SIGNAL RECOVERY FROM NOISY SAMPLES VIA MULTIPLE OBSERVATIONS AND THRESHOLDING

L. V. Nguyen and M. Pawlak

Department of Electrical and Computer Engineering
University of Manitoba,Winnipeg, Manitoba R3T 5V6, Canada
Emails: {nvluu, pawlak}@ee.umanitoba.ca

ABSTRACT

The problem of sampling and recovering bandlimited signals in the presence of noise is studied. A new technique based on acquisition of multiple sets of sampled values at the Nyquist sampling rate and thresholding-based reconstruction is proposed. The exact formula and upper bound for mean integrated squared error of the proposed scheme is established. The obtained results show that the proposed technique gives better reconstruction accuracy than the common reconstruction method using oversampling and post-filtering.

1. INTRODUCTION AND PRELIMINARIES

Let signal $x(t)$ belong to a class of $L_2(R)$ signals with the property that its Fourier transform $X(\omega)$ vanishes outside the finite interval $(-\Omega,\Omega)$, i.e. $X(\omega) = 0$ for $|\omega| > \Omega$. The finite number Ω is called the signal's bandwidth and the class of signals with this property is often referred to as the class of bandlimited signals, which we will denote $BL(\Omega)$ in the subsequent discussions.

The Whittaker-Shannon (W-S) sampling theorem says that every $x(t) \in BL(\Omega)$ can be reconstructed exactly from its discrete samples $x(k\tau)$, $k = 0,\pm1,\pm2,...\pm\infty$, by:

$$x(t) = \sum_{k=-\infty}^{\infty} x(k\tau)\cdot \text{sinc}\left(\frac{t}{\tau}-k\right), \tag{1.1}$$

provided that $\tau \le \tau_q = \pi/\Omega$, where $\text{sinc}(t) = \sin(\pi t)/\pi t$.

We refer to [1, 2, 3] for an extensive overview of the theory and applications of (1.1) and its extensions.

In this paper, we consider the problem of reconstructing bandlimited signals given the following observation model:

$$y_k = x(k\tau)+\varepsilon_k = x_k+\varepsilon_k, \qquad |k| \le n$$

where $\{\varepsilon_k\}$ is an additive noise process.

Our motivation comes from the fact that in many practical applications, noise is often contained in the observed samples. For instance, measurement and/or communication channel noises could be potential noise sources.

The classical W-S interpolation scheme would replace $\{x(k\tau)\}$ in (1.1) by $\{y_k\}$ yielding the following estimate:

$$\tilde{x}(t) = \sum_{|k|\le n} y_k\cdot \text{sinc}\left(\frac{t}{\tau}-k\right). \tag{1.2}$$

It has been pointed out in [4] that the above estimate does not have noise-diminishing property since it interpolates noise, i.e., $\tilde{x}(k\tau) = y_k$. The common method to overcome this problem is using oversampling and post-filtering. Consider the following estimate [5, 6]:

$$\check{x}(t) = \sum_{|k|\le n} y_k\cdot \text{sinc}\left(\frac{t}{\tau}-k\right) * h(t), \tag{1.3}$$

where, $\tau \le \tau_q = \pi/\Omega$ and $h(t) = \tau_q^{-1}\text{sinc}(t/\tau_q)$ is the impulse response of the ideal low-pass filter with cutoff frequency Ω. The block diagram of this method is illustrated in Figure 1.

Figure 1: The oversampling/post-filtering method

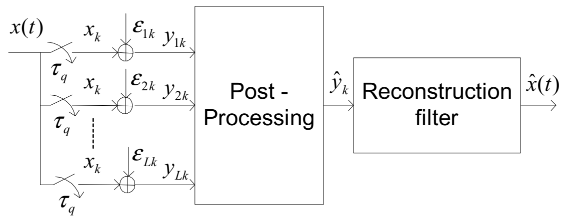

Figure 2: The proposed sampling/reconstruction scheme

Since the estimate in (1.3) relies on the use of high sampling rate for noise reduction, its realization requires costly high-speed electronic circuits. There also exists the potential problem of increasing the correlation between successive samples as the sampling rate increases, which leads to unstable reconstruction.

In this paper, instead of oversampling, we propose a new technique that collects multiple sets of the sampled values acquired at the Nyquist sampling rate and uses thresholding-based reconstruction algorithm recently proposed in [7] for recovering the signal. The block diagram of the proposed technique is illustrated in Figure 2. According to the figure,

$$y_{lk} = x_k+\varepsilon_{lk}, \qquad k = 0,\pm1,...,\pm n_q \qquad l = 1,2,\cdots,L \tag{1.4}$$

where, ε_{lk} are independent random variables with zero-mean and variance σ^2. Here, $x_k = x(k\tau_q)$ and L is the number of sets of sampled values acquired at the Nyquist rate. Since the signal is observed within a fixed internal, n_q relates to n in (1.3) by $(2n+1)\tau = (2n_q+1)\tau_q$, or equivalently, $(2n+1)/(2n_q+1) = \tau_q/\tau$.

This work was supported in part by TR*Labs* and NSERC

We propose the following 3-step procedure for reconstruction of $x(t)$ from $\{y_{lk}\}$:

Step 1: Calculate the sample means

$$\bar{y}_k = \frac{1}{L}\sum_{l=1}^{L} y_{lk}$$

Step 2: Thresholding $\bar{y}_k$ to obtain:

$$\hat{y}_k = \begin{cases} \bar{y}_k & \text{if } |\bar{y}_k| > T \\ 0 & \text{if } |\bar{y}_k| \le T, \end{cases} \tag{1.5}$$

where, T is the threshold level. Its value is to be determined later.

Step 3: Using $\hat{y}_k$ founded above to form the following estimate:

$$\hat{x}(t) = \sum_{|k|\le n_q} \hat{y}_k \cdot \mathrm{sinc}\left(\frac{t}{\tau_q} - k\right). \tag{1.6}$$

The paper is organized as follows. In Section 2, we give the exact formula for computing mean integrated squared error (MISE) of the proposed reconstruction scheme. In Section 3, we investigate the influence of the choice of threshold level T on the reconstruction accuracy. Section 4 provides some simulation results. Section 5 concludes the paper.

2. ERROR ANALYSIS

Let us consider MISE as a measure of performance of $\hat{x}(t)$ and other estimates examined in the paper.

$$MISE(\hat{x}) \triangleq E\int_{-\infty}^{\infty} (\hat{x}(t) - x(t))^2 dt.$$

The exact formula for MISE of the proposed reconstruction scheme is presented in the following theorem. Due to the space limitation, we do not give the proof of the theorem in this paper.

Theorem 2.1. *Consider the observation model in (1.4), where ε_{lk} are independent Gaussian random variables with zero mean and variance σ^2. The MISE of the estimate $\hat{x}(t)$ in (1.6) is given by:*

$$MISE(\hat{x}) = \tau_q\sigma_1^2(2n_q+1) + \tau_q\sum_{|k|>n_q} x_k^2 - \tau_q\sigma_1^2 D, \tag{2.1}$$

with,

$$\begin{aligned} D = f(x_k,\sigma_1^2,T) = & \\ & \sum_{\substack{|k|\le n_q \\ x_k<-T}} \frac{1}{\sqrt{\pi}}\left[\gamma\left(\frac{3}{2},\frac{(x_k-T)^2}{2\sigma_1^2}\right) - \gamma\left(\frac{3}{2},\frac{(x_k+T)^2}{2\sigma_1^2}\right)\right] \\ + & \sum_{\substack{|k|\le n_q \\ |x_k|\le T}} \frac{1}{\sqrt{\pi}}\left[\gamma\left(\frac{3}{2},\frac{(x_k-T)^2}{2\sigma_1^2}\right) + \gamma\left(\frac{3}{2},\frac{(x_k+T)^2}{2\sigma_1^2}\right)\right] \\ + & \sum_{\substack{|k|\le n_q \\ x_k>T}} \frac{1}{\sqrt{\pi}}\left[\gamma\left(\frac{3}{2},\frac{(x_k+T)^2}{2\sigma_1^2}\right) - \gamma\left(\frac{3}{2},\frac{(x_k-T)^2}{2\sigma_1^2}\right)\right] \\ - & \sum_{|k|\le n_q} \frac{x_k^2}{\sigma_1^2}\left[\frac{1}{2}\mathrm{erf}\left(\frac{x_k+T}{\sigma_1\sqrt{2}}\right) - \frac{1}{2}\mathrm{erf}\left(\frac{x_k-T}{\sigma_1\sqrt{2}}\right)\right], \end{aligned} \tag{2.2}$$

where, $\sigma_1^2 = \sigma^2/L$; $\gamma(\alpha,x)$ and erf(x) *are incomplete gamma function and error function, respectively [8].*

Theorem 2.2. *For a given signal $x(t)$ and noise variance σ^2, there always exists the optimum threshold level T_{opt} so that $MISE(\hat{x})$ is minimum. This is the value of T, at which the term D defined in theorem 2.1 is maximum.*

Proof. $MISE(\hat{x})$ can be decomposed into the integrated variance (IVAR) and integrated bias (IBIAS) components as follows.

$$MISE(\hat{x}) = IVAR(\hat{x}) + IBIAS(\hat{x})$$

$$IBIAS(\hat{x}) \triangleq \int_{-\infty}^{\infty} (E\hat{x}(t) - x(t))^2 dt$$

$$IVAR(\hat{x}) \triangleq E\int_{-\infty}^{\infty} (\hat{x}(t) - E\hat{x}(t))^2 dt.$$

The term $IBIAS(\hat{x})$ describes the truncation error and the error due to thresholding operation. On other hand, the term $IVAR(\hat{x})$ reflects the existence of random errors in the observed samples.

Parseval's formula yields the following decompositions of the errors [4]:

$$IBIAS(\hat{x}) = \tau_q\sum_{|k|\le n_q}(E\hat{y}_k - x_k)^2 + \tau_q\sum_{|k|>n_q} x_k^2$$

$$IVAR(\hat{x}) = \tau_q\sum_{|k|\le n_q}\mathrm{var}(\hat{y}_k).$$

Taking (1.5) into account, we can write:

$$\begin{aligned} IBIAS(\hat{x}) &= \tau_q E\Big[\sum_{k:\ |\bar{y}_k|>T}(E\bar{y}_k - x_k)^2 + \sum_{k:\ |\bar{y}_k|\le T}(0 - x_k)^2\Big] \\ &\quad + \tau_q\sum_{|k|>n_q} x_k^2 \\ &= \tau_q E\sum_{k\in I_d} x_k^2 + \tau_q\sum_{|k|>n_q} x_k^2 \end{aligned} \tag{2.3}$$

$$IVAR(\hat{x}) = \tau_q E\Big[\sum_{k:\ |\bar{y}_k|>T}\mathrm{var}(\bar{y}_k)\Big] = \tau_q\sigma_1^2 E|I_p|, \tag{2.4}$$

where, $I_p = \{k:\ |\bar{y}_k| > T;\ |k| \le n_q\}$ and $I_d = \{k:\ |\bar{y}_k| \le T;\ |k| \le n_q\}$ are two random index sets; $|I_p|$ is the volume (size) of I_p.

As T increases, $E|I_p|$ decreases and $IVAR(\hat{x})$ decreases accordingly. On other hand, as T increases, the first term in the RHS of (2.3) increases and $IBIAS(\hat{x})$ increases accordingly. Since $IVAR(\hat{x})$ is a monotonically decreasing function of T, whereas $IBIAS(\hat{x})$ is a monotonically increasing function of T, the sum $MISE(\hat{x}) = IVAR(\hat{x}) + IBIAS(\hat{x})$ has the minimum at the value of T, for which $IVAR(\hat{x}) = IBIAS(\hat{x})$. This value of T is called the optimal threshold level.

Since the first two terms in the RHS of (2.1) do not depend on T, we conclude that the optimal threshold level is the value of T, at which the term D defined in (2.2) is maximum. The proof is complete. □

Remark 2.1. Let assume that the proposed reconstruction scheme and the reconstruction scheme defined in (1.3) use the same number of sampled values for signal reconstruction, i.e., $2n+1 = (2n_q+1)L$. Since we have assumed that the duration of observation is fixed, i.e., $\tau_q(2n_q+1) = \tau(2n+1)$, we have $L = \tau_q/\tau$. It can be shown that the first term in the RHS of (2.1) equals to $IVAR(\check{x})$ [5].

3. CHOICE OF THRESHOLD LEVEL

From theorem 2.2, we know that in order to minimize $MISE(\widehat{x})$, one should choose T such that the term D defined in (2.2) is maximum. The term D can be pre-computed during dada acquisition process and the optimal threshold level T_{opt} can be pre-determined accordingly. T_{opt} is considered as the side information and can be made available for signal reconstruction by various methods. For example, with $L = 4$ and the testing signal in (4.1), the optimal threshold levels at various SNRs are given in the table 1. It is observed from

Table 1: The optimal threshold levels at various E_0/σ^2 ($L = 4$)

E_0/σ^2 (dB)	-25	-20	-15	-10
T_{opt}	48.04	46.28	27.00	15.94
E_0/σ^2 (dB)	-5	0	5	10
T_{opt}	1.01	0.107	0.055	0.001

this table that the optimal threshold level increases as noise variance increases, and decreases as noise variance decreases. If $\sigma^2 \to 0$, then $T_{opt} \to 0$.

Reconstruction of the signal using the optimal threshold level requires pre-determination of T_{opt} and a mechanism for delivering T_{opt} to the receiver. Therefore, we suggest a sub-optimal method, in which threshold level T is set adaptively to the noise variance. For instance, one may set $T = \sigma_1\sqrt{a}$, where a is the parameter used to control the reconstruction accuracy. In order to investigate the dependence of $MISE(\widehat{x})$ on parameter a, the following theorem is established. Due to the space limitation, we do not give the proof of this theorem here.

Theorem 3.1. *Let the assumptions of theorem 2.1 remain in force. The MISE of the estimate $\widehat{x}(t)$ in (1.6) is upper bounded by:*

$$MISE(\widehat{x}) \leq \tau_q\sigma_1^2(2n_q+1) + \tau_q \sum_{|k|>n_q} x_k^2 - \tau_q\sigma_1^2(2n_q+1)\cdot$$
$$\cdot\left[\frac{2}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{T^2}{2\sigma_1^2}\right) - \frac{\sum_{|k|\leq n_q} x_k^2}{(2n_q+1)\sigma_1^2}\mathrm{erf}\left(\frac{T}{\sigma_1\sqrt{2}}\right)\right].$$

The above inequality can be re-written as follows:

$$MISE(\widehat{x}) \leq \tau_q\sigma_1^2(2n_q+1) + \tau_q \sum_{|k|>n_q} x_k^2 - \tau_q\sigma_1^2(2n_q+1) f(G_{nq},T),$$

where,

$$f(G_{nq},T) = \frac{2}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{T^2}{2\sigma_1^2}\right) - G_{nq}\mathrm{erf}\left(\frac{T}{\sigma_1\sqrt{2}}\right)$$

and,

$$G_{nq} \triangleq \frac{\frac{1}{2n_q+1}\sum_{|k|\leq n_q} x_k^2}{\sigma_1^2},$$

Equivalently, we have:

$$\Delta \triangleq MISE(\widehat{x})|_{T=0} - MISE(\widehat{x}) \geq \tau_q\sigma_1^2(2n_q+1) f(G_{nq},T).$$

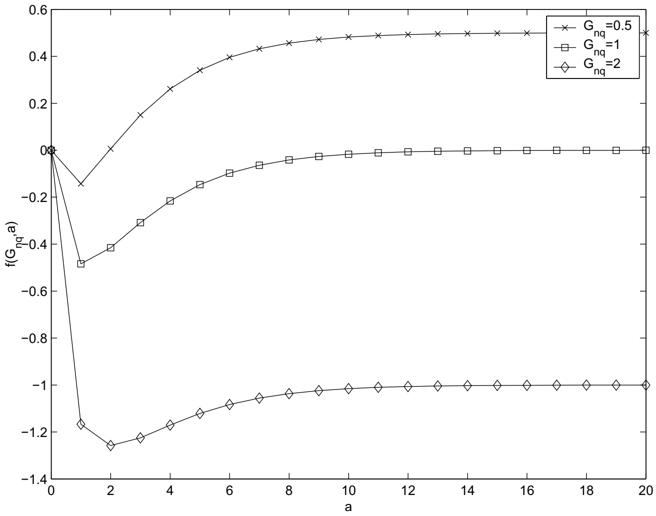

Figure 3: $f(G_{nq},a)$ versus a

Consequently, to maximize Δ, we have to choose T such that $f(G_{nq},T)$ is maximum. Figure 3 plots $f(G_{nq},a)$ as the function of a for various G_{nq}, where a relates to T by $T = \sigma_1\sqrt{a}$. The following conclusions can be drawn from the plot:

- For $G_{nq} \geq 1$, one should select $T = 0$.
- For $G_{nq} < 1$, $T = 0$ is no longer the best choice. In fact, one should select $3\sigma_1 \leq T \leq 5\sigma_1$.

Taking the above observations into account, we further propose the following joint detection/estimation scheme for recovering signals from noisy data:

Step 1: Estimating $P \triangleq \sum_{|k|\leq n_q} x_k^2$ from data. Computing G_{nq} using the estimated P

Step 2: If $G_{nq} \geq 1$, setting $T = 0$. Otherwise, setting T within $[3\sigma_1, 5\sigma_1]$

Step 3: Following 3-step procedure introduced in section 1 to reconstruct the original signal

Remark 3.1. For sufficiently large n_q and/or for fast decay functions, we do not need to estimate P in order to determine G_{nq}. In this case we have $\sum_{|k|\leq n_q} x_k^2 \approx E_0/\tau_q$, which implies $G_{nq} \approx E_0/(\tau_q\sigma_1^2(2n_q+1))$. We assume that the signal-to-noise ratio E_0/σ^2 is known in advance.

4. SIMULATION RESULTS

The following signal is used as the testing signal in simulations:

$$x(t) = \frac{1}{\sqrt{\pi\Omega}}\cdot\frac{\sin\Omega t}{t}. \quad (4.1)$$

The above signal is the unit energy $BL(\Omega)$ signal, where $\Omega = 2\pi f_{max}$ is the radian frequency. In our simulation, we set $f_{max} = 3900$ Hz. This signal has the smallest rate of decay as $t \to \infty$ among all bandlimited signals having the same bandwidth. Given the fact that the proposed scheme works especially well for fast-decay signals, our use of this signal as the testing signal is justified.

The experiment is repeated $M = 100$ times for various realization of random errors $\{\varepsilon_{lk}\}$. The empirical counterpart of MISE, further called EMISE, is calculated according to

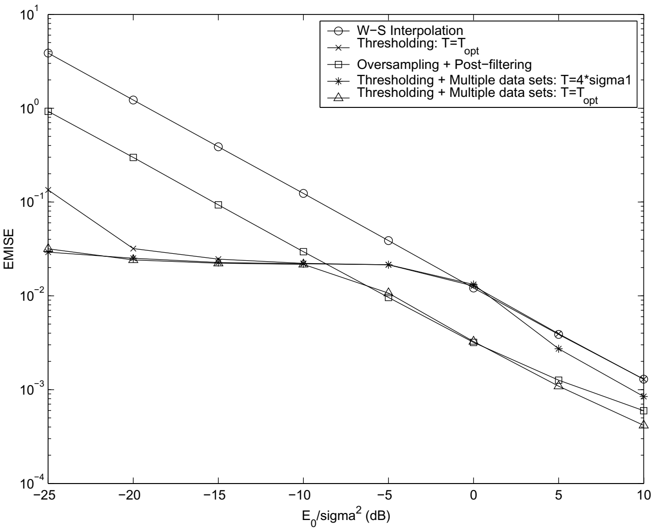

Figure 4: $EMISE(\hat{x})$ versus E_0/σ^2 (OSF=4)

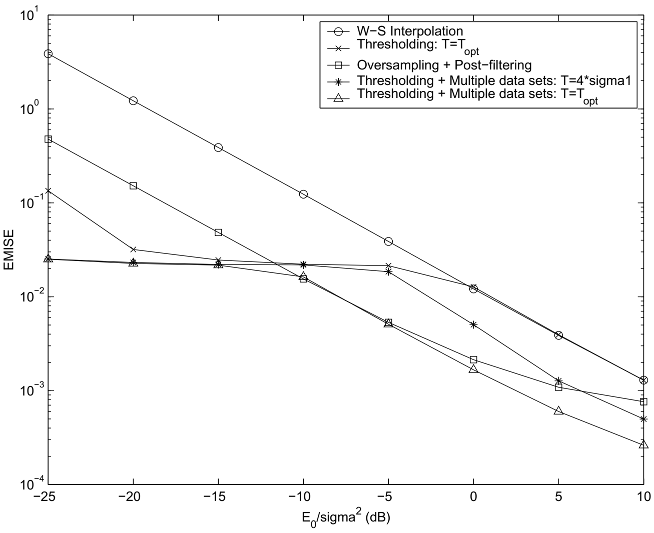

Figure 5: $EMISE(\hat{x})$ versus E_0/σ^2 (OSF=8)

the following formula:

$$EMISE(\hat{x}) = \frac{\bar{\tau}}{M} \sum_{j=1}^{M} \sum_{k=-n}^{n} [\hat{x}(k\bar{\tau}) - x(k\bar{\tau})]^2,$$

where $\bar{\tau} \ll \tau_q$ is the simulation sampling period.

Figure 4 and Figure 5 plot the EMISE of the proposed sampling and reconstruction scheme when $L = 4$ and 8, respectively. In this simulation, $2n_q = 100$ and $\tau_q = 1.25 \cdot 10^{-4}$ seconds (sampling frequency is just slightly larger than the Nyquist rate). For comparison, the EMISE of the thresholding-based reconstruction scheme proposed in [7] and the EMISE of the oversampling/post-filtering reconstruction scheme defined in (1.3) are also plotted. As being indicated in the plots, when threshold level T is optimally chosen, the proposed scheme gives better reconstruction accuracy than the others for every values of SNR. However, the significant improvement in the reconstruction accuracy is observed at the region of low SNR. If threshold level T is chosen according to the sub-optimal criteria suggested in the previous section, i.e. $T = 4\sigma_1$, the reconstruction accuracy of the proposed scheme does not seem to be effected at the region of low SNR, but degradation in the reconstruction accuracy is noticed at the region of high SNR.

5. CONCLUSIONS

A new method for acquisition and reconstruction of band-limited signals in the presence of noise was presented. The proposed scheme acquires multiple sets of the signal's samples at the Nyquist sampling rate and uses thresholding-based reconstruction algorithm introduced in [7] for recovering the signal. We showed that threshold level can be optimally chosen to yield the minimum reconstruction error. We also suggested a sub-optimal criteria for setting threshold level, which eliminates the need for delivering the side information (i.e. the optimal threshold level) to the receiver. The obtained results showed that the proposed technique gives better reconstruction accuracy than the oversampling/post-filtering method.

The proposed technique has a potential application in designing signal acquisition and reconstruction systems operating in highly noisy environments. Since the proposed technique does not use oversampling, it can be applied to acquire ultra wideband signals. It also has the applications in multi-sensor information systems.

REFERENCES

[1] A.J. Jerri, "The Shannon sampling theorem - its various extensions and applications: A tutorial review," *Proceedings of the IEEE*, vol. 65, pp. 1565–1596, Nov. 1977.

[2] R.J. Marks II, *Introduction to Shannon Sampling and Interpolation Theory*, Springer-Verlag, New York, NY, 1991.

[3] M. Unser, "Sampling-50 years after Shannon," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.

[4] M. Pawlak and E. Rafajlowicz, "On restoring band-limited signals," *IEEE Transactions on Information Theory*, vol. 40, pp. 1490–1503, Sep. 1994.

[5] M. Pawlak, E. Rafajlowicz, and A. Krzyzak, "Postfiltering versus prefiltering for signal recovering from noisy samples," *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3195–3212, Dec. 2003.

[6] M. Pawlak and U. Stadmuller, "Statistical aspects of sampling for noisy and grouped data," *In J. Benedetto and P. Ferreira, editors, Advances in Shannon Sampling Theory: Mathematics and Applications*, Boston, TX, pp. 317-342, 2001.

[7] Vu-Luu Nguyen and M. Pawlak, "Reconstruction of bandlimited signals from noisy data by thresholding," *Proc. International Conference on Acoustics, Speech, and Signal Processing*, pp. 169–172, Apr. 2003.

[8] I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, New York, 4th edition, 1980.