Signal Reconstruction from Noisy Random Projections

Jarvis Haupt and Robert Nowak
Department of Electrical and Computer Engineering
University of Wisconsin-Madison

March, 2005; Revised May, 2006

Abstract

Recent results show that a relatively small number of random projections of a signal can contain most of its salient information. It follows that if a signal is compressible in some orthonormal basis, then a very accurate reconstruction can be obtained from random projections. This "compressive sampling" approach is extended here to show that signals can be accurately recovered from random projections contaminated with noise. A practical iterative algorithm for signal reconstruction is proposed, and potential applications to coding, A/D conversion, and remote wireless sensing are discussed.

Index Terms

complexity regularization, data compression, denoising, Rademacher chaos, random projections, sampling, wireless sensor networks

I. Introduction

Recent theory informs us that, with high probability, a relatively small number of random projections of a signal can contain most of its relevant information. For example, the groundbreaking work in [1] has shown that k random Fourier projections contain enough information to reconstruct piecewise smooth signals at a distortion level nearly equivalent to that attainable from k optimally selected observations. Similar results hold for random Gaussian and Rademacher projections (i.e., projections consisting of independent and identically distributed Gaussian or Rademacher random variables) [2], [3]. The results presented in these works can be roughly summarized as follows. Assume that a signal $f \in \mathbb{R}^n$ is "compressible" in some orthonormal basis in the following sense. Let $f^{(m)}$ denote the best m-term approximation of f in terms of this basis and suppose that the average squared error obeys

$$\frac{\|f - f^{(m)}\|^2}{n} = \frac{1}{n}\sum_{i=1}^{n}(f_i - f_i^{(m)})^2 \leq C_A m^{-2\alpha}$$

for some $\alpha \geq 0$ and some constant $C_A > 0$. The parameter α governs the degree to which f is compressible with respect to the basis. In a noiseless setting, it can be shown that an approximation of such a signal can be recovered from k random projections with an average squared error that is upper bounded by a constant times $(k/\log n)^{-2\alpha}$, nearly as good as the best k-term approximation error. For this reason, these procedures are often referred to as "compressive sampling" since the number of samples required is directly related to the sparsity or compressibility of the signal, rather than its temporal/spatial extent or bandwidth.

This paper takes the investigation of compressive sampling a step further by considering the performance of sampling via random projections in noisy conditions. We show that if the projections are contaminated with zero-mean Gaussian noise, then compressible signals can be reconstructed with an expected average squared error that is upper bounded by a constant times $(k/\log n)^{\frac{-2\alpha}{2\alpha+1}}$. For truly sparse signals (with only a small number of non-zero terms) a stronger result is obtained; the expected average squared reconstruction error is upper bounded by a constant times $(k/\log n)^{-1}$. These bounds demonstrate a remarkable capability of compressive sampling – accurate reconstructions can be obtained even when the signal dimension n greatly exceeds the number of samples k *and* the samples themselves are contaminated with significant levels of noise.

This research was supported in part by the NSF under grants CCR-0310889 and CCR-0325571, and by the Office of Naval Research, grant N00014-00-1-0966.

This effect is highlighted by the following "needle in a haystack" problem. Suppose the signal f^* is a vector of length n with one nonzero entry of amplitude $\sqrt{n}$. If we sample the vector at k random locations (akin to conventional sampling schemes), then the probability of completely missing the non-zero entry is $(1 - 1/n)^k$, which is very close to 1 when k is significantly smaller than n. This implies that the expected average squared error may be almost 1, or larger if noise is present. On the other hand, by sampling with randomized projections our results guarantee that the expected average squared error will be no larger than a constant times $(k/\log n)^{-1}$, which can be close to 0 even when $k \ll n$, provided $k > \log n$.

A closely related problem is the reconstruction of signals with sparse Fourier spectra from a relatively small number of non-uniform time samples (e.g., random samples in time) [4]–[7]. Most of this work concerns noiseless situations, but [5] addresses the problem of reconstruction from noise-corrupted samples. Another area of work related to our results concerns the reconstruction of signals with finite degrees of freedom using a small number of non-traditional samples [8], [9]. A special instance of this setup is the case of signals that are sparse in time (the dual of the spectrally sparse case). Reconstruction from noise-corrupted samples is the focus of [9]. In a sense, the sampling and reconstruction problems addressed in the papers above are special cases of the class of problems considered here, where we allow signals that are sparse in some arbitrary domain. Again, this more universal perspective is precisely the focus of [2], [3], which consider signal reconstruction from noiseless random projections. An interesting line of similar work concerns the related problem of signal reconstruction from random projections corrupted by an unknown but bounded perturbation [10], [11]. In this paper we consider unbounded, Gaussian noise contamination in the sampling process. Finally, while this paper was under review a related investigation was reported in [12] pertaining to the statistical estimation of sparse signals from underdetermined and noisy observations. That work proposes a linear program to obtain an estimator with quantitative bounds for sparse signal reconstruction similar to ours, but is based on a uniform uncertainty principle rather than randomized designs as here.

The paper is organized as follows. In Section II we state the basic problem and main theoretical results of the paper. In Section III we derive bounds on the accuracy of signal reconstructions from noisy random projections. In Section IV we specialize the bounds to cases in which the underlying signal is compressible in terms of a certain orthonormal basis. In Section V we propose a simple iterative algorithm for signal reconstruction. In Section VI we discuss applications to encoding, A/D conversion, and wireless sensing, and we make concluding remarks in Section VII. Detailed derivations are relegated to the Appendix.

II. Main Results

Consider a vector $f^* = [f_1^* \; f_2^* \ldots f_n^*]^T \in \mathbb{R}^n$ and assume that $\sum_{i=1}^n (f_i^*)^2 \equiv \|f^*\|^2 \leq nB^2$ for a known constant $B > 0$. The assumption simply implies that the average per element energy is bounded by a constant. This is a fairly weak restriction since it permits a very large class of signals, including signals with peak magnitudes as large as $O(\sqrt{n})$. Now suppose that we are able to make k measurements of f^* in the form of noisy, random projections. Specifically, let $\Phi = \{\phi_{i,j}\}$ be an $n \times k$ array of bounded, i.i.d. zero-mean random variables of variance $E[\phi_{i,j}^2] = 1/n$. Samples take the form

$$y_j = \sum_{i=1}^n \phi_{i,j} f_i^* + w_j, \quad j = 1, \ldots, k \tag{1}$$

where $w = \{w_j\}$ are i.i.d. zero-mean random variables, independent of $\{\phi_{i,j}\}$, with variance σ^2. The goal is to recover an estimate of f^* using $\{y_j\}$ and $\{\phi_{i,j}\}$.

Define the risk of a candidate reconstruction f to be

$$R(f) \;=\; \frac{\|f^* - f\|^2}{n} + \sigma^2$$

where the norm is the Euclidean distance. Next assume that both $\{\phi_{i,j}\}$ and $\{y_j\}$ are available. Then we can compute the empirical risk

$$\widehat{R}(f) = \frac{1}{k}\sum_{j=1}^k \left(y_j - \sum_{i=1}^n \phi_{i,j} f_i\right)^2 .$$

It is easy to verify that $E[\widehat{R}(f)] = R(f)$ using the facts that $\{\phi_{i,j}\}$ and $\{w_j\}$ are independent random variables and $E[\phi_{i,j}^2] = 1/n$. Thus, $\widehat{R}(f)$ is an unbiased estimator of $R(f)$. We will use the empirical risk to obtain an estimator $\widehat{f}$ of f^*, and bound the resulting error $E[\|\widehat{f} - f^*\|^2]$. The estimator is based on a complexity-regularized empirical risk minimization, and we use the Craig-Bernstein concentration inequality to control the estimation error of the reconstruction process. That inequality entails the verification of certain moment conditions, which depend on the nature of Φ and w. In this paper we focus on (normalized) Rademacher projections, in which case each $\phi_{i,j}$ is $\pm 1/\sqrt{n}$ with equal probability, and assume that w is a sequence of zero-mean Gaussian noises. Generalizations to other random projections and noise models may be possible following our approach; this would only require one to verify the moment conditions required by the Craig-Bernstein inequality.

Suppose that we have a countable collection $\mathcal{F}$ of candidate reconstruction functions and a non-negative number $c(f)$ assigned to each $f \in \mathcal{F}$ such that $\sum_{f \in \mathcal{F}} 2^{-c(f)} \leq 1$. Furthermore, assume that each $f \in \mathcal{F}$ satisfies $\|f\|^2 \leq nB^2$. Select a reconstruction according to the complexity-regularized empirical risk minimization

$$\widehat{f}_k \ = \ \arg\min_{f \in \mathcal{F}} \left\{ \widehat{R}(f) + \frac{c(f) \log 2}{k\epsilon} \right\}$$

where $\epsilon > 0$ is a constant that depends on B and σ. Then we have the following oracle inequality.

Theorem 1 *Let $\epsilon = 1/(50(B+\sigma)^2)$, then*

$$E\left[\frac{\|\widehat{f}_k - f^*\|^2}{n}\right] \leq C_1 \min_{f \in \mathcal{F}} \left\{ \frac{\|f - f^*\|^2}{n} + \frac{c(f)\log 2 + 4}{k\epsilon} \right\},$$

where the constant C_1 is given by

$$C_1 = \frac{(27-4e)S^2 + (50 - 4\sqrt{2})S + 26}{(23-4e)S^2 + (50-4\sqrt{2})S + 24}$$

with $S = B/\sigma$, the signal-to-noise ratio.

An important point regarding the constants above is that they depend only on σ^2 and B^2, the noise power and the average signal power, respectively. More specifically, note that $1.08 \leq C_1 \leq 1.33$ and that ϵ is inversely proportional to σ^2, so the bound degrades gracefully as the noise level σ increases, in the manner one would expect. The dependence of ϵ on the average signal power is also not surprising, since there is a small chance of "missing" significant components of the signal in our random samples and this source of error scales with B^2.

If f^* is compressible with respect to some known orthonormal basis, then we can obtain explicit bounds on the reconstruction error in terms of the number of random projections k and the degree to which f^* is compressible. Let $f^{(m)}$ denote the best m-term approximation of f^* in the basis. That is, if f^* has a representation $f^* = \sum_{i=1}^n \theta_i \psi_i$ in the basis $\{\psi_i\}$, then $f^{(m)} = \sum_{i=1}^m \theta_{(i)} \psi_{(i)}$, where coefficients and basis functions are ordered such that $|\theta_{(1)}| \geq |\theta_{(2)}| \geq \cdots \geq |\theta_{(n)}|$. Assume that the average squared error $\|f^* - f^{(m)}\|^2/n \equiv \frac{1}{n}\sum_{i=1}^n (f_i^* - f_i^{(m)})^2$ satisfies

$$\frac{\|f^* - f^{(m)}\|^2}{n} \leq C_A m^{-2\alpha}$$

for some $\alpha \geq 0$ and some constant $C_A > 0$. Power-law decays like this arise quite commonly in applications. For example, smooth and piecewise smooth signals as well as signals of bounded variation exhibit this sort of behavior [2], [3]. It is also unnecessary to restrict our attention to orthonormal basis expansions. Much more general approximation strategies can be accommodated [3], but to keep the presentation as simple as possible we will not delve further into such extensions.

Take $\mathcal{F}_c(B, \alpha, C_A) = \{f : \|f\|^2 < nB^2, \ \|f - f^{(m)}\|^2 \leq nC_A m^{-2\alpha}\}$ to be the class of functions to which f^* belongs, and let $\mathcal{F}$ to be a suitably quantized collection of functions represented in terms of the basis $\{\psi_i\}$ (the construction of $\mathcal{F}$ is discussed in Section IV). We have the following error bound.

Theorem 2 *If*

$$c(f) = 2\log(n) \times \\ (\text{\# non-zero coefficients of } f \text{ in the basis } \{\psi_i\})$$

then there exists a constant $C_2 = C_2(B, \sigma, C_A) > 0$ *such that*

$$\sup_{f^* \in \mathcal{F}_c(B,\alpha,C_A)} E\left[\frac{\|\widehat{f}_k - f^*\|^2}{n}\right] \\ \leq \quad C_1 C_2 \left(\frac{k}{\log n}\right)^{-2\alpha/(2\alpha+1)},$$

where C_1 *is as given in Theorem 1.*

Note that the exponent $-2\alpha/(2\alpha+1)$ is the usual exponent governing the rate of convergence in nonparametric function estimation.

A stronger result is obtained if the signal is sparse (*i.e.* belonging to the class $\mathcal{F}_s(B, m) = \{f : \|f\|^2 < nB^2,\ \|f\|_0 \leq m\}$) as stated in the following Corollary.

Corollary 1 *Suppose that* $f^* \in \mathcal{F}_s(B, m)$*. Then there exists a constant* $C_2' = C_2'(B, \sigma) > 0$ *such that*

$$\sup_{f^* \in \mathcal{F}_s(B,m)} E\left[\frac{\|\widehat{f}_k - f^*\|^2}{n}\right] \leq C_1 C_2' \left(\frac{k}{m \log n}\right)^{-1},$$

where C_1 *is as given in Theorem 1.*

Similar results hold if the signal is additionally contaminated with noise prior to the random projection process, as described in the following Corollary.

Corollary 2 *Suppose observations take the form*

$$y_j = \sum_{i=1}^{n} \phi_{i,j}\,(f_i^* + \eta_i) + w_j, \quad j = 1, \ldots, k$$

where $\{\eta_i\}$ *are i.i.d. zero-mean Gaussian random variables with variance* σ_s^2 *that are independent of* $\{\phi_{i,j}\}$ *and* $\{w_j\}$*. Then Theorems 1 and 2 and Corollary 1 hold with slightly different constants* C_1*,* C_2*,* C_2'*, and* ϵ*.*

It is important to point out that all the results above hold for arbitrary signal lengths n, and the constants do not depend on n. The fact that the rate depends only logarithmically on n is significant and illustrates the scalability of this approach. One can interpret these bounds as good indicators of the exceptional performance of random projection sampling in large-n regimes. The dependence on k is shown to be polynomial. In analogy with nonparametric estimation theory (e.g., estimating smooth functions from random point samples), the polynomial rate in k is precisely what one expects in general, and thus we believe the upper bounds are tight (up to constant and logarithmic factors).

To drive this point home, let us again consider the "needle in a haystack" problem, this time in a bit more detail. Suppose the signal f^* is a vector of length n with one nonzero entry of amplitude $\sqrt{n}$ such that $\|f^*\|^2/n = 1$. First, consider random spatial point sampling where observations are noise-free (i.e., each sample is of the form $y_j = f^*(t_j)$, where t_j is selected uniformly at random from the set $\{1, \ldots, n\}$). The squared reconstruction error is 0 if the spike is located and 1 otherwise, and the probability of not finding the spike in k trials is $(1 - 1/n)^k$, giving an average squared error of $(1 - 1/n)^k \cdot 1 + (k/n) \cdot 0 = (1 - 1/n)^k$. If n is large, we can approximate this by $(1 - 1/n)^k \approx e^{-k/n}$, which is very close to 1 when k is significantly smaller than n. On the other hand, randomized Rademacher projections (corrupted with noise) yield an average squared reconstruction error bound of $C_2'\,(k/\log n)^{-1}$, as given above in Corollary 1. This bound may be close to 0 even when $k \ll n$, provided $k > \log n$. This shows that even given the advantage of being noiseless, the reconstruction error from spatial point sampling may be far greater than that resulting from random projections.

III. Oracle Inequality

In this section we prove Theorem 1. For ease of notation, we adopt the shorthand notation $\phi_j = [\phi_{1,j}\ \phi_{2,j}\ \ldots\ \phi_{n,j}]^T$ for the vector corresponding to the j^{th} projection. The empirical risk of a vector f can now be written as

$$\widehat{R}(f) = \frac{1}{k}\sum_{j=1}^{k}(y_j - \phi_j^T f)^2.$$

We will bound $r(f, f^*) \equiv R(f) - R(f^*)$, the "excess risk" between a candidate reconstruction f and the actual function f^*, using the complexity-regularization method introduced in [13]. Note that $r(f, f^*) = \|f - f^*\|^2/n$.

Define the empirical excess risk $\widehat{r}(f, f^*) \equiv \widehat{R}(f) - \widehat{R}(f^*)$. Then

$$\begin{aligned}\widehat{r}(f, f^*) &= -\frac{1}{k}\sum_{j=1}^{k}\left[(y_j - \phi_j^T f^*)^2 - (y_j - \phi_j^T f)^2\right] \\ &= -\frac{1}{k}\sum_{j=1}^{k} U_j\end{aligned}$$

where $U_j = \left[(y_j - \phi_j^T f^*)^2 - (y_j - \phi_j^T f)^2\right]$ are i.i.d. for $j = 1 \ldots k$. Notice that $r(f, f^*) - \widehat{r}(f, f^*) = \frac{1}{k}\sum_{j=1}^{k}(U_j - E\left[U_j\right])$. We will make use of the Craig-Bernstein inequality [14], which states that the probability of the event

$$\frac{1}{k}\sum_{j=1}^{k}(U_j - E\left[U_j\right]) \geq \frac{t}{k\epsilon} + \frac{\epsilon k\ var\left(\frac{1}{k}\sum_{j=1}^{k} U_j\right)}{2(1-\zeta)}$$

is less than or equal to e^{-t} for $0 < \epsilon h \leq \zeta < 1$ and $t > 0$, provided the variables U_j satisfy the moment condition

$$E\left[|U_j - E[U_j]|^k\right] \ \leq\ \frac{k!\, var(U_j)\, h^{k-2}}{2}$$

for some $h > 0$ and all integers $k \geq 2$. If we consider vectors f^* and estimates f where $\|f^*\|^2 \leq nB^2$ and $\|f\|^2 \leq nB^2$, Rademacher projections, and Gaussian noises with variance σ^2, then the moment condition is satisfied with $h = 16B^2 e + 8\sqrt{2}B\sigma$, as shown in the Appendix. Alternative forms of random projections and noises can also be handled using the approach outlined next, provided the moment conditions are satisfied.

To use the Craig-Bernstein inequality we also need a bound on the variance of U_j itself. Defining $g = f - f^*$, we have

$$var(U_j) = E[(\phi^T g)^4] - \left(\frac{\|g\|^2}{n}\right)^2 + \frac{4\sigma^2\|g\|^2}{n}.$$

As shown in the Appendix, for integers $k \geq 1$

$$\left(\frac{\|g\|^2}{n}\right)^k \leq E[(\phi^T g)^{2k}] \leq (2k)!!\left(\frac{\|g\|^2}{n}\right)^k,$$

where $(2k)!! \equiv (1)(3)\ldots(2k-1)$. Thus we can bound the variance of U_j by

$$var(U_j) \leq \left(2\frac{\|g\|^2}{n} + 4\sigma^2\right)\frac{\|g\|^2}{n}.$$

Since g satisfies $\|g\|^2 \leq 4nB^2$ and $r(f, f^*) = \|f - f^*\|^2/n = \|g\|^2/n$, the bound becomes

$$var(U_j) \leq \left(8B^2 + 4\sigma^2\right) r(f, f^*).$$

So, we can replace the term in the Craig-Bernstein inequality that depends on the variance by

$$\begin{aligned}k\ var\left(\frac{1}{k}\sum_{j=1}^{k} U_j\right) &= \frac{1}{k}\sum_{j=1}^{k} var(U_j) \\ &\leq \left(8B^2 + 4\sigma^2\right) r(f, f^*).\end{aligned}$$

For a given function f, we have that the probability of the event

$$r(f,f^*) - \widehat{r}(f,f^*) > \frac{t}{k\epsilon} + \frac{\left(8B^2+4\sigma^2\right)\epsilon\ r(f,f^*)}{2(1-\zeta)}$$

is less than or equal to e^{-t}, or by letting $\delta = e^{-t}$, the probability of

$$r(f,f^*) - \widehat{r}(f,f^*) > \frac{\log\left(\frac{1}{\delta}\right)}{k\epsilon} + \frac{\left(8B^2+4\sigma^2\right)\epsilon\ r(f,f^*)}{2(1-\zeta)}$$

is less than or equal to δ. Now assign to each $f \in \mathcal{F}$ a non-negative penalty term $c(f)$ such that the penalties satisfy the Kraft Inequality [15]

$$\sum_{f\in\mathcal{F}} 2^{-c(f)} \leq 1$$

and let $\delta(f) = 2^{-c(f)}\delta$. Then by applying the union bound we have for all $f \in \mathcal{F}$ and for all $\delta > 0$

$$\begin{aligned} r(f,f^*) - \widehat{r}(f,f^*) \leq& \frac{c(f)\log 2 + \log\left(\frac{1}{\delta}\right)}{k\epsilon} \\ &+ \quad \frac{\left(8B^2+4\sigma^2\right)\epsilon\ r(f,f^*)}{2(1-\zeta)} \end{aligned}$$

with probability at least $1-\delta$. Set $\zeta = \epsilon h$, and define

$$a \equiv \frac{\left(8B^2+4\sigma^2\right)\epsilon}{2(1-\zeta)}.$$

Choose

$$\epsilon < \frac{1}{(4+16e)B^2 + 8\sqrt{2}B\sigma + 2\sigma^2},$$

and notice that $a < 1$ and $\zeta < 1$ by choice of ϵ. Then

$$(1-a)r(f,f^*) \leq \widehat{r}(f,f^*) + \frac{c(f)\log 2 + \log\left(\frac{1}{\delta}\right)}{k\epsilon}$$

holds with probability at least $1-\delta$ for all $f \in \mathcal{F}$ and any $\delta > 0$.

For the given training samples, we can minimize the upper bound by choosing

$$\widehat{f}_k = \arg\min_{f\in\mathcal{F}} \left\{ \widehat{r}(f,f^*) + \frac{c(f)\log 2}{k\epsilon} \right\}$$

which is equivalent to

$$\widehat{f}_k = \arg\min_{f\in\mathcal{F}} \left\{ \widehat{R}(f) + \frac{c(f)\log 2}{k\epsilon} \right\}$$

since we can ignore $\widehat{R}(f^*)$ when performing the optimization. If we define

$$f_k^* \equiv \arg\min_{f\in\mathcal{F}} \left\{ R(f) + \frac{c(f)\log 2}{k\epsilon} \right\}$$

then with probability at least $1-\delta$

$$\begin{aligned} &(1-a)r(\widehat{f}_k,f^*) \\ &\leq \quad \widehat{r}(\widehat{f}_k,f^*) + \frac{c(\widehat{f}_k)\log 2 + \log\left(\frac{1}{\delta}\right)}{k\epsilon} \\ &\leq \quad \widehat{r}(f_k^*,f^*) + \frac{c(f_k^*)\log 2 + \log\left(\frac{1}{\delta}\right)}{k\epsilon} \end{aligned} \tag{2}$$

since $\widehat{f}_k$ minimizes the complexity-regularized empirical risk criterion. Using the Craig-Bernstein inequality again to bound $\widehat{r}(f_k^*,f^*) - r(f_k^*,f^*)$ (with the same variance bound as before) we get that with probability at least $1-\delta$

$$\widehat{r}(f_k^*,f^*) - r(f_k^*,f^*) \leq a\, r(f_k^*,f^*) + \frac{\log\left(\frac{1}{\delta}\right)}{k\epsilon}. \tag{3}$$

We want both (2) and (3) to hold simultaneously, so we use the union bound to obtain

$$\begin{aligned} r(\widehat{f}_k, f^*) \le \left(\frac{1+a}{1-a}\right) r(f_k^*, f^*) \\ + \quad \frac{1}{1-a}\left(\frac{c(f_k^*)\log 2 + 2\log\left(\frac{1}{\delta}\right)}{k\epsilon}\right) \end{aligned}$$

holding with probability at least $1-2\delta$.

We will convert this probability deviation bound into a bound on the expected value using the fact that, for positive random variables X, $E[X] = \int_0^\infty P(X > t)dt$. Let $\delta = e^{-k\epsilon t(1-a)/2}$ to obtain

$$\begin{aligned} P\left(r(\widehat{f}_k, f^*) - \left(\frac{1+a}{1-a}\right) r(f_k^*, f^*) - \frac{c(f_k^*)\log 2}{k\epsilon(1-a)} \ge t\right) \\ \le \quad 2e^{-k\epsilon t(1-a)/2}. \end{aligned}$$

Integrating this relation gives

$$E\left[r(\widehat{f}_k, f^*)\right] \le \left(\frac{1+a}{1-a}\right) r(f_k^*, f^*) + \frac{c(f_k^*)\log 2 + 4}{k\epsilon(1-a)}.$$

Now, since a is positive,

$$\begin{aligned} E\left[\frac{\|\widehat{f}_k - f^*\|^2}{n}\right] &= E\left[r(\widehat{f}_k, f^*)\right] \\ &\le \left(\frac{1+a}{1-a}\right) r(f_k^*, f^*) + \frac{c(f_k^*)\log 2 + 4}{k\epsilon(1-a)} \\ &\le \left(\frac{1+a}{1-a}\right) r(f_k^*, f^*) + (1+a)\frac{c(f_k^*)\log 2 + 4}{k\epsilon(1-a)} \\ &= \left(\frac{1+a}{1-a}\right)\left\{R(f_k^*) - R(f^*) + \frac{c(f_k^*)\log 2 + 4}{k\epsilon}\right\} \\ &\le \left(\frac{1+a}{1-a}\right)\min_{f\in\mathcal{F}}\left\{R(f) - R(f^*) + \frac{c(f)\log 2 + 4}{k\epsilon}\right\} \\ &= C_1 \min_{f\in\mathcal{F}}\left\{\frac{\|f - f^*\|^2}{n} + \frac{c(f)\log 2 + 4}{k\epsilon}\right\}, \end{aligned}$$

where $C_1 = (1+a)/(1-a)$.

Typical values of C_1 can be determined by approximating the constant

$$\epsilon < \frac{1}{(4+16e)B^2 + 8\sqrt{2}B\sigma + 2\sigma^2}.$$

Upper bounding the denominator guarantees that the condition is satisfied, so let $\epsilon = 1/\left(50(B+\sigma)^2\right)$. Now

$$\begin{aligned} a &= \frac{\left(8B^2 + 4\sigma^2\right)\epsilon}{2(1-\zeta)} \\ &= \frac{2B^2 + \sigma^2}{(25-4e)B^2 + (50-4\sqrt{2})B\sigma + 25\sigma^2}. \end{aligned}$$

If we denote the signal to noise ratio by $S^2 = B^2/\sigma^2$ then

$$a = \frac{2S^2 + 1}{(25-4e)S^2 + (50-4\sqrt{2})S + 25}$$

for which the extremes are $a_{min} = 1/25$ and $a_{max} = 2/(25-4e)$, giving constants C_1 in the range of $[13/12, (27-4e)/(23-4e)] \approx [1.08, 1.33]$.

IV. Error Bounds for Compressible Signals

In this section we prove Theorem 2 and Corollary 1. Suppose that f^* is compressible in a certain orthonormal basis $\{\psi_i\}_{i=1}^n$. Specifically, let $f^{(m)}$ denote the best m-term approximation of f^* in terms of $\{\psi_i\}$, and assume that the error of the approximation obeys

$$\frac{\|f^* - f^{(m)}\|^2}{n} \;\leq\; C_A\, m^{-2\alpha}$$

for some $\alpha \geq 0$ and a constant $C_A > 0$. Let $\mathcal{F}_c(B, \alpha, C_A) = \{f : \|f\|^2 < nB^2,\ \|f - f^{(m)}\|^2 \leq nC_A m^{-2\alpha}\}$ so $f^* \in \mathcal{F}_c(B, \alpha, C_A)$.

Let us use the basis $\{\psi_i\}$ for the reconstruction process. Any vector f can be expressed in terms of the basis $\{\psi_i\}$ as $f = \sum_{i=1}^n \theta_i \psi_i$, where $\theta = \{\theta_i\}$ are the coefficients of f in this basis. Let T denote the transform that maps coefficients to functions, so that $f = T\theta$. Define $\Theta = \{\theta : \|T\theta\|^2 \leq nB^2, \theta_i \text{ uniformly quantized to } n^p \text{ levels}\}$ to be the set of candidate solutions in the basis $\{\psi_i\}$, so that $\mathcal{F} = \{f : f = T\theta,\ \theta \in \Theta\}$. The penalty term $c(f)$ written in terms of the basis $\{\psi_i\}$ is $c(f) = c(\theta) = (1+p)\log(n)\sum_{i=1}^n I_{\theta_i \neq 0} = (1+p)\log(n)\|\theta\|_0$. It is easily verified that $\sum_{f \in \mathcal{F}} 2^{-c(f)} \leq 1$ by noting that each $\theta \in \Theta$ can be uniquely encoded via a prefix code consisting of $(1+p)\log n$ bits per non-zero coefficient ($\log n$ bits for the location and $p \log n$ bits for the quantized value) in which case the codelengths $c(f)$ must satisfy the Kraft inequality [15].

The oracle inequality

$$E\left[\frac{\|\widehat{f}_k - f^*\|^2}{n}\right] \leq C_1 \min_{f \in \mathcal{F}} \left\{\frac{\|f - f^*\|^2}{n} + \frac{c(f)\log 2 + 4}{k\epsilon}\right\}$$

can also be written as

$$E\left[\frac{\|\widehat{f}_k - f^*\|^2}{n}\right] \leq C_1 \min_{\theta \in \Theta} \left\{\frac{\|\theta - \theta^*\|^2}{n} + \frac{c(\theta)\log 2 + 4}{k\epsilon}\right\}$$

where $f^* = T\theta^*$. For each integer $m \geq 1$, let $\theta^{(m)}$ denote the coefficients corresponding to the best m-term approximation of f^* and let $\theta_q^{(m)}$ denote the nearest element in Θ. The maximum possible dynamic range for the coefficient magnitudes, $\pm\sqrt{n}B$, is quantized to n^p levels, giving $\|\theta_q^{(m)} - \theta^{(m)}\|^2 \leq 4B^2/n^{2p-2} = C_Q/n^{2p-2}$. Now insert $\theta_q^{(m)}$ in place of θ in the oracle bound. The first term can be expanded as

$$\begin{aligned}
&\|\theta_q^{(m)} - \theta^*\|^2 \\
&= \|\theta_q^{(m)} - \theta^{(m)} + \theta^{(m)} - \theta^*\|^2 \\
&\leq \|\theta_q^{(m)} - \theta^{(m)}\|^2 + \\
&\quad 2\|\theta_q^{(m)} - \theta^{(m)}\| \cdot \|\theta^{(m)} - \theta^*\| + \\
&\quad \|\theta^{(m)} - \theta^*\|^2 \\
&\leq \frac{C_Q}{n^{2p-2}} + 2m^{-\alpha}\sqrt{\frac{nC_A C_Q}{n^{2p-2}}} + C_A n m^{-2\alpha}.
\end{aligned}$$

Now notice that $c(\theta_q^{(m)}) = (1+p)m\log n$, so

$$\begin{aligned}
&E\left[\frac{\|\widehat{f}_k - f^*\|^2}{n}\right] \\
&\leq C_1 \min_m \left\{\frac{C_Q}{n^{2p-1}} + \frac{2m^{-\alpha}\sqrt{C_A C_Q}}{n^{p-1/2}}\right. \\
&\quad + \left. C_A m^{-2\alpha} + \frac{(1+p)m\log n \log 2}{k\epsilon} + \frac{4}{k\epsilon}\right\}.
\end{aligned}$$

The quantization error terms decay exponentially in p, so they can be made arbitrarily small while incurring only a modest (linear) increase in the complexity term. Balancing the third and fourth terms gives

$$m = \left(\frac{(1+p)\log 2}{\epsilon C_A}\right)^{\frac{-1}{2\alpha+1}} \left(\frac{k}{\log n}\right)^{\frac{1}{2\alpha+1}}$$

so

$$C_A m^{-2\alpha} = C_A \left(\frac{(1+p)\log 2}{\epsilon C_A}\right)^{\frac{2\alpha}{2\alpha+1}} \left(\frac{k}{\log n}\right)^{\frac{-2\alpha}{2\alpha+1}},$$

and since

$$\frac{1}{k} < \left(\frac{\log n}{k}\right)^{\frac{2\alpha}{2\alpha+1}}$$

when $k > 1$ and $n > e$, then

$$E\left[\frac{\|\widehat{f}_k - f^*\|^2}{n}\right] \leq C_1 C_2 \left(\frac{k}{\log n}\right)^{\frac{-2\alpha}{2\alpha+1}}$$

holds for every $f^* \in \mathcal{F}_c(B, \alpha, C_A)$ as claimed in the Theorem, where

$$C_2 = \left\{2C_A \left(\frac{(1+p)\log 2}{\epsilon C_A}\right)^{\frac{2\alpha}{2\alpha+1}} + \frac{4}{\epsilon}\right\}.$$

Suppose now that $f^* \in \mathcal{F}_s(B, m)$ where $\mathcal{F}_s(B, m) = \{f : \|f\|^2 < nB^2,\ \|f\|_{l_0} \leq m\}$. In this case,

$$\|\theta_q^{(m)} - \theta^*\|^2 \leq \frac{C_Q}{n^{2p-2}}$$

since $\|\theta^{(m)} - \theta^*\| = 0$. Now the penalty term dominates in the oracle bound and

$$E\left[\frac{\|\widehat{f}_k - f^*\|^2}{n}\right] \leq C_1 C_2' \left(\frac{k}{m \log n}\right)^{-1},$$

holds for every $f^* \in \mathcal{F}_s(B, m)$ where

$$C_2' = \left\{\frac{(1+p)\log 2 + 4}{\epsilon}\right\}.$$

V. Optimization Scheme

Although our optimization is non-convex, it does permit a simple, iterative optimization strategy that produces a sequence of reconstructions for which the corresponding sequence of complexity-regularized empirical risk values is non-increasing. This algorithm, which is described below, has demonstrated itself to be quite effective in similar denoising and reconstruction problems [16]–[18]. A possible alternative strategy might entail "convexifying" the problem by replacing the l_0 penalty with an l_1 penalty. Recent results show that often the solution to this convex problem coincides with or approximates the solution to the original non-convex problem [19].

Let us assume that we wish to reconstruct our signal in terms of the basis $\{\psi_i\}$. Using the definitions introduced in the previous section, the reconstruction

$$\widehat{f}_k = \arg\min_{f \in \mathcal{F}} \left\{\widehat{R}(f) + \frac{c(f)\log 2}{k\epsilon}\right\}$$

is equivalent to $\widehat{f}_k = T\widehat{\theta}_k$ where

$$\widehat{\theta}_k = \arg\min_{\theta \in \Theta} \left\{\widehat{R}(T\theta) + \frac{c(\theta)\log 2}{k\epsilon}\right\}$$

Thus, the optimization problem can then be written as

$$\widehat{\theta}_k = \arg\min_{\theta \in \Theta} \left\{\|y - PT\theta\|^2 + \frac{2\log(2)\log(n)}{\epsilon}\|\theta\|_0\right\}$$

where $P = \Phi^T$, the transpose of the $n \times k$ projection matrix Φ, y is a column vector of the k observations, and $\|\theta\|_0 = \sum_{i=1}^n I_{\{\theta_i \neq 0\}}$.

To solve this, we use an iterative bound-optimization procedure, as proposed in [16]–[18]. This procedure entails a two-step iterative process that begins with an initialization $\theta^{(0)}$ and computes:

$$1. \qquad \varphi^{(t)} \;=\; \theta^{(t)} + \frac{1}{\lambda}(PT)^T(y - PT\theta^{(t)})$$

$$2. \qquad \widehat{\theta}_i^{(t+1)} \;=\; \begin{cases} \varphi_i^{(t)} & \text{if } |\varphi_i^{(t)}| \geq \sqrt{\frac{2\log(2)\log(n)}{\lambda\epsilon}} \\ 0 & \text{otherwise} \end{cases}$$

where λ is the largest eigenvalue of $P'P$. This procedure is desirable since the second step, in which the complexity term plays its role, involves a simple coordinate-wise thresholding operation. It is easy to verify that the iterations produce a monotonically non-increasing sequence of complexity-regularized empirical risk values [18]. Thus, this procedure provides a simple iteration that tends to minimize the original objective function, and appears to give good results in practice [16]. The iterations can be terminated when the entries uniformly satisfy $|\widehat{\theta}_i^{(t+1)} - \widehat{\theta}_i^{(t)}| \leq \delta$, for a small positive tolerance δ.

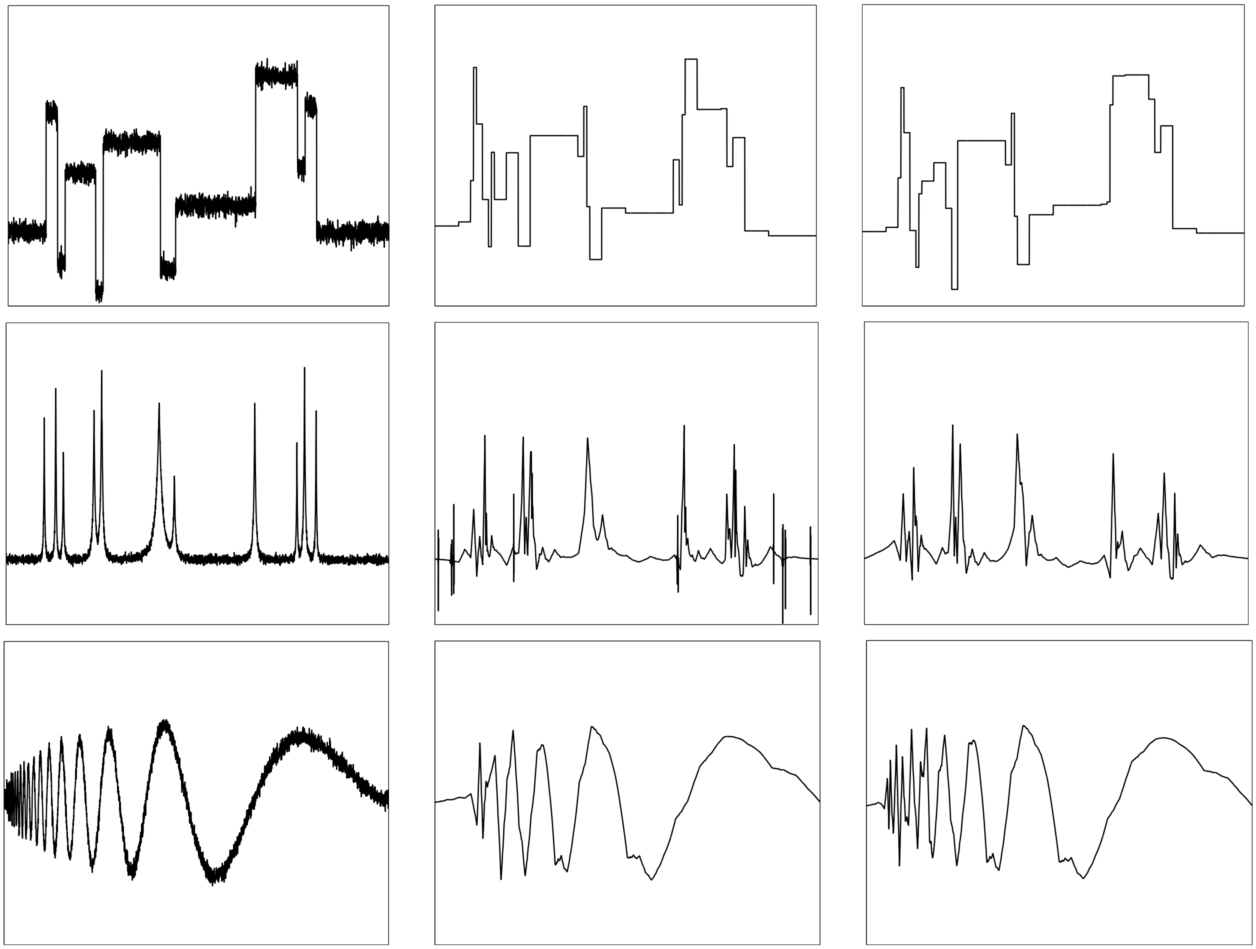
The computational complexity of the above procedure is quite appealing. Each iteration requires only $O(nk)$ operations, assuming that the transform T can be computed in $O(n)$ operations. For example, the discrete wavelet of Fourier transforms can be computed in $O(n)$ and $O(n \log n)$ operations, respectively. Multiplication by P is the most intensive operation, requiring $O(nk)$ operations. The thresholding step is carried out independently in each coordinate, and this step requires $O(n)$ operations as well. Of course, the number of iterations required is problem-dependent and difficult to predict, but in our experience in this application and others [16], [18] algorithms of this sort tend to converge in a reasonably small number of iterations, even in very high dimensional cases.

One point worthy of mention relates to the factor $1/\epsilon = 50(B+\sigma)^2$ in the penalty. As is often the case with conservative bounds of this type, the theoretical penalty is larger than what is needed in practice to achieve good results. Also, a potential hurdle to calibrating the algorithm is that it depends on knowledge of B and σ, neither of which may be known a priori. Strictly speaking, these values do not need to be known *independently* but rather we need only estimate $(B+\sigma)^2$. To that end, notice that each observation is a random variable with variance equal to $\|f\|^2/n + \sigma^2$. Let $B = \sqrt{\|f\|^2/n}$, which is the minimum B satisfying the stated bound $\|f\|^2 \leq nB^2$. Then the variance of each observation is $B^2 + \sigma^2$. Further, it is easy to verify that $2(B^2+\sigma^2) \geq (B+\sigma)^2$. So, a scheme could be developed whereby the sample variance is used as a surrogate for the unknown quantities in the form in which they appear in the parameter ϵ. This would entail using another concentration inequality to control the error between the sample variance and its mean value, and propagating this additional error through the derivation of the oracle inequality. While this is relatively straightforward, we omit a complete derivation here.

To illustrate the performance of the algorithm above, in Figure 1 we consider three standard test signals, each of length $n = 4096$. Rademacher projection samples (contaminated with additive white Gaussian noise) are taken for the Blocks, Bumps, and Doppler test signals. The algorithm described above is employed for reconstruction, with one slight modification. Since the theoretical penalty can be a bit too conservative in practice, the threshold used in this example is about 20% of the theoretical value (i.e., a threshold of $\sqrt{2\log(2)\log(n)/(\lambda\epsilon)}/4.6$ was used). The SNR, defined as $SNR = 10\log_{10}\frac{B^2}{\sigma^2}$ where $B^2 = \|f\|^2/n$, is 21dB for each test signal. To convey a sense of the noise level, column (a) of Figure 1 shows the original signals contaminated with the same level of noise (i.e., the signal resulting from conventional point samples contaminated with noise of the same power). Column (b) shows the reconstructions obtained from 600 projections; reconstructions from 1200 projections are shown in column (c). The Blocks signal (top row) was reconstructed using the Haar wavelet basis (Daubechies-2), well-suited to the piecewise constant nature of the signal. The Bumps and Doppler signals (middle and bottom row, respectively) were reconstructed using the Daubechies-6 wavelet basis. Of course, the selection of the "best" reconstruction basis is a separate matter that is beyond the scope of this paper.

VI. Applications

One immediate application of the results and methods above is to signal coding and A/D conversion. In the noiseless setting, several authors have suggested the use of random projection sampling for such purposes [1]–[3].



(a) (b) (c)

Fig. 1. Simulation examples using Blocks, Bumps, and Doppler test signals of length 4096. Column (a) shows the original signals with an equivalent level of additive per-pixel noise. Columns (b) and (c) show reconstructions from 600 and 1200 projections, respectively.

Our results indicate how such schemes might perform in the presence of noise. Suppose that we have an array of n sensors, each making a noisy measurement. The noise could be due to the sensors themselves or environmental factors. The goal of encoding and A/D conversion is to represent the n sensor readings in a compressed form, suitable for digital storage or transmission. Our results suggest that k random Rademacher projections of the n sensor readings can be used for this purpose, and the error bounds suggest guidelines for how many projections might be required for a certain level of precision.

Our theory and method can also be applied to wireless sensing as follows. Consider the problem of sensing a distributed field (e.g., temperature, light, chemical) using a collection of n wireless sensors distributed uniformly over a region of interest. Such systems are often referred to as *sensor networks*. The goal is to obtain an accurate, high-resolution reconstruction of the field at a remote destination. One approach to this problem is to require each sensor to digitally transmit its measurement to the destination, where field reconstruction is then performed. Alternatively, the sensors might collaboratively process their measurements to reconstruct the field themselves and then transmit the result to the destination (i.e., the nodes collaborate to compress their data prior to transmission). Both approaches pose significant demands on communication resources and infrastructure, and it has recently been suggested that non-collaborative analog communication schemes may offer a more resource-efficient alternative [20]–[22].

Assume that the sensor data is to be transmitted to the destination over an additive white Gaussian noise channel. Suppose the destination broadcasts (perhaps digitally) a random seed to the sensors. Each node modifies this seed in a unique way known to only itself and the destination (e.g., this seed could be multiplied by the node's address or geographic position) and uses the seed to generate a pseudorandom Rademacher sequence. The sequences can also be reconstructed at the destination. The nodes then transmit the random projections to the destination *phase-coherently* (*i.e.*, beamforming). This is accomplished by requiring each node to simply multiply its reading by an element of its random sequence in each projection/communication step and transmit the result to the destination via amplitude modulation. If the transmissions from all n sensors can be synchronized so that they all arrive in phase at the destination, then the averaging inherent in the multiple access channel computes the desired inner product. After receiving k projections, the destination can employ the reconstruction algorithm above using a basis of choice (e.g., wavelet). The communications procedure is completely non-adaptive and potentially very simple to implement. The collective functioning of the wireless sensors in this process is more akin to an ensemble of phase-coherent emitters than it is to conventional networking operations. Therefore, we prefer the term *sensor ensemble* instead of sensor network in this context.

A remarkable aspect of the sensor ensemble approach is that the power required to achieve a target distortion level can be very minimal. Let σ_s^2 and σ_c^2 denote the noise variance due to sensing and communication, respectively. Thus, each projection received at the destination is corrupted by a noise of total power $\sigma_s^2 + \sigma_c^2$. The sensing noise variance is assumed to be a constant and the additional variance due to the communication channel is assumed to scale like the inverse of the total received power

$$\sigma_c^2 \propto \frac{1}{nP}$$

where P is the transmit power per sensor. In order to achieve the distortion decay rates given by our upper bounds, it is sufficient that the variance due to the communication channel behaves like a constant. Therefore, we require only that $P \propto n^{-1}$. This results in a rather surprising conclusion. Ideal reconstruction is possible at the destination with per sensor transmit power P tending to zero as the density of sensors increases. If conventional spatial point samples were taken instead (e.g., if a single sensor is selected at random in each step and transmits its measurement to the destination), then the power required per sensor would be a constant, since only one sensor would be involved in such a transmission. Thus, it appears that random projection sampling may be more desirable in wireless sensing applications. Thorough treatments of compressive sampling via random projections in sensor networking applications can be found in our recent work [23], [24].

VII. Conclusions and Future Work

We have shown that compressible signals can be accurately recovered from random projections contaminated with noise. The squared error bounds for compressible signals are $O((k/\log n)^{\frac{-2\alpha}{2\alpha+1}})$, which is within a logarithmic factor of the usual nonparametric estimation rate, and $O((k/\log n)^{-1})$ for sparse signals. We also proposed a practical iterative algorithm for signal reconstruction. One of the most promising potential applications of our theory and method is to wireless sensing, wherein one realizes a large transmission power gain by random projection sampling as opposed to conventional spatial point sampling.

The role of the noise variance in the rates we presented is worthy of further attention. As the noise variance tends to zero, one intuitively expects to attain the fast rates that are known to be achievable in the noiseless setting. Our theory is based in the noisy regime and it does not directly imply the previously established bounds in the noiseless setting. Simply put, our analysis assumes a noise variance strictly greater than zero.

Let us comment briefly on the tightness of the upper bounds given by our theory. In analogy with classical nonparametric estimation theory (e.g., estimating smooth functions from random point samples), the polynomial rate in k is precisely what one expects in general, and thus we believe the upper bounds are tight (up to constant and logarithmic factors). Moreover, in the special case of sparse signals with m non-zero terms, we obtain an error bound of m/k (ignoring constant and logarithmic factors). Standard parametric statistical analysis suggests that one should not expect a rate of better than m/k (degrees-of-freedom/ sample-size) in such cases, which again supports our intuition regarding the tightness of the bounds (in terms of the convergence rate). However, to our knowledge explicit minimax lower bounds have not been established in the context of this problem, and the determination of such bounds is one of our future research directions.

Although we considered only the case of Gaussian noise in the observation model (1), the same results could be achieved for any zero-mean, symmetrically distributed noise that is independent of the projection vector elements and satisfies

$$E[w_j^{2k}] \leq (2k)!!var(w_j)h^{k-2}$$

for some constant h not depending on k, a result that follows immediately using the lemmas presented in the Appendix. Another extension would be the consideration of other random projections instead of the Rademacher projections considered here. Most of our basic approach would go through in such cases; one would only need to verify the moment conditions of the Craig-Bernstein inequality for particular cases.

ACKNOWLEDGEMENT

The authors would like to thank Rui Castro for his assistance in all stages of the work presented here. The authors would also like to thank Emmanuel Candès and the anonymous reviewers for their careful readings of the paper and helpful comments.

APPENDIX

A. The Craig-Bernstein Moment Condition

The central requirement of the Craig-Bernstein inequality is the satisfaction of the moment condition

$$E\left[|X - E[X]|^p\right] \leq \frac{p!\, var(X) h^{p-2}}{2}$$

for integers $p \geq 2$ with some positive constant h that does not depend on p, a condition that is trivially verified for $p = 2$ with any finite value of h. For larger p this condition can be very difficult to verify for several reasons, not the least of which is the absolute value present in the moments. Complexity regularization methods based on the Craig-Bernstein inequality often assume that the observations are bounded [13], a condition that makes verification of the moment condition relatively simple, but which does not hold in our application due to the Gaussian noise model we work with. Moreover, even if we adopted a bounded noise model, the boundedness of the observations alone is not sufficient to obtain the rates of convergence we establish. Indeed, even in the absence of noise, the observations in our set-up may only be pointwise bounded by the inequality $|y_i| \leq \sqrt{n}B$, yielding a constant h that would grow proportionally to n. With that motivation, we develop a framework under which the moments and a bounding constant h (that does not depend on n) can be determined more directly.

First, observe that the moment condition is usually easier to verify for the even powers because the absolute value need not be dealt with directly. This is sufficient to guarantee the moment condition is satisfied for all integers $p \geq 2$, as proved in the following lemma.

Lemma 1 *Suppose the Craig-Bernstein moment condition holds for all even integers greater than or equal to* 2*, that is there exists an* $h > 0$ *such that*

$$E\left[|X - E[X]|^{2k}\right] \leq \frac{(2k)! var(X) h^{2k-2}}{2}$$

for $k \geq 2$ *since the* $k = 1$ *case is satisfied trivially for any* h*. Then the condition holds also for the odd absolute moments,*

$$E\left[|X - E[X]|^{2k-1}\right] \leq \frac{(2k-1)! var(X) \tilde{h}^{2k-3}}{2}$$

for $k \geq 2$ *with* $\tilde{h} = 2h$*. Thus*

$$E\left[|X - E[X]|^p\right] \leq \frac{p! var(X)(2h)^{p-2}}{2}, \quad p \geq 2.$$

Proof: For ease of notation, let $Z = X - E[X]$. Hölder's Inequality states, for any random variables A and B,

$$E[|AB|] \leq E[|A|^p]^{1/p} E[|B|^q]^{1/q}$$

where $1 < p, q < \infty$ and $1/p + 1/q = 1$. Take $A = Z$, $B = Z^{2k-2}$, and $p = q = 2$ to get

$$E[|Z|^{2k-1}] \leq \sqrt{E[Z^2] E[Z^{4k-4}]}$$

where the absolute values inside the square root have been dropped because the exponents are even. Now

$$E[Z^{4k-4}] \leq \frac{(4k-4)! E[Z^2] h^{4k-6}}{2}$$

by assumption, so

$$\begin{aligned} E[|Z|^{2k-1}] &\leq \sqrt{\frac{(4k-4)!(E[Z^2])^2 h^{4k-6}}{2}} \\ &\leq \sqrt{\frac{(4k-4)!}{2}} E[Z^2] h^{2k-3}. \end{aligned}$$

We want to satisfy the following inequality by choice of $\tilde{h}$

$$E[|Z|^{2k-1}] \leq \frac{(2k-1)! E[Z^2] \tilde{h}^{2k-3}}{2},$$

which means $\tilde{h}$ must satisfy

$$\frac{(2k-1)!}{2} \tilde{h}^{2k-3} \geq \sqrt{\frac{(4k-4)!}{2}} h^{2k-3}.$$

If we choose

$$\tilde{h} \geq \max_{k \geq 2} \left\{ \left(\frac{\sqrt{2(4k-4)!}}{(2k-1)!} \right)^{\frac{1}{2k-3}} \right\} h$$

then the moment condition will be satisfied for the odd exponents $2k-1$. An upper bound for the term in brackets is 2, as shown here.

For $k \geq 2$, the bound $(2k)! \leq 2^{2k}(k!)^2$ holds and can be verified by induction on k. This implies

$$\left(\frac{\sqrt{2(4k-4)!}}{(2k-1)!} \right)^{\frac{1}{2k-3}} \leq 2 \left(\frac{2\sqrt{2}}{2k-1} \right)^{\frac{1}{2k-3}}. \tag{4}$$

Now, the term in parentheses on the right hand side of (4) is always less than 1 for $k \geq 2$. The final step is to show that

$$\lim_{k \to \infty} \left(\frac{2\sqrt{2}}{2k-1} \right)^{\frac{1}{2k-3}} = 1,$$

which is verified by noting that

$$\begin{aligned} &\lim_{k \to \infty} \log \left(\frac{2\sqrt{2}}{2k-1} \right)^{\frac{1}{2k-3}} \\ &= \lim_{k \to \infty} \frac{1}{2k-3} \left(\log(2\sqrt{2}) - \log(2k-1) \right) = 0. \end{aligned}$$

Thus, the moment condition is satisfied for odd moments with $\tilde{h} = 2h$. Also, if the moment condition is satisfied for a given h, it is also satisfied for any $\tilde{h} \geq h$ so

$$E[|Z|^p] = E[|X - E[X]|^p] \leq \frac{p! var(X)(2h)^{p-2}}{2}$$

holds for all integers $p \geq 2$ as claimed. ■

We will also need results for how sums and products of random variables behave with respect to the moment condition. For that, we have the following two lemmas.

Lemma 2 *Let $Z = A + B$ be the sum of two zero-mean random variables with variances $var(A) = E[A^2]$ and $var(B) = E[B^2]$, not both zero, and such that $E[AB] \geq 0$. Suppose both A and B satisfy the moment condition for a given integer $p \geq 3$ with positive constants h_A and h_B, respectively. That is*

$$E[|A|^p] \leq \frac{p! var(A) h_A^{p-2}}{2}, \quad E[|B|^p] \leq \frac{p! var(B) h_B^{p-2}}{2}.$$

Then

$$E[|Z|^p] \leq \frac{p! var(Z) h_S^{p-2}}{2}$$

where $h_S = 2^{1/(p-2)}(h_A + h_B)$.

Proof: First, define $V_A = var(A)/(var(A)+var(B))$, $V_B = var(B)/(var(A)+var(B))$, $H_A = h_A/(h_A + h_B)$, and $H_B = h_B/(h_A + h_B)$. Use Minkowski's Inequality to write

$$\begin{aligned}
E[|A+B|^p] &\leq \left[E[|A|^p]^{1/p} + E[|B|^p]^{1/p}\right]^p \\
&\leq \frac{p!}{2}\left[\left(var(A)h_A^{p-2}\right)^{\frac{1}{p}} + \left(var(B)h_B^{p-2}\right)^{\frac{1}{p}}\right]^p \\
&= \frac{p!(var(A)+var(B))}{2} \times \\
&\qquad \left[\left(V_A h_A^{p-2}\right)^{\frac{1}{p}} + \left(V_B h_B^{p-2}\right)^{\frac{1}{p}}\right]^p \\
&= \frac{p!(var(A)+var(B))\,(h_A+h_B)^{p-2}}{2} \times \\
&\qquad \left[\left(V_A H_A^{p-2}\right)^{\frac{1}{p}} + \left(V_B H_B^{p-2}\right)^{\frac{1}{p}}\right]^p \\
&\leq \frac{p!var(A+B)\,(h_A+h_B)^{p-2}}{2} \times \\
&\qquad \left[\left(V_A H_A^{p-2}\right)^{\frac{1}{p}} + \left(V_B H_B^{p-2}\right)^{\frac{1}{p}}\right]^p,
\end{aligned}$$

where the last step follows from the assumption that $E[AB] \geq 0$, implying $var(A) + var(B) \leq var(A+B)$. Showing that

$$\left[\left(V_A H_A^{p-2}\right)^{\frac{1}{p}} + \left(V_B H_B^{p-2}\right)^{\frac{1}{p}}\right]^p \leq C^{p-2},$$

or

$$\left[\left(V_A H_A^{p-2}\right)^{\frac{1}{p}} + \left(V_B H_B^{p-2}\right)^{\frac{1}{p}}\right]^{\frac{p}{p-2}} \leq C,$$

where $C = 2^{1/(p-2)}$ will complete the proof. Since $V_B = 1 - V_A$ and $H_B = 1 - H_A$, the result follows by maximizing

$$\left[\left(V_A H_A^{p-2}\right)^{\frac{1}{p}} + \left((1-V_A)(1-H_A)^{p-2}\right)^{\frac{1}{p}}\right]^{\frac{p}{p-2}}$$

by choice of H_A, V_A, and p. The same values of H_A and V_A will maximize

$$\left[\left(V_A H_A^{p-2}\right)^{\frac{1}{p}} + \left((1-V_A)(1-H_A)^{p-2}\right)^{\frac{1}{p}}\right],$$

and simple calculus shows that, for $p \neq 1$, the maximum occurs when $H_A = V_A = 1/2$. Thus

$$\left[\left(V_A H_A^{p-2}\right)^{\frac{1}{p}} + \left(V_B H_B^{p-2}\right)^{\frac{1}{p}}\right]^{\frac{p}{p-2}} \leq 2^{\frac{1}{p-2}}.$$

Choosing $h_S = C(h_A + h_B) = 2^{1/(p-2)}(h_A + h_B)$ gives

$$E[|Z|^p] \leq \frac{p!var(Z)h_S^{p-2}}{2},$$

concluding the proof. ∎

Lemma 3 *Let* $Z = AB$ *be the product of* A *and* B*, two independent zero-mean random variables satisfying general moment conditions for a given integer* $p \geq 3$*. That is,*

$$E\left[|A|^p\right] \leq C_p^A var(A)h_A^{p-2}, \; E\left[|B|^p\right] \leq C_p^B var(B)h_B^{p-2}$$

for some positive constants h_A and h_B and positive numbers C_p^A and C_p^B possibly depending on p. Then

$$E\left[|Z|^p\right] \leq C_p^A C_p^B var(Z) h_P^{p-2}$$

where $h_P = h_A h_B$.

Proof: Because A and B are independent, we can write

$$E[|AB|^p] = E[|A|^p]E[|B|^p].$$

Substituting in the given bounds and noting that $var(A)var(B) = var(AB)$ by independence, we get

$$E[|Z|^p] \leq C_p^A C_p^B var(Z) h_P^{p-2}$$

where $h_P = h_A h_B$. ∎

B. Determination of the Bounding Constant h for Noisy Randomized Projection Encoding

Equipped with the previous lemmas, we are now ready to determine the bounding constant h for the Randomized Projection Encoding setup with binary basis elements $\{\phi_{i,j}\}$ taking values $\pm 1/\sqrt{n}$ with equal probability and additive Gaussian noise. For that we examine the moments of random variables of the form

$$U_j - E[U_j] = \frac{\|g\|^2}{n} - (\phi_j^T g)^2 - 2(\phi_j^T g)w_j,$$

where $g = f - f^*$. For ease of notation, let $\tilde{Z} = U_j - E[U_j]$. Since we are dealing with the moments in absolute value, it suffices to consider

$$-\tilde{Z} = (\phi_j^T g)^2 - \frac{\|g\|^2}{n} + 2(\phi_j^T g)w_j.$$

To further simplify notation we drop the subscripts, let $Z = -\tilde{Z}$, and consider random variables of the form

$$Z = (\phi^T g)^2 - \frac{\|g\|^2}{n} + 2(\phi^T g)\tilde{w}$$

where $\tilde{w}$ is a zero-mean Gaussian random variable independent of ϕ having variance σ^2. Let $Z_1 = (\phi^T g)^2 - \frac{\|g\|^2}{n}$, $Z_2 = (\phi^T g)$, and $Z_3 = 2\tilde{w}$, then $Z = Z_1 + Z_2 Z_3$.

Our procedure for determining the bounding constant for Z will be the following. First, we will find the bounding constants for the even moments of each of the components of Z (Z_1, Z_2, and Z_3). Then, we will apply earlier lemmas to determine the bounding constants for the even moments of $Z_2 Z_3$, and finally $Z = Z_1 + Z_2 Z_3$. The last step will be to extend this final bounding constant so that it is valid for all moments.

Instead of explicitly stating the bounding constants in terms of B and noise variance σ^2 here, we will derive relationships between the even moments of the $Z_i,\ i = 1, 2, 3$, and their respective variances. Later, we will use these relationships to obtain the explicit bounding constants that hold for all candidate functions f.

First, since Z_3 is Gaussian, we have

$$E[Z_3^{2k}] = (2k)!!(var(Z_3))^k$$

where $(2k)!! \equiv (1)(3)(5)\ldots(2k-1)$ for integers $k \geq 1$ and $0!! \equiv 1$. Now we define the multinomial coefficient as

$$\begin{aligned} \binom{N}{K_1 \ldots K_{n-1}} &\equiv \binom{N}{K_1}\binom{K_1}{K_2}\cdots\binom{K_{n-2}}{K_{n-1}} \\ &= \frac{N!}{K_1! K_2! \ldots K_{n-1}!(N-K_{n-1})!} \end{aligned}$$

so that for $n = 2$ this is just the binomial coefficient $\binom{N}{K_1}$. Using this, we can write the even powers of Z_2 as

$$\begin{aligned} Z_2^{2k} = (\phi^T g)^{2k} = \sum_{i_1=0}^{2k}\sum_{i_2=0}^{i_1}\cdots\sum_{i_{n-1}=0}^{i_{n-2}} \Bigg[& \binom{2k}{i_1 \ldots i_{n-1}} \times \\ & [\phi_1 g_1]^{2k-i_1}[\phi_2 g_2]^{i_1-i_2}\ldots[\phi_n g_n]^{i_{n-1}} \Bigg]. \end{aligned}$$

All ϕ_i raised to an odd power are zero in expectation while $E[\phi_i^m] = (1/\sqrt{n})^m$ if m is even. Using this fact and taking the expectation, we get

$$E[(\phi^T g)^{2k}] = \sum_{i_1=0}^{k} \sum_{i_2=0}^{i_1} \cdots \sum_{i_{n-1}=0}^{i_{n-2}} \Bigg[\binom{2k}{2i_1 \ldots 2i_{n-1}} \times n^{-k} g_1^{2k-2i_1} g_2^{2i_1-2i_2} \ldots g_n^{2i_{n-1}} \Bigg]. \quad (5)$$

We would like to "collapse" this sum back into a closed form expression, but for that we need another lemma.

Lemma 4 *The multinomial coefficient satisfies the following equality*

$$\binom{2k}{2i_1 \ldots 2i_{n-1}} = \binom{k}{i_1 \ldots i_{n-1}} \times \frac{(2k)!!}{(2i_1)!!(2i_2)!! \ldots (2i_{n-1})!!(2k-2i_{n-1})!!}.$$

Proof: For even integers $2j$ where j is an integer satisfying $j \geq 0$, we can write $(2j)!! = \frac{(2j)!}{j! 2^j}$. Substituting this into the above equation for all terms $(2j)!!$ we see that the right hand side becomes

$$\binom{k}{i_1 \ldots i_{n-1}} \frac{\binom{2k}{2i_1 \ldots 2i_{n-1}}}{\binom{k}{i_1 \ldots i_{n-1}}} = \binom{2k}{2i_1 \ldots 2i_{n-1}}.$$

■

Using Lemma 4 we can rewrite the sum in (5) as

$$E[(\phi^T g)^{2k}] = \sum_{i_1=0}^{k} \sum_{i_2=0}^{i_1} \cdots \sum_{i_{n-1}=0}^{i_{n-2}} \Bigg[\binom{k}{i_1 \ldots i_{n-1}} \times \frac{(2k)!! n^{-k}}{(2i_1)!! \ldots (2k-2i_{n-1})!!} g_1^{2k-2i_1} g_2^{2i_1-2i_2} \ldots g_n^{2i_{n-1}} \Bigg].$$

Now observe that the following bound holds

$$1 \leq \frac{(2k)!!}{(2i_1)!! \ldots (2k-2i_{n-1})!!} \leq (2k)!!,$$

so we obtain bounds on the even moments of $\phi^T g$,

$$\left(\frac{\|g\|^2}{n} \right)^k \leq E[(\phi^T g)^{2k}] \leq (2k)!! \left(\frac{\|g\|^2}{n} \right)^k.$$

Since $E[(\phi^T g)^2] = \big(\|g\|^2/n\big) = var(\phi^T g)$, we have

$$E[Z_2^{2k}] \leq (2k)!! \left(var(Z_2)\right)^k.$$

Finally, we need the bounding constant for $Z_1 = (\phi^T g)^2 - \|g\|^2/n$. We can write

$$Z_1 = \sum_{i=1}^{n} \sum_{j=1}^{n} g_i g_j (\phi_i \phi_j - E[\phi_i \phi_j]).$$

Notice that if $i = j$ that term in the sum is zero, and we have symmetry such that the $i = k, j = l$ term is equal to the $i = l, j = k$ term so we can write

$$Z_1 = \sum_{i<j}^{n} 2\phi_i \phi_j g_i g_j.$$

The random variable Z_1 is a Rademacher chaos of order 2. Bounds for the absolute moments of such random variables were derived in [25], from which the following result is obtained. We refer the reader to that work for the proof.

Lemma 5 (Bonami's Inequality) *Let F be a Rademacher chaos of order d. That is, F is of the form*

$$F = \sum_{i_1 < \cdots < i_d}^{n} a_{i_1 \ldots i_d} \phi_{i_1} \cdots \phi_{i_d}$$

where $d \geq 1$, $a_{i_1 \ldots i_d}$ are real or complex numbers, and $\{\phi_{i_j}\}$ are i.i.d. random variables taking the values ± 1 each with probability $1/2$. Then

$$E\left[|F|^p\right] \leq (p-1)^{pd/2} \left(E\left[F^2\right]\right)^{p/2}$$

holds for $p \geq 2$.

We apply this result by letting $a_{ij} = 2g_i g_j / n$ and $p = 2k$. Then

$$\begin{aligned} E\left[Z_1^{2k}\right] &\leq (2k-1)^{2k} \left(E\left[Z_1^2\right]\right)^k \\ &= (2k-1)^{2k} \left(var(Z_1)\right)^k . \end{aligned} \tag{6}$$

We will now collect the results derived above and determine bounding constants that hold for all candidate functions f. Our random variable of interest is $Z = Z_1 + Z_2 Z_3$ where $Z_1 = (\phi^T g)^2 - \|g\|^2/n$, $Z_2 = (\phi^T g)$, and $Z_3 = 2\tilde{w}$. The variance of Z_1 can be obtained from [25] or by direct calculation and is given in closed form by

$$\begin{aligned} var(Z_1) = E[Z_1^2] &= E[(\phi^T g)^4] - \left(\frac{\|g\|^2}{n}\right)^2 \\ &= \frac{4}{n^2} \sum_{i<j} (g_i g_j)^2 \equiv \beta_g^2 . \end{aligned}$$

To obtain the bounding constant for this term we simplify notation by letting $2k = p$, and upper bound the right-hand side of (6) by choice of h. That is,

$$(p-1)^p \beta_g^p \leq \frac{p!}{2} \beta_g^2 h^{p-2},$$

for which h must satisfy

$$h \geq \beta_g \max_{p \geq 3} \left\{ \left(\frac{2(p-1)^p}{p!} \right)^{\frac{1}{p-2}} \right\} .$$

It is straightforward to verify, using Stirling's Formula, that

$$\max_{p \geq 3} \left\{ \left(\frac{2(p-1)^p}{p!} \right)^{\frac{1}{p-2}} \right\} = e$$

and so $h = \beta_g e$, where e is the base of the natural logarithm.

To find a constant h_1 such that the moment condition is satisfied for Z_1 for any g, we solve

$$h_1 = \max_g \{\beta_g\} e = \max_g \left\{ \sqrt{\frac{4}{n^2} \sum_{i<j} (g_i g_j)^2} \right\} e$$

subject to the energy constraint $\|g\|^2 \leq 4nB^2$. A straightforward application of Lagrange Multipliers shows that for a fixed nonzero energy $\|g\|^2 = C$, the variance of Z_1 is maximized when $|g_1| = |g_2| = \cdots = |g_n|$. Thus we let $g_i^2 = 4B^2$ to obtain

$$\begin{aligned} h_1 &= \max_g \left\{ \sqrt{\frac{4}{n^2} \sum_{i<j} (g_i g_j)^2} \right\} e \\ &= \max \left\{ \sqrt{\frac{4}{n^2} \frac{n(n-1)}{2} 16B^4} \right\} e \\ &= 4\sqrt{2} B^2 e, \end{aligned}$$

and so

$$E[Z_1^{2k}] \leq \frac{(2k)!}{2} var(Z_1) h_1^{2k-2}.$$

The next term, Z_2, satisfies

$$\begin{aligned} E[Z_2^{2k}] &\leq (2k)!! var(Z_2) \max_g \left\{ \sqrt{\left(\frac{\|g\|^2}{n} \right)} \right\}^{2k-2} \\ &\leq (2k)!! var(Z_2)(2B)^{2k-2} = (2k)!! var(Z_2) h_2^{2k-2} \end{aligned}$$

where $h_2 = 2B$, and Z_3 satisfies

$$E[Z_3^{2k}] \leq (2k)!! var(Z_3) \, (2\sigma)^{2k-2} = (2k)!! var(Z_3) h_3^{2k-2}$$

where $h_3 = 2\sigma$.

Notice first that Z_2 and Z_3 are independent, zero-mean, and both satisfy moment conditions as shown above. Applying Lemma 3 gives

$$\begin{aligned} E[(Z_2 Z_3)^{2k}] &\leq ((2k)!!)^2 var(Z_2 Z_3)(4B\sigma)^{2k-2} \\ &\leq \frac{(2k)!}{2} var(Z_2 Z_3)(4B\sigma)^{2k-2} \end{aligned}$$

for all integers $k \geq 1$, where the last step follows by inspection:

$$\begin{aligned} \frac{(2k)!}{2} = (1)(1)(3)(4)\dots(2k-1)(2k) &\geq ((2k)!!)^2 \\ &= (1)(1)(3)(3)\dots(2k-1)(2k-1). \end{aligned}$$

Now since $E[Z_1(Z_2 Z_3)] = 0$ we can apply Lemma 2 to get

$$\begin{aligned} & E[(Z_1 + Z_2 Z_3)^{2k}] \\ &\leq \frac{(2k)!}{2} var(Z_1 + Z_2 Z_3)[\sqrt{2}(4\sqrt{2}B^2 e + 4B\sigma)]^{2k-2} \end{aligned}$$

for integers $k \geq 1$. Finally, to extend this result to all moments we use Lemma 1 to obtain

$$\begin{aligned} & E[|Z|^p] = E[|Z_1 + Z_2 Z_3|^p] \\ &\leq \frac{p!}{2} var(Z_1 + Z_2 Z_3)[2\sqrt{2}(4\sqrt{2}B^2 e + 4B\sigma)]^{2k-2} \end{aligned}$$

for all integers $p \geq 2$, so the constant $h = 8\sqrt{2}B(\sqrt{2}Be + \sigma) = 16B^2 e + 8\sqrt{2}B\sigma$.

C. Working with Projected Gaussian Noise

In this section, we prove Corollary 2. Suppose the signal is contaminated with additive zero-mean Gaussian noise, prior to projective sampling. In this case, the observations are

$$y_j = \sum_{i=1}^{n} \phi_{i,j} \left(f_i^* + \eta_i \right) + w_j, \quad j = 1, \dots, k$$

where the $\{\eta_i\}$ are i.i.d. zero-mean Gaussian random variables with variance σ_s^2 and assumed to be independent $\{\phi_{i,j}\}$ and $\{w_j\}$. With respect to the moment condition, the random variable of interest becomes

$$Z = (\phi^T g)^2 - \frac{\|g\|^2}{n} + 2(\phi^T g)(\phi^T \eta) + 2(\phi^T g) w,$$

where the subscripts j have been dropped for ease of notation. This is equivalent in distribution to

$$Z_{eq} = (\phi^T g)^2 - \frac{\|g\|^2}{n} + 2(\phi^T g)\tilde{w},$$

where $\tilde{w}$ is a zero-mean Gaussian random variable, independent of $\{\phi_i\}$, with variance $\sigma^2 + \sigma_s^2$, as shown below.

Lemma 6 *The random variables $(\phi^T g)$ and $(\phi^T \eta)$ are independent and thus $(\phi^T \eta)$ is independent of $\{\phi_i\}$.*

Proof: To prove independence, we will show that the joint characteristic function of $(\phi^T g)$ and $(\phi^T \eta)$ factorizes. The characteristic function is $M_{(\phi^T g),(\phi^T \eta)} = E[e^{(j\nu_1(\phi^T g)+j\nu_2(\phi^T \eta))}]$. Note that $E[e^{j\nu_2(\phi^T \eta)} \mid \phi] = e^{-\nu_2^2 \sigma_s^2/2}$, which does not depend on ϕ. The result follows from this observation. Furthermore, note that $e^{-\nu_2^2 \sigma_s^2/2}$ is the characteristic function of a zero-mean Gaussian random variable with variance σ_s^2, implying that $\phi^T \eta$ is Gaussian distributed.

■

This proof immediately gives the following corollary.

Corollary 3 *The random variables $(\phi^T g)(\phi^T \eta)$ and $(\phi^T g)\eta_i$ are equivalent in distribution.*

By the above results, we see that our random variable of interest

$$Z = (\phi^T g)^2 - \frac{\|g\|^2}{n} + 2(\phi^T g)(\phi^T \eta) + 2(\phi^T g)w$$

is equivalent in distribution to

$$Z_{eq} = (\phi^T g)^2 - \frac{\|g\|^2}{n} + 2(\phi^T g)(\eta_i + w).$$

Let $\tilde{w} = (\eta_i + w)$ and notice that $\tilde{w}$ is Gaussian with mean zero and variance $\sigma^2 + \sigma_s^2$ and independent of $\{\phi_i\}$. Then

$$Z = (\phi^T g)^2 - \frac{\|g\|^2}{n} + 2(\phi^T g)(\phi^T \eta) + 2(\phi^T g)w$$

is equivalent in distribution to

$$Z_{eq} = (\phi^T g)^2 - \frac{\|g\|^2}{n} + 2(\phi^T g)\tilde{w},$$

and thus Theorems 1 and 2 and Corollary 1 apply in this situation, as well.

References

[1] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[2] E. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?" IEEE Trans. Inform. Theory, 2004, submitted for publication.

[3] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[4] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," in *Proc. IEEE int. Conf. Acoust. Speech and Sig. Proc.*, vol. 3, Atlanta, GA, May 1996, pp. 1689–1692.

[5] M. Gastpar and Y. Bresler, "On the necessary density for spectrum-blind nonuniform sampling subject to quantization," in *Proc. IEEE int. Conf. Acoust. Speech and Sig. Proc.*, vol. 1, Istanbul, June 2000, pp. 348–351.

[6] A. C. Gilbert, S. Guha, P. Indyk, S. Muthukrishnan, and M. J. Strauss, "Near-optimal sparse fourier representations via sampling," in *Proc. of the 2002 ACM Symposium on Theory of Computing (STOC)*, 2002, pp. 152–161.

[7] J. Zou, A. Gilbert, M. Strauss, and I. Daubechies, "Theoretical and experimental analysis of a randomized algorithm for sparse fourier transform analysis," *Journal of Computational Physics*, vol. 211, no. 2, pp. 572–595, Jan. 2006.

[8] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Processing*, vol. 50, no. 6, pp. 1417–1428, June 2002.

[9] I. Maravic and M. Vetterli, "Sampling and reconstruction of signals with finite rate of innovation in the presence of noise," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2788–2805, Aug. 2005.

[10] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[11] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," Communications on Pure and Applied Mathematics, 2005, submitted for publication.

[12] E. Candès and T. Tao, "The dantzig selector: statistical estimation when p is much larger than n," Annals of Statistics, 2005, submitted for publication.

[13] A. R. Barron, "Complexity regularization with application to artificial neural networks," in *Nonparametric Functional Estimation and Related Topics*. Kluwer Academic Publishers, 1991, pp. 561–576.

[14] C. Craig, "On the Tchebychef inequality of Bernstein," *The Annals of Mathematical Statistics*, vol. 4, no. 2, pp. 94–102, May 1933.

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[16] M. Figueiredo and R. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Processing*, vol. 12, no. 8, pp. 906–916, Aug. 2003.

[17] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004.

[18] M. Figueiredo and R. Nowak, "A bound optimization approach to wavelet-based image deconvolution," in *Proc. Intl. Conf. Image Processing*, vol. 2, Sept. 2005, pp. 782–785.

[19] D. L. Donoho, "For most large underdetermined systems of linear equations, the minimal l1-norm near-solution approximates the sparsest near-solution," *Comm. Pure Appl. Math*, vol. 59, no. 7, pp. 907–934, July 2006.

[20] M. Gastpar and M.Vetterli, "Source-channel communication in sensor networks," in *Proc. Inform. Processing in Sensor Networks*, Apr. 2003, pp. 162–177.

[21] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.

[22] W. Bajwa, A. Sayeed, and R. Nowak, "Matched source-channel communication for field estimation in wireless sensor networks," in *Proc. Inform. Processing in Sensor Networks*, Apr. 2005, pp. 332–339.

[23] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive wireless sensing," in *Proc. Inform. Processing in Sensor Networks*, Nashville, TN, Apr. 2006.

[24] M. Rabbat, J. Haupt, A. Singh, and R. Nowak, "Decentralized compression and predistribution via randomized gossip," in *Proc. Inform. Processing in Sensor Networks*, Nashville, TN, Apr. 2006.

[25] A. Bonami, "Étude des coefficients de Fourier des fonctions de $L^p(g)$," *Annales de l'institut Fourier*, vol. 20, no. 2, pp. 335–402, 1970.