/\partial W = 0$ or, equivalently, from (13)-(15)

$$\alpha_1RW + \alpha_2P = 0. \tag{16}$$

Clearly, if $\alpha_1 = 0$ and $\alpha_2 \neq 0$, W_{opt} does not exist, if $\alpha_1 = 0$ and $\alpha_2 = 0$, (16) is trivially satisfied, and for $\alpha_1 \neq 0$ we get

$$W_{opt} = \alpha R^{-1}P$$

where

$$\alpha = -\frac{\alpha_2}{\alpha_1}. \tag{17}$$

□

Assuming the W_{opt} exists we may substitute (8) in (4) and (5) to get

$$p(y, d) = \frac{1}{2\pi|r|^{1/2}}\exp\left\{-\frac{1}{2|r|}(R_dy^2 - 2\alpha\beta yd + \alpha^2\beta d^2)\right\} \tag{18}$$

and

$$r = \begin{bmatrix}\alpha^2\beta & \alpha\beta \\ \alpha\beta & R_d\end{bmatrix} \tag{19}$$

where

$$\beta = P^TR^{-1}P. \tag{20}$$

Denote

$$\gamma = \frac{R_d}{\beta}$$

then we have

Theorem 2: For (8) to be the solution of (7) α must satisfy the equation

$$E\{[\gamma y^2 - \alpha yd - \alpha^2\beta(\gamma - 1)]f(y, d)\} = 0. \tag{21}$$

Proof: By straightforward calculation of $\partial^2 p(y, d)/\partial y^2$, $\partial^2 p(y, d)/\partial y\partial d$, and substitution in (14), (15) and (17) will result in (21). □

Corollary 1: Assuming $f(y, d)$ is twice differentiable then for (8) to be the solution of (7) α must satisfy the equation

$$\alpha h_1(\alpha) + h_2(\alpha) = 0 \tag{22}$$

where

$$h_1(\alpha) = E\left\{\frac{\partial^2 f(y, d)}{\partial y^2}\right\} \tag{23}$$

$$h_2(\alpha) = E\left\{\frac{\partial^2 f(y, d)}{\partial y\partial d}\right\}. \tag{24}$$

Proof: Since we have assumed that $f(y, d)$ increases at most exponentially, clearly $f(y, d)p(y, d)$ and $\partial f(y, d)/\partial y\; p(y, d)$ both go to zero as either $|y|$ or $|d|$ go to ∞. Thus, (13) can be integrated by parts to result in

$$\frac{\partial J(W)}{\partial W} = E\left\{RW\frac{\partial^2 f(y, d)}{\partial y^2} + P\frac{\partial^2 f(y, d)}{\partial y\partial d}\right\}.$$

Then, substituting $\partial J(W)/\partial W = 0$ and (8) we get (22) which completes the proof. □

In a typical perceptron the function $f(y, d)$ has the form

$$f(y, d) = [g_1(d) - g_2(y)]^2 \tag{25}$$

where $g_1(\cdot)$ and $g_2(\cdot)$ are some nonlinear functions.

Assuming $g_i(\cdot)$ is twice differentiable (22) becomes

$$E\{\alpha([g_2'(y)]^2 - [g_1(d) - g_2(y)]g_2''(y)) - g_1'(d)g_2'(y)\} = 0 \tag{26}$$

where $g_i'(x) = dg_i(x)/dx$.

Let us now consider some choices of the functions $g_i(\cdot)$.

Case 1.

$$g_1(x) = g_2(x) = \mathrm{sgn}\,(x) = \begin{cases}1, & \text{for } x > 0\\ 0, & \text{for } x = 0\\ -1, & \text{for } x < 0.\end{cases}$$

Here $g_i(\cdot)$ are not differentiable so neither is $f(y, d)$. Hence we will use Theorem 2 to find the needed α of the optimal solution in this case.

Straight calculation shows that here

$$f(y, d) = 2[1 - \mathrm{sgn}\,(y)\,\mathrm{sgn}\,(d)]$$

and since

$$E\{\mathrm{sgn}\,(y)\,\mathrm{sgn}\,(d)\} = \frac{2}{\pi}\arcsin\left(\frac{\mathrm{sgn}\,(\alpha)}{\sqrt{\gamma}}\right)$$

$$E\{y^2 \operatorname{sgn}(y) \operatorname{sgn}(d)\} = \frac{2\alpha^2\beta}{\pi}\left[\operatorname{sgn}(\alpha)\frac{\sqrt{\gamma-1}}{\gamma} + \arcsin\left(\frac{\operatorname{sgn}(\alpha)}{\sqrt{\gamma}}\right)\right]$$

$$E\{|y||d|\} = \frac{2\alpha\beta}{\pi}\left[\operatorname{sgn}(\alpha)\sqrt{\gamma-1} + \arcsin\left(\frac{\operatorname{sgn}(\alpha)}{\sqrt{\gamma}}\right)\right]$$

substitution in (21) results in an identity $0 = 0$. This means that in this case, every scalar α satisfies (21). However, since

$$J(W)|_{W=\alpha R^{-1}P} = 2\left[1 - \frac{2}{\pi}\arcsin\left(\operatorname{sgn}(\alpha)\sqrt{\frac{P^T R^{-1} P}{R_d}}\right)\right]$$

clearly, *any* $\alpha > 0$ will provide an optimal solution, namely, for this case

$$W_{\text{opt}} = \alpha R^{-1}P \qquad \text{for any } \alpha > 0.$$

The negative α corresponds to the maximal $J(W)$.

Case 2. $g_1(x) = g_2(x) = Kx$ (the linear case).

For this case $g_i'(x) = K$ and $g_i''(x) = 0$, so (26) becomes

$$E\{\alpha K^2 - K^2\} = 0$$

the solution of which is $\alpha = 1$. Namely, $W_{\text{opt}} = R^{-1}P$ as is well known.

Case 3. $g_i(x) = g_2(x) = x^2$.

Here, $g_i'(x) = 2x$ and $g_i''(x) = 2$, so (26) becomes

$$E\{\alpha(4y^2 - 2(d^2 - y^2)) - 4dy\} = 0$$

or

$$\alpha(3\alpha^2 - \gamma - 2) = 0$$

with the solutions $\alpha_1 = 0$ and $\alpha_2 = \sqrt{(\gamma + 2)/3}$. It can readily be shown that α_2 corresponds to the optimal solution we seek, while α_1 corresponds to the maximal value for $J(W)$.

Case 4. $g_1(x) = g_2(x) = x^n$.

In this case, which generalizes cases 2 and 3, we have $g_i'(x) = nx^{n-1}$ and $g_i''(x) = n(n-1)x^{n-2}$ so (26) becomes

$$E\{\alpha[(2n-1)y^{2(n-1)} - (n-1)d^n y^{n-2}] - nd^{n-1}y^{n-1}\} = 0.$$

After calculating the needed expectations we find that this equation has two solutions only, $\alpha_1 = 0$ and

$$\alpha_2 = \sqrt[n]{\frac{f_2(n, \gamma)}{f_1(n)}}$$

where

$$f_1(n) = \frac{(2n-1)!}{2^{n-1}(n-1)!}$$

$$f_2(n, \gamma) = n! + (n!)^2 \sum_{j=1}^{[n/2]} \frac{\gamma^j}{2^{2j}(n-2j)!(j!)^2}$$

(by $[\cdot]$ we denote the largest integer smaller than the number in the brackets).

III. Conclusion

The general form of the optimal solution in a single perceptron with almost arbitrary nonlinear function has been introduced. It has been shown that this solution is always collinear with the Wiener solution obtained in the linear case. Conditions for the existence of this solution have been presented as well. Extension to more complex structures of perceptrons (such as multiple layers) is currently being investigated.

References

[1] R. P. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Mag.*, vol. 4, pp. 4-22, Apr. 1987.

[2] B. Widrow, R. G. Winter, and R. A. Baxter, "Layered neural nets for pattern recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1109-1118, July 1988.

[3] J. J. Shynk and S. Roy, "Convergence properties and stationary points of a perceptron learning algorithm," *Proc. IEEE* (Special Issue on Neural Networks), vol. 78, pp. 1599-1604, Oct. 1990.

[4] R. Price, "A useful theorem for nonlinear devices having Gaussian inputs," *IRE Trans. Inform. Theory*, vol. IT-4, pp. 69-72, June 1958.

A Note on the *PQ* Theorem and the Extrapolation of Signals

Irwin W. Sandberg

***Abstract*—The problem of determining a band-limited function from its values on a finite interval is ill conditioned in the sense that although the pertinent inverse map exists, it is discontinuous at every point. We show that whenever certain closely related general problems are well conditioned in the sense that the inverse operator is continuous, they can be solved using a special case of a known algorithm. In particular, attention is directed to the relation between the *PQ* theorem, its Hilbert space projection-operator setting, and later work.**

I. Introduction and Ill Conditioning

A familiar extrapolation problem in the area of signal processing is that of determining a band-limited signal from its values on a finite interval. More specifically, the problem is to find v given u where

$$Wv = u,$$

v belongs to the subset B of $L_2(\mathfrak{R})$ consisting of all functions band limited to some band $[-\beta, \beta]$, W denotes the windowing defined by

$$(Wf)(t) = f(t), \qquad |t| \le t_0$$
$$= 0, \qquad \text{otherwise}$$

Manuscript received February 25, 1992; revised November 30, 1992. The associate editor coordinating the review of this correspondence and approving it for publication was Dr. Barry Sullivan.

The author is with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712.

IEEE Log Number 9208212.