14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

CONSISTENT SIGNAL PARAMETER ESTIMATION WITH 1-BIT DITHERED SAMPLING

Onkar Dabeer, Aditya Karnik

School of Technology and Computer Science
Tata Institute of Fundamental Research
Colaba, Mumbai 400 005 India
onkar@tcs.tifr.res.in

Dept. of Electrical and Computer Engg.
University of Waterloo
Waterloo, N2L 3G1, Canada
karnik@ece.waterloo.ca

ABSTRACT

We consider the problem of estimating a parameter θ of a signal $s(x;\theta)$ corrupted by noise when only 1-bit precision samples are allowed. We propose and analyze a new estimator based on dithered 1-bit samples. Our estimate is consistent and satisfies an asymptotic CLT for a wide class of dither distributions. In particular, uniformly distributed dither leads to only a logarithmic rate loss compared to the case of full precision samples.

1. INTRODUCTION

This paper is concerned with the problem of estimating a parameter of a signal corrupted by noise when only its 1-bit precision samples are available. This problem is particularly motivated by the possibilities that have opened up with the development of miniature sensing devices which can compute and communicate autonomously. Networks of such "smart" sensors could be used for detecting and estimating characteristics of spatio-temporal processes in several situations ([4]). When physical models for such processes are known, these problems can often be cast as those of estimation of parameters of a signal; for example, estimating the velocity of a wave propagating through some medium. The constraint posed by this exciting sensing technology, however, is that being miniature and low cost, a typical sensor will be a low precision device; hence, any estimation scheme in this context can employ only low precision, noisy measurements available from each sensor.

Formally, the problem is to estimate a deterministic but unknown parameter $\theta_0 \in \mathbb{R}$ of a real-valued signal $s(x;\theta_0)$ corrupted with additive noise. In this paper, we consider only one dimensional space and no time variation. We consider bounded observation space: $x \in [0,1]$. Results similar to ours can be obtained for the general case with minor notational changes. To make the problem well-posed, we assume throughout that this problem has a unique solution when $s(x;\theta_0)$ is observed without any noise, that is, $s(x;\theta) \neq s(x;\theta')$ whenever $\theta \neq \theta'$.

When full precision samples with i.i.d. Gaussian noise are available at N locations, then the maximum-likelihood (ML) estimate of θ_0 is well analyzed. Let the observations be denoted by

$$Z(n) = s\left(\frac{n}{N};\theta_0\right) + \sigma_t W_t(n)$$

where $\{W_t(n), n = 1,\ldots,N\}$ is the i.i.d. thermal/ambient noise in the measurements with common distribution $\mathcal{N}(0,1)$. The maximum likelihood (ML) estimator of θ_0 seeks a least-squares fit to the observations and the ML estimate is a solution of the log-likelihood equation (see [7, pp. 187])

$$\sum_{n=1}^{N} s'\left(\frac{n}{N};\theta\right)\left[Z(n) - s\left(\frac{n}{N};\theta\right)\right] = 0. \qquad (1)$$

Under mild regularity conditions, the ML estimate is consistent (in probability as $N \to \infty$), satisfies a central limit theorem (CLT), and the variance in the CLT attains the Cramer-Rao lower bound.

In contrast when only 1-bit precision samples of observations ,i.e., $\mathrm{sign}(Z(n))$, are available, the ML equation is complicated and little is known about the performance of the ML estimate for the general case. For example, when $s(x;\theta) = \theta$, then in the full precision case the ML estimate is the sample mean, but no closed form solution is known for the ML estimate in the 1-bit sampling case. In this paper, we employ *dithered* 1-bit sampling and propose an estimate that requires solving an equation of the form (1). In dithered sampling, a random noise is added to $Z(n)$ before quantizing it. Thus, our method leverages the idea of obtaining a low precision measurement at each of the possibly large number of sensors and combining them to form an estimate of the parameter[1]. Moreover, our estimator requires no more computation than the ML estimate in the full precision case, and for the above example of $s(x;\theta) = \theta$, our estimate just involves the averaging of dithered samples. Our estimate is consistent and asymptotically normal for a wide class of dither distributions. In particular, uniform dithering leads to only a logarithmic rate loss compared to the full precision case. Due to space constraint, we do not present the full mathematical derivations in this paper (these can be found in [3]). But we discuss the intuition behind our results with examples and simulations.

1.1 Related Prior Work

While a theory of estimation using full precision sampling is now well-established (see for example, [7, Section IV.E.2]), the case of low precision sampling seems to have been considered only for particular cases. For example, the problem of frequency estimation using 1-bit ADC has been studied in [5], [10]. In [9], the problem of choosing the quantization threshold for the case of signal amplitude estimation is investigated.

To the best of our knowledge, the problem of estimating θ_0 for a general signal $s(x;\theta_0)$ with only 1-bit ADC has

[1] We do not discuss strategies for distributed implementation of this scheme in this paper.

not been addressed before in the literature. However, a consistent estimator using only 1-bit ADC can be derived from reference [6], which addresses the problem of estimating the signal $s(x;\theta_0)$ from the observations $\{\text{sign}(Z(n))\}$. Our estimator is an improvement over this estimator: a) it requires less computation, and b) it has a better rate for the CLT (see Section 3.1 for a discussion).

A key element in our results is dithering. Dithering has been used previously by several researchers in different contexts; for example, [6] uses it to estimate a smooth signal using only 1-bit ADC, [2] employs a deterministic dither signal and oversampling to recover a band-limited function from finite precision samples. However, in the context of signal parameter estimation, dithering does not seem to have been exploited before.

2. OBSERVATION MODEL AND ASSUMPTIONS

The observations $\{Y(n),\ n = 1,...,N\}$ are given by

$$Y(n) = \text{sign}\left(s\left(X_n;\theta_0\right) + \sigma_t W_t(n) + a(N)W_d(n)\right) \qquad (2)$$

The random variables $\{W_d(n)\}$ constitute the dither signal, while $\{X_n\}$ are the sampling locations. We estimate θ_0 by solving $T_N(\theta) = 0$, where,

$$T_N(\theta) = \frac{1}{N}\sum_{n=1}^{N} s'\left(X_n;\theta\right)\left[c_d(N)Y(n) - s\left(X_n;\theta\right).\right] \qquad (3)$$

The value of $c_d(N)$ depends on the distribution of the dither and is specified below. Recall that N denotes the number of space samples or simply the number of sensors. Below we collect together most of the assumptions we need for analyzing the estimator and we also discuss their implications.

Noise: The usual assumption about the noise is:

GN) $\{W_t(n), n \geq 1\}$ are i.i.d. $\mathcal{N}(0,1)$.

We assume that the observation noise is dominated by the thermal noise in the sensor circuitry which justifies our i.i.d. noise assumption. (If the noise at different sensors is correlated, then the situation is best modeled as estimation of parameter of a stochastic process. This is a tough problem that we do not address here.)

Signal: The signal is assumed to satisfy the following regularity constraints.

A1) The signal $s(x;\theta)$ is thrice differentiable in θ and once in x.

A2) The signal $|s(x;\theta)| \leq b < \infty$ for all $x \in [0,1]$ and $\theta \in \Theta$. Similarly, its three derivatives w.r.t. θ and the one derivative w.r.t. x are bounded by a constant that does not depend on x and θ.

A3) The function

$$J(\theta;\theta_0) := \int_0^1 s'(x;\theta)[s(x;\theta_0) - s(x;\theta)]dx$$

has a unique zero-crossing at $\theta = \theta_0$.

The assumption A3) ensures that the parameter estimation problem is well-posed. It basically says that

$$\int_0^1 [s(x;\theta_0) - s(x;\theta)]^2 dx$$

has a unique minimum at θ_0, that is, if there is no noise and we collect full precision data over the entire space $[0,1]$, then we can estimate the parameter uniquely. For example, when $s(x;\theta) = \theta p(x)$, then $J(\theta;\theta_0) = (\theta_0 - \theta)\int_0^1 p^2(x)dx$ and A3) is seen to be satisfied.

Sampling: Our results can be established for deterministic as well as random spatial sampling. Here we report our results only under:

DS) Uniform deterministic sampling: $X_n = n/N$, $1 \leq n \leq N$.

Other sampling designs are also of importance and can also be accommodated. If a non-uniform deterministic sampling is given by $X_n = B(n/N)$, where $B : [0,1] \to [0,1]$, then our results under DS) can be applied by redefining the signal to be $s(B(x);\theta)$.

Dithering: For the dither signal, we present results under the following two assumptions.

UD) The dither signal $\{W_d(n), 1 \leq n \leq N\}$ is i.i.d. with uniform distribution on $[-1,1]$ and it is independent of the thermal noise $\{W_t(n), 1 \leq n \leq N\}$. The dither magnitude is taken to be

$$a(N) = \beta(\log(N))^{(1+\eta)/2} \text{ for some } \beta > 0, \eta > 0$$

The constant $c_d(N)$ in (3) is taken to be $a(N)$.

NUD) The dither signal $\{W_d(n), 1 \leq n \leq N\}$ is i.i.d. with distribution $F(t)$ and it is independent of the thermal noise $\{W_t(n), 1 \leq n \leq N\}$. The distribution $F(t)$ has the following expansion for some $\varepsilon > 0$, $|t| < \varepsilon$ and some $q \in \{2,3,...\}$

$$F(t) = \frac{1}{2} + \frac{b_1}{2}t + \frac{b_q}{2}t^q + R(t)$$

where $b_1 \neq 0$, $b_q \neq 0$, and $|R(t)| \leq \text{constant} \cdot |t|^{q+1}$ for $|t| < \varepsilon$. The dither magnitude $a(N) \to \infty$ and $a^2(N)/N \to 0$ as $N \to \infty$. The constant $c_d(N)$ in (3) is taken to be $a(N)/b_1$.

The assumption UD) is of special importance because it leads to best rates in the CLT for the parameter estimate. In practice, uniformly distributed dithering can be implemented by varying the threshold of the quantizer by a sawtooth wave. As long as the phases of the sawtooth wave at the different sampling locations are independent and uniformly distributed, the i.i.d. random dither assumption is justified.

3. MAIN RESULTS

3.1 Consistency and CLT for the Estimate Under Uniform Dithering

Our first main result is the following.

Theorem 1 *Suppose the observations are as in* (2) *and the conditions GN), A1)-A3), DS), UD) are satisfied. Then as $N \to \infty$, there exists a sequence of estimates $\{\hat{\theta}_N\}$ such that*

1. $P(T_N(\hat{\theta}_N) = 0) \to 1$;
2. *the estimate $\hat{\theta}_N \to \theta_0$ almost surely;*
3. *the following CLT holds*

$$\sqrt{\frac{N}{(\log N)^{(1+\eta)}}}\left(\hat{\theta}_N - \theta_0\right) \Longrightarrow \mathcal{N}\left(0, \frac{\beta^2}{\int_0^1 (s'(x;\theta))^2 dx}\right)$$

where $\Longrightarrow$ denotes convergence in distribution.

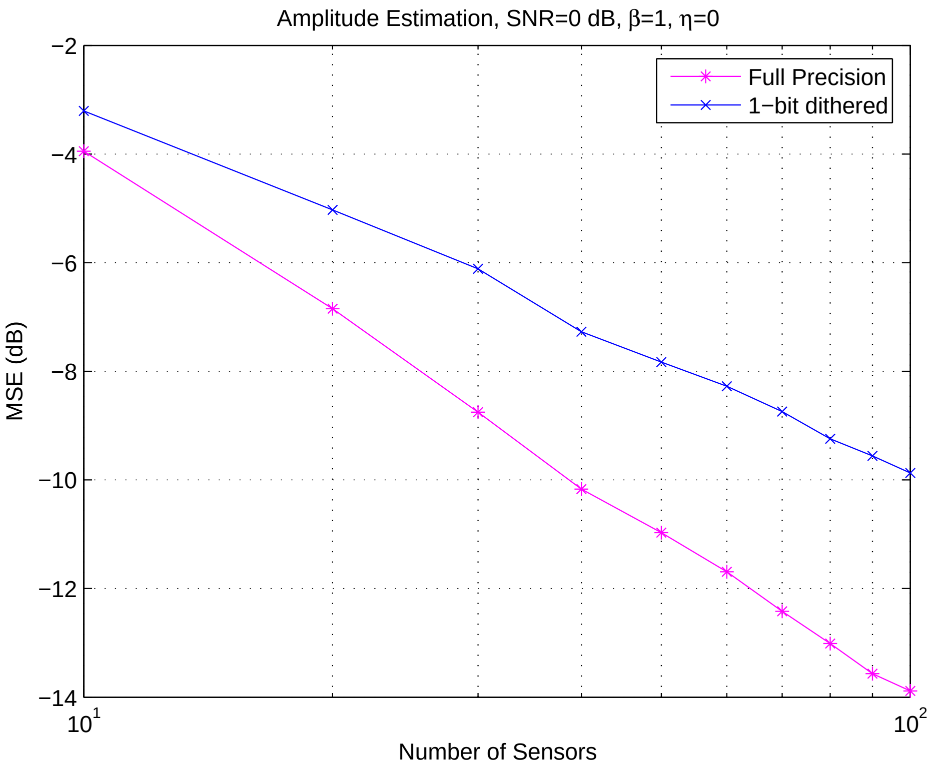

Figure 1: 1-bit quantization with uniform dithering leads to only logarithmic rate loss.

The proof uses elementary facts from probability theory to first establish the almost sure convergence of $T_N(\theta)$ to $J(\theta;\theta_0)$. The consistency and CLT is then established using techniques similar to those in the analysis of ML estimators for i.i.d. observations (see [7, Chapter IV]). We do not present the proof here due to space constraints; it can be found in the longer version [3]. However, we discuss the intuition behind this result and present simulation results below.

The main conclusion of the above theorem is that our estimator only leads to logarithmic loss with respect to the best possible rate of $1/N$. In other words, there is little loss in performance with respect to full precision samples. We are not aware of any work where an estimator of θ_0 under a general signal model based on 1-bit precision samples is analyzed. Hence, we compare our result with an estimator that can be derived from the results in [6]. In [6], an estimator of the signal $s(x;\theta_0)$ based on $\{Y(n)\}$ with $a(N)=0$ are given. The estimate of the signal is obtained by local linear filtering of the observations followed by a memoryless non-linear transformation. An estimate of θ_0 may be obtained from the signal estimate using a least-squares fit. Note that this is the same as solving (3) but using the signal estimate in place of $\{c_d Y(n)\}$. It can be shown (using a proof similar to that of Theorem 1) that such an estimate is consistent (in probability) and a CLT holds. The variance in the CLT decays like $1/N^{2/3}$, which is a significant loss compared with the rate $1/N$ for the full precision case. In comparison, we see that our estimator requires less computation and by choosing uniformly distributed dither, only a logarithmic rate loss is incurred.

The reason for the success of our estimator is not immediately clear. To obtain more insight, consider the simple case when the observation noise is absent ($\sigma_t=0$), $s(x;\theta)=\theta$, and $\theta\in[-0.5,0.5]$. In the absence of dithering, 1-bit quantization gives only one observation value $\operatorname{sign}(\theta_0)$ and we can do no better than estimating the sign of θ_0. However, with dithering (choosing $a(N)=1$) the estimate is given by

$$\hat{\theta}_N=\frac{1}{N}\sum_{n=1}^{N}Y(n)=\frac{1}{N}\sum_{n=1}^{N}\operatorname{sign}(\theta_0+W_d(n)).$$

Since $\{W_d(n)\}$ are i.i.d. uniform on $[-1,1]$, by the law of large numbers we get that $\hat{\theta}_N\to\theta_0$ almost surely and a CLT also holds. In other words, due to dithering, we get a look at θ_0 from a family of 1-bit quantizers and this 'diversity' helps in obtaining consistency. In the general case, the diversity provided by dithering causes $T_N(\theta)$ to converge to $J(\theta;\theta_0)$ (as is the case for full precision samples). But from assumption A3), we know that $J(\theta;\theta_0)=0$ has a unique zero crossing at $\theta=\theta_0$. This leads to existence and consistency of the estimate. The increasing dither magnitude $a(N)$ is required to remove bias introduced in the estimate by the thermal noise. We do not present a proof of the main result due to space constraints, but we illustrate it in Figure 1 with the help of a simulation. In Figure 1, we plot the mean-square estimation error as a function of N for the case when $s(x;\theta)=\theta$ and $\sigma_t^2=1$. The consistency as well as the logarithmic rate loss compared to full precision case are evident.

We have not been able to obtain the best possible rate in the CLT for uniform dither in the general case. However, for a specific example, we identify the rate below.

Proposition 1 *Consider the problem of estimating* $\theta_0 \in (0,1]$ *using samples* $\{Y(n),\ 1 \le n \le N\}$ *where,*

$$Y(n) = sign\left(\theta_0 + a(N)W_d(n) + \sigma_t W_t(n)\right)$$

$\{W_d(n)\}$ *and* $\{W_t(n)\}$ *are as in Theorem 1 above. We note that in this case* $T_N(\theta) = 0$ *has a unique solution for every N given by,*

$$\hat{\theta}_N = \frac{a(N)}{N}\sum_{n=1}^{N} Y(n).$$

If $a(N) \to \infty$ *and* $a^2(N)/N \to 0$*, then* $\hat{\theta}_N \to \theta_0$ *in m.s.s. The best possible rate for the decay of the mean-square error is obtained by choosing* $a(N)$ *such that*

$$\lim_{N\to\infty} \frac{N}{a^2(N)} \exp\left(-\frac{(a(N)-\theta_0)^2}{\sigma_t^2}\right) = 1. \qquad (4)$$

and in this case

$$\frac{\sqrt{N}}{a(N)}\left(\hat{\theta}_N - \theta_0\right) \implies \mathcal{N}(\mu, \sigma^2)$$

for constants $\mu \neq 0$ *and* σ^2.

We omit the proof here. We, however, note that for a choice of $a(N)$ as in Theorem 1, the limit in (4) is zero. Thus the rate in Proposition 1 is strictly better than that in Theorem 1.

3.2 Optimality of Uniform Dithering

Now we address the question of whether a different dither distribution can improve the rate of convergence even further. Theorem 2 shows that we cannot do better than uniform dithering for a class of distributions.

Theorem 2 *Suppose the conditions GN), A1)-A3), DS) and NUD) are true. Then there exists* $\{\hat{\theta}_N\}$ *such that as* $N \to \infty$

1. $P(T_N(\hat{\theta}_N) = 0) \to 1$;
2. *the estimate* $\hat{\theta}_N \to \theta_0$ *in probability;*
3. *the following CLT holds*
   - *If* $a(N) = \beta N^{1/(2q)}$*, then,*

     $$\frac{\sqrt{N}}{a(N)}\left(\hat{\theta}_N - \theta_0\right) \implies \mathcal{N}(\mu, \sigma^2)$$

     where μ *and* σ^2 *are specified in [3].*
   - *If* $a(N)N^{-1/(2q)} \to 0$*, then*

     $$(a(N))^{q-1}\left(\hat{\theta}_N - \theta_0\right) \implies \delta_\mu$$

     which is the unit mass at μ.
   - *If* $a(N)N^{-1/(2q)} \to \infty$*, then*

     $$\frac{\sqrt{N}}{a(N)}\left(\hat{\theta}_N - \theta_0\right) \implies \mathcal{N}(0, \sigma^2).$$

The main idea in the proof is similar to that in the proof of Theorem 1. We do not give the proof here.

We see that the best rate in the CLT is obtained in the first case. In this case the variance in the CLT behaves like $1/N^{(q-1)/q}$ which is strictly worse than the rate in Theorem 1. In particular, if the Gaussian dither is chosen, then $q = 3$ and we get the rate is $N^{-2/3}$. The higher the value of q, the better is the rate. Note that $q \to \infty$ implies that the dither distribution is getting closer to the uniform distribution. However, the case of uniform distribution is not covered by Theorem 2.

The existence of an optimal rate for the variance in part 3) of Theorem 2 is a consequence of a basic principle: increasing $a(N)$ increases variance but decreases bias. By increasing $a(N)$, we are expanding the family of 1-bit quantizers being used for observations and this diversity helps in reducing the bias. However, dithering also makes the observations more noisy and the variance of the estimate increases as $a(N)$ increases. Hence there is an optimum rate for $a(N)$, where the gains of diversity due to dithering, and the ill effects of the noise due to dithering are balanced.

4. CONCLUSIONS

We proposed a parameter estimate for a general signal model based on dithered 1-bit samples. This estimator is motivated by the need to design low cost low precision sensors and possibility of combining measurements from a multitude of such sensors. The main idea is that the full precision samples can be replaced by suitably scaled 1-bit dithered samples and hence the estimate needs no more computation than the ML estimate in the full precision case. The estimate is consistent for a wide class of dither distributions. But the uniform distribution leads to the best rate in the CLT amongst a broad class of dither distributions. Several other properties of the estimate are also established in a longer version of this paper ([3]). In particular, due to the hard-limiting operation involved in 1-bit sampling, the estimate is also robust and remains consistent even for noise with infinite variance. We have also analyzed the effect of random sampling, inaccurate knowledge of sampling locations, and unreliable communication of observations in [3].

Acknowledgement

We are thankful to the reviewers for their comments and for suggesting reference [1].

REFERENCES

[1] I. Bilinskis and A. Mikelsons, “Randomized Signal Processing,” Prentice Hall International Series, London, 1992.

[2] Z. Cvetkovic and I. Daubechies, “Single-bit Oversampled A/D Conversion with Exponential Accuracy in the Bit-Rate,” in *Proceedings DCC*, Utah, USA, March 2000, pp. 343–352.

[3] Onkar Dabeer and Aditya Karnik, “Signal paramter estimation with 1-bit dithered quantization,” submitted for journal publication, 2005.

[4] IEEE Journal on Selected Areas in Communications, April 2005, Special Issue on Self-Organizing Distributed Collaborative Sensor Networks.

[5] A. Host-Madsen and P. Handel, "Effects of Sampling and Quantization on Single-Tone Frequency Estimation," *IEEE Transactions on Signal Processing*, vol. 48, no. 3, pp. 650–662, March 2000.

[6] E. Masry, "The Reconstruction of Analog Signals from the Sign of Their Noisy Samples," *IEEE Transactions on Information Theory*, vol. 27, no. 6, pp. 735–745, November 1981.

[7] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer, New York, second edition, 1994.

[8] J. G. Proakis, *Digital Communications*, McGraw Hill, New York, third edition, 1995.

[9] D. Rousseau, G. V. Anand, and F. Chapeau-Blondeau, "Nonlinear Estimation from Quantized Signals: Quantizer Optimization and Stochastic Resonance," in *Third International Symposium on Physics in Signal and Image Processing*, Grenoble, France, January 2003, pp. 89–92.

[10] T. Andersson, M. Skoglund, and P. Handel, "Frequency Estimation by 1-bit Quantization and Table Look-Up Processing," in *Proceedings of European Signal Processing Conference*, September 2000.

[11] H. L. Van Trees, *Detection, Estimation, and Modulation Theory - Part I*, Wiley, New York, 1968.