

ASSOCIATING NONLINEAR EFFECTS IN NLMS ADAPTATION WITH DYNAMIC WEIGHT BEHAVIOR

A. A. (Louis) Beex & James R. Zeidler

DSPRL – ECE 0111
 Virginia Tech
 Blacksburg, VA 24061-0111, USA

Communications & Information Systems, 28505
 SPAWAR Systems Center
 San Diego, CA 92152, USA

ABSTRACT

Nonlinear effects, or non-Wiener behavior, have been observed in several adaptive filtering applications, such as adaptive equalization of wideband communication signals contaminated by narrowband interference, adaptive narrowband noise cancellation, and adaptive linear prediction of narrowband processes. The nonlinear effects have been especially prominent when narrowband processes have been involved, and revealed by performance better than expected from the conventional Wiener filter for that scenario. Furthermore, in the adaptive noise-canceling scenario, almost deterministic semi-periodic dynamic behavior of the adaptive filter weights has been observed. In the adaptive equalization and prediction scenarios the weight behavior is also dynamic, albeit of a different nature. The dynamic weight behavior associated with nonlinear effects is shown to be the result of how the conventional NLMS adaptive filter tracks the equivalent model for the desired signal.

1. INTRODUCTION

Non-linear effects have been demonstrated in adaptive noise canceling [1, 2], interference contaminated adaptive equalization [1, 3], and adaptive linear prediction [4]. Often these nonlinear effects are strongest when involving narrowband processes and when adaptive filter stepsizes are relatively large. The nonlinear, or non-Wiener, effects are characterized by performance better than that of the Wiener filter of the same structure, and by dynamic or time-varying behavior of the adaptive filter weights.

As was shown for the noise canceling, equalizer, and prediction contexts [1 - 4], nonlinear effects arise in LMS as well as NLMS adaptation. We concentrate here on use of the NLMS algorithm. The nonlinear effects were attributed to the error signal feedback, which is used in the NLMS weight update, and led to the establishment of a bound for the performance under nonlinear effects [2].

We show that the mechanism, by which the error feedback results in the observed nonlinear effects, is associated with the instantaneous information it carries about the deviation between the desired signal and its NLMS modeled version. The structure for the desired signal will be related to the underlying optimal estimator,

which may have a structure that does not fit the conventional NLMS tapped delay line model. When the structure underlying the desired data differs from the NLMS model that is being adapted, the potential for non-Wiener effects is created. We show that NLMS adaptation, by virtue of its inherent tracking capability, can often form a time-varying tapped delay line model that reasonably well approximates an underlying, time-varying Wiener filter equivalent.

2. NLMS ADAPTATION

The NLMS adaptation algorithm is as follows:

$$\begin{aligned}\hat{d}_n &= \mathbf{w}_n^H \mathbf{u}_n \\ e_n &= d_n - \hat{d}_n \\ \mathbf{w}_{n+1} &= \mathbf{w}_n + \bar{\mu} \frac{e_n^*}{\mathbf{u}_n^H \mathbf{u}_n} \mathbf{u}_n\end{aligned}\quad (1)$$

To emphasize the situations under which nonlinear effects are prominent, a large stepsize is chosen, $\bar{\mu} = 1$, so that adaptation is the fastest. Furthermore, it can then be shown that the *a posteriori* error equals zero [5], i.e.

$$\begin{aligned}\varepsilon_n &= d_n - \mathbf{w}_{n+1}^H \mathbf{u}_n \\ &= 0\end{aligned}\quad (2)$$

An additional property of the NLMS algorithm is that the *a posteriori* weight vector is such that the norm of the weight vector increment is minimized, while (2) serves as the constraint. Note from (1) that the change from \mathbf{w}_n to \mathbf{w}_{n+1} is by necessity in the direction of \mathbf{u}_n . If the desired signal d_n has the same structure as that used in the modeling process, i.e. $d_n = \mathbf{w}_o^H \mathbf{u}_n$ for some fixed weight vector \mathbf{w}_o , then the NLMS adaptation converges to that weight vector (it produces *a posteriori* errors of zero and weight vector increment norms of zero).

From (1) we see that the current estimate for the desired signal depends on the current weight vector and the current input, while the current weight vector depends

on the previous weight vector, the previous error, and the previous input, etc. Assuming that we operate in steady state, so that the effects of the initial weight vector choice have vanished, the current estimate for the desired signal is some function of all past desired signal samples and all current and past samples of the input vector elements.

$$\hat{d}_n = \text{fct} \left[\left\{ d_m \right\}_{m=-\infty}^{n-1}, \left\{ u_m \right\}_{m=-\infty}^n \right] \quad (3)$$

This NLMS estimate is not necessarily optimal.

3. WIENER FILTERS & DATA STRUCTURES

If we assume that the processes in (3) are jointly Gaussian and wide-sense stationary then the optimal (Wiener) estimate for the desired signal is time-invariant and linear.

$$\hat{d}_{n,opt} = \sum_{k=1}^{\infty} h_k^d d_{n-k} + \sum_{k=0}^{\infty} h_k^u u_{n-k} \quad (4)$$

If the processes are such that only a finite memory depth is required, say of L and M samples respectively, then the optimal estimator takes on the following structure.

$$\hat{d}_{n,opt} = \sum_{k=1}^L h_k^d d_{n-k} + \sum_{k=0}^{M-1} h_k^u u_{n-k} \quad (5)$$

We note that the latter can be rewritten as follows:

$$\begin{aligned} \hat{d}_{n,opt} &= \begin{bmatrix} \mathbf{h}_{opt}^{dH} & \mathbf{h}_{opt}^{uH} \end{bmatrix} \begin{bmatrix} \mathbf{d}_{n-1} \\ \mathbf{u}_n \end{bmatrix} \\ &= \mathbf{h}_{opt}^H \tilde{\mathbf{u}}_n \end{aligned} \quad (6)$$

The consequence of this development is that we now have an underlying structure for the desired data itself.

$$\begin{aligned} d_n &= \hat{d}_{n,opt} + \varepsilon_{n,opt} \\ &= \mathbf{h}_{opt}^H \tilde{\mathbf{u}}_n + \varepsilon_{n,opt} \end{aligned} \quad (7)$$

This structure shows that the desired data can be represented with a two-channel LTI model, and that the corresponding representation error equals the two-channel Wiener filter error.

The model inherent in the NLMS algorithm in (1) is of the form of the structure in (6). The distinction lies in whether \mathbf{u}_n or $\tilde{\mathbf{u}}_n$ is used. The difference resides in the importance of the past of the desired signal samples. If there is no information in $\{d_m\}_{m=-\infty}^{n-1}$ about d_n , then effectively \mathbf{u}_n is the same as $\tilde{\mathbf{u}}_n$, and NLMS – that is to say \mathbf{w}_n – converges to (a neighborhood of) \mathbf{h}_{opt} . NLMS then approaches the Wiener filter in performance.

The difference between the conventional and two-channel cases lies in the definition for the NLMS input vector \mathbf{u}_n . For a general two-channel case, the NLMS input vector can be defined as follows.

$$\mathbf{u}_n = \begin{bmatrix} \mathbf{x}_n \\ \mathbf{r}_n \end{bmatrix} \quad (8)$$

The vectors \mathbf{x}_n and \mathbf{r}_n are referred to as the auxiliary and the reference channel respectively. The conventional NLMS algorithm uses only the so-called reference channel, i.e. $\mathbf{u}_n \equiv \mathbf{r}_n$. In order to match the structure of the optimal estimator in (6), the NLMS input would contain an auxiliary channel $\mathbf{x}_n \equiv \mathbf{d}_{n-1}$. The auxiliary channel definition depends on the application.

The corresponding Wiener filter (WF) designs follow from the Wiener-Hopf equation.

$$E \left\{ \begin{bmatrix} \mathbf{x}_n \\ \mathbf{r}_n \end{bmatrix} \begin{bmatrix} \mathbf{x}_n^H & \mathbf{r}_n^H \end{bmatrix} \right\} \mathbf{w}_{xr} = E \left\{ \begin{bmatrix} \mathbf{x}_n \\ \mathbf{r}_n \end{bmatrix} d_n^H \right\} \quad (9)$$

The performance of the resulting WF, using L auxiliary taps and M reference taps, is given by

$$MMSE_{WF(L,M)} = E \left\{ |d_n|^2 \right\} - E \left\{ \begin{bmatrix} \mathbf{x}_n \\ \mathbf{r}_n \end{bmatrix} d_n^* \right\} \mathbf{w}_{xr}^H \quad (10)$$

The nonlinear effect scenarios usually corresponded to (jointly) wide-sense stationary processes, in which case the resulting WF solutions are linear time-invariant (LTI) filters. The two-channel Wiener filter design and performance evaluation are based on auto- and cross-correlation information, and can be done efficiently [6, 7].

4. APPLICATION SCENARIOS

For each application scenario, the two-channel version is reflected in the figures below. For each of the input channels a tapped delay line exists internal to the NLMS structure. Conventional NLMS does not use \mathbf{x}_n , i.e., the auxiliary channel partition of the input is inactive.

The adaptive noise-canceling (ANC) scenario is reflected in Fig. 1. Its aim is to predict a desired signal from a reference input, which presumably contains information about the desired signal because it is correlated therewith. In many cases performance is increased when the desired signal is estimated from its own immediate past, rather than from a different signal. The choice for the auxiliary signal in ANC is therefore the immediate past of the desired signal, i.e. $\mathbf{x}_n \equiv \mathbf{d}_{n-1}$.

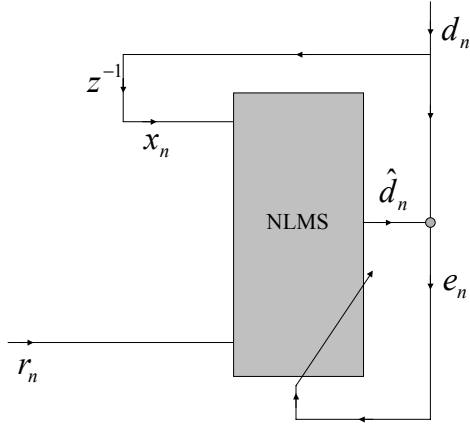


Fig. 1. ANC scenario.

The interference contaminated adaptive equalization (AEQ) scenario, in training mode, is reflected in Fig. 2.

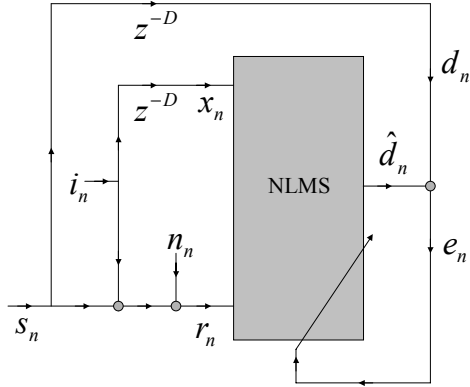


Fig. 2. AEQ scenario.

Nonlinear effects were reported for scenarios where the interference i_n was strong relative to the signal s_n , which itself was strong relative to the noise n_n . In this case we choose i_{n-D} , the interference at the center tap where the signal estimate is desired, as the auxiliary channel.

The adaptive linear prediction (ALP) scenario is reflected in Fig. 3. In this case we are interested in predicting the desired signal from its far past, i.e. Δ and more samples in the past. Generally a signal can be predicted much better from its recent past than from its far past, so that the immediate past is chosen here as the auxiliary channel.

From the Wiener filtering point of view, the reference signal is related to the desired signal in each of the above cases. In ANC we assume there is statistical correlation between the two, in AEQ the desired signal is a component of the reference signal, and in ALP the reference signal is the far past of the desired signal.

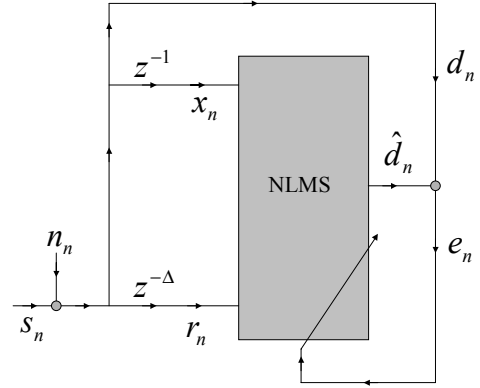


Fig. 3. ALP scenario.

5. TIME-VARYING WIENER FILTERS

Strong nonlinear effects have been observed for narrowband processes, which are modeled well with an AR(1) model. For these processes, together with high signal to noise ratios, the Wiener filter of (6) – for increasing L and M – very quickly saturates to the Wiener filter of (4). The latter happens for $L = 1$ in ANC, ALP and AEQ [8]. Consequently, the desired signal has an underlying structure as given in (7), where – furthermore – the dimension of the auxiliary channel vector is one. The corresponding data structure can be written as follows.

$$d_n = \mathbf{h}_{opt}^H \tilde{\mathbf{u}}_n + \varepsilon_{n,opt} \quad (11)$$

$$= \begin{bmatrix} h_{opt}^{x*} & \mathbf{h}_{opt}^{rH} \end{bmatrix} \begin{bmatrix} x_n \\ \mathbf{r}_n \end{bmatrix} + \varepsilon_{n,opt}$$

We use the concept of linking sequences $\rho_n^{(i)}$ [8],

$$\rho_{n-i}^{(i)} = \frac{x_n}{r_{n-i}} \quad (12)$$

to write the data structure in (11) in equivalent form.

$$d_n = h_{opt}^{x*} \rho_{n-i}^{(i)} r_{n-i} + \mathbf{h}_{opt}^{rH} \mathbf{r}_n + \varepsilon_{n,opt} \quad (13)$$

$$= \begin{bmatrix} h_{opt}^{x*} \rho_{n-i}^{(i)} \mathbf{1}_{i+1}^H + \mathbf{h}_{opt}^{rH} \end{bmatrix} \mathbf{r}_n + \varepsilon_{n,opt}$$

The vector $\mathbf{1}_{i+1}$ has a 1 in the $(i+1)^{st}$ position as its only non-zero element. The auxiliary signal can be linked with any of the elements in the reference vector, so that the following affine combination is the most general equivalent data structure.

$$d_n = \left[\sum_{i=0}^{M-1} \alpha_i h_{opt}^{x*} \rho_{n-i}^{(i)} \mathbf{1}_{i+1}^H + \mathbf{h}_{opt}^{rH} \right] \mathbf{r}_n + \varepsilon_{n,opt} \quad (14)$$

The bracketed term represents Wiener filter coefficients operating on the reference vector only, while still incurring the error of the optimal two-channel Wiener filter. This reference-only Wiener filter equivalent has a time-varying component, due to the presence of the linking sequences, in addition to a time-invariant component, corresponding to the reference partition of the two-channel WF weights.

The model inherent in conventional NLMS, as in (1), is of the same form as the first term in (14). The latter affine combination represents the manifold of weight vector solutions from which NLMS takes its *a posteriori* weight vector, while minimizing the norm of the change in the weight vector. If the NLMS weights could capture the behavior of the bracketed term in (14), NLMS performance would equal that of the two-channel (optimal) Wiener filter. In practice, NLMS can capture a time-varying weight vector *a posteriori* at best, so that NLMS *a priori* performance is reduced – relative to the optimal two-channel WF performance – by tracking error as well as estimation error. Nevertheless, in many scenarios, conventional NLMS may be able to track to the extent that its performance is better than that of the corresponding conventional Wiener filter.

6. DYNAMIC WEIGHT BEHAVIORS

For high SNR conditions, and in scenarios as indicated in the previous section, the bracketed term in (14) forms the target weight vector for NLMS. This implies that the weight behavior of NLMS, in particular for a stepsize $\bar{\mu} = 1$, follows the behavior indicated above. This behavior depends on the linking sequences, which in turn depend on the particular application scenarios.

We now illustrate the dynamic weight behavior for the ANC scenario with narrowband AR(1) signals (with desired and reference complex poles at p_d and p_r , respectively), for $(L, M) = (1, 2)$. For the bracketed term in (14) we then find [8]:

$$\mathbf{h}_n^{TV} = h_{opt}^{x*} \rho_{n-1}^{(0)*} \begin{bmatrix} \alpha_0 p_d^* p_r^{-*} \\ \alpha_1 p_d^* \end{bmatrix} + \begin{bmatrix} \eta_n^d \\ \eta_n^r \end{bmatrix} + \mathbf{h}_{opt}^r \quad (15)$$

where the η_n^d and η_n^r terms are results due to the driving noises of the corresponding AR(1) processes. For the remaining linking sequence the following holds:

$$\begin{aligned} \rho_{n-1}^{(0)*} &= p_d^* p_r^{-*} \rho_{n-2}^{(0)*} + \eta_n^p \\ &= \frac{|p_d|}{|p_r|} e^{-j(\omega_d - \omega_r)} \rho_{n-2}^{(0)*} + \eta_n^p \end{aligned} \quad (16)$$

Apart from the stochastic terms due to AR(1) driving noises and the time-invariant term corresponding to the

reference partition of the optimal two-channel WF, the weight vector target for NLMS is seen to consist of a time-varying term that rotates from sample to sample by the difference of the center frequencies of the desired and reference processes. Further note from (15) that the second weight vector element is rotated with respect to the first weight vector element by the complex conjugate of the reference pole.

To illustrate the dynamic NLMS weight behavior, for the ANC case, let $SNR = 80\text{dB}$, $p_d = 0.99e^{j0.7\pi}$, and $p_r = 0.99e^{j0.5\pi}$. Fig. 4 shows the absolute error in dB, during iterations 4700-5000, which is in steady state.

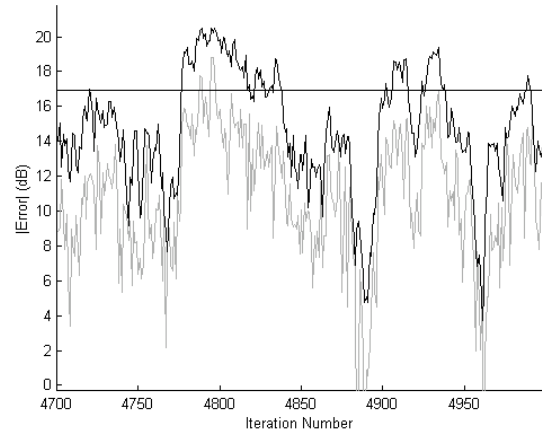


Fig. 4. Steady-state NLMS (light) and WF (dark) error.

We observe that NLMS outperforms the LTI WF, almost on a sample-for-sample basis. Over the steady-state interval, WF(0,2) – the implementation of the optimal WF – realizes MSE of 15.82 dB, while NLMS(0,2) realizes MSE of 12.04 dB. The theoretical min MSE is 16.92 dB; WF(0,2) is about 1 dB less because the signal power in this realization is about 1 dB lower than its expected value.

Fig. 5 shows the dynamic weight behavior of NLMS.

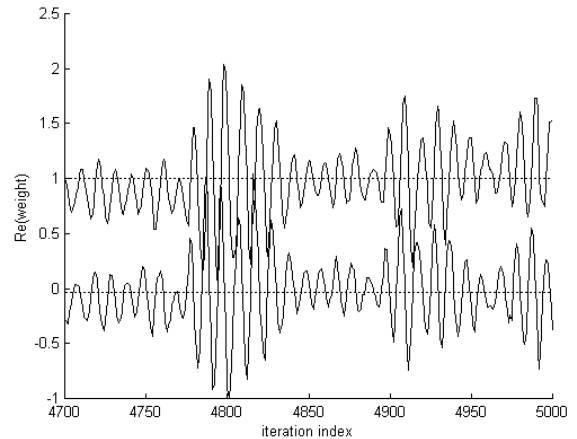


Fig. 5. NLMS and WF (constant) weight behavior.

The behavior of the imaginary part is similar. We observe that the weight behavior is semi-periodic, with a period of 10 as expected from (16) when $\omega_d - \omega_r = 0.2\pi$. We also observe that the second weight element is shifted relative to the first one by 0.5π , corresponding to ω_r .

Comparing Figs. 4 and 5 we observe that the dynamic weight behavior is strongly periodic when signal levels are relatively high (iterations 4800-4819), and less so when signal levels are lower (iterations 4700-4719). When signal levels are high, the driving noise related terms in (16) are less influential and therefore periodicity is more pronounced and more clearly centered about the reference partition of the WF(1,2) filter. The latter is further illustrated by the dynamic weight behavior of NLMS(0,2) during iterations 4700-4719, in Fig. 6, and during iterations 4800-4819, in Fig. 7.

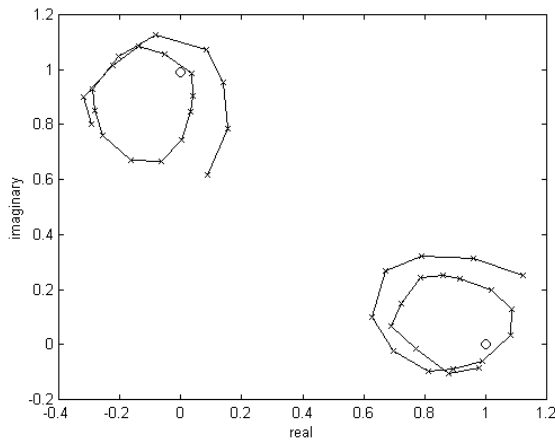


Fig. 6. NLMS and WF weights during 4700-4719.

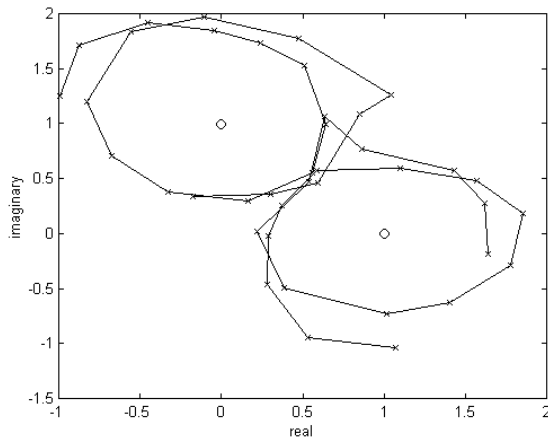


Fig. 7. NLMS and WF weights during 4800-4819.

Recall that NLMS is of necessity one step behind in tracking its time-varying equivalent target. Figs. 6 and 7 illustrate that the weight vector one sample off is generally still closer to the target than the WF(0,2) weights are.

6. SUMMARY

We have illustrated that a conventional NLMS adaptive filter, using a tapped delay line of reference values, can have a time-varying Wiener filter equivalent as its target. As a result of its tracking ability, the conventional NLMS adaptive filter can achieve better performance than the corresponding conventional Wiener filter for a given wide-sense stationary scenario. A noise-canceling example shows semi-periodic weight dynamics in tune with the behavior of the time-varying Wiener filter equivalent.

7. ACKNOWLEDGEMENT

The present work was funded by the Independent Laboratory Research Program at SPAWAR Systems Center, San Diego, and – in part – by the National Research Council, when the first author was a Senior Research Associate at SPAWAR Systems Center, San Diego, during his Fall 2001 sabbatical semester from Virginia Tech.

8. REFERENCES

- [1] M. Reuter, K. Quirk, J. Zeidler, and L. Milstein, Nonlinear effects in LMS adaptive filters, *Proc. Symp. 2000 on Adaptive Systems for Signal Processing, Communications and Control*, 141-146, Lake Louise, Alberta, October 2000.
- [2] K. J. Quirk, L. B. Milstein, and J. R. Zeidler, "A performance bound of the LMS estimator," *IEEE Trans. Information Theory*, 46, 1150-1158, May 2000.
- [3] M. Reuter and J. R. Zeidler, Nonlinear effects in LMS adaptive equalizers, *IEEE Trans. Signal Processing*, 47, 1570-1579, June 1999.
- [4] J. Han, J. R. Zeidler, and W. H. Ku, Nonlinear Effect of the LMS Adaptive Predictor for Chirped Input Signals, *EUROSIP Journal on Applied Signal Processing*, Special Issue on Nonlinear Signal and Image Processing, Part II, pp. 21-29, January 2002.
- [5] G. C. Goodwin and K. S. Sin, *Adaptive Filtering, Prediction, and Control*, Prentice-Hall, 1984.
- [6] J-P. Dugré, A. A. (Louis) Beex, and L. L. Scharf, Generating Covariance Sequences and the Calculation of Quantization and Rounding Error Variances in Digital Filters, *IEEE Trans. Acoustics, Speech, and Signal Processing*, 28, 102-104, February 1980.
- [7] A. A. (Louis) Beex, Efficient Generation of ARMA Cross-Covariance Sequences, *Proc. Int'l Conf. On Acoustics, Speech, and Signal Processing*, 327-330, Tampa FL, 26-29 March 1985.
- [8] A. A. (Louis) Beex and James R. Zeidler, "Nonlinear Effects in Adaptive Filters," in *Advances in LMS Adaptive Filters*, eds. S. Haykin and B. Widrow, John Wiley & Sons, to appear 2002.