Fast Iterative Methods for Reconstruction from Non-Uniform Samples

Chris B Dock

Abstract

The problem of non-uniform sampling arises out of necessity in astronomy, seismology, tomography, and physics. Without access to the Shannon Whittaker explicit reconstruction formula for uniformly sampled band limited signals, an iterative approach is needed. By employing an adaptive weights scheme and exploiting the structure of Toeplitz matrices, the authors of [2] claim a substantial improvement over so called "first generation" iterative schemes for the problem. The adaptive weights scheme provides an efficient pre-conditioner which makes the approach applicable to a wider range of problems (with worse condition number), and the structure of Toeplitz matrices provides fast matrix multiplication and also makes the conjugate gradient algorithm particularly effective. In what follows I present the key arguments of their theoretical framework and investigate their claims numerically.

I. Sampling Theory and Setup

The statement of the problem is as follows. Given $f \in L^2(\mathbb{R})$ that is band-limited (ie there exists $\Omega > 0$ such that $\text{supp}\hat{f} \subset [-\Omega, -\Omega]$), and a collection of sampling points $\{t_i\}_{i\in\mathbb{Z}}$, we we would like to reconstruct f from the collection of samples $\{f(t_i)\}_{i\in\mathbb{Z}}$. When this is possible, we would prefer the reconstruction to be efficient and numerically stable, as well as to obey good theoretical error bounds. If the samples are uniform with $t_i = iT$ with $1/2\Omega$ then the Shannon Whitaker sampling theorem would give

$$f(t) = \sum_{i=-\infty}^{\infty} f(iT)\text{sinc}(\frac{t-iT}{T}) \tag{1}$$

where $\text{sinc}(x) = \sin(x)/x$. If, however, the sampling points are non-uniform then a different approach is needed. As it stands, the problem presented is infinite dimensional and therefore unlikely to be amenable to numerical strategies. Fortunately, we can employ the discrete Fourier transform to create an N dimensional analog of the problem that, owing to the accuracy of the DFT in approximating the Fourier transform, suffices. Namely, we define our band-limited space as

$$\mathcal{B}_M = \{s \in \mathbb{C}^N | \hat{s}(k) = 0 \text{ for } |k \mod N| > M\} \tag{2}$$

Where we take $M < N/2$ and

$$\hat{s}(k) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} s(n)e^{-2\pi ikn/N} \tag{3}$$

is the discrete Fourier transform. Our reconstruction problem then becomes to recover s from a collection of samples $\{s(n_i)\}_{i=1}^N$. This is of course possible in principle whenever the $N \times 2M+1$ matrix

$$Q = \frac{1}{\sqrt{N}} \left[e^{2\pi in_ij} \right]_{i,j} \tag{4}$$

is full rank, since in that case

$$\begin{bmatrix} \hat{s}(-M) \\ \vdots \\ \hat{s}(M) \end{bmatrix} = (Q^TQ)^{-1}Q^T \begin{bmatrix} s(n_1) \\ \vdots \\ s(n_N) \end{bmatrix} \tag{5}$$

at which point the Fourier inversion formula can be applied to find s. The difficulty comes in performing this reconstruction efficiently. To this end, the authors of [2] find it convenient to rephrase the problem slightly. For $s \in \mathcal{B}_M$ we have

$$s(n) = \frac{1}{\sqrt{N}} \sum_{k=-M}^{M} \hat{s}(k)e^{2\pi ikn/N} \tag{6}$$

Hence the authors interpret discrete signals of length N and bandwidth M as the restrictions of degree M trigonometric polynomials of period 1 to arithmetic sequences in $[0,1)$. They define

$$p(t) = \frac{1}{\sqrt{N}} \sum_{k=-M}^{M} \hat{s}(k)e^{2\pi ikt} \tag{7}$$

so that $s(n) = p(n/N)$. Passing from uniform sampling to non uniform sampling then amounts to dropping the "arithmetic" requirement on the sampling the sequence $0 \leq t_1 < \cdots < t_r \leq 1$. If

$$\mathcal{P}_M = \{p | p(t) = \sum_{k=-M}^{M} a_k e^{2\pi i k t}\} \tag{8}$$

then the problem becomes to reconstruct an arbitrary $p \in \mathcal{P}_M$ from a collection of sampled values $\{p(t_i)\}_{i=1}^{r}$.

II. Richardson Iteration Approach

The authors next connect the problem to the stability of a particular frame operator and comment on the ubiquitously used "first generation" approach to solving the above reconstruction problem. Note that so long as the sample points $t_i \in [0,1)$ are distinct the aforementioned matrix Q is full rank whenever $r \geq 2M+1$. More is true, however. The authors give the following

Lemma 1. *Let $0 \leq t_1 < \cdots < t_r$ be r distinct sampling points in $[0,1)$ and let $p \in \mathcal{P}_M$. Then p is uniquely determined by its samples if and only if $r \geq 2M+1$, and in this case there exist $0 < A \leq B$ so that*

$$A \int_0^1 |p(t)|^2 dt \leq \sum |p(t_i)|^2 \leq B \int_0^1 |p(t)|^2 dt \tag{9}$$

Proof. Obviously the matrix Q cannot possibly be full rank if $r < 2M+1$. On the other hand, if $r \geq 2M+!$ then we use the fact that $e^{iMt} p(t) = \sum_{k=0}^{2M} a_k e^{2\pi i k t}$ is the restriction of a complex polynomial of degree $2M$ to the unit circle, hence any distinct $2M+1$ points determine $e^{iMt}p(t)$ and thus p. Moreover, in this case the map

$$\begin{aligned} &\Psi : (\mathcal{P}_M, ||\cdot||_{L^2(\mathbb{T})}) \to (\mathbb{C}^r, ||\cdot||_{l^2}) \\ &\Psi(p) = \begin{bmatrix} p(t_1) \\ \vdots \\ p(t_r) \end{bmatrix} \end{aligned} \tag{10}$$

is linear, one to one, and invertible on its range. The fact that the dimension is finite gives the existence of A and B.□

The authors next employ the following result from classical harmonic analysis that if

$$D_M(t) = \sum_{k=-M}^{M} e^{2\pi i k t} = \frac{\sin((M+\frac{1}{2})2\pi t)}{\sin(\pi t)} \tag{11}$$

is the Dirichlet kernel then

$$\begin{aligned} \langle p, \tau_t D_M \rangle_{L^2(\mathbb{T})} &= \int_0^1 \sum_{k=-M}^{M} a_k e^{2\pi i k u} \overline{\sum_{j=-M}^{M} e^{2\pi i j (u-t)}} du \\ &= \sum_{k,j=-M}^{M} a_k e^{2\pi i j t} \int_0^1 e^{2\pi i (k-j) u} du = \sum_{k,j=-M}^{M} a_k e^{2\pi i j t} \delta_{kj} \\ &= \sum_{k=-M}^{M} a_k e^{2\pi i k t} = p(t) \end{aligned} \tag{12}$$

Where $\tau_t f(u) = f(u-t)$ is the translation operator on $L^2(\mathbb{T})$. Thus the map Ψ is given by

$$\Psi(p) = \begin{bmatrix} p(t_1) \\ \vdots \\ p(t_r) \end{bmatrix} = \begin{bmatrix} \langle p, \tau_{t_1} D_M \rangle_{L^2(\mathbb{T})} \\ \vdots \\ \langle p, \tau_{t_r} D_M \rangle_{L^2(\mathbb{T})} \end{bmatrix} \tag{13}$$

or in other words precisely the analysis map for the frame $\{\tau_{t_r} D_M\}_{i=1}^{r}$ of $\mathcal{P}_M$. The corresponding synthesis map Ψ^* is

$$\begin{aligned} &\Psi^* : \mathbb{C}^r \to \mathcal{P}_M \\ &\Psi^*(w) = \sum_{i=1}^{r} w_i D_M(t - t_i) \end{aligned} \tag{14}$$

So that the frame operator for $\{\tau_{t_r} D_M\}_{i=1}^{r}$ is

$$\begin{aligned} &S : \mathcal{P}_M \to \mathcal{P}_M \\ &Sp(t) = \Psi^* \Psi p(t) = \sum_{i=1}^{r} p(t_i) D_M(t - t_i) \end{aligned} \tag{15}$$

Borrowing from frame theory then we note that we have the following reconstruction formula for p

$$p(t) = \sum_{i=1}^{r} \langle p, \tau_{t_i} D_M \rangle_{L^2(\mathbb{T})} S^{-1} \tau_{t_i} D_M(t) \tag{16}$$

Thus the hope is to iteratively invert the frame operator S. Duffin and Schaeffer of [1] employ the Richardson iteration $x_(k+1) = x_k + \lambda(b - Sx_k)$ to obtain

Lemma 2. *Fix $M \in \mathbb{N}$ and suppose $r \geq 2M+1$ and $\lambda < \frac{1}{B}$ where B is the upper frame bound in Lemma 1. Define iteratively $p_0 = 0$,*

$$p_n = p_{n-1} + \lambda S(p - p_{n-1}) \tag{17}$$

Then $\lim_{n\to\infty} p_n = p$ for $p \in \mathcal{P}_M$ and moreover

$$||p - p_n||_{L^2(\mathbb{T})} \leq \gamma^n ||p||_{L^2(\mathbb{T})} \tag{18}$$

where $\gamma = \max\{|1-\lambda A|, |1-\lambda B|\} < 1$. Since $Sp(t) = \sum_{i=1}^{r} p(t_i) D_M(t - t_i)$ this is in fact a reconstruction from samples only.

Proof. The authors note that $\langle Sp, p\rangle_{L^2(\mathbb{T})} = \sum_{i=1}^{r} |p(t_i)|^2$, showing that S is a positive operator on $\mathcal{P}_M$. Moreover by Lemma 1 we have that

$$(1-\lambda B)||p||^2_{L^2(\mathbb{T})} \leq \langle (\mathbb{I} - \lambda S)p, p\rangle_{L^2(\mathbb{T})} \leq (1-\lambda A)||p||^2_{L^2(\mathbb{T})} \tag{19}$$

Hence, given the definition of the iteration, we have that

$$||p - p_n||_{L^2(\mathbb{T})} = ||(\mathbb{I} - \lambda S)(p - p_{n-1}||_{L^2(\mathbb{T})} \leq \gamma ||(p - p_{n-1}||_{L^2(\mathbb{T})} \leq \cdots \leq \gamma^n ||p - p_0||_{L^2(\mathbb{T})} = \gamma^n ||p||_{L^2(\mathbb{T})} \tag{20}$$

This proof is of course exactly that of the convergence result for Richardson iteration in the particular case of the operator S. □

The authors point out that Lemma 2 gives good convergence estimates only when one has explicit estimates for the constants A and B and moreover when $(B-A)/(B+A)$ is small. For non-uniform sampling, this is often woefully inadequate since it is difficult to derive explicit expressions for A and B and moreover in the presence of clustered sampling points the problem can be very poorly conditioned, to the tune of $\kappa \approx 10^{12} - 10^{15}$. The authors suggest that the reason the above algorithm is still used so frequently is its similarity to the Shannon-Whittaker sampling theorem, the discrete version of which says precisely that

$$p(t) = \frac{1}{N} \sum_{k=0}^{N-1} p(\frac{k}{N}) D_M(t - \frac{k}{N}) \tag{21}$$

for any $p \in \mathcal{P}_M$ and for any $N > 2M$. They note that this is precisely the frame reconstruction formula in the case where the sample points are uniform.

The operator S being somewhat abstract, the authors note that Lemma 2 can be put on a more numerical footing by "changing basis", or rather frame, via

$$p_n(t) \equiv \sum_{i=1}^{r} w_i^{(n)} D_M(t - t_i) \tag{22}$$

In which case if $w_i^{(0)} \equiv p(t_i)$ then we have $Sp(t) = \sum_{i=1}^{r} p(t_i) D_M(t - t_i) = \sum_{i=1}^{r} w_i^{(0)} D_M(t - t_i)$ and furthermore

$$\begin{aligned} Sp_{n-1}(t) &= S \sum w_i^{(n-1)} D_M(t - t_i) \\ &= \sum_{i=1}^{r} w_i^{(n-1)} S D_M(t - t_i) \\ &= \sum_{i=1}^{r} w_i^{(n-1)} \sum_{j=1}^{r} D_M(t_j - t_i) D_M(t - t_j) \\ &= \sum_{j=1}^{r} (\sum_{i=1}^{r} D_M(t_j - t_i) w_i^{(n-1)}) D_M(t - t_j) \end{aligned} \tag{23}$$

So that if the matrix $D \in \mathbb{R}^{r\times r}$ has entries $D_{jk} = D_M(t_j - t_k) = \frac{\sin(2\pi(M+\frac{1}{2})(t_j - t_k)}{\sin(\pi(t_j - t_k))}$ then the iteration in Lemma 2 becomes

$$w^{(n)} = w^{(n-1)} + \lambda(w^{(0)} - Dw^{(n-1)}) \tag{24}$$

and the conclusion of Lemma 2 is that $w^{(i)}$ converges to $w \in \mathbb{C}^r$ such that $p(t) = \sum_{i=1}^r w_i D_M(t-t_i)$. The authors point out that this makes clear the non-optimality of this approach, since the dimension of the relevant matrix grows with the number of sampling points – making the problem harder with greater redundancy rather than easier.

III. Toeplitz Structure

To address this clear fault in the approach shown above, the authors propose to look at the action of the frame operator S specificically on trigonemtric polynomials. This will result in a reformulation of the problem in terms of Toeplitz matrices which can be exploited to reduce the dimension of the problem. Recall that

Definition 1. *A matrix $T \in \mathbb{C}^{n\times n}$ is called Toeplitz if for all i and j $A_{ij} = A_{(i+1),(j+1)} \equiv a_{i-j}$, that is if each descending diagonal of the matrix is constant.*

If $e^{2\pi t M}p$ is a degree $2M$ trigonometric polynomial so that $p(t) = \sum_{k=-M}^M a_k e^{2\pi ikt}$ then

$$\begin{aligned} Sp(t) &= \sum_{l=1}^r p(t_l) D_M(t-t_l) \\ &= \sum_{l=1}^r (\sum_{k=-M}^M a_k e^{2\pi ikt_l})(\sum_{j=-M}^M e^{2\pi ij(t-t_l)}) \\ &= \sum_{j=-M}^M (\sum_{k=-M}^M (\sum_{l=1}^r e^{2\pi i(k-j)t_l}) a_k) e^{2\pi ijt} \end{aligned} \tag{25}$$

Thus if we define the $2M+1 \times 2M+1$ Toeplitz matrix T via

$$T_{jk} = T_{l-k} = \sum_{l=1}^r e^{2\pi i(k-j)t_l} \tag{26}$$

then $Sp \in \mathcal{P}_M$ has a vector of coefficients given by Ta where a is the vector of coefficients for the trigonometric polynomial p. Precisely as was done in obtaining, (24) we may therefore change basis for the iteration scheme in Lemma 2 to find

Lemma 3. *Let $p(t) = \sum_{k=-M}^M a_k e^{2\pi ikt} \in \mathcal{P}_M$ with vector of coefficients $a = (a_k)_{k=-M}^M \in \mathbb{C}^{2M+1}$ and let $0 \le t_1 < \cdots < t_r < 1$ be an arbitrary sequence of distinct sample points with $r > 2M+1$. Let $b \in \mathbb{C}^{2M+1}$ be given by*

$$b_k = \sum_{j=1}^r p(t_j) e^{-2\pi ikt_j} \tag{27}$$

Then for λ small enough and $a^{(0)} = 0$ the iteration

$$a^{(k)} = a^{(k-1)} + \lambda(b - Ta^{(k-1)}) \tag{28}$$

converges to $a \in \mathbb{C}^{2M+1}$. Since $a^{(1)} = \lambda b$ requires only knowledge of the sample points, this is a reconstruction of p.

Proof. The proof is identical to the reformulation of Lemma 2 that lead to (24). Namely, we note that

$$\begin{aligned} Sp(t) &= \sum_{k=-M}^M (Ta)_k e^{2\pi ikt} \\ &= \sum_{j=1}^r p(t_j) D_M(t-t_j) \\ &= \sum_{k=-M}^M (\sum_{j=1}^r p(t_j) e^{-2\pi ikt_j}) e^{2\pi ikt} = \sum_{k=-M}^M b_k e^{2\pi ikt} \end{aligned} \tag{29}$$

If we then define $a_k^{(n)}$ via $p_n(t) \equiv \sum_{k=-M}^M a_k^{(n)} e^{2\pi ikt}$ then (25) with p replaced by p_{n-1} tells us that

$$Sp_{n-1}(t) = \sum_{j=-M}^M (Ta^{(n-1)})_j e^{2\pi ijt} \tag{30}$$

so that the Richardson iteration in Lemma 2 becomes precisely (28). □

This form is far preferable, since the size of T depends on the band limiting number $2M+1$ rather than on the number of samples r. That said, however, one should note that (29) implies a further simplification is possible. Namely,

Lemma 4. *Let $p(t) = \sum_{k=-M}^{M} a_k e^{2\pi ikt} \in \mathcal{P}_M$ with vector of coefficients $a = (a_k)_{k=-M}^{M} \in \mathbb{C}^{2M+1}$ and let $0 \leq t_1 < \cdots < t_r < 1$ be an arbitrary sequence of distinct sample points with $r > 2M+1$. Let $b \in \mathbb{C}^{2M+1}$ be as in the previous lemma. Then*

$$a = T^{-1}b \tag{31}$$

provides a reconstruction of $p(t)$.

The authors note that numerous efficient Toeplitz solvers exist which could be deployed at this point, but they claim that the approach is further improved by an adaptive weights algorithm Grochenig developed in [3].

IV. Adaptive Weights Preconditioner

The authors cite the following proposition derived by Grochenig in [3]

Proposition 1. *Suppose that $0 \leq t_1 < \cdots < t_r < 1$ and define the maximal gap δ as*

$$\delta \equiv \max(t_{i+1} - t_i) < \frac{1}{2M} \tag{32}$$

Where $t_0 \equiv t_r - 1$ and $t_{r+1} \equiv t_1 + 1$. Then

$$(1 - 2\delta M)^2 ||p||_{L^2(\mathbb{T})}^2 \leq \sum_{i=1}^{r} |p(t_i)|^2 \frac{t_{i+1} - t_{i-1}}{2} \leq (1 + 2\delta M)^2 ||p||_{L^2(\mathbb{T})}^2 \tag{33}$$

holds for all $p \in \mathcal{P}_M$.

Proof. The authors of [3] consider an increasing sequence of points $\{x_j\}_{j=1}^{r} \in [1, 0)$ with a maximum separation of δ and $\{y_j\}_{j=1}^{r} \in [1, 0)$ their midpoints. They then construct an operator on $\mathcal{P}_M$ of the form

$$Ap = \Psi(\sum_{i=1}^{r} p(x_j)\chi_j) \tag{34}$$

where $\chi_j = \mathbb{1}_{[y_{j-1}, y_j]}$ and Ψ is the projection operator from earlier. The authors note that

$$||p - Ap||_2^2 \leq \sum_{j=1}^{r} \int_{y_{j-1}}^{y_j} |p(x) - p(x_j)|^2 dx \tag{35}$$

They then apply the following version of the discrete Wirtinger inequality

Lemma 5. *Assume that $s(1) = 0$. Then for $d > 0$ we have*

$$\sum_{n=1}^{d} |s(n)|^2 \leq (4 \sin^2 \frac{\pi}{2(2d-1)})^{-1} \sum_{n=1}^{d-1} |\Delta s(n)|^2 \tag{36}$$

Where $\Delta s(n) = s(n+1) - s(n)$

Using this and the fact that $|x_j - y_j| < \delta/2$, they obtain

$$\sum_{j=1}^{r} \int_{y_{j-1}}^{y_j} |p(x) - p(x_j)|^2 dx \leq \frac{\delta^2}{\pi^2} \sum_{j=1}^{r} \int |p'(x)|^2 dx = \frac{\delta^2}{\pi^2} ||p'||_2^2 \tag{37}$$

Employing the discrete version of Bernstein's inequality, the authors observe that for $p \in \mathcal{P}_M$ we have $||p'||_2 \leq 2\pi M ||p||_2$ so that finally

$$||p - Ap||_2^2 \leq 2\delta M ||p||_2 \tag{38}$$

Note finally that this implies $||\mathbb{I} - A||_{\text{op}} \leq 2\delta M$ so that

$$\begin{aligned} ||A^{-1}||_{\text{op}} &= ||\sum_{n=0}^{\infty} (\mathbb{I} - A)^n||_{\text{op}} \\ &\leq \sum_{n=0}^{\infty} (2\delta M)^n = (1 - 2\delta M)^{-1} \end{aligned} \tag{39}$$

Thus A is invertible with bounded inverse on $\mathcal{P}_M$ and we deduce that

$$(1 - 2\delta M)^2 ||p||_2^2 \leq ||\sum_{j=1}^{r} p(x_j)\chi_j||_2^2 = \sum_{j=1}^{r} |p(x_j)|^2 \frac{x_{j+1} - x_{j-1}}{2} \leq (1 + 2\delta M)^2 ||p||_2^2 \tag{40}$$

This concludes the proof. □

The authors of [2] note that this weighting "compensates local variations in sampling density". Namely, they have the following proposition:

Proposition 2. *Suppose that $\delta < 1/2M$ and let $p \in \mathcal{P}_M$. Set $b \in \mathbb{C}^{2M+1}$ with entries*

$$b_k = \sum_{j=1}^{r} p(t_j) w_j e^{-2\pi i k t_j} \tag{41}$$

for $|k| \leq M$ and $w_j = (t_{j+1} - t_{j-1})/2$. Then create the Toeplitz matrix

$$(T_w)_{lk} = (T_w)_{l-k} = \sum_{j=1}^{r} w_j e^{2\pi i (l-k) t} \tag{42}$$

for $|l|, |k| \leq M$ and calculate $a = T_w^{-1} b$. Then $\sum_{k=-M}^{M} a_k e^{2\pi i k t}$ is the desired reconstruction of p from samples and moreover the condition number of T_w satisfies

$$cond(T_w) \leq (\frac{1 + 2\delta M}{1 - 2\delta M})^2 \tag{43}$$

Finally, the numerical inversion algorithm the authors propose to use is conjugate gradient descent. This provides a substantial acceleration and is applicable since the relevant Toeplitz matrices are positive definite. They cite the following familiar proposition

Proposition 3. *Let A be a positive definite $N \times N$ matrix with smallest eigenvalue λ and largest eigenvalue Λ. Then let $x_0 \in \mathbb{C}^N$ be arbitrary, $r_0 = q_0 = b - A x_0$. For $n \geq 1$ set*

$$\begin{aligned} x_n &= x_{n-1} + \frac{\langle r_{n-1}, q_{n-1} \rangle}{\langle A q_{n-1}, q_{n-1} \rangle} q_{n-1} \\ r_n &= r_{n-1} - \frac{\langle r_{n-1}, q_{n-1} \rangle}{\langle A q_{n-1}, q_{n-1} \rangle} A q_{n-1} \\ q_n &= r_n - \frac{\langle r_n, A q_{n-1} \rangle}{\langle A q_{n-1}, q_{n-1} \rangle} q_{n-1} \end{aligned} \tag{44}$$

Then x_n converges in at most N iterations to the exact solution of $Ax = b$. For $n < N$ the error is at most

$$||x - x_n||_A \leq 2(\frac{\sqrt{\Lambda} + \sqrt{\lambda}}{\sqrt{\Lambda} - \sqrt{\lambda}})^2 \tag{45}$$

V. Algorithm

The above sections lead immediately to the following algorithm for the reconstruction of band-limited functions from non-uniform samples

Theorem 1. *Let M be the size of the spectrum and let $0 \leq t_1 < \cdots < t_r < 1$ be an arbitrary sequence of sampling points with $r > 2M + 1$. Set $t_0 = t_r - 1$ and $t_{r+1} = t_1 + 1$ and let $w_j = \frac{1}{2}(t_{j+1} - t_{j-1})$ and compute*

$$\gamma_k = \sum_{j=1}^{r} e^{2\pi i k t_j} w_j \tag{46}$$

so that the associated Toeplitz matrix has $(T_w)_{lk} = \gamma_{l-k}$ for $|k|, |l| \leq M$. To reconstruct a trigonometric polynomial $p \in \mathcal{P}_M$ from its samples $p(t_j)$ compute first

$$b_k = \sum_{j=1}^{r} p(t_j) w_j e^{-2\pi i k t_j} \tag{47}$$

for $|k| \leq M$ and set $r_0 = q_0 = b \in \mathbb{C}^{2M+1}$. Compute iteratively for $n \geq 1$

$$a_n = a_{n-1} + \frac{\langle r_{n-1}, q_{n-1} \rangle}{\langle T_w q_{n-1}, q_{n-1} \rangle} q_{n-1} \tag{48}$$

$$r_n = r_{n-1} - \frac{\langle r_{n-1}, q_{n-1} \rangle}{\langle T_w q_{n-1}, q_{n-1} \rangle} T_w q_{n-1} \tag{49}$$

$$q_n = r_n - \frac{\langle r_n, T_w q_{n-1} \rangle}{\langle T_w q_{n-1}, q_{n-1} \rangle} q_{n-1} \tag{50}$$

Then a_n converges in at most $2M+1$ steps to a vector $a \in \mathbb{C}^{2M+1}$ solving $T_w a = b$. The reconstruction of p is given by $p_n(t) = \sum_{k=-M}^{M} a_{n,k} e^{2\pi i k t} \in \mathcal{P}_M$ denotes the reconstruction of after n iterations. Moreover

$$(\sum_{j=1}^{r} |p(t_j) - p_n(t_j)|w_j)^{1/2} \leq 2(2\delta M)^n (\sum_{j=1}^{r} |p(t_j)|^2 w_j)^{1/2} \tag{51}$$

Proof. Most of the proof is contained in the sections above, the only part that remains is (51). The authors note that

$$\sum_{j=1}^{r} |p(t_j)|^2 w_j = \langle S_w p, p\rangle = \langle T_w a, a\rangle = ||a||_{T_w}^2 \tag{52}$$

The conjugate gradient error bound will give the result since (40) implies

$$(1 - 2\delta M)^2 \leq \lambda \leq \Lambda \leq (1 + 2\delta M)^2 \tag{53}$$

VI. Numerics

I used the following matlab code to approach testing the algorithm above:

```
% Setup

r=300;
M=20;
t = (1:r)/r+ randn(1,r)/r;
aa = randn(1,2*M+1);
s = arrayfun(@(x) testpoly(x,M,aa), t);



% M is the bandwidth, r is the number of samples
% t are our sample points
% s are our samples
% aa are the true coefficients
% w are the preconditioner weights
% Tw is the relevant Toeplitz matrix
% b is the initialization of conjugate gradient descent
w = zeros(1,r);
for k =2:(r-1)
    w(k)=.5*(t(k+1)-t(k-1));
end
w(1)=.5*(t(2)-t(r))+.5;
w(r)=.5*(t(1)-t(r-1))+.5;
Tw =zeros(2*M+1, 2*M+1);
for l=-M:M
   for j = -M:M
      Tw(l+M+1,j+M+1) = dot(w,exp(-2*pi*1i*(l-j)*t));
   end
end

b=zeros(1,2*M+1);
for k=-M:M
   b(k+M+1)=dot(w.*s, exp(2*pi*1i*k*t));
end
r0=b;
q0=b;
a0=randn(1,2*M+1);
for n=1:(2*M+1)
    a = a0 + q0*dot(r0,q0)/dot(q0*Tw,q0);
    r = r0 - (q0*Tw)*dot(r0,q0)/dot(q0*Tw,q0);
    q = r - q0*dot(r, q0*Tw)/dot(q0*Tw,q0);
    a0=a; r0=r;q0=q;
end
disp(dot(a0-aa,(a0-aa)*Tw))

function p = testpoly(t,M,aa)
    p = dot(aa, exp(2*pi*1i*(-M:M)*t));
end
```

The algorithm presented herein managed to quite accurately reconstruct the randomly generated band-limited functions I produced across a wide range of randomly generated sample points. It also did so much more reliably than when I replaced the conjugate gradient section of the code with Richardson iteration or, interestingly, direct inversion use Matlab's $b \setminus T_w$ functionality. I used the T_w norm to estimate the error between a and the output of conjugate gradient descent on b. Moreover, the time the algorithm took did not depend strongly on the number of sampled points. Interestingly, I found that the algorithm maintained a baseline level of error even as the number of sampling points increased. I suspect that this is due to to the extreme sensitivity of trigonometric polynomials to changes in their coefficients, which were slightly off due to floating point errors. This phenomenon can be seen quite clearly in Figure 3, in which $M = 10$ is fixed and r increases. Figure 1 contains some signals and their reconstructions using the above algorithm.

There are essentially three parameters that affect the performance of the above algorithm. Two of them, the number of samples r and the bandwidth M, can be controlled and tested thoroughly with some ease. The third, namely the *distribution* of sample points (or more directly the spectrum of the matrix T_w) is more difficult to investigate completely. Figure 2 shows the dependence of the reconstruction performance on r and M varying together – essentially on the complexity of the problem. Figure 2 appears to show therefore that the error grows, approximately and on average, as the square root of the "size" of the reconstruction problem. Meanwhile, somewhat oddly, Figure 3 appears to show that increasing the number of samples essentially does not noticeably increase the performance of the algorithm so long as $r > 2M + 1$, at least not in this "close to uniformly distributed" sample points case. This would appear to somewhat limit the utility of the algorithm in practical contexts, given its relative complexity. Finally, in Figure 4, r is held constant at 20 sample points and the bandwidth M is increased – evidently leading to a linear increase in the error. Interestingly, the error doesn't demonstrate a jump as M passes through $(r - 1)/2$ – the theoretical critical point indicated by Theorem 1.

In order to investigate the dependence of the algorithm in Theorem 1 on the distribution of the sample points, I essentially created an ensemble of distributions that that interpolated between the "close to uniform sampling" distribution $t_i = i/r + \delta_i$ where δ_i is uniformly distributed in $[-r/2, r/2]$ and a uniform distribution on $[0, 1]$ (confusingly, this situation is far from uniform sampling). Figure 5 shows the average error over 10000 runs for each distribution in the aforementioned interpolative ensemble. The result depicted in **??** actually makes a lot of sense – the average performance of the algorithm over many many runs is essentially invariant as we shift from "close to uniformly sampled" to "jointly uniform on [0,1]". The reason for this is that the two distributions are the same in expectation, indicating that the outliers (in which case uniform on [0,1] looks very different from uniformly sampled) don't have a significant effect on the error. This indicates that the algorithm has a degree of robustness with regards to changes in the distribution of sampling points, at least probabilistically and along this particular axis of change. Obviously, if the sampling were made "maximally non-uniform" by putting all of the samples in esssentially the sampe place, the performance would be terrible.

References

[1] Richard J Duffin and Albert C Schaeffer. A class of nonharmonic fourier series. *Transactions of the American Mathematical Society*, 72(2):341–366, 1952.

[2] Hans G Feichtinger, Karlheinz Gr, Thomas Strohmer, et al. Efficient numerical methods in non-uniform sampling theory. *Numerische Mathematik*, 69(4):423–440, 1995.

[3] Karlheinz Gröchenig. A discrete theory of irregular sampling. *Linear Algebra and its applications*, 193:129–150, 1993.

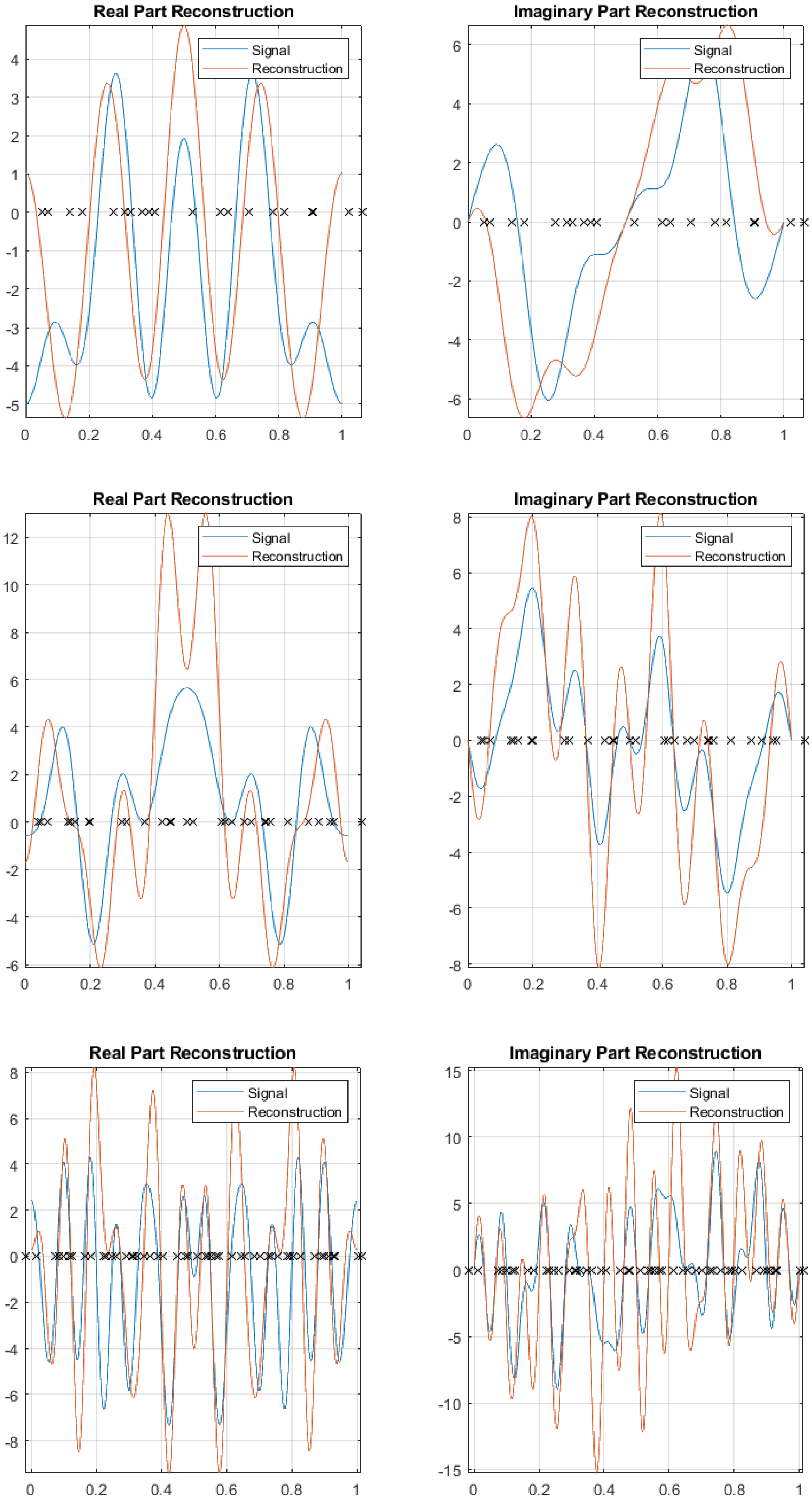

Fig. 1. Sample points are marked with a cross. For the above plots the bandwidtsh are $M = 5$, $M = 8$, and $M = 15$ respectively and the number of sample points are $r = 20$, $r = 30$, and $r = 50$ respectively. Only the fundamental period of the signals is shown.

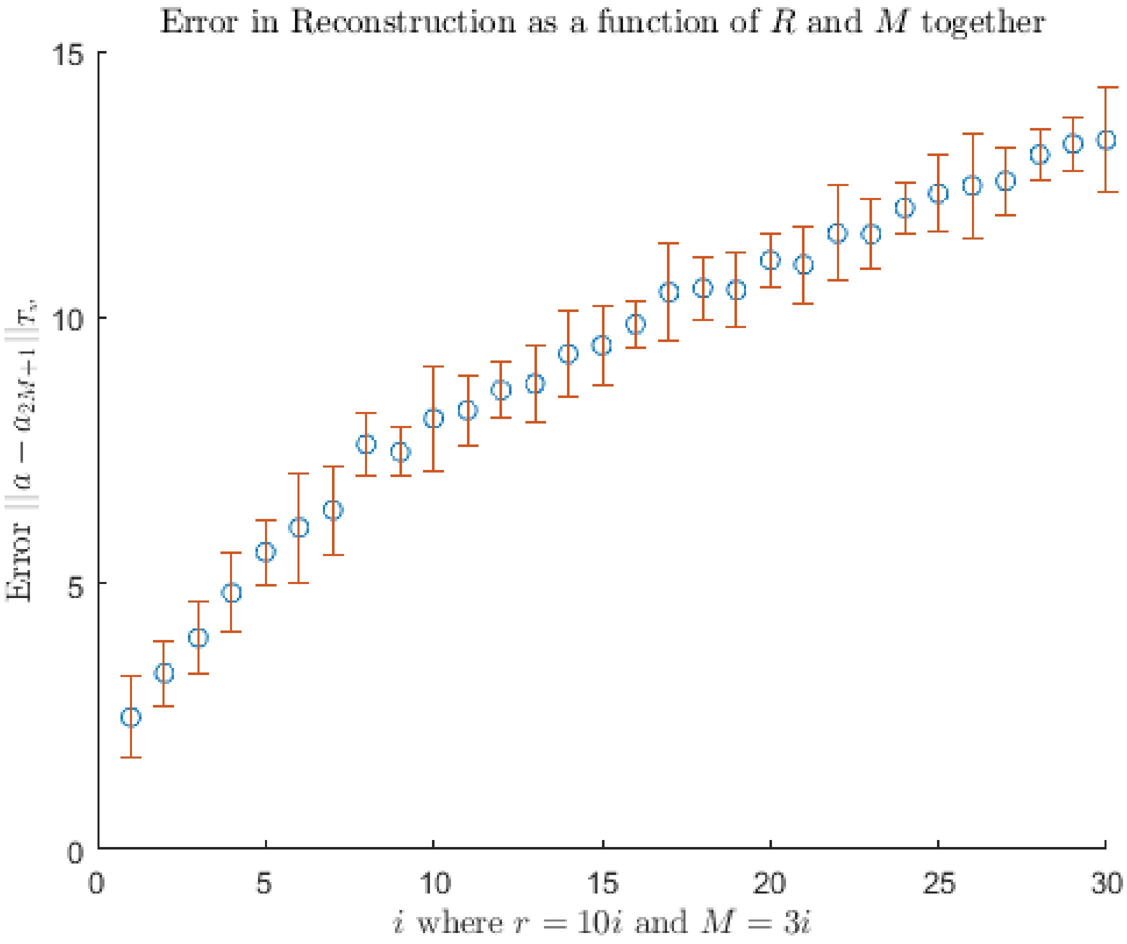

Fig. 2. Each value of i represents 20 runs through the reconstruction algorithm with uniformly distributed trigonometric polynomial coefficients and sampling points distributed according to $t_i = i/r + \delta_i$ where δ_i is uniformly distributed in $[-r/2, r/2]$. The average error over this distribution is plotted, as well as the standard deviation. In each trial, $M = 3i$ and $r = 10i$.

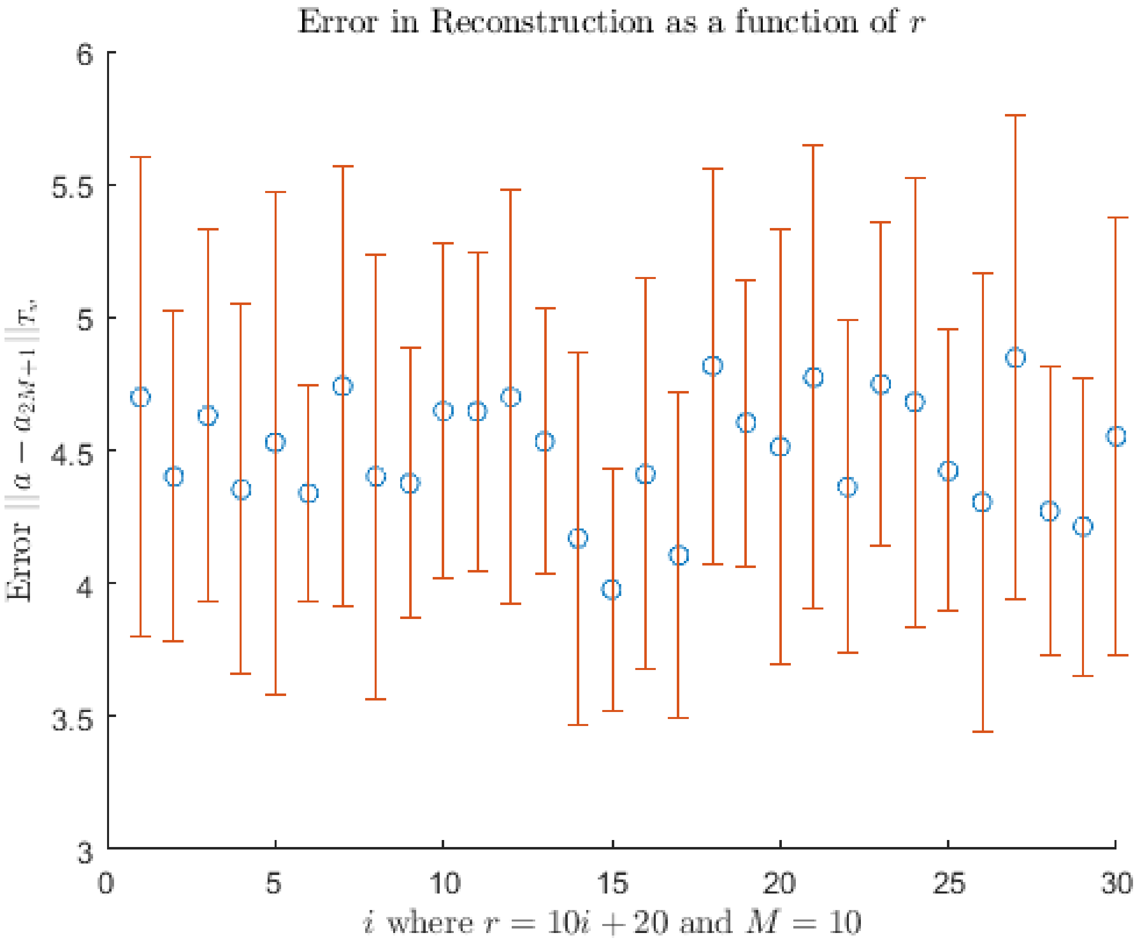

Fig. 3. Each value of i represents 20 runs through the reconstruction algorithm with uniformly distributed trigonometric polynomial coefficients and sampling points distributed according to $t_i = i/r + \delta_i$ where δ_i is uniformly distributed in $[-r/2, r/2]$. The average error over this distribution is plotted, as well as the standard deviation. In each trial, $M = 10$ and $r = 10i + 20$.

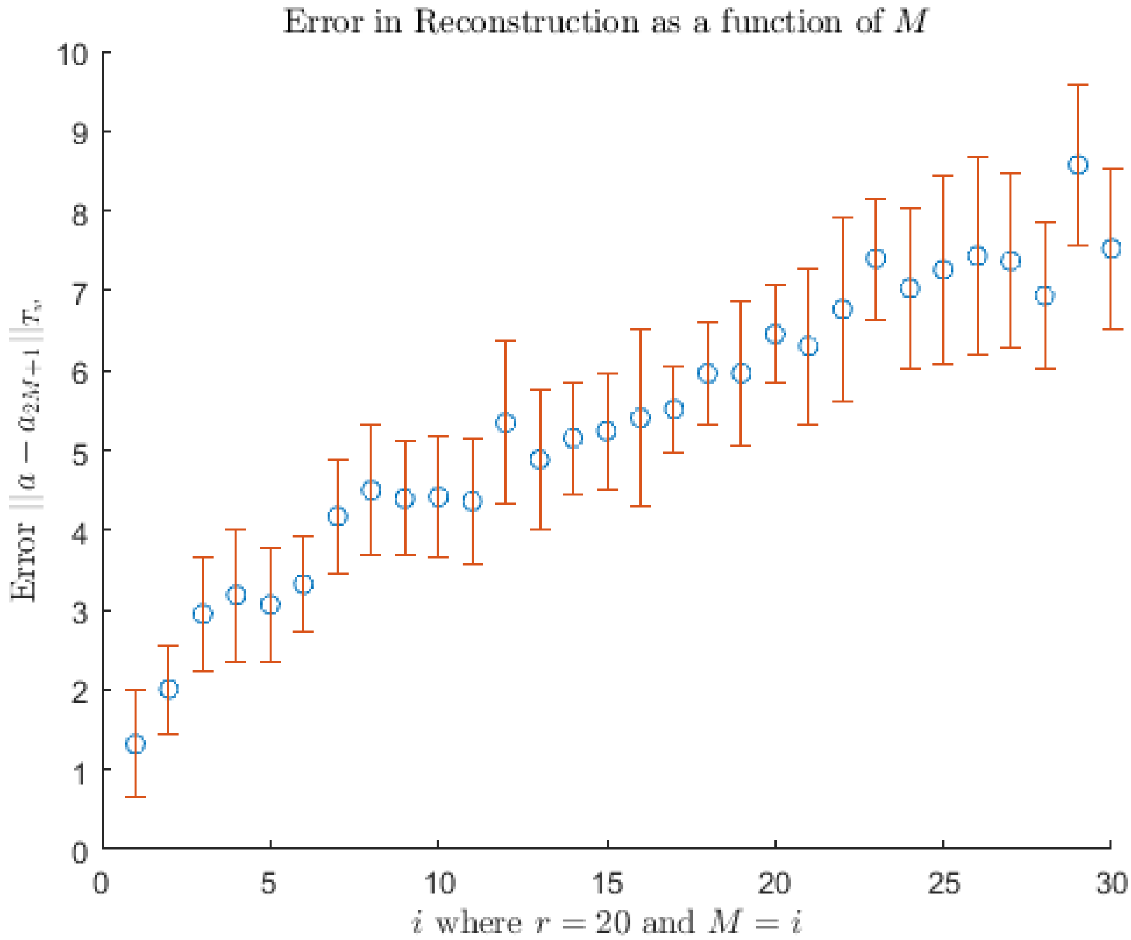

Fig. 4. Each value of i represents 20 runs through the reconstruction algorithm with uniformly distributed trigonometric polynomial coefficients and sampling points distributed according to $t_i = i/r + \delta_i$ where δ_i is uniformly distributed in $[-r/2, r/2]$. The average error over this distribution is plotted, as well as the standard deviation. In each trial, $M = i$ and $r = 20$.

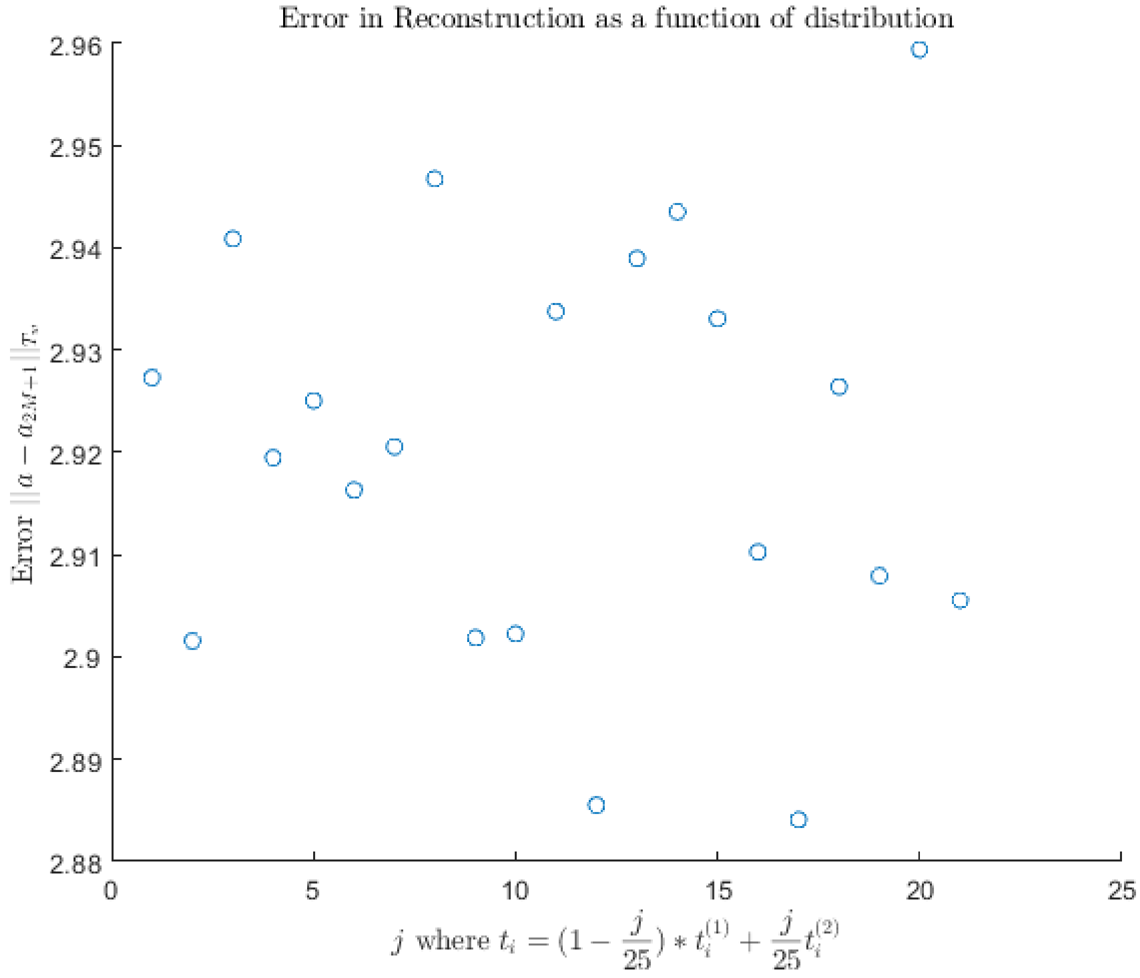

Fig. 5. Each value of j represents 10000 runs through the reconstruction algorithm with uniformly distributed trigonometric polynomial coefficients and sampling points distributed according to the interpolated distribution $t_i = (1 - \frac{j}{25})t_i^{(1)} + \frac{j}{25}t_i^{(2)}$ and $t_i^{(1)}$ is the almost uniform sampling distribution used earlier and $t_i^{(2)}$ is the uniform distribution on $[0, 1]$. Here $M = 5$ and $r = 20$ for all twenty five thousand trials.