Sobolev Duals of Random Frames

C. Sinan Güntürk,[1] Mark Lammers,[2] Alex Powell,[3] Rayan Saab,[4] Özgür Yılmaz [4]
[1]Courant Institute of Mathematical Sciences, New York University, NY, USA.
[2]University of North Carolina, Wilmington, NC, USA.
[3]Vanderbilt University, Nashville, TN, USA.
[4]University of British Columbia, Vancouver BC, Canada.

***Abstract*—Sobolev dual frames have recently been proposed as optimal alternative reconstruction operators that are specifically tailored for Sigma-Delta ($\Sigma\Delta$) quantization of frame coefficients. While the canonical dual frame of a given analysis (sampling) frame is optimal for the white-noise type quantization error of Pulse Code Modulation (PCM), the Sobolev dual offers significant reduction of the reconstruction error for the colored-noise of $\Sigma\Delta$ quantization. However, initial quantitative results concerning the use of Sobolev dual frames required certain regularity assumptions on the given analysis frame in order to deduce improvements of performance on reconstruction that are similar to those achieved in the standard setting of bandlimited functions. In this paper, we show that these regularity assumptions can be lifted for (Gaussian) random frames with high probability on the choice of the analysis frame. Our results are immediately applicable in the traditional oversampled (coarse) quantization scenario, but also extend to compressive sampling of sparse signals.**

I. Introduction

A. Background on $\Sigma\Delta$ quantization

Methods of quantization have long been studied for oversampled data conversion. Sigma-Delta ($\Sigma\Delta$) quantization has been one of the dominant methods in this setting, such as A/D conversion of audio signals [9], and has received significant interest from the mathematical community in recent years [5], [6]. Oversampling is typically exploited to employ very coarse quantization (e.g., 1 bit/sample), however, the working principle of $\Sigma\Delta$ quantization is applicable to any quantization alphabet. In fact, it is more natural to consider $\Sigma\Delta$ quantization as a "noise shaping" method, for it seeks a quantized signal (q_j) by a recursive procedure to push the quantization error signal $y - q$ towards an unoccupied portion of the signal spectrum. In the case of bandlimited signals, this spectrum corresponds to high-frequency bands.

The governing equation of a standard rth order $\Sigma\Delta$ quantization scheme with input $y = (y_j)$ and output $q = (q_j)$ is

$$(\Delta^r u)_j = y_j - q_j, \quad j = 1, 2, \dots, \tag{1}$$

where Δ stands for the difference operator defined by $(\Delta u)_j := u_j - u_{j-1}$, and the $q_j \in \mathcal{A}$ are chosen according to some *quantization rule* that incorporates quantities available in the recursion; the quantization rule typically takes the form

$$q_j = Q_{\mathcal{A}}(u_{j-1}, u_{j-2}, \dots, y_j, y_{j-1}, \dots). \tag{2}$$

$\mathcal{A}$ is called the *quantization alphabet* and is typically a finite arithmetic progression of a given spacing δ, and symmetric about the origin.

Not all $\Sigma\Delta$ quantization schemes are presented (or implemented) in the above canonical form, but they all can be rewritten as such for an appropriate choice of r and u. For the noise shaping principle to work, it is crucial that the solution u remains bounded. The smaller the size of the alphabet $\mathcal{A}$ gets relative to r, the harder it is to guarantee this property. The extreme case is 1-bit quantization, i.e., $|\mathcal{A}| = 2$, which is also the most challenging setting. In this paper, we will assume that $Q_{\mathcal{A}}$ is such that the sequence u remains uniformly bounded for all inputs y such that $\|y\|_\infty \le \mu$ for some $\mu > 0$, i.e., that there exists a constant $C(\mu, r, Q_{\mathcal{A}}) < \infty$ such that

$$|u_j| \le C(\mu, r, Q_{\mathcal{A}}), \quad \forall j. \tag{3}$$

In this case, we say that the $\Sigma\Delta$ scheme or the quantization rule is *stable*.

A standard quantization rule is the so called "greedy rule" which minimizes $|u_j|$ given $u_{j-1}, \dots, u_{j-r}$ and y_j, i.e.,

$$q_j = \arg\min_{a \in \mathcal{A}} \Big| \sum_{i=1}^{r} (-1)^{i-1} \binom{r}{i} u_{j-i} + y_j - a \Big|. \tag{4}$$

The greedy rule is not always stable for large values of r. However, it is stable if $\mathcal{A}$ is sufficiently big. It can be checked that if

$$(2^r - 1)\delta/2 + \mu \le |\mathcal{A}|\delta/2, \tag{5}$$

then $\|u\|_\infty \le \delta/2$. However, this requires that $\mathcal{A}$ contain at least 2^r levels. With more stringent, fixed size quantization alphabets, the best constant $C(\mu, r, Q_{\mathcal{A}})$ in (3) has to be significantly larger than this bound. In fact, it is known that for any quantization rule with a 1-bit alphabet, $C(\mu, r, Q_{\mathcal{A}})$ is $\Omega(r^r)$, e.g., see [5], [6].

The significance of stability in a $\Sigma\Delta$ quantization scheme has to do with the reconstruction error analysis. In standard oversampled quantization, it is shown that the reconstruction error incurred after low-pass filtering of the quantized output is mainly controlled by $\|u\|_\infty \lambda^{-r}$, where λ is the *oversampling ratio*, i.e., the ratio of the sampling frequency to the bandwidth of the reconstruction filter [5]. This error bound relies on the specific structure of the space of bandlimited functions and the associated sampling theorem.

B. $\Sigma\Delta$ quantization and finite frames

Various authors have explored analogies of $\Sigma\Delta$ quantization and the above error bound in the setting of finite frames, especially tight frames [1], [2], [4]. Recall that in a finite dimensional inner product space, a frame is simply a spanning set of (typically not linearly independent) vectors. In this setting, if a frame of consisting of m vectors is employed in a space of k dimensions, then a corresponding oversampling ratio can be defined as $\lambda = m/k$. If a $\Sigma\Delta$ quantization algorithm is used to quantize frame coefficients, then it would be desirable to obtain reconstruction error bounds that have the same nature as in the specific infinite dimensional setting of bandlimited functions.

In finite dimensions, the equations of sampling, $\Sigma\Delta$ quantization, and reconstruction can all be phrased using matrix equations, which we shall describe next. For the simplicity of our analysis, it will be convenient for us to set the initial condition of the recursion in (1) equal to zero. With $u_{-r+1} = \cdots = u_0 = 0$, and $j = 1, \ldots, m$, the difference equation (1) can be rewritten as a matrix equation

$$D^r u = y - q, \tag{6}$$

where D is defined by

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ & & \ddots & \ddots & \ddots & \\ 0 & 0 & \cdots & -1 & 1 & 0 \\ 0 & 0 & \cdots & 0 & -1 & 1 \end{bmatrix} \tag{7}$$

Suppose that we are given an input signal x and an analysis frame $(e_i)_1^m$ of size m in $\mathbb{R}^k$. We can represent the frame vectors as the rows of an an $m \times k$ matrix E, the sampling operator. The sequence y will simply be the frame coefficients, i.e., $y = Ex$. Similarly, let us consider a corresponding synthesis frame $(f_j)_1^m$. We stack these frame vectors along the columns of a $k \times m$ matrix F, the reconstruction operator, which is then a left inverse of E, i.e., $FE = I$. The $\Sigma\Delta$ quantization algorithm will replace the coefficient sequence y with its quantization given by q, which will then represent the coefficients of an approximate reconstruction $\hat{x}$ using the synthesis frame. Hence, $\hat{x} = Fq$. Since $x = Fy$, the reconstruction error is given by

$$x - \hat{x} = FD^r u. \tag{8}$$

With this expression, $\|x - \hat{x}\|$ can be bounded for any norm $\|\cdot\|$ simply as

$$\|x - \hat{x}\| \le \|u\|_\infty \sum_{j=1}^{m} \|(FD^r)_j\|. \tag{9}$$

Here $(FD^r)_j$ is the jth column of FD^r. In fact, when suitably stated, this bound is also valid in infinite dimensions, and has been used extensively in the mathematical treatment of oversampled A/D conversion of bandlimited functions.

For $r = 1$, and $\|\cdot\| = \|\cdot\|_2$, the summation term on the right hand side of (9) motivated the study of the so-called *frame variation* defined by

$$V(F) := \sum_{j=1}^{m} \|f_j - f_{j+1}\|_2, \tag{10}$$

where one defines $f_{m+1} = 0$. Higher-order frame variations to be used with higher-order $\Sigma\Delta$ schemes are defined similarly, see [1], [2]. It is clear that for the frame variation to be small, the frame vectors must follow a smooth path in $\mathbb{R}^k$. Frames (analysis as well as synthesis) that are obtained via uniform sampling a smooth curve in $\mathbb{R}^k$ (so-called *frame path*) are typical in many settings. However, the above "frame variation bound" is useful in finite dimensions when the frame path terminates smoothly. Otherwise, it does not necessarily provide higher-order reconstruction accuracy (i.e., of the type λ^{-r}) due to the presence of boundary terms. On the other hand, designing smoothly terminating frames can be technically challenging, e.g., [4].

C. Sobolev duals

Recently, a more straightforward approach was proposed for the design of (alternate) duals of finite frames for $\Sigma\Delta$ quantization [8], [3]. Here, one instead considers the operator norm of FD^r on ℓ_2^m and the corresponding bound

$$\|x - \hat{x}\|_2 \le \|FD^r\|_{\text{op}} \|u\|_2. \tag{11}$$

(Note that this bound is not available in the infinite dimensional setting of bandlimited functions due to the fact that u is typically not in $\ell_2(\mathbb{N})$.) With this bound, it is now natural to minimize $\|FD^r\|_{\text{op}}$ over all dual frames of a given analysis frame E. These frames have been called *Sobolev duals*, in analogy with classical L_2-type Sobolev (semi)norms.

$\Sigma\Delta$ quantization algorithms are normally designed for analog circuit operation, so they control $\|u\|_\infty$, which would control $\|u\|_2$ only in a suboptimal way. However, it turns out that there are important advantages in working with the ℓ_2 norm in the analysis. The first advantage is that Sobolev duals are readily available by an explicit formula. The solution $F_{\text{sob},r}$ of the optimization problem

$$\min_F \|FD^r\|_{\text{op}} \text{ subject to } FE = I \tag{12}$$

is given by the matrix equation

$$F_{\text{sob},r} D^r = (D^{-r}E)^\dagger, \tag{13}$$

where † stands for the Moore-Penrose inversion operator, which, in our case, is given by $E^\dagger := (E^*E)^{-1}E^*$. Note that for $r = 0$ (i.e., no noise-shaping, or PCM), one simply obtains $F = E^\dagger$, the canonical dual frame of E.

As for the reconstruction error bound, plugging (13) into (11), it immediately follows that

$$\|x - \hat{x}\|_2 \le \|(D^{-r}E)^\dagger\|_{\text{op}} \|u\|_2 = \frac{1}{\sigma_{\min}(D^{-r}E)} \|u\|_2, \tag{14}$$

where $\sigma_{\min}(D^{-r}E)$ stands for the smallest singular value of $D^{-r}E$.

D. Main result

The main advantage of the Sobolev dual approach is that highly developed methods are present for spectral norms of matrices, especially in the random setting. Minimum singular values of random matrices with i.i.d. entries have been studied extensively in the mathematical literature. For an $m \times k$ random matrix E with i.i.d. entries sampled from a sub-Gaussian distribution with zero mean and unit variance one has

$$\sigma_{\min}(E) \geq \sqrt{m} - \sqrt{k} \tag{15}$$

with high probability [11]. However, note that $D^{-r}E$ would not have i.i.d. entries. A naive approach would be to split; $\sigma_{\min}(D^{-r}E)$ is bounded from below by $\sigma_{\min}(D^{-r})\sigma_{\min}(E)$. However (see Lemma II.1), $\sigma_{\min}(D^{-r})$ satisfies

$$\sigma_{\min}(D^{-r}) \asymp_r 1, \tag{16}$$

and therefore this naive product bound yields no improvement in r on the reconstruction error for $\Sigma\Delta$-quantized measurements. As can be expected, the true behavior of $\sigma_{\min}(D^{-r}E)$ turns out to be drastically different, and is described as part of Theorem A, our main result in this paper.

Theorem A. *Let E be an $m \times k$ random matrix whose entries are i.i.d. Gaussian random variables with zero mean and unit variance. For any $\alpha \in (0,1)$ and positive integer r, there are constants $c, c', c'' > 0$ such that if $\lambda := m/k \geq c(\log m)^{1/(1-\alpha)}$, then with probability at least $1 - \exp(-c'm\lambda^{-\alpha})$, the smallest singular value of $D^{-r}E$ satisfies*

$$\sigma_{\min}(D^{-r}E) \geq c''\lambda^{\alpha(r-\frac{1}{2})}\sqrt{m}. \tag{17}$$

In this event, if a stable rth order $\Sigma\Delta$ quantization scheme is used in connection with the rth order Sobolev dual frame of E, then the resulting reconstruction error is bounded by

$$\|x - \hat{x}\|_2 \lesssim \lambda^{-\alpha(r-\frac{1}{2})}. \tag{18}$$

Previously, the only setting in which this type of approximation accuracy could be achieved (with or without Sobolev duals) was the case of highly structured frames (e.g. when the frame vectors are found by sampling along a piecewise smooth frame path). Theorem A shows that such accuracy is generically available, provided the reconstruction is done via Sobolev duals.

II. Sketch of Proof of Theorem A

Below we give a sketch of the proof of Theorem A as well as describe the main ideas in our approach. The full argument is given in [7].

In what follows, $\sigma_j(A)$ will denote the jth largest singular value of the matrix A. Similarly, $\lambda_j(B)$ will denote the jth largest eigenvalue of the Hermitian matrix B. Hence, we have $\sigma_j(A) = \sqrt{\lambda_j(A^*A)}$. We will also use the notation $\Sigma(A)$ for the diagonal matrix of singular values of A, with the convention $(\Sigma(A))_{jj} = \sigma_j(A)$. All matrices in our discussion will be real valued and the Hermitian conjugate reduces to the transpose.

We have seen that the main object of interest for the reconstruction error bound is $\sigma_{\min}(D^{-r}E)$ for a random frame E. Let H be a square matrix. The first observation we make is that when E is i.i.d. Gaussian, the distribution of $\Sigma(HE)$ is the same as the distribution of $\Sigma(\Sigma(H)E)$. To see this, let $U\Sigma(H)V^*$ be the singular value decomposition of H where U and V are unitary matrices. Then $HE = U\Sigma(H)V^*E$. Since the unitary transformation U does not alter singular values, we have $\Sigma(HE) = \Sigma(\Sigma(H)V^*E)$, and because of the unitary invariance of the i.i.d. Gaussian measure, the matrix $\tilde{E} := V^*E$ has the same distribution as E, hence the claim. Therefore it suffices to study the singular values of $\Sigma(H)E$. In our case, $H = D^{-r}$ and we first need information on the deterministic object $\Sigma(D^{-r})$.

A. Singular values of D^{-r}

It turns out that we only need an approximate estimate of the singular values of D^{-r}:

Lemma II.1. *Let r be any positive integer and D be as in (7). There are positive numerical constants $c_1(r)$ and $c_2(r)$, independent of m, such that*

$$c_1(r)\left(\frac{m}{j}\right)^r \leq \sigma_j(D^{-r}) \leq c_2(r)\left(\frac{m}{j}\right)^r, \quad j = 1, \ldots, m. \tag{19}$$

Here we shall only give a simple heuristic argument. First, it is convenient to work with the singular values of D^r instead, because D^r is a banded (Toeplitz) matrix whereas D^{-r} is full. Note that because of our convention of descending ordering of singular values, we have

$$\sigma_j(D^{-r}) = \frac{1}{\sigma_{m+1-j}(D^r)}, \quad j = 1, \ldots, m. \tag{20}$$

For $r = 1$, an explicit formula is available [12]. Indeed, we have

$$\sigma_j(D) = 2\cos\left(\frac{\pi j}{2m+1}\right), \quad j = 1, \ldots, m, \tag{21}$$

which implies

$$\sigma_j(D^{-1}) = \frac{1}{2\sin\left(\frac{\pi(j-1/2)}{2(m+1/2)}\right)}, \quad j = 1, \ldots, m. \tag{22}$$

For $r > 1$, the first observation is that $\sigma_j(D^r)$ and $(\sigma_j(D))^r$ are different, because D and D^* do not commute. However, this becomes insignificant as the size $m \to \infty$. In fact, the asymptotic distribution of $(\sigma_j(D^r))_{j=1}^m$ as $m \to \infty$ is rather easy to find using standard results in the theory of Toeplitz matrices: D is a banded Toeplitz matrix whose symbol is $f(\theta) = 1 - e^{i\theta}$, hence the symbol of D^r is $(1 - e^{i\theta})^r$. It then follows by Parter's extension of Szegö's theorem [10] that for any continuous function ψ, we have

$$\lim_{m\to\infty} \frac{1}{m}\sum_{j=1}^{m} \psi(\sigma_j(D^r)) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \psi(|f(\theta)|^r)\, d\theta. \tag{23}$$

We have $|f(\theta)| = 2\sin|\theta|/2$ for $|\theta| \leq \pi$, hence the distribution of $(\sigma_j(D^r))_{j=1}^m$ is asymptotically the same as that of $2^r \sin^r(\pi j/2m)$, and consequently, we can think of $\sigma_j(D^{-r})$ roughly as $\left(2^r \sin^r(\pi j/2m)\right)^{-1}$ which behaves, up to r-dependent constants, as $(m/j)^r$. Moreover, we know that $\|D^r\|_{\text{op}} \leq \|D\|_{\text{op}}^r \leq 2^r$, hence $\sigma_{\min}(D^{-r}) \geq 2^{-r}$.

In [7], the above heuristic is turned into a formal statement via a standard perturbation theorem (namely, Weyl's theorem) on the spectrum of Hermitian matrices. Here, the important observation is that $D^{*r}D^r$ is a perturbation of $(D^*D)^r$ of rank at most $2r$.

B. Lower bound for $\sigma_{\min}(D^{-r}E)$

In light of the above discussion, the distribution of $\sigma_{\min}(D^{-r}E)$ is the same as that of

$$\inf_{\|x\|_2=1} \|\Sigma(D^{-r})Ex\|_2. \tag{24}$$

We replace $\Sigma(D^{-r})$ with an arbitrary diagonal matrix S with $S_{jj} =: s_j > 0$. For the next set of results, we will work with Gaussian variables with variance equal to $1/m$.

Lemma II.2. *Let E be an $m \times k$ random matrix whose entries are i.i.d. $\mathcal{N}(0, \frac{1}{m})$. For any $\Theta > 1$, consider the event*

$$\mathcal{E} := \left\{ \|SE\|_{\ell_2^k \to \ell_2^m} \leq 2\sqrt{\Theta}\|s\|_\infty \right\}.$$

Then

$$\mathbb{P}(\mathcal{E}^c) \leq 5^k \Theta^{m/2} e^{-(\Theta-1)m/2}.$$

The above lemma follows easily from the bound $\|S\| \cdot \|E\|$ on $\|SE\|$ and the corresponding standard concentration estimates for $\|E\|$. Likewise, the proof of the next lemma also follows from standard methods.

Lemma II.3. *Let $\xi \sim \mathcal{N}(0, \frac{1}{m}\mathrm{I}_m)$, r be a positive integer, and $c_1 > 0$ be such that*

$$s_j \geq c_1 \left(\frac{m}{j}\right)^r, \qquad j = 1, \ldots, m. \tag{25}$$

Then for any $\Lambda \geq 1$ and $m \geq \Lambda$,

$$\mathbb{P}\left(\sum_{j=1}^m s_j^2 \xi_j^2 < c_1^2 \Lambda^{2r-1}\right) < (60m/\Lambda)^{r/2} e^{-m(r-1/2)/\Lambda}.$$

The next lemma is what we will need to prove the main result.

Lemma II.4. *Let E be an $m \times k$ random matrix whose entries are i.i.d. $\mathcal{N}(0, \frac{1}{m})$, r be a positive integer, and assume that the entries s_j of the diagonal matrix S satisfy*

$$c_1 \left(\frac{m}{j}\right)^r \leq s_j \leq c_2 m^r, \qquad j = 1, \ldots, m. \tag{26}$$

Let $\Lambda \geq 1$ be any number and assume $m \geq \Lambda$. Consider the event

$$\mathcal{F} := \left\{ \|SEx\|_2 \geq \frac{1}{2} c_1 \Lambda^{r-1/2} \|x\|_2, \ \forall x \in \mathbb{R}^k \right\}.$$

Then

$$\mathbb{P}(\mathcal{F}^c) \leq 5^k e^{-m/2} + 8^r \left(\frac{17c_2\sqrt{\Lambda}}{c_1}\right)^k \left(\frac{m}{\Lambda}\right)^{r(k+\frac{1}{2})} e^{-m(r-\frac{1}{2})\Lambda}.$$

Proof: Consider a ρ-net $\tilde{Q}$ of the unit sphere of $\mathbb{R}^k$ with $\#\tilde{Q} \leq \left(\frac{2}{\rho}+1\right)^k$ where the value of $\rho < 1$ will be chosen later. Let $\tilde{\mathcal{E}}(\tilde{Q})$ be the event $\left\{\|SEq\|_2 \geq c_1 \Lambda^{r-1/2}, \quad \forall q \in \tilde{Q}\right\}$. By Lemma II.3, we know that

$$\mathbb{P}\left(\tilde{\mathcal{E}}(\tilde{Q})^c\right) \leq \left(\frac{2}{\rho}+1\right)^k \left(\frac{60m}{\Lambda}\right)^{r/2} e^{-m(r-1/2)/\Lambda}. \tag{27}$$

Let $\mathcal{E}$ be the event in Lemma II.2 with $\Theta = 4$. Let E be any given matrix in the event $\mathcal{E} \cap \tilde{\mathcal{E}}(\tilde{Q})$. For each $\|x\|_2 = 1$, there is $q \in \tilde{Q}$ with $\|q - x\|_2 \leq \rho$, hence by Lemma II.2, we have

$$\|SE(x-q)\|_2 \leq 4\|s\|_\infty \|x-q\|_2 \leq 4c_2 m^r \rho.$$

Choose

$$\rho = \frac{c_1 \Lambda^{r-1/2}}{8c_2 m^r} = \frac{c_1}{8c_2\sqrt{\Lambda}} \left(\frac{\Lambda}{m}\right)^r.$$

Hence

$$\begin{aligned} \|SEx\|_2 &\geq \|SEq\|_2 - \|SE(x-q)\|_2 \\ &\geq c_1 \Lambda^{r-1/2} - 4c_2 m^r \rho \\ &= \frac{1}{2} c_1 \Lambda^{r-1/2}. \end{aligned}$$

This shows that $\mathcal{E} \cap \tilde{\mathcal{E}}(\tilde{Q}) \subset \mathcal{F}$. Clearly, $\rho \leq 1/8$ by our choice of parameters and hence $\frac{2}{\rho} + 1 \leq \frac{17}{8\rho}$. The result follows by using the probability bounds of Lemma II.2 and (27). ∎

The following theorem is now a direct corollary of the above estimate.

Theorem II.5. *Let E be an $m \times k$ random matrix whose entries are i.i.d. $\mathcal{N}(0, \frac{1}{m})$, r be a positive integer, D be the difference matrix defined in (7), and the constant $c_1 = c_1(r)$ be as in Lemma II.1. Let $0 < \alpha < 1$ be any number. Assume that*

$$\lambda := \frac{m}{k} \geq c_3 (\log m)^{1/(1-\alpha)}, \tag{28}$$

where $c_3 = c_3(r)$ is an appropriate constant. Then

$$\mathbb{P}\left(\sigma_{\min}(D^{-r}E) \geq c_1 \lambda^{\alpha(r-1/2)}\right) \geq 1 - 2e^{-c_4 m^{1-\alpha} k^\alpha} \tag{29}$$

for some constant $c_4 = c_4(r) > 0$.

Proof: Set $\Lambda = \lambda^\alpha$ in Lemma II.4. We only need to show that

$$\begin{aligned} &\max\left[5^k e^{-m/2}, 8^r \left(\frac{17c_2\sqrt{\Lambda}}{c_1}\right)^k \left(\frac{m}{\Lambda}\right)^{r(k+\frac{1}{2})} e^{-m(r-\frac{1}{2})/\Lambda}\right] \\ &\leq e^{-c_4 m^{1-\alpha} k^\alpha}. \end{aligned}$$

This condition is easily verified once we notice that $m/\Lambda = m^{1-\alpha}k^\alpha$ and $k \log m \leq c_5 m/\Lambda$ for a sufficiently small c_5 which follows from our assumption (28) on λ by setting $c_5 = 1/c_3^{1-\alpha}$. ∎

Remark. Replacing E in Theorem II.5 with $\sqrt{m}E$, we obtain Theorem A. Also, a closer inspection of the proof shows that if an $m \times N$ Gaussian matrix Φ is given where $k < m < N$ and

$$\lambda := \frac{m}{k} \geq c_6 (\log N)^{1/(1-\alpha)}, \tag{30}$$

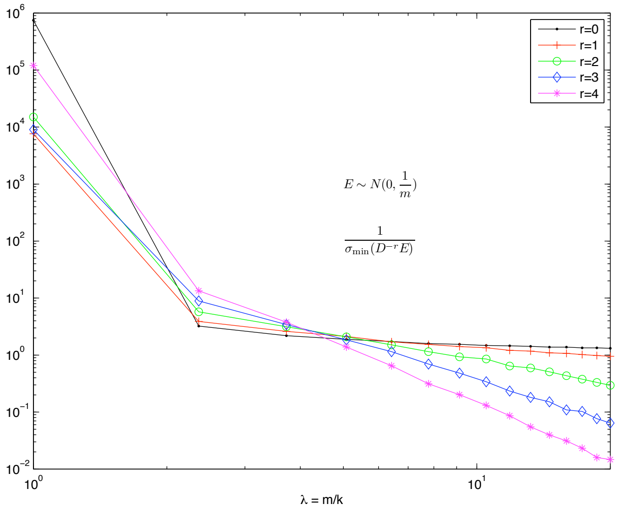

Fig. 1. Numerical behavior (in log-log scale) of $1/\sigma_{\min}(D^{-r}E)$ as a function of $\lambda = m/k$, for $r = 0, 1, 2, 3, 4$. In this figure, $k = 50$ and $1 \leq \lambda \leq 20$. For each problem size, the largest value of $1/\sigma_{\min}(D^{-r}E)$ among 50 realizations of a random $m \times k$ matrix E sampled from the Gaussian ensemble $\mathcal{N}(0, \frac{1}{m} I_m)$ was recorded.

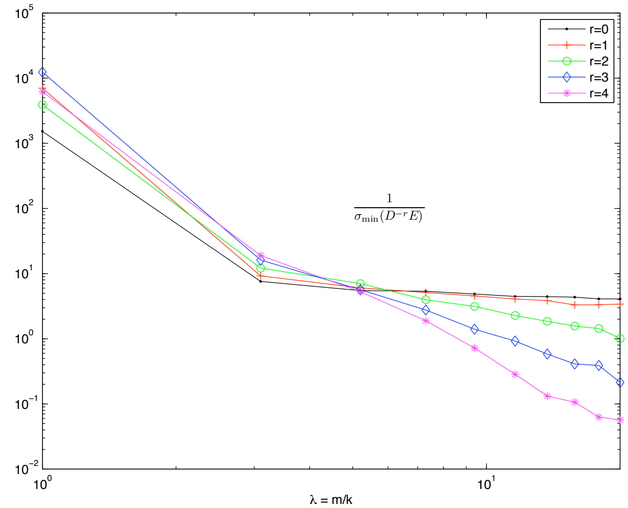

Fig. 2. Numerical behavior (in log-log scale) of $1/\sigma_{\min}(D^{-r}E)$ as a function of $\lambda = m/k$, for $r = 0, 1, 2, 3, 4$. In this figure, $k = 10$ and $1 \leq \lambda \leq 20$. For each r and problem size (m, k), a random $m \times k$ submatrix E of the $N \times N$ Discrete Cosine Transform matrix, where $N = 10m$, was selected. After 50 such random selections, the largest value of $1/\sigma_{\min}(D^{-r}E)$ was recorded.

then the same probabilistic estimate on $\sigma_{\min}(D^{-r}E)$ holds for every $m \times k$ submatrix E of Φ. This extension has powerful implications on compressed sensing. The details are given in [7].

III. Numerical experiments for Gaussian and other random ensembles

In order to test the accuracy of Theorem II.5, our first numerical experiment concerns the minimum singular value of $D^{-r}E$ as a function of $\lambda = m/k$. In Figure 1, we plot the worst case (the largest) value, among 50 realizations, of $1/\sigma_{\min}(D^{-r}E)$ for the range $1 \leq \lambda \leq 20$, where we have kept $k = 50$ constant. As predicted by this theorem, we find that the negative slope in the log-log scale is roughly equal to $r - 1/2$, albeit slightly less, which seems in agreement with the presence of our control parameter α. As for the size of the r-dependent constants, the function $5^r \lambda^{-r+1/2}$ seems to be a reasonably close numerical fit, which also explains why we observe the separation of the individual curves after $\lambda > 5$.

It is natural to consider other random matrix ensembles when constructing random frames and their Sobolev duals. In the setting of compressed sensing, the restricted isometry property has been extended to measurement matrices sampled from the Bernoulli and general sub-Gaussian distributions. Structured random matrices, such as random Fourier samplers, have also been considered now in some detail. Our numerical findings for the Bernoulli case are almost exactly the same as those for the Gaussian case, hence we omit. For structured random matrices, we give the results of one of our numerical experiments in Figure 2. Here E is a randomly chosen $m \times k$ submatrix of a much larger $N \times N$ Discrete Cosine Transform matrix. We find the results somewhat similar, but as expected, perhaps less reliable than the Gaussian ensemble.

Acknowledgment

The authors would like to thank Ronald DeVore for valuable discussions, and the American Institute of Mathematics and the Banff International Research Station for hosting two meetings where this work was initiated.

References

[1] J.J. Benedetto, A.M. Powell, and Ö. Yılmaz. Sigma-delta ($\Sigma\Delta$) quantization and finite frames. *IEEE Transactions on Information Theory*, 52(5):1990–2005, May 2006.

[2] J.J. Benedetto, A.M. Powell, and Ö. Yılmaz. Second order sigma–delta ($\Sigma\Delta$) quantization of finite frame expansions. *Appl. Comput. Harmon. Anal*, 20:126–148, 2006.

[3] J. Blum, M. Lammers, A.M. Powell, and Ö. Yılmaz. Sobolev duals in frame theory and Sigma-Delta quantization. *Journal of Fourier Analysis and Applications*. Published online in October 2009.

[4] B.G. Bodmann, V.I. Paulsen, and S.A. Abdulbaki. Smooth Frame-Path Termination for Higher Order Sigma-Delta Quantization. 13(3):285–307, 2007.

[5] I. Daubechies and R. DeVore. Approximating a bandlimited function using very coarsely quantized data: A family of stable sigma-delta modulators of arbitrary order. *Ann. of Math.*, 158(2):679–710, September 2003.

[6] C.S. Güntürk. One-bit sigma-delta quantization with exponential accuracy. *Comm. Pure Appl. Math.*, 56(11):1608–1630, 2003.

[7] C.S. Güntürk, A. Powell, R. Saab, and Ö. Yılmaz, Sobolev Duals for Random Frames and Sigma-Delta Quantization of Compressed Sensing Measurements. arXiv:1002.0182.

[8] M. Lammers, A.M. Powell, and Ö. Yılmaz. Alternative dual frames for digital-to-analog conversion in Sigma-Delta quantization. *Advances in Computational Mathematics*, Vol. 32 (1) pp. 73-102, 2008.

[9] S.R. Norsworthy, R.Schreier, and G.C. Temes, editors. *Delta-Sigma Data Converters*. IEEE Press, 1997.

[10] S.V. Parter. On the distribution of the singular values of Toeplitz matrices. *Linear Algebra Appl.*, 80:115–130, 1986.

[11] M. Rudelson and R. Vershynin. Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.*, 62(12):1595–1739, 2009.

[12] G. Strang. The discrete cosine transform. *SIAM review*, pages 135–147, 1999.