

JITTER COMPENSATION IN SAMPLING VIA POLYNOMIAL LEAST SQUARES ESTIMATION

Daniel S. Weller and Vivek K Goyal

Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science
Email: {dweller, vgoyal}@mit.edu

ABSTRACT

Sampling error due to jitter, or noise in the sample times, affects the precision of analog-to-digital converters in a significant, nonlinear fashion. In this paper, a polynomial least squares (PLS) estimator is derived for an observation model incorporating both independent jitter and additive noise, as an alternative to the linear least squares (LLS) estimator. After deriving this estimator, its implementation is discussed, and it is simulated using Matlab. In simulations, the PLS estimator is shown to improve the mean squared error performance by up to 30 percent versus the optimal linear estimator.

Index Terms— Sampling, jitter, estimation, polynomial least squares, analog-to-digital converters

1. INTRODUCTION

Samples generated by an analog-to-digital converter (ADC) are typically corrupted by errors in both the amplitudes of the samples and in the sample times. Additive noise has been studied thoroughly in many situations, and effective algorithms are well-known. However, jitter differs from additive noise in several respects: the effects are always signal-dependent and vary over time, and jitter nonlinearly affects the observations. While these characteristics make jitter more difficult to mitigate effectively, it can have just as deleterious an effect on the quality of the samples as additive noise: when quality is limited by jitter, doubling the standard deviation of the jitter σ_z reduces the effective number of bits (ENOB) by one [1]. Thus, using the speed-power-accuracy tradeoff in [2], doubling σ_z , without any mitigation, increases the power consumption required to achieve the original accuracy by a factor of four.

Jitter is studied extensively in the literature. The optimal linear interpolator is developed for signals with independent or correlated jitter in [3]; polynomial interpolators are investigated briefly, but mixed data terms are not considered. More recently, [4] presents two block-processing algorithms for reconstructing uniform samples from jittered observations, whose sample times are distributed on a dense (discrete) grid. The jitter problem appears similar to the phase-offset problem encountered in time-interleaved ADCs; see [5]. However, estimating the offsets between time-interleaved ADC is an easier problem than the problem addressed in this paper because the offsets remain fixed over a large number of samples.

Several iterative approaches to jitter mitigation, including a Gibbs sampler, are explored in [6]. The Gibbs sampler provides

This work was supported in part by the Office of Naval Research through a National Defense Science and Engineering Graduate (NDSEG) fellowship, NSF CAREER Grant CCF-0643836, and the Texas Instruments Leadership University Consortium Program.

substantial improvement in jitter tolerance—thus exhibiting the potential for jitter mitigation—but it does so at a high computational cost. The polynomial least squares (PLS) estimator, also known as a polynomial minimum mean square error (PMMSE) estimator, developed here attempts to improve upon the basic linear least squares (LLS) estimator, but with only a minor increase in on-line computational complexity as compared to LLS.

In Section 2, the observation model for the problem of jitter mitigation is provided. Then, the PLS estimator is derived in Section 3, and one possible implementation of the estimator is presented in Section 4. In Section 5, simulation results are summarized demonstrating the effectiveness of the new estimator. The paper closes with conclusions drawn from these results and a brief discussion of future directions for this work.

2. OBSERVATION MODEL

A signal in the span of a finite basis, parameterized by $\mathbf{x} = [x_0, \dots, x_{K-1}]^T$, is oversampled by a factor of M in the presence of both additive noise w_n and jitter z_n . The resulting $N = KM$ observations $\mathbf{y} = [y_0, \dots, y_{N-1}]^T$ are modeled by the equation

$$\mathbf{y} = \mathbf{H}(\mathbf{z})\mathbf{x} + \mathbf{w}, \quad (1)$$

where $\mathbf{H}(\mathbf{z})$ reflects the choice of finite frame. While many signal classes and parameterizations could be used, this paper focuses on \mathbf{x} being the Nyquist-rate samples of a periodic bandlimited signal and \mathbf{y} being the time-domain representation with oversampling factor M . With Nyquist rate normalized to 1 Hz without loss of generality,

$$[\mathbf{H}(\mathbf{z})]_{n,k} = \text{psinc}_K\left(\frac{n}{M} + z_n - k\right) \quad (2)$$

where

$$\text{psinc}_K(t) \triangleq \frac{\sin(\pi t)}{K \sin(\pi t/K)}. \quad (3)$$

Let the jitter z_n be zero-mean, white Gaussian noise with variance σ_z^2 , and let the additive noise w_n be zero-mean, white Gaussian noise with variance σ_w^2 . These variances are relative to the scale of the model: σ_z is relative to the critical sampling period, and σ_w is relative to the scale of the parameters \mathbf{x} . Both noise sources are independent of each other and independent of the input signal parameters \mathbf{x} . To complete the Bayesian setup, the prior on the parameters is chosen to minimize prior information, which could bias an algorithm's performance. To this end, the Uniform prior, which is the maximum entropy prior, is favorable because it weighs equally all possible inputs over a predetermined range. In addition, it is assumed that each input parameter is independent of the others for the same reason.

3. POLYNOMIAL LEAST SQUARES ESTIMATION

The PLS estimator is the member of the class of polynomial functions in \mathbf{y} up to a certain total degree P that minimizes the mean squared error (MSE) of the estimate $\hat{\mathbf{x}}$. Let $\mathbf{y}^{(0:P)}$ represent a vector of all monomials of \mathbf{y} up to degree P , including cross-terms. For instance, $\mathbf{y}^{(0:2)} = [1, y_0, \dots, y_{N-1}, y_0^2, \dots, y_0 y_{N-1}, \dots, y_{N-2} y_{N-1}, y_{N-1}^2]^T$. Then, the PLS estimator is defined to be

$$\hat{\mathbf{x}}_{\text{PLS}}(\mathbf{y}) = \mathbf{A}^{(0:P)} \mathbf{y}^{(0:P)}, \quad (4)$$

where

$$\mathbf{A}^{(0:P)} = \arg \min_{\mathbf{A}} \mathbb{E} \left[\|\mathbf{A} \mathbf{y}^{(0:P)} - \mathbf{x}\|_2^2 \right]. \quad (5)$$

Solving the optimization problem by taking the derivative with respect to \mathbf{A} yields the optimal coefficients

$$\mathbf{A}^{(0:P)} = \mathbb{E} \left[\mathbf{x} \mathbf{y}^{(0:P)T} \right] \left(\mathbb{E} \left[\mathbf{y}^{(0:P)} \mathbf{y}^{(0:P)T} \right] \right)^{-1}. \quad (6)$$

The PLS estimator is closely related to the P th-order Volterra filter, which is discussed extensively in [7], and the problem of finding the optimal coefficients is almost equivalent to finding the Volterra kernel coefficients $a(n_1, \dots, n_j)$ for the filter

$$x_k = \sum_{j=0}^P \sum_{n_1, \dots, n_j} a(n_1, \dots, n_j) y_{kM-n_1} \cdots y_{kM-n_j}. \quad (7)$$

The difference here is that the PLS estimator block-processes the entire set of observations to determine all the parameters, whereas the Volterra filter estimates the parameters one-by-one. However, one could easily adapt the method for computing the coefficients in $\mathbf{A}^{(0:P)}$ to finding the Volterra kernel coefficients.

For the observation model in (1), the odd moments of x_k and w_n are zero:

$$m_x^{(q)} \triangleq \mathbb{E} [x_k^q] = \begin{cases} 1/(q+1), & q \text{ even;} \\ 0, & q \text{ odd.} \end{cases} \quad (8)$$

$$m_w^{(q)} \triangleq \mathbb{E} [w_n^q] = \begin{cases} (1 \cdot 3 \cdots (q-1)) \sigma_w^q, & q \text{ even;} \\ 0, & q \text{ odd.} \end{cases} \quad (9)$$

Since $y_n = \mathbf{h}_n(z_n)^T \mathbf{x} + w_n$ is the weighted sum of parameters x_k and additive noise term w_n , the expectations $\mathbb{E} [y_{n_1}^{p_1} \cdots y_{n_L}^{p_L}]$ that have total degree $p_1 + \cdots + p_L$ odd equal zero. Similarly, those expectations $\mathbb{E} [x_k y_{n_1}^{p_1} \cdots y_{n_L}^{p_L}]$ with total degree $p_1 + \cdots + p_L$ even also equal zero. Therefore, by splitting $\mathbf{y}^{(0:P)}$ into the terms with odd and even total degree, it is easy to see that the optimal coefficients for the even total degree terms must be zero. In addition, the expectation $\mathbb{E} [\mathbf{y}^{(0:P)}]$ is zero for the terms with nonzero coefficients. Thus, $\mathbb{E} [\hat{\mathbf{x}}_{\text{PLS}}(\mathbf{y})] = \mathbf{0} = \mathbb{E} [\mathbf{x}]$, so the PLS estimator is unbiased. (Unbiasedness can also be easily shown using orthogonality, even in the general case, since the PLS optimization is a result of the Projection Theorem.)

4. IMPLEMENTATION

Finding the coefficients in (6) involves solving a system of linear equations, of which the known matrices are hard to compute directly. Here, expressions for the elements of these matrices are derived.

Consider the expectation $\mathbb{E} [y_{n_1}^{p_1} \cdots y_{n_L}^{p_L}]$, where $p_1 + \cdots + p_L$ is even:

$$\begin{aligned} \mathbb{E} \left[\prod_{\ell=1}^L \left(\sum_{k=0}^{K-1} h_{n_\ell}(z_{n_\ell})_k x_k + w_{n_\ell} \right)^{p_\ell} \right] = \\ \mathbb{E} \left[\prod_{\ell=1}^L \left(\sum_{q_{0,\ell} + \cdots + q_{K-1,\ell} + q_{w,\ell} = p_\ell} \binom{p_\ell}{q_{0,\ell} \cdots q_{K-1,\ell} q_{w,\ell}} \right. \right. \\ \left. \left. \cdot \left(\prod_{k=0}^{K-1} (h_{n_\ell}(z_{n_\ell})_k x_k)^{q_{k,\ell}} \right) w_{n_\ell}^{q_{w,\ell}} \right) \right]. \quad (10) \end{aligned}$$

Combining the summations into one large summation for all $\ell = 1, \dots, L$, (10) becomes

$$\begin{aligned} \mathbb{E} \left[\sum_{\substack{q_{0,\ell} + \cdots + q_{K-1,\ell} + q_{w,\ell} = p_\ell, \\ \forall \ell=1, \dots, L}} \prod_{\ell=1}^L \binom{p_\ell}{q_{0,\ell} \cdots q_{K-1,\ell} q_{w,\ell}} \right. \\ \left. \cdot \left(\prod_{k=0}^{K-1} (h_{n_\ell}(z_{n_\ell})_k x_k)^{q_{k,\ell}} \right) w_{n_\ell}^{q_{w,\ell}} \right]. \quad (11) \end{aligned}$$

Distributing the x_k out of the innermost product in (11) and collecting terms outside the product over ℓ yields

$$\begin{aligned} \sum_{\substack{q_{0,\ell} + \cdots + q_{K-1,\ell} + q_{w,\ell} = p_\ell, \\ \forall \ell=1, \dots, L}} \left(\prod_{\ell=1}^L \binom{p_\ell}{q_{0,\ell} \cdots q_{K-1,\ell} q_{w,\ell}} \right) \\ \cdot \mathbb{E} \left[\prod_{k=0}^{K-1} h_{n_\ell}(z_{n_\ell})_k^{q_{k,\ell}} m_w^{(q_{w,\ell})} \right] \left(\prod_{k=0}^{K-1} m_x^{(q_{x,k})} \right) \quad (12) \end{aligned}$$

where $q_{x,k} = \sum_{\ell=1}^L q_{k,\ell}$, and $m_x^{(q_{x,k})}$ and $m_w^{(q_{w,\ell})}$ are defined in (8) and (9). Using Gauss-Hermite quadrature [8],

$$\mathbb{E} \left[\prod_{k=0}^{K-1} h_{n_\ell}(z_{n_\ell})_k^{q_{k,\ell}} \right] = \sum_{i=1}^I w_i \prod_{k=0}^{K-1} h_{n_\ell}(z_i)_k^{q_{k,\ell}}. \quad (13)$$

The computational complexity of the computation in (13) is $O(KI)$. Thus, the computational complexity for one choice of $q_{0,1}, \dots, q_{K-1,1}, q_{w,1}, \dots, q_{0,L}, \dots, q_{K-1,L}, q_{w,L}$ is $O(KIL)$. By elementary combinatorics, for each $\ell = 1, \dots, L$, there are $\binom{p_\ell + K}{K}$ ways to choose $q_{0,\ell}, \dots, q_{K-1,\ell}$, and $q_{w,\ell}$ such that they sum to p_ℓ . An expectation of the form $\mathbb{E} [x_k y_{n_1}^{p_1} \cdots y_{n_L}^{p_L}]$ is computed in an identical manner to the above. Clearly, the number of computations required is prohibitive when P , and thus L and p_ℓ , becomes large.

A few tricks can be used to reduce the number of expectations that need to be evaluated. The most obvious is that the matrix $\mathbb{E} [\mathbf{y}^{(0:P)} \mathbf{y}^{(0:P)T}]$ is symmetric, so only half of the matrix needs to be computed. Since the most highly correlated cross-terms can be expected to be those adjacent to each other, the number of terms in $\mathbf{y}^{(0:P)}$ can be reduced to be linear in K and M without sacrificing much information. Another way to reduce the number of computations is to take advantage of the periodicity of the psinc function:

Theorem 1 Assume w_n and z_n are both L th-order stationary random processes. Then, y_n is an L th-order cyclostationary random process with period M , and for all indexes n_1, \dots, n_L , powers p_1, \dots, p_L , and offsets $n = Mk$,

$$\mathbb{E} [y_{n_1}^{p_1} \cdots y_{n_L}^{p_L}] = \mathbb{E} [y_{n_1+n}^{p_1} \cdots y_{n_L+n}^{p_L}]. \quad (14)$$

Proof. Express $p(y_{n_1}, \dots, y_{n_L})$ in terms of the observation model likelihood function and priors on \mathbf{x} and z_{n_1}, \dots, z_{n_L} :

$$\begin{aligned} p(y_{n_1}, \dots, y_{n_L}) &= \int \cdots \int p(y_{n_1}, \dots, y_{n_L} | \mathbf{x}, z_{n_1}, \dots, z_{n_L}) \\ &\quad p(\mathbf{x})p(z_{n_1}, \dots, z_{n_L}) d\mathbf{x} dz_{n_1} \cdots dz_{n_L}. \end{aligned} \quad (15)$$

Consider the observation model in (1). This system is time-invariant for shifts of multiples of M . Therefore, for shifts of $n = kM$,

$$\begin{aligned} p(y_{n_1}, \dots, y_{n_L}) &= \int \cdots \int p(y_{n_1+n}, \dots, y_{n_L+n} | \mathbf{x}, z_{n_1+n}, \dots, z_{n_L+n}) \\ &\quad p(\mathbf{x})p(z_{n_1+n}, \dots, z_{n_L+n}) d\mathbf{x} dz_{n_1+n} \cdots dz_{n_L+n}. \end{aligned} \quad (16)$$

As a result, y_n is L th-order stationary for shifts of multiples of M :

$$p(y_{n_1}, \dots, y_{n_L}) = p(y_{n_1+kM}, \dots, y_{n_L+kM}), \quad \forall k \in \mathbb{Z}. \quad (17)$$

Thus, y_n can be thought of as the interleaving of M L th-order stationary random processes; such a process is called L th-order cyclostationary with period M . The equivalence of expectation (14) holds as a direct result. \square

5. SIMULATION RESULTS

To measure the performance of the PLS estimator, the LLS estimator is used as a baseline:

$$\hat{\mathbf{x}}_{\text{LLS}}(\mathbf{y}) = \mathbb{E}[\mathbf{H}(\mathbf{z})]^T \left(\mathbb{E}[\mathbf{H}(\mathbf{z})^T \mathbf{H}(\mathbf{z})] + \frac{\sigma_w^2}{\sigma_x^2} \mathbf{I} \right)^{-1} \mathbf{y}. \quad (18)$$

Using Matlab, both algorithms are implemented and tested for various choices of oversampling factor M , jitter variance σ_z^2 , and additive noise variance σ_w^2 . The MSE of the LLS estimator has a closed form:

$$\begin{aligned} \mathbb{E} [\|\hat{\mathbf{x}}_{\text{LLS}}(\mathbf{y}) - \mathbf{x}\|^2] &= \sigma_x^2 \text{tr} \left(\mathbf{I} - \mathbb{E}[\mathbf{H}(\mathbf{z})^T] \right. \\ &\quad \left. \left(\mathbb{E}[\mathbf{H}(\mathbf{z})^T \mathbf{H}(\mathbf{z})] \right)^{-1} \mathbb{E}[\mathbf{H}(\mathbf{z})] \right). \end{aligned} \quad (19)$$

Since the PLS estimator can be thought of as an LLS estimator with an augmented data set, the MSE of the PLS estimator is similar:

$$\begin{aligned} \mathbb{E} [\|\hat{\mathbf{x}}_{\text{PLS}}(\mathbf{y}) - \mathbf{x}\|^2] &= \text{tr} \left(\sigma_x^2 \mathbf{I} - \mathbb{E}[\mathbf{x}\mathbf{y}^{(0:P)T}] \right. \\ &\quad \left. \left(\mathbb{E}[\mathbf{y}^{(0:P)} \mathbf{y}^{(0:P)T}] \right)^{-1} \mathbb{E}[\mathbf{y}^{(0:P)} \mathbf{x}^T] \right). \end{aligned} \quad (20)$$

In Figure 1, the MSE improvements of three flavors of the PLS estimator relative to the LLS estimator are compared for different levels of oversampling.

As expected, augmenting the data set improves the estimator, so all of the PLS estimators depicted are better than the optimal linear

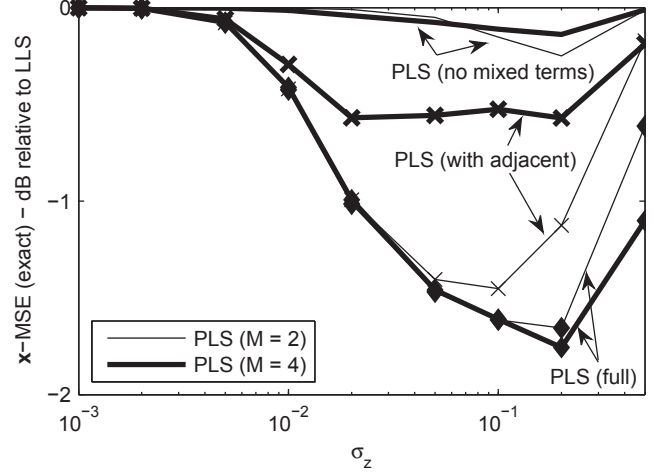


Fig. 1. Three 3rd-order PLS estimators (full, adjacent mixed terms only, and no mixed terms) are compared against the LLS estimator for $K = 3$, $\sigma_w = 0.01$ and different values of M and σ_z .

estimator. However, as the graph shows, the greatest improvement is achieved when the full set of mixed terms are incorporated into the data set. In addition, the graph suggests that when M increases, the full PLS still gives considerable MSE improvement, but the PLS estimator using only adjacent terms, and the PLS estimator using only adjacent terms, give up most of their performance gains. While for a standard bandlimited signal, the samples associated with one parameter are orthogonal to the samples for the other parameters, the fact that jitter offsets the sample times away from the zero-crossings of the psinc basis functions increases the correlation among the samples. This effect is more pronounced for the psinc function because unlike the regular sinc interpolator, the psinc does not decay.

The contribution of each value of the full augmented data set $\mathbf{y}^{(0:P)}$ can be compared by measuring the 2-norm of the associated column vector of coefficients in $\mathbf{A}^{(0:P)}$. The members of the augmented data set are divided into four categories: linear terms y_n , cubic terms y_n^3 , adjacent mixed terms $y_\ell y_m y_n$, and other mixed terms $y_\ell y_m y_n$, and the 2-norms of all the associated column vectors for each category are added together. For several choices of σ_z and M , Table 1 describes the 2-norms for each category as a fraction of the overall total. Note that in the case when $\sigma_z \gg \sigma_w$, the PLS terms account for more than half of the total. Also, the contribution of adjacent terms decreases as M increases; this explains the relatively poorer performance of adjacent terms PLS for $M = 4$ in Figure 1. The relative contributions of the mixed terms are shown in Figure 2 for the case of oversampling by $M = 4$ and $\sigma_z = 0.1$ much greater than $\sigma_w = 0.01$. As expected, the majority of the contribution is from those terms with y_ℓ , y_m , and y_n relatively close together.

How does the PLS estimator compare against the hybrid Gibbs and slice sampler proposed in [6] to approximate the Bayes least squares (BLS) estimator? In general, the BLS estimator is optimal in minimizing the MSE; however, the method for computing the BLS is an approximate one. Hampered by the random nature of the Gibbs sampler, a large number of iterations may be required to ensure convergence; the resulting run-time of the algorithm is highly undesirable for real time applications. In comparison, the PLS estimator has a relatively low online computational complexity; the majority of the work is done off-line, before any data is obtained, to generate the coefficients of the estimator. In Figure 3, the PLS estimator with

Table 1. Fraction (%) of the contribution to the estimate $\hat{\mathbf{x}}_{\text{PLS}}$ by each category of observations in the augmented data $\mathbf{y}^{(0:P)}$, as measured by the 2-norm of the column vector of coefficients in $\mathbf{A}^{(0:P)}$. As in Figure 1, these values are for the $K = 3$ parameter case, with low additive noise $\sigma_w = 0.01$.

σ_z	M	Linear	PLS (no mixed terms)	PLS (adjacent)	PLS (other)
0.1	2	43.73	10.95	20.22	25.10
0.1	4	37.40	21.14	4.23	37.23
0.02	2	55.06	9.16	11.16	24.62
0.02	4	52.95	15.87	2.91	28.27
0.005	2	91.79	0.42	1.40	6.40
0.005	4	94.42	0.03	0.81	4.73

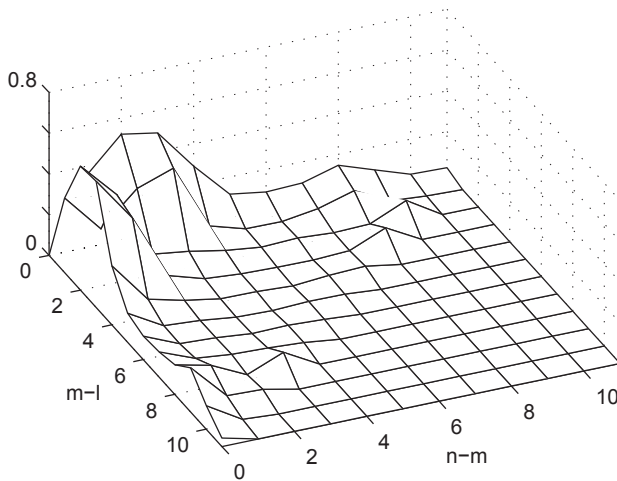


Fig. 2. The relative contributions of the mixed terms $y_\ell y_m y_n$ of the full 3rd-order PLS estimator are shown according to the displacements $m - l$ and $n - m$, for $K = 3$, $M = 4$, $\sigma_z = 0.1$, and $\sigma_w = 0.01$. The contribution for $n = m = \ell$ is omitted for clarity; this contribution is equal to 0.44.

adjacent mixed terms is compared against the BLS estimator, with mixed results.

For low levels of oversampling, the BLS estimator is not very effective for high σ_z , whereas the PLS estimator provides consistently reduced MSE over the LLS estimator. However, as the oversampling level increases, the BLS estimator improves, while the PLS estimator does only marginally better than the LLS estimator. Figure 1 suggests that the full third order PLS estimator may perform better with higher oversampling, but the number of terms in $\mathbb{E}[\mathbf{y}^{(0:P)} \mathbf{y}^{(0:P)T}]$ is on the order of $O(M^{2P})$, so for even $M = 16$, the matrix does not fit in 1 GB of memory.

6. CONCLUSION

Simulation results demonstrate the usefulness of the PLS estimator in reducing the MSE in the presence of jitter for low oversampling. It remains to be seen if higher orders would yield substantial improvements for higher levels of oversampling. Although the on-line computational cost is comparable to that of the LLS estimator, computing the coefficients off-line is extremely computationally intensive

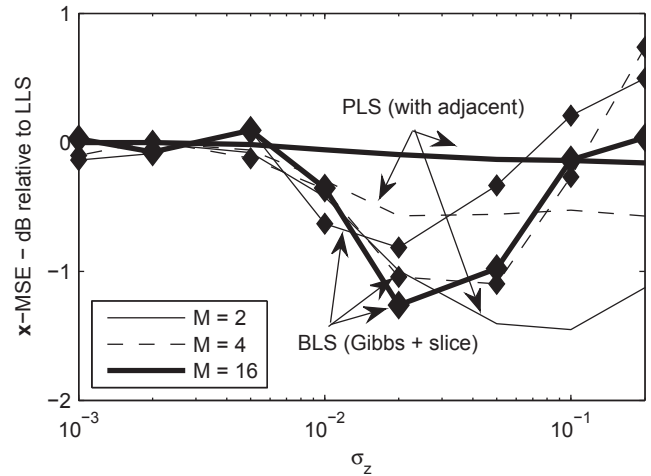


Fig. 3. The 3rd-order PLS estimator (adjacent mixed terms only) is compared against the BLS estimator (Gibbs + slice sampler), using the LLS estimator as a baseline, for $K = 3$, $\sigma_w = 0.01$, and $M = 2, 4, 16$. The MSE for the LLS and PLS estimators are exact, while the BLS estimator's MSE is approximated using Monte Carlo simulation with 1000 trials.

for low orders, and essentially intractable for higher orders. Further work remains in simplifying the coefficient computation process for higher orders. Perhaps, orthogonalization techniques used to simplify computing the coefficients of Volterra filters may be applied similarly here. However, discounting a faster method for computing the PLS estimator coefficients, the design procedure for the BLS estimators described in [6] is more easily scalable to achieve greater accuracy.

7. REFERENCES

- [1] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.
- [2] K. Uyttenhove and M. S. J. Steyaert, "Speed-power-accuracy tradeoff in high-speed CMOS ADCs," *IEEE Trans. Circuits Syst. II*, vol. 49, no. 4, pp. 280–287, Apr. 2002.
- [3] A. V. Balakrishnan, "On the problem of time jitter in sampling," *IRE Trans. Inform. Th.*, vol. 8, no. 3, pp. 226–236, Apr. 1962.
- [4] T. E. Tuncer, "Block-based methods for the reconstruction of finite-length signals from nonuniform samples," *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 530–541, Feb. 2007.
- [5] V. Divi and G. Wornell, "Signal recovery in time-interleaved analog-to-digital converters," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, May 2004, vol. 2, pp. 593–596.
- [6] D. S. Weller and V. K. Goyal, "Nonlinear digital post-processing to mitigate jitter in sampling," arXiv:0809.4244v1 [stat.AP], Sept. 2008.
- [7] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*, Krieger, Malabar, Florida, 2006.
- [8] P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration*, Academic Press, Orlando, 1984.