International Scholarly Research Network
ISRN Applied Mathematics
Volume 2012, Article ID 630702, 7 pages
doi:10.5402/2012/630702

Research Article

Sign Data Derivative Recovery

L. M. Houston,[1] G. A. Glass,[2] and A. D. Dymnikov[1]

[1] *Louisiana Accelerator Center, The University of Louisiana at Lafayette, Lafayette, LA 70504-4210, USA*
[2] *Ion Beam Modification and analysis Laboratory, Department of Physics, University of North Texas, Denton, TX 76203, USA*

Correspondence should be addressed to L. M. Houston, houston@louisiana.edu

Received 2 November 2011; Accepted 29 November 2011

Academic Editors: J. Shen and F. Zirilli

Given only the signs of signal plus noise added repetitively or sign data, signal amplitudes can be recovered with minimal variance. However, discrete derivatives of the signal are recovered from sign data with a variance which approaches infinity with decreasing step size and increasing order. For industries such as the seismic industry, which exploits amplitude recovery from sign data, these results place constraints on processing, which includes differentiation of the data. While methods for smoothing noisy data for finite difference calculations are known, sign data requires noisy data. In this paper, we derive the expectation values of continuous and discrete sign data derivatives and we explicitly characterize the variance of discrete sign data derivatives.

1. Introduction

Sign-bit recording systems discard all information on the detailed motion of the geophone and ask only whether its output is positive or negative, whether it is going up or coming down. In a sign-bit system, therefore, the signal waveform is converted into a square wave. All amplitude information is lost [1].

It is well known that, for a range of signal-to-noise ratios between about 0.1 and 1, the final result of sign-bit recording, after stacking, correlating, and other processing, looks no less good, to the eye, than the result from full-fidelity recording. This is considered to be as intriguing as it is surprising [1]. Alternatively, what we present in this paper is evidence that the processing of sign-bit data (i.e., sign data) can be limited for certain cases relative to the processing of the full-bandwidth data.

Model signal appears as a one-dimensional function, $f(v)$, and noise as a random variable, X. In industries like the seismic industry, measurements of signal, $f(v) : \mathbb{R} \to \mathbb{R}$

and noise, $X : \Omega \to \mathbb{R}$, $f(v) + X$ are recorded for multiple iterations of the noise. The average of the measurement (i.e., the expectation E) recovers the signal

$$E\big(f(v) + X\big) = f(v). \tag{1.1}$$

If the noise is chosen to be uniform, where $\rho(x)$ is the density function such that

$$\rho(x) = \begin{cases} \dfrac{1}{2a}, & -a \le x \le a \\ 0, & \text{else}, \end{cases} \tag{1.2}$$

then the variance, $E(f(v) + X)^2 - (E(f(v) + X))^2$, reduces to

$$\operatorname{Var}\big(f(v) + X\big) = \frac{1}{3}a^2. \tag{1.3}$$

As reported by O'Brien et al. [2], it was empirically discovered that the average of the signs of signal plus noise recovers the signal if the signal-to-noise ratio is less than or equal to one. This can be shown mathematically [3] using the signum function [4], $\operatorname{sgn}(x) = +1, x > 0$, $\operatorname{sgn}(x) = -1, x < 0$, $\operatorname{sgn}(0) = 0$,

$$E\big(\operatorname{sgn}\big(f(v) + X\big)\big) = \int_{-\infty}^{\infty} \operatorname{sgn}\big(f(v) + x\big)\rho(x)dx = \int_{-f}^{\infty} \rho(x)dx - \int_{-\infty}^{-f} \rho(x)dx. \tag{1.4}$$

Because $\rho(x)$ is even and equals

$$\int_{-f}^{f} \rho(x)dx \tag{1.5}$$

$$E\big(\operatorname{sgn}\big(f(v) + X\big)\big) = \frac{f(v)}{a}, \quad f \in [-a, a]. \tag{1.6}$$

The variance is $E(\operatorname{sgn}(f(v) + X))^2 - (E(\operatorname{sgn}(f(v) + X)))^2$, reducing to

$$\operatorname{Var}\big(\operatorname{sgn}\big(f(v) + X\big)\big) = 1 - \left(\frac{f(v)}{a}\right)^2. \tag{1.7}$$

Consequently, the error is minimal when the signal-to-noise ratio is near unity.

The advantage of retaining only the signs of signal plus noise is the requirement of approximately 1 bit to record the information as opposed to requiring 16 to 20 bits to record full amplitude data [2].

The goal of this paper is to examine the recovery of derivatives from sign data in uniform noise. The issue is that recovery of signal from sign data can be extended to recovery of derivatives of the signal through the use of finite differences and that recovery is constrained by the size of the variance. In this paper, we first examine sign data derivatives for both

the discrete and continuous case. We follow with a derivation of variance. We conclude our analysis with a computational test, which lists the true variance versus the variance estimate derived statistically for a test function for selected step sizes.

2. Sign Data Derivatives

Let the signal $f(v)$ be an nth order differentiable function. Based on signal recovery from sign data, it can be shown that derivatives of the signal are also recoverable. Using the linearity of the expectation value,

$$E\left(\frac{\Delta_v^n}{(\Delta v)^n}\operatorname{sgn}(f(v)+X)\right)=\frac{\Delta_v^n}{(\Delta v)^n}E(\operatorname{sgn}(f(v)+X)), \tag{2.1}$$

where Δ_v^n is the nth order finite difference operator with respect to the variable v [5]. In this case, a nonunit step size, Δv, is used (e.g., [6]).

In detail, we can write

$$\frac{\Delta_v^n}{(\Delta v)^n}\operatorname{sgn}(f(v)+X)=\frac{1}{(\Delta v)^n}\sum_{i=0}^{n}(-1)^i\binom{n}{i}\operatorname{sgn}(f(v+(n-i)\Delta v+X_i)), \tag{2.2}$$

where the notation $\binom{n}{i}$ represents the binomial coefficient $n!/i!(n-i)!$ and where $X_i = X_0$, $X_1,\ldots$ are independent representations of the random variable, X.

Substituting from (1.6) into (2.1) yields

$$E\left(\frac{\Delta_v^n}{(\Delta v)^n}\operatorname{sgn}(f(v)+X)\right)=\frac{1}{a}\frac{\Delta_v^n f(v)}{(\Delta v)^n}. \tag{2.3}$$

In the limit of infinitesimal step size, this becomes a continuous derivative

$$\lim_{\Delta v\to 0}E\left(\frac{\Delta_v^n}{(\Delta v)^n}\operatorname{sgn}(f(v)+X)\right)=\frac{1}{a}\frac{d^n f(v)}{dv^n} \tag{2.4}$$

or

$$E\left(\frac{d^n}{dv^n}\operatorname{sgn}(f(v)+X)\right)=\frac{1}{a}\frac{d^n f(v)}{dv^n}. \tag{2.5}$$

Equation (2.4) presents an alternative solution to direct integration. For example, using the rule, $\int f(x)\delta^{(n)}(x)dx = -\int(\partial f/\partial x)\delta^{(n-1)}(x)dx$, [7], the integral

$$E\left(\frac{d^3}{dv^3}\operatorname{sgn}(f(v)+X)\right)=\int_{-\infty}^{\infty}\left(2\frac{d^2\delta}{du^2}\left(\frac{df}{dv}\right)^3+6\frac{d\delta}{du}\frac{df}{dv}\frac{d^2f}{dv^2}+2\delta\frac{d^3f}{dv^3}\right)\rho(x)dx \tag{2.6}$$

loses all terms with derivatives of the delta functional, reducing to

$$= 2\rho(-f)\frac{d^3 f}{dv^3}\bigg|_{f=-x}. \tag{2.7}$$

In general,

$$E\left(\frac{d^n}{dv^n}\operatorname{sgn}(f(v)+X)\right) = 2\rho(-f)\frac{d^n f}{dv^n}\bigg|_{f=-x} = \frac{1}{a}\frac{d^n f}{dv^n}. \tag{2.8}$$

It follows that the noise is restricted such that $a \geq |f|$.

3. The Variance of Sign Data Derivatives

Letting $S_n \equiv (\Delta_v^n/(\Delta v)^n)\operatorname{sgn}(f(v)+X)$, compute the variance, $E(S_n^2)-(E(S_n))^2$. From (2.3), it follows that $(E(S_n))^2 = (\Delta_v^n f/a(\Delta v)^n)^2$. $E(S_n^2)$ can be found by inductively generalizing from $n = 2$:

$$\begin{aligned} E\left(S_2^2\right) &= \frac{1}{\Delta v^4}\left(b_0\operatorname{sgn}(f_0+X_0)+b_1\operatorname{sgn}(f_1+X_1)+b_2\operatorname{sgn}(f_2+X_2)\right)^2 \\ &= \frac{1}{\Delta v^4}\left(b_0^2+b_1^2+b_2^2+2b_0b_1\left(\frac{f_0f_1}{a^2}\right)+2b_0b_2\left(\frac{f_0f_2}{a^2}\right)+2b_1b_2\left(\frac{f_1f_2}{a^2}\right)\right), \end{aligned} \tag{3.1}$$

where $f_i = f(v+(n-i)\Delta v)$, $f_k = f(v+(n-k)\Delta v)$, and $b_i = (-1)^i\binom{n}{i}$.

These results generalize to

$$\begin{aligned} \operatorname{Var}(S_n) &= \frac{1}{(\Delta v)^{2n}}\sum_{i=0}^{n}\binom{n}{i}^2 \\ &\quad + \frac{2}{(\Delta v)^{2n}}\sum_{i\neq k}^{n}(-1)^{i+k}\binom{n}{i}\binom{n}{k}\left(\frac{f_if_k}{a^2}\right)-\left(\frac{\Delta_v^n f}{a(\Delta v)^n}\right)^2. \end{aligned} \tag{3.2}$$

Since f is differentiable, $|(\Delta_v^n f/(\Delta v)^n)-(d^n f/dv^n)| < \varepsilon$ and, thus, $\Delta_v^n f/(\Delta v)^n$ is finite. Based on definition, $\operatorname{Var}(S_n) > 0$.

Consequently, $\lim_{\Delta v\to 0}\operatorname{Var}(S_n) = +\infty$. Similarly, $\lim_{n\to\infty}\operatorname{Var}(S_n) = +\infty, 0 < \Delta v < 1$. The variance of a discrete sign derivative approaches infinity with decreasing step size and increasing order. In addition, since $\lim_{\Delta v\to 0}(S_n) = (d^n/dv^n)\operatorname{sgn}(f(v)+X)$, $\operatorname{Var}((d^n/dv^n)\operatorname{sgn}(f(v)+X)) = +\infty$, so in the case of the continuous derivatives (2.5) the variance is infinite.

Use (3.2) to find the variance of the first discrete sign derivative by letting $n = 1$:

$$\operatorname{Var}(S_1) = \frac{1}{(\Delta v)^2}\left(2-\frac{(f_1^2+f_0^2)}{a^2}\right). \tag{3.3}$$

Table 1: True variance, $\mathrm{Var}(S_1)$, versus the variance estimate, $\mathrm{Var}_N(S_1)$, for the function $f = \sin(v)$, with the number of iterations $N = 1000$, $a = 1$, and $v = 3$.

Δv	$\mathrm{Var}(S_1)$	$\mathrm{Var}_N(S_1)$
1.0	1.4073	1.4513
0.5	7.4281	7.3562
0.2	49.4169	49.3202
0.1	197.8356	199.5195
0.04	1231.1	1283.2
0.02	4913.4	4739.8
0.01	19629	20401

Table 2: True variance, $\mathrm{Var}(S_2)$, versus the variance estimate, $\mathrm{Var}_N(S_2)$, for the function $f = \sin(v)$, with the number of iterations $N = 1000$, $a = 1$, and $v = 3$.

Δv	$\mathrm{Var}(S_2)$	$\mathrm{Var}_N(S_2)$
1.0	2.7695	2.7192
0.5	78.6422	80.2276
0.2	3688.2	3752.3
0.1	59698	59000
0.04	2.3×10^6	2.3×10^6
0.02	3.69×10^7	3.685×10^7
0.01	5.89×10^8	5.988×10^8

The variance of the second discrete sign derivative ($n = 2$) is similarly computed as

$$\mathrm{Var}(S_2) = \frac{1}{(\Delta v)^4}\left(6 - \frac{1}{a^2}\left(f_0^2 + 4f_1^2 + f_2^2\right)\right). \tag{3.4}$$

4. Computational Tests

These results can be tested computationally. Variance can be estimated for N iterations with

$$\mathrm{Var}_N(S_n) = \frac{1}{N}\sum_{m=1}^{N}(S_n(m) - E(S_n)), \tag{4.1}$$

where the index m designates the sample number.

Consider the test function $f = \sin(v)$. Using the first-order sign data derivative ($n = 1$), compare $\mathrm{Var}(S_1)$ to $\mathrm{Var}_N(S_1)$, and using the second-order sign data derivative ($n = 2$), compare $\mathrm{Var}(S_2)$ to $\mathrm{Var}_N(S_2)$ for $N = 1000$, $a = 1$, and $v = 3$. The results are shown in Tables 1 and 2.

We illustrate the change in variance in Figure 1, which shows three curves, each consisting of $N = 1000$ iterations. The first curve in blue shows the sign data recovery of the function $f = \sin(v)$ or $E(S_0)$ for $a = 1$ and $\Delta v = 0.5$. The second curve in green shows the sign data recovery $E(S_1)$, which approximates f' for $a = 1$ and $\Delta v = 0.5$. The third curve in red shows the sign data recovery $E(S_2)$, which approximates f'' for $a = 1$ and $\Delta v = 0.5$.

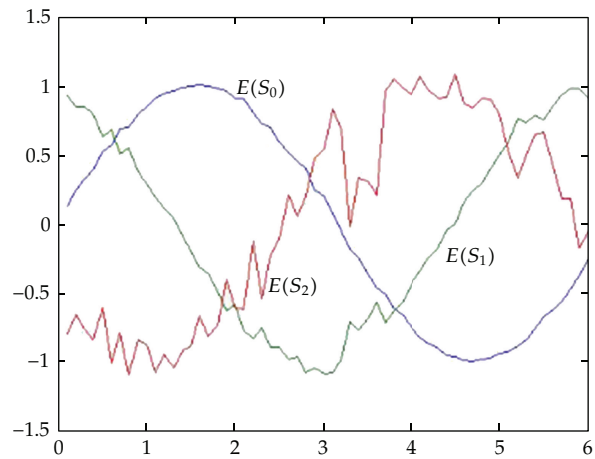

Figure 1: The expectation value curves for $S_n = (1/(\Delta v)^n)\sum_{i=0}^{n}(-1)^i\binom{n}{i}\operatorname{sgn}(f(v+(n-i)\Delta v+X_i))$ or $E(S_n)$ for $n = 0,\ 1,\ 2,\ f(v) = \sin(v),\quad a = 1$, and $\Delta v = 0.5$. The number of iterations in the expectation values is $N = 1000$. $E(S_0)$ corresponds to the blue curve and approximates f, $E(S_1)$ corresponds to the green curve and approximates f', and $E(S_2)$ corresponds to the red curve and approximates f''.

5. Conclusions

Recovery of signal from the signs of signal plus noise incurs a variance, which only depends on the noise amplitude, while recovery of discrete derivatives from the signs of signal plus noise (i.e., sign data) incurs a variance which grows infinite for infinitesimal step size and infinite order.

The application problem is that sign data can be used in the seismic industry in processes which may differentiate the data. In such cases, if the step size or order of the finite difference is not constrained, the process will incur large variance and convergence of the process will be minimized. While methods for smoothing noisy data for finite difference calculations are known, sign data requires noisy data. In this paper, we have characterized the problem by explicitly evaluating the variance of discrete sign data derivatives.

Appendix

Clarification of $E(S_2^2)$

$$\begin{aligned}
E\left(S_2^2\right) &= \frac{1}{\Delta v^4}\left(b_0\operatorname{sgn}(f_0+X_0)+b_1\operatorname{sgn}(f_1+X_1)+b_2\operatorname{sgn}(f_2+X_2)\right)^2 \\
&= \frac{1}{\Delta v^4}\Big(b_0^2\operatorname{sgn}^2(f_0+X_0)+2b_0\operatorname{sgn}(f_0+X_0)b_1\operatorname{sgn}(f_1+X_1)+b_1^2\operatorname{sgn}^2(f_1+X_1) \\
&\qquad +2b_2\operatorname{sgn}(f_2+X_2)b_0\operatorname{sgn}(f_0+X_0)+2b_2\operatorname{sgn}(f_2+X_2)b_1\operatorname{sgn}(f_1+X_1) \\
&\qquad +b_2^2\operatorname{sgn}^2(f_2+X_2)\Big).
\end{aligned} \tag{A.1}$$

This simply reduces to

$$E\left(S_2^2\right) = \frac{1}{\Delta v^4}\Big(b_0^2 + 2b_0b_1E\big(\operatorname{sgn}(f_0 + X_0)\operatorname{sgn}(f_1 + X_1)\big) + b_1^2 + 2b_2b_0E\big(\operatorname{sgn}(f_2 + X_2)\operatorname{sgn}(f_0 + X_0)\big) + 2b_2b_1E\big(\operatorname{sgn}(f_2 + X_2)\operatorname{sgn}(f_1 + X_1)\big) + b_2^2\Big). \tag{A.2}$$

In order to compute (A.2), we must compute an integral of the form

$$E\big(\operatorname{sgn}(f_i + X)\operatorname{sgn}(f_k + X)\big) = \iint_{-\infty}^{\infty} \operatorname{sgn}(f_i + x_i)\operatorname{sgn}(f_k + x_k)\rho(x_i)\rho(x_k)dx_i\,dx_k. \tag{A.3}$$

The probability densities are both uniform:

$$\rho(x_i) = \rho(x_k) = \begin{cases} \dfrac{1}{2a}, & -a \le x \le a, \\ 0, & \text{else} \end{cases} \tag{A.4}$$

and using the results of (1.6),

$$E\big(\operatorname{sgn}(f_i + X)\operatorname{sgn}(f_k + X)\big) = \frac{f_i f_k}{a^2}. \tag{A.5}$$

Consequently, (A.2) reduces to

$$E\left(S_2^2\right) = \frac{1}{\Delta v^4}\left(b_0^2 + b_1^2 + b_2^2 + 2b_0b_1\left(\frac{f_0f_1}{a^2}\right) + 2b_0b_2\left(\frac{f_0f_2}{a^2}\right) + 2b_1b_2\left(\frac{f_1f_2}{a^2}\right)\right). \tag{A.6}$$

Acknowledgment

Thanks are due to Gwendolyn Houston for advice and proofreading.

References

[1] N. A. Anstey, *Seismic Prospecting Instruments*, Gebruder Borntraeger, Berlin, Germany, 2nd edition, 1981.
[2] J. T. O'Brien, W. P. Kamp, and G. M. Hoover, "Sign-bit amplitude recovery with applications to seismic data," *Geophysics*, vol. 47, no. 11, pp. 1527–1539, 1982.
[3] L. M. Houston and B. A. Richard, "The Helmholtz-Kirchoff 2.5D integral theorem for sign-bit data," *Journal of Geophysics and Engineering*, vol. 1, no. 1, pp. 84–87, 2004.
[4] R. A. Gabel and R. A. Roberts, *Signals and Linear Systems*, Wiley, New York, NY, USA, 3rd edition, 1987.
[5] W. G. Kelley and A. C. Peterson, *Difference Equations*, Academic Press, Boston, Mass, USA, 1991.
[6] D. M. Dubois, "Computing anticipatory systems with incursion and hyperincursion, computing anticipatory systems," in *Proceedings of the 1st International Conference on Computing Anticipatory Systems (CASYS '98)*, vol. 437 of *AIP Conference Proceedings*, pp. 3–29, The American Institute of Physics, 1998.
[7] G. Arfken, *Mathematical Methods for Physicists*, Academic Press, New York, NY, USA, 1966.