Bandlimited Signal Reconstruction from Nonuniform Samples

Santhosh Karnik, Justin Romberg, and Mark A. Davenport
Georgia Institute of Technology, School of Electrical and Computer Engineering
skarnik1337@gatech.edu, jrom@ece.gatech.edu, mdav@gatech.edu

***Abstract*—From infinitely many equispaced samples, a bandlimited signal can be reconstructed as a linear combination of shifted sinc functions, where the coefficients are the sample values. If instead of receiving infinitely many equispaced samples, we receive M nonuniformly spaced samples, the optimal reconstruction is still a linear combination of shifted sincs, but one must solve an ill-conditioned system of equations to determine the coefficients. For large scale problems, this can be infeasible. By using the structural and spectral properties of the system of equations, we can obtain the weights in slightly worse than linear time in the number of samples.**

I. Introduction

In this paper we develop an efficient numerical technique for solving the symmetric system of equations that arises when a signal is being reconstructed from a finite number of nonuniform samples. Reconstructing from nonuniform samples is a classic problem in signal processing [1], [2], as well as frame theory [3]–[5]. Applications of nonuniform sampling are manifold, arising in seismic imaging, radar imaging, level-crossing analog-to-digital converters, and array processing.

The key computation is solving a symmetric positive definite system of equations of the form $\boldsymbol{y} = (\boldsymbol{G} + \delta \mathbf{I})\boldsymbol{z}$. Our analysis below will show if the signal is bandlimited to W and if we collect $M \sim 2WT\log(2WT)$ samples in a time interval of length T, then almost all of the eigenvalues of this system are $\sim \frac{M}{2WT}$ or $\approx \delta$. We then show that this approximately two-tiered eigenvalue structure means that the conjugate gradients algorithm for solving this system is guaranteed to converge in a small number of iterations, logarithmic in all the parameters of the problem. Moreover, since $\boldsymbol{G}+\delta\mathbf{I}$ is a *generalized Cauchy-like matrix* (see Section III), it can be approximately applied in linear time. Combing these results means that the system above can be solved in time roughly linear in the number of samples M.

The solution to the system above parameterizes a continuous-time bandlimited signal. We also show that computing a batch of uniformly spaced samples of this signal can be done in linear time, again by exploiting the Cauchy structure of the forward operator.

This work is similar to [6], which studies reconstructing a trigonometric polynomial from nonuniform samples. The resulting system of equations for that problem is Toeplitz, and thus, each iteration of the conjugate gradients algorithm takes $O(M \log M)$ operations. This work is also similar to [7], which studies reconstructing a bandlimited signal from nonuniform samples, but doesn't explore fast methods for solving the resulting system of equations when the number of samples is large.

II. Problem Formulation

Suppose we have a continuous time bandlimited signal

$$x_c(t) = \int_{-W}^{W} X(f) e^{j2\pi f t}\, df,$$

and we observe a vector $\boldsymbol{y} \in \mathbb{C}^M$ of M (possibly noisy) samples of the signal at nonuniformly spaced times $t_1, \dots, t_M \in [-\frac{T}{2}, \frac{T}{2}]$ i.e.

$$\boldsymbol{y}[m] = x_c(t_m) + \text{noise for } m = 1, \dots, M.$$

We can treat recovering $x_c(t)$ as a linear inverse problem where we try to estimate the sequence $\boldsymbol{x}_d \in \ell_2(\mathbb{Z})$ of Nyquist-rate samples

$$\boldsymbol{x}_d[n] = x_c\left(\tfrac{n}{2W}\right) \text{ for } n \in \mathbb{Z}.$$

For any time $t \in \mathbb{R}$, define a shifted sinc sequence $\boldsymbol{a}_t \in \ell_2(\mathbb{Z})$ by

$$\boldsymbol{a}_t[n] = \frac{\sin[\pi(n - 2Wt)]}{\pi(n - 2Wt)} \text{ for } n \in \mathbb{Z}.$$

Then, using the Whittaker-Shannon interpolation formula, we can write

$$\begin{aligned}\boldsymbol{y}[m] &= x_c(t_m) + \text{noise} \\ &= \sum_{n=-\infty}^{\infty} \frac{\sin[\pi(n - 2Wt_m)]}{\pi(n - 2Wt_m)} x_c\left(\frac{n}{2W}\right) + \text{noise} \\ &= \langle \boldsymbol{a}_{t_m}, \boldsymbol{x}_d \rangle + \text{noise}.\end{aligned}$$

Let $\mathcal{A} : \ell_2(\mathbb{Z}) \to \mathbb{C}^M$ be the operator which sinc-interpolates a sequence of samples $\boldsymbol{x} \in \ell_2(\mathbb{Z})$ at times $t_1, \dots, t_M$, i.e.

$$(\mathcal{A}\boldsymbol{x})[m] = \langle \boldsymbol{a}_{t_m}, \boldsymbol{x} \rangle \text{ for } m = 1, \dots, M$$

For a given vector of measurements $\boldsymbol{y} \in \mathbb{C}^M$ and a regularization parameter $\delta > 0$, we can estimate the Nyquist rate samples $\boldsymbol{x}_d$ by solving the Tikhonov regularization problem

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x} \in \ell_2(\mathbb{Z})}{\arg\min} \|\boldsymbol{y} - \mathcal{A}\boldsymbol{x}\|_2^2 + \delta \|\boldsymbol{x}\|_{\ell_2(\mathbb{Z})}^2.$$

The solution to this Tikhonov regularization problem can be written as

$$\begin{aligned}\widehat{\boldsymbol{x}} &= \mathcal{A}^*(\mathcal{A}\mathcal{A}^* + \delta\mathcal{I})^{-1}\boldsymbol{y} \\ &= \mathcal{A}^*[(\boldsymbol{G} + \delta\mathbf{I})^{-1}\boldsymbol{y}]\end{aligned}$$

where the adjoint $\mathcal{A}^* : \mathbb{C}^M \to \ell_2(\mathbb{Z})$ is given by

$$\mathcal{A}^*\boldsymbol{z} = \sum_{m=1}^{M} \boldsymbol{z}[m]\boldsymbol{a}_{t_m},$$

and the Gram matrix $\boldsymbol{G} \in \mathbb{C}^{M\times M}$ is given by

$$\boldsymbol{G}[m, m'] = \left\langle \boldsymbol{a}_{t_m}, \boldsymbol{a}_{t_{m'}} \right\rangle = \frac{\sin[2\pi W(t_m - t_{m'})]}{2\pi W(t_m - t_{m'})}.$$

In other words, the reconstruction of the Nyquist-rate samples is given by

$$\widehat{\boldsymbol{x}}_d[n] = \sum_{m=1}^{M} \boldsymbol{z}[m]\frac{\sin[\pi(n - 2Wt_m)]}{\pi(n - 2Wt_m)}.$$

Thus, the corresponding continuous time signal

$$\widehat{x}_c(t) = \sum_{m=1}^{M} \boldsymbol{z}[m]\frac{\sin[2\pi W(t - t_m)]}{2\pi W(t - t_m)},$$

is a linear combination of sincs shifted by the sample times. This is very similar to the classic Whittaker–Shannon interpolation formula, except the weights $\boldsymbol{z}[m]$ satisfy the system of equations $(\boldsymbol{G} + \delta\mathbf{I})\boldsymbol{z} = \boldsymbol{y}$ instead of being the samples themselves.

After solving the system $(\boldsymbol{G} + \delta\mathbf{I})\boldsymbol{z} = \boldsymbol{y}$, we can evaluate $\widehat{x}_c(t)$ at N sample times via a matrix-vector multiply

$$\begin{bmatrix}\widehat{x}_c(s_1) & \cdots & \widehat{x}_c(s_N)\end{bmatrix}^T = \boldsymbol{H}\boldsymbol{z}$$

where the matrix $\boldsymbol{H} \in \mathbb{R}^{N\times M}$ has entries

$$\boldsymbol{H}[n, m] = \frac{\sin[2\pi W(s_n - t_m)]}{2\pi W(s_n - t_m)}.$$

For large M and N, explicitly forming $\boldsymbol{G} + \delta\mathbf{I}$ and solving $(\boldsymbol{G} + \delta\mathbf{I})\boldsymbol{z} = \boldsymbol{y}$ takes $O(M^{2.37369})$ operations [8] via an improvement over the Coppersmith–Winograd algorithm [9][1]. for practical values of M. Also, explicitly computing $\boldsymbol{H}\boldsymbol{z}$ takes $O(MN)$ operations.

In this work, we note that the matrices $\boldsymbol{G} + \delta\mathbf{I}$ and $\boldsymbol{H}$ have a special structure which allows us to multiply them by an $M \times 1$ vector in $O(M \log \frac{1}{\alpha})$ and $O((N + M) \log \frac{1}{\alpha})$ operations respectively, where α is the approximation tolerance. Furthermore, if we have at least $M \gtrsim 3WT \log\left(\frac{4WT}{\beta}\right)$ samples which are chosen i.i.d. Uniform$[-\frac{T}{2}, \frac{T}{2}]$, then with probability at least $1 - \beta$, the eigenvalues of $\boldsymbol{G} + \delta\mathbf{I}$ exhibit a special clustering property which can be used to prove that conjugate gradient descent (CGD) will only need $O(\text{polylog}(2WT, \frac{M}{2WT}, \frac{1}{\delta}, \frac{1}{\epsilon}))$ iterations to return a solution $\widehat{\boldsymbol{z}}$ satisfying $\|\widehat{\boldsymbol{z}} - (\boldsymbol{G} + \delta\mathbf{I})^{-1}\boldsymbol{y}\|_2 \le \epsilon\|\boldsymbol{y}\|_2$. Hence, we can solve this nonuniform sinc interpolation problem in complexity that is only polylogarithmically worse than linear time in M and N.

[1]It should be noted that Strassen's algorithm, which has a runtime of $O(M^{2.807})$ is often used in practice over the Coppersmith–Winograd algorithm as it is faster for practical values of M [10].

III. Structured Matrices

Our first objective is to demonstrate that the matrices $\boldsymbol{G}+\delta\mathbf{I}$ and $\boldsymbol{H}$ can be applied to a vector efficiently, even if M and N are large. We start by defining a couple types of structured matrices for which there are efficient matrix-vector multiplication methods.

A *Cauchy matrix* is an $N \times M$ matrix whose entries are of the form

$$\boldsymbol{C}[n, m] = \frac{1}{\sigma_n - \tau_m}$$

where $\sigma_1, \ldots, \sigma_N$ and $\tau_1, \ldots, \tau_M$ are real numbers such that $\sigma_n \neq \tau_m$ for all indices n, m. Using the fast multipole method [11], [12], it is possible to apply an $N \times M$ Cauchy matrix to an $M \times 1$ vector in $O((N + M) \log \frac{1}{\alpha})$ operations, where α is the desired level of precision. See [13] for details on using the fast multipole method for Cauchy matrices.

A *Cauchy-like matrix* is an $M \times M$ matrix whose entries are of the form

$$\boldsymbol{K}[m, m'] = \begin{cases} \dfrac{1}{\tau_m - \tau_{m'}} & \text{if } m \neq m' \\ 0 & \text{if } m = m' \end{cases}$$

where $\tau_1, \ldots, \tau_M$ are real numbers such that $\tau_m \neq \tau_{m'}$ if $m \neq m'$. Again, we can use the fast multipole method to apply an $M \times M$ Cauchy-like matrix to an $M \times 1$ vector in $O(M \log \frac{1}{\alpha})$ operations, where α is the desired level of precision.

A *generalized Cauchy matrix* is an $N \times M$ matrix whose entries are of the form

$$\widetilde{\boldsymbol{C}}[n, m] = \sum_{\ell=1}^{r} \frac{\boldsymbol{p}_\ell[n]\overline{\boldsymbol{q}_\ell[m]}}{\sigma_n - \tau_m}$$

where $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_r \in \mathbb{C}^N$, $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_r \in \mathbb{C}^M$, and $\sigma_1, \ldots, \sigma_N$ and $\tau_1, \ldots, \tau_M$ are real numbers such that $\sigma_n \neq \tau_m$ for all indices n, m. Note that we can write[2]

$$\widetilde{\boldsymbol{C}} = \sum_{\ell=1}^{r} \boldsymbol{D}_{\boldsymbol{p}_\ell}\boldsymbol{C}\boldsymbol{D}^*_{\boldsymbol{q}_\ell}$$

where $\boldsymbol{C}$ is a Cauchy matrix. Applying each term $\boldsymbol{D}_{\boldsymbol{p}_\ell}\boldsymbol{C}\boldsymbol{D}^*_{\boldsymbol{q}_\ell}$ to an $M \times 1$ vector takes $O((N + M) \log \frac{1}{\alpha})$ operations via two diagonal matrix multiplies and a Cauchy matrix multiply. Hence, applying the sum of r matrices of that form takes $O(r(N + M) \log \frac{1}{\alpha})$ operations.

A *symmetric generalized Cauchy-like matrix* is an $M \times M$ matrix whose entries are of the form

$$\widetilde{\boldsymbol{K}}[m, m'] = \begin{cases} \displaystyle\sum_{\ell=1}^{r} \frac{\boldsymbol{p}_\ell[m]\overline{\boldsymbol{q}_\ell[m']} - \boldsymbol{q}_\ell[m]\overline{\boldsymbol{p}_\ell[m']}}{\tau_m - \tau_{m'}} & \text{if } m \neq m' \\ d_m & \text{if } m = m' \end{cases}$$

[2]For a vector $\boldsymbol{v}$, we use $\boldsymbol{D}_{\boldsymbol{v}}$ to denote a diagonal matrix whose diagonal entries match the entries of the vector $\boldsymbol{v}$, i.e. $\boldsymbol{D}_{\boldsymbol{v}}[m, m] = \boldsymbol{v}[m]$.

where $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_r \in \mathbb{C}^M$, $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_r \in \mathbb{C}^M$, and $\tau_1, \ldots, \tau_M$ are real numbers such that $\tau_m \neq \tau_{m'}$ if $m \neq m'$. Note that we can write

$$\widetilde{\boldsymbol{K}} = \boldsymbol{D_d} + \sum_{\ell=1}^{r} \boldsymbol{D_{p_\ell}} \boldsymbol{K} \boldsymbol{D}^*_{\boldsymbol{q}_\ell} - \boldsymbol{D_{q_\ell}} \boldsymbol{K} \boldsymbol{D}^*_{\boldsymbol{p}_\ell}$$

where $\boldsymbol{K}$ is a Cauchy-like matrix. Applying each term $\boldsymbol{D_{p_\ell}} \boldsymbol{K} \boldsymbol{D_{q_\ell}}$ or $\boldsymbol{D_{q_\ell}} \boldsymbol{K} \boldsymbol{D}^*_{\boldsymbol{p}_\ell}$ to an $M \times 1$ vector takes $O(M \log \frac{1}{\alpha})$ operations via two diagonal matrix multiplies and a Cauchy matrix multiply. Applying $\boldsymbol{D_d}$ to a vector takes $O(M)$ operations via a diagonal matrix multiply. Hence, applying the above sum to an $M \times 1$ vector takes $O(rM \log \frac{1}{\alpha})$ operations.

The matrix $\boldsymbol{H}$ defined in the previous section is a generalized Cauchy matrix. To see this, note that the entries of $\boldsymbol{H}$ can be written as

$$\begin{aligned}\boldsymbol{H}[n,m] &= \frac{\sin[2\pi W(s_n - t_m)]}{2\pi W(s_n - t_m)} \\ &= \frac{\sin(2\pi W s_n)\cos(2\pi W t_m) - \cos(2\pi W s_n)\sin(2\pi W t_m)}{2\pi W s_n - 2\pi W t_m}\end{aligned}$$

Hence, $\boldsymbol{H}$ fits the form of a generalized Cauchy matrix with $\sigma_n = 2\pi W s_n$, $\tau_m = 2\pi W t_m$, $r = 2$, $\boldsymbol{p}_1[n] = \sin(2\pi W s_n)$, $\boldsymbol{p}_2[n] = \cos(2\pi W s_n)$, $\boldsymbol{q}_1[m] = \cos(2\pi W t_m)$, $\boldsymbol{q}_2[m] = -\sin(2\pi W t_m)$. Therefore, $\boldsymbol{H}$ can be applied to an $M \times 1$ vector in $O((N+M)\log \frac{1}{\alpha})$ operations.

In a similar manner, the matrix $\boldsymbol{G} + \delta \mathbf{I}$ defined in the previous section is a symmetric generalized Cauchy-like matrix. The diagonal entries are all $1+\delta$, and the off-diagonal entries are

$$\begin{aligned}\boldsymbol{G}[m,m'] &= \frac{\sin[2\pi W(t_m - t_{m'})]}{2\pi W(t_m - t_{m'})} \\ &= \frac{\sin(2\pi W t_m)\cos(2\pi W t_{m'}) - \cos(2\pi W t_m)\sin(2\pi W t_{m'})}{2\pi W t_m - 2\pi W t_{m'}}\end{aligned}$$

Hence, $\boldsymbol{G} + \delta \mathbf{I}$ fits the form of a generalized Cauchy-like matrix with $\tau_m = 2\pi W t_m$, $r = 1$, $\boldsymbol{p}_1[m] = \sin(2\pi W t_m)$, $\boldsymbol{q}_1[m] = \cos(2\pi W t_m)$, and $d_m = 1 + \delta$. Therefore, $\boldsymbol{G} + \delta \mathbf{I}$ can be applied to an $M \times 1$ vector in $O(M \log \frac{1}{\alpha})$ operations.

IV. Spectral Properties

A symmetric generalized Cauchy-like matrix $\widetilde{\boldsymbol{K}}$ satisfies a low-rank displacement equation

$$\boldsymbol{D_\tau} \widetilde{\boldsymbol{K}} - \widetilde{\boldsymbol{K}} \boldsymbol{D_\tau} = \boldsymbol{P}\boldsymbol{Q}^* - \boldsymbol{Q}\boldsymbol{P}^*,$$

where $\boldsymbol{P} = \begin{bmatrix}\boldsymbol{p}_1 & \cdots & \boldsymbol{p}_r\end{bmatrix}$ and $\boldsymbol{Q} = \begin{bmatrix}\boldsymbol{q}_1 & \cdots & \boldsymbol{q}_r\end{bmatrix}$. Also, the entries of $\widetilde{\boldsymbol{K}}$ can be recovered from the parameters $\tau_1, \ldots, \tau_M$, $d_1, \ldots, d_M$, and the "generators" $\boldsymbol{P}$ and $\boldsymbol{Q}$.

Furthermore, if $\widetilde{\boldsymbol{K}}$ is invertible, then $\widetilde{\boldsymbol{K}}^{-1}$ also satisfies a low-rank displacement equation

$$\boldsymbol{D_\tau} \widetilde{\boldsymbol{K}}^{-1} - \widetilde{\boldsymbol{K}}^{-1} \boldsymbol{D_\tau} = \widetilde{\boldsymbol{K}}^{-1} \boldsymbol{Q}\boldsymbol{P}^* \widetilde{\boldsymbol{K}}^{-1} - \widetilde{\boldsymbol{K}}^{-1} \boldsymbol{P}\boldsymbol{Q}^* \widetilde{\boldsymbol{K}}^{-1},$$

and thus, $\widetilde{\boldsymbol{K}}^{-1}$ is also a symmetric generalized Cauchy-like matrix.

This fact has been exploited to yield recursive methods for inverting symmetric generalized Cauchy-like matrices [14], [15]. These methods partition the generalized Cauchy-like matrix $\widetilde{\boldsymbol{K}}$ into a 2×2 block matrix, compute the generators of the $(1,1)$-block and its Schur complement via recursion, and then determine the "generators" of $\widetilde{\boldsymbol{K}}$. This Schur recursion takes $O(rM \log M \log \frac{1}{\alpha})$ operations to compute the generators of $\widetilde{\boldsymbol{K}}^{-1}$. Unfortunately, for our problem, when the number of samples exceeds $2WT + O(\log(2WT))$, the matrix $\boldsymbol{G}$ becomes numerically rank deficient. As such, these recursive methods for inverting $\boldsymbol{G}$ are unstable if more than a few recursive stages are used.

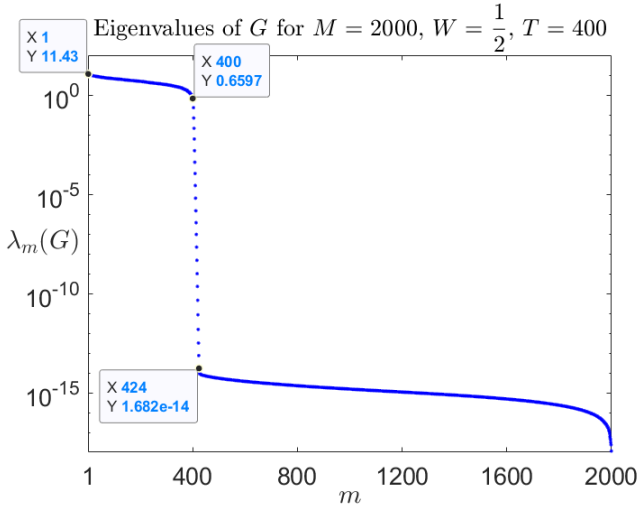

Fig. 1. A plot of the eigenvalues of $\boldsymbol{G}$. The largest $2WT = 400$ eigenvalues are all between $\lambda_1(\boldsymbol{G}) \approx 11.43$ and $\lambda_{400}(\boldsymbol{G}) \approx 0.6597$. The smallest 1577 eigenvalues are all between $\lambda_{424}(\boldsymbol{G}) \approx 1.68 \times 10^{-14}$ and 0. Only 24 eigenvalues fail to fit in one of those ranges.

Figure 1 shows a plot of the eigenvalues of $\boldsymbol{G}$ in descending order where we have chosen $M = 2000$, $W = \frac{1}{2}$, $T = 400$, and $t_1, \ldots, t_M$ are i.i.d. Uniform$[-\frac{T}{2}, \frac{T}{2}]$. It can be seen that the first $\approx 2WT = 400$ eigenvalues are all roughly the same order of magnitude, and for $m > 2WT$, the eigenvalues $\lambda_m(\boldsymbol{G})$ decay exponentially towards zero as m increases.

We remark that this behavior is very similar to that of the eigenvalues of the prolate spheroidal wave functions (PSWF) [16]–[19]. The first $\approx 2WT$ PSWF eigenvalues are ≈ 1, and the rest exponentially decay towards zero. Also, in the specific case where the sample times are uniformly spaced, the matrix $\boldsymbol{G}$ becomes the so called prolate matrix [20] whose eigenvalues also exhibit the same clustering behavior as the PSWF eigenvalues.

V. Conjugate Gradient Descent

Conjugate gradient descent (CGD) is an iterative algorithm which aims to solve the system of equations $\boldsymbol{A}\boldsymbol{z} = \boldsymbol{y}$ for a positive definite matrix $\boldsymbol{A} \in \mathbb{C}^{M \times M}$ and a vector $\boldsymbol{y} \in \mathbb{C}^M$. If the solution to the system of equations is $\widehat{\boldsymbol{z}} = \boldsymbol{A}^{-1}\boldsymbol{y}$, and we initialize CGD to start at $\boldsymbol{z}^{(0)} = \mathbf{0}$, then the output of CGD at the k-th iteration is

$$\boldsymbol{z}^{(k)} = \underset{\boldsymbol{z} \in \mathcal{K}_k(\boldsymbol{A}, \boldsymbol{y})}{\arg\min} \ (\boldsymbol{z} - \widehat{\boldsymbol{z}})^* \boldsymbol{A} (\boldsymbol{z} - \widehat{\boldsymbol{z}})$$

where $\mathcal{K}_k(\boldsymbol{A}, \boldsymbol{y}) = \text{span}\{\boldsymbol{y}, \boldsymbol{A}\boldsymbol{y}, \boldsymbol{A}^2\boldsymbol{y}, \ldots, \boldsymbol{A}^{k-1}\boldsymbol{y}\}$ is the order-k Krylov subspace generated by $\boldsymbol{A}$ and $\boldsymbol{y}$.

As a result, it can be shown that the error after k iterations

$$\|\boldsymbol{z}^{(k)} - \widehat{\boldsymbol{z}}\|_{\boldsymbol{A}}^2 := (\boldsymbol{z}^{(k)} - \widehat{\boldsymbol{z}})^* \boldsymbol{A} (\boldsymbol{z}^{(k)} - \widehat{\boldsymbol{z}})$$

satisfies the bound

$$\|\boldsymbol{z}^{(k)} - \widehat{\boldsymbol{z}}\|_{\boldsymbol{A}}^2 \leq \|\widehat{\boldsymbol{z}}\|_{\boldsymbol{A}}^2 \cdot \min_{\substack{\text{polynomials } P \\ \deg P = k \\ P(0)=1}} \left[\max_{\lambda \in \text{Spec}(\boldsymbol{A})} |P(\lambda)|^2 \right].$$

For a general matrix $\boldsymbol{A}$, this bound is often simplified by first relaxing the maximum over $\lambda \in \text{Spec}(\boldsymbol{A})$ to the maximum over $\lambda \in [\lambda_{\min}(\boldsymbol{A}), \lambda_{\max}(\boldsymbol{A})]$, and then using properties of Chebyshev polynomials to get

$$\|\boldsymbol{z}^{(k)} - \widehat{\boldsymbol{z}}\|_{\boldsymbol{A}} \leq \|\widehat{\boldsymbol{z}}\|_{\boldsymbol{A}} \cdot 2 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^k,$$

where $\kappa = \lambda_{\max}(\boldsymbol{A}) / \lambda_{\min}(\boldsymbol{A})$ is the condition number of $\boldsymbol{A}$. Hence, CGD returns a vector $\boldsymbol{z}^{(k)}$ which satisfies $\|\boldsymbol{z}^{(k)} - \widehat{\boldsymbol{z}}\|_{\boldsymbol{A}} \leq \epsilon \|\widehat{\boldsymbol{z}}\|_{\boldsymbol{A}}$ in at most $\left\lceil \frac{1}{2}\sqrt{\kappa} \log \frac{2}{\epsilon} \right\rceil$ iterations. A more detailed discussion regarding CGD can be found in [21].

For our matrix $\boldsymbol{G} + \delta \mathbf{I}$, the largest eigenvalue is $\sim \frac{M}{2WT}$ and the smallest eigenvalue is $\approx \delta$. Hence, the condition number is roughly $\kappa \sim \frac{M}{2WT\delta}$. Typically, the regularization parameter δ will be chosen to be small (values of 10^{-2} to 10^{-5} are typical), and thus, κ will be rather large. Hence, the bound of $\left\lceil \frac{1}{2}\sqrt{\kappa} \log \frac{2}{\epsilon} \right\rceil$ iterations is worrisome. It is possible to get a better bound if we exploit the clustering behavior of the eigenvalues of $\boldsymbol{G} + \delta \mathbf{I}$.

By using the fact that $\boldsymbol{G} = \mathcal{A}\mathcal{A}^*$ has the same non-zero eigenvalues as $\mathcal{A}^*\mathcal{A} = \sum_{m=1}^M \boldsymbol{a}_{t_m} \boldsymbol{a}_{t_m}^*$, matrix concentration inequalities from [22], and bounds on the number of prolate spheroidal wave function eigenvalues in the so called "plunge region" [23], we can get the following result.

Lemma 1. *If the sample times $t_1, \ldots, t_M$ are i.i.d. Uniform$[-\frac{T}{2}, \frac{T}{2}]$, then there exist constants $C_1, C_2, C_3 > 0$ and indices m_1 and m_2 such that*

$$m_2 - m_1 \leq C_3 \log(2WT) \log \left(\frac{M^2}{\delta^{9/4} \epsilon^{1/4}} \right)$$

and the following bounds hold simultaneously with probability at least $1 - 4WTe^{-\frac{M}{3WT}}$:

$$\lambda_1(\boldsymbol{G}) \leq C_1 \frac{M}{2WT},$$

$$\lambda_{m_1}(\boldsymbol{G}) \geq C_2 \frac{M}{2WT},$$

$$\lambda_{m_2}(\boldsymbol{G}) \leq 2\delta(\delta\epsilon)^{1/8}.$$

Furthermore, the following lemma gives a bound of the number of CGD iterations required when working with a matrix with similar eigenvalue clustering behavior.

Lemma 2. *Let $\boldsymbol{A} \in \mathbb{C}^{M \times M}$ be a positive-semidefinite matrix. Let $\boldsymbol{y} \in \mathbb{C}^M$ and define $\widehat{z} = \boldsymbol{A}^{-1}\boldsymbol{y}$. Let $\boldsymbol{z}^{(k)} \in \mathbb{C}^M$ be the CGD iterates with initial point $\boldsymbol{z}^{(0)} = \mathbf{0}$, and let $\epsilon > 0$ be the desired CGD tolerance. Suppose there exist real numbers $\delta, a, b, c_1, c_2, \ldots, c_L$ with $\delta + 2\delta(\delta\epsilon)^{1/8} < c_1 < c_2 < \cdots < c_L < a < b$ such that*

$$\textit{Spec}(\boldsymbol{A}) \subseteq \left[\delta, \delta + 2\delta(\delta\epsilon)^{1/8}\right] \bigcup \{c_\ell\}_{\ell=1}^L \bigcup [a, b].$$

Then, after

$$k = \left\lceil \frac{(L+d)\ln\left(\frac{b-\delta}{\delta}\right) + \ln\left(\frac{2}{\delta\epsilon}\right)}{\ln\left(\frac{\sqrt{b/a}+1}{\sqrt{b/a}-1}\right)} \right\rceil + L + 8$$

iterations, we have $\|\boldsymbol{z}^{(k)} - \widehat{\boldsymbol{z}}\|_2 \leq \epsilon \|\boldsymbol{y}\|_2$.

The proof of this lemma invokes properties of Chebyshev polynomials to explicitly construct a polynomial $P(\lambda)$ such that $P(0) = 1$, $P(c_\ell) = 0$ for $\ell = 1, \ldots, L$, and $|P(\lambda)| \leq \delta\epsilon$ for $\lambda \in [\delta, \delta + 2\delta(\delta\epsilon)^{1/8}]$ and $\lambda \in [a, b]$. Then, the degree of the polynomial is a bound on the number of CGD iterations needed for convergence.

If we apply lemma 2 to the matrix $\boldsymbol{G} + \delta\mathbf{I}$ along with the eigenvalue bounds from lemma 1, we get that CGD will only need $O(\text{polylog}(2WT, \frac{M}{2WT}, \frac{1}{\delta}, \frac{1}{\epsilon}))$ iterations to return a solution $\widehat{\boldsymbol{z}}$ satisfying $\|\widehat{\boldsymbol{z}} - (\boldsymbol{G} + \delta\mathbf{I})^{-1}\boldsymbol{y}\|_2 \leq \epsilon \|\boldsymbol{y}\|_2$. Due to space constraints, the proofs of these lemmas will be deferred to a future publication.

VI. Extension to Multiband Signals

The problem setup has a fairly straightforward extension to multiband signals, i.e. signals of the form

$$x_c(t) = \sum_{\ell=1}^L \int_{f_\ell - W_\ell}^{f_\ell + W_\ell} X(f) e^{j2\pi f t} \, df,$$

where the intervals $[f_\ell - W_\ell, f_\ell + W_\ell]$ are disjoint. The reconstructed signal is again a linear combination of functions $\phi(t)$ that are each shifted by the sample times

$$\widehat{x}_c(t) = \sum_{m=1}^M \boldsymbol{z}[m] \phi(t - t_m).$$

However, instead of a sinc function, $\phi(t)$ is a sum of sincs that are modulated to the center frequency of each interval, i.e.

$$\phi(t) = \sum_{\ell=1}^L \frac{\sin(2\pi W_\ell t)}{\pi t} e^{j2\pi f_\ell t}.$$

As in the single band case, we have to solve the system of equations

$$(\boldsymbol{G} + \delta\mathbf{I})\boldsymbol{z} = \boldsymbol{y}$$

to get the coefficients $\boldsymbol{z}[m]$ where $\boldsymbol{G}$ is an $M \times M$ matrix defined by

$$\boldsymbol{G}[m, m'] = \phi(t_m - t_{m'}).$$

After obtaining the coefficients $\boldsymbol{z}[m]$, we can evaluate the reconstruction at times $s_1, \ldots, s_N$ via

$$\begin{bmatrix} \widehat{x}_c(s_1) & \cdots & \widehat{x}_c(s_N) \end{bmatrix}^T = \boldsymbol{H}\boldsymbol{z}$$

where $\boldsymbol{H}$ is an $N \times M$ matrix defined by

$$\boldsymbol{H}[n, m] = \phi(s_n - t_m).$$

Similarly to the single band case, we can show that $\boldsymbol{G}+\delta\mathbf{I}$ is a generalized Cauchy-like matrix with $r = L$, and $\boldsymbol{H}$ is a generalized Cauchy matrix with $r = 2L$. As such, $\boldsymbol{G}+\delta\mathbf{I}$ and $\boldsymbol{H}$ can be applied to an $M \times 1$ vector in $O(LM \log \frac{1}{\alpha})$ operations and $O(L(N + M) \log \frac{1}{\alpha})$ operations respectively where α is the approximation tolerance. Unfortunately, we do not yet have a theoretical bound on the number of CGD iterations needed for convergence in the multiband case. However, we hypothesis the number of iterations to be $O(L\text{polylog}(2WT, \frac{M}{2WT}, \frac{1}{\delta}, \frac{1}{\epsilon}))$ with high probability.

VII. Experiments

We run a synthetic experiment to test the efficiency of our proposed method for multiband signal reconstruction as the number of samples M gets large. We first generate a multiband signal $x_c(t)$ whose Fourier transform is supported on $f \in [-0.9, -0.6] \cup [0.1, 0.2] \cup [0.9, 1.0]$ by summing several sinusoids at random frequencies in those bands. For several values of M between 2^{10} and 2^{18}, we pick T such that $M \approx 2W_{\text{total}} T \log\left(\frac{2W_{\text{total}} T}{0.01}\right)$ where $W_{\text{total}} = 0.5$ is the total occupied bandwidth, and then draw M random sample times $t_1, \ldots, t_M$ i.i.d. Uniform$[-\frac{T}{2}, \frac{T}{2}]$. This choice of T ensures that the spectrum of $\boldsymbol{G}$ is very likely to have the clustering behavior described in section IV. We then set $\delta = 10^{-4}$ and attempt reconstruct the signal on a grid of $N = M$ uniformly spaced sample times in $[-\frac{T}{2}, \frac{T}{2}]$ using three methods:

- Use CGD along with the fast method for applying $\boldsymbol{G}+\delta\mathbf{I}$ to solve $(\boldsymbol{G} + \delta\mathbf{I})\boldsymbol{z} = \boldsymbol{y}$. Then, use the fast method for computing $\boldsymbol{H}\boldsymbol{z}$ to evaluate $\widehat{x}_c(t)$ at the uniformly spaced times.
- Use CGD to solve $(\boldsymbol{G} + \delta\mathbf{I})\boldsymbol{z} = \boldsymbol{y}$, but explicitly form $\boldsymbol{G} + \delta\mathbf{I}$. Then, explicitly form $\boldsymbol{H}$ to evaluate $\widehat{x}_c(t)$ at the uniformly spaced times.
- Solve the system $(\boldsymbol{G} + \delta\mathbf{I})\boldsymbol{z} = \boldsymbol{y}$ using MATLAB's backslash operator. Then, explicitly form $\boldsymbol{H}$ to evaluate $\widehat{x}_c(t)$ at the uniformly spaced times.

Note that due to memory constraints, we were only able to test the 2nd and 3rd methods for $M < 2^{15}$. For each value of M, we repeat this experiment 10 times, to get an accurate average result. The average time to compute the reconstructed signal at the uniform grid of sample times versus the number of samples is shown in figure 2 and the average relative RMS error of the reconstructed uniform samples versus the number of samples is shown in figure 3. All three methods achieve nearly identical reconstruction errors. For $M > 2^{11}$, our proposed method is noticeably faster than the methods which don't take advantage of the structure of $\boldsymbol{G}+\delta\mathbf{I}$ and $\boldsymbol{H}$. Also, the total computation time needed for our method scales roughly linearly with the number of samples. The average number of CGD iterations vs. the number of samples M is shown in figure 4. By using the fast and approximate method for applying $\boldsymbol{G} + \delta\mathbf{I}$, CGD takes slightly more iterations to converge.

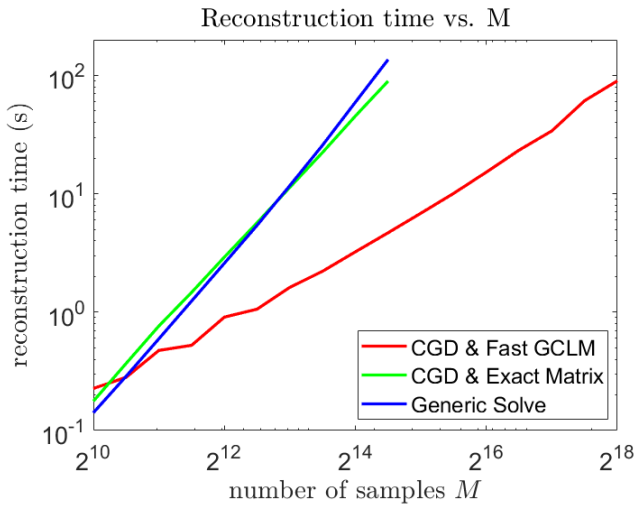

Fig. 2. Plot of the time needed for each of the three methods to compute the reconstructed signal on a uniformly spaced grid of M points from the M nonuniformly spaced samples.

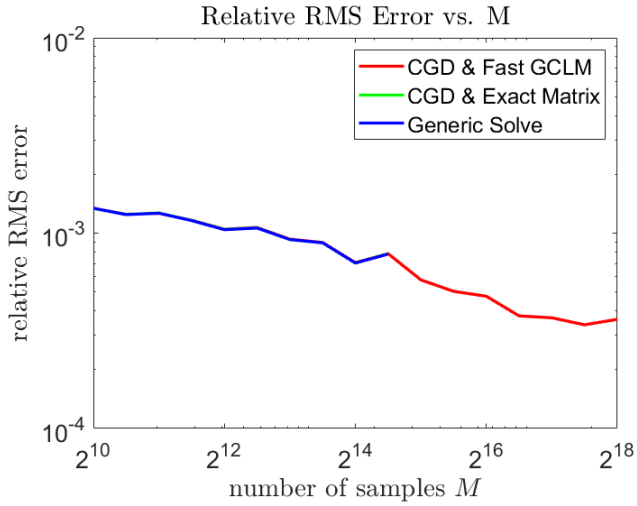

Fig. 3. Plot of the relative RMS error of the reconstructed signal for each of the three methods. All three methods yield a nearly identical reconstruction error for the values of M for which all of them could be tested.

Acknowledgements

This work was supported, in part, by NSF grants CCF-1350616 and CCF-1409406, a grant from Lockheed Martin, and a gift from the Alfred P. Sloan Foundation.

References

[1] S. Maymon and A. V. Oppenheim, "Sinc interpolation of nonuniform samples," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4745–4758, 2011.

[2] F. Marvasti, *Nonuniform sampling: theory and practice*. Springer Science & Business Media, 2012.

[3] J. J. Benedetto and W. Heller, "Irregular sampling and the theory of frames, i," *Note di Matematica*, vol. 10, no. suppl. 1, pp. 103–125, 1990.

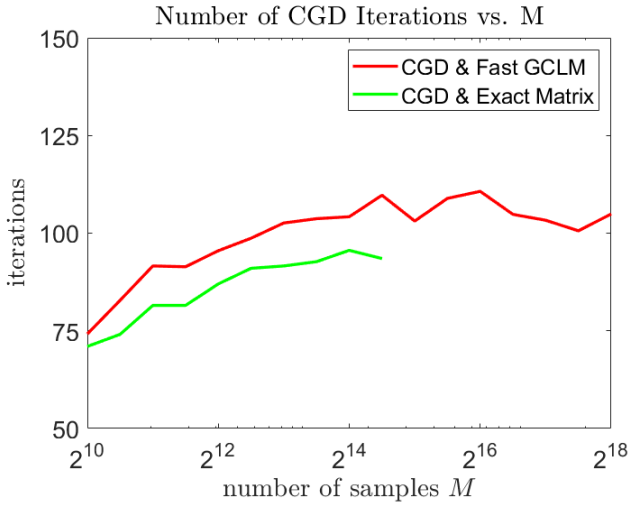

Fig. 4. Plot of the number of CGD iterations needed for the first two methods versus the number of samples M.

[4] J. J. Benedetto and S. Scott, *Frames, Irregular Sampling, and a Wavelet Auditory Model*. Boston, MA: Springer US, 2001, pp. 585–617. [Online]. Available: https://doi.org/10.1007/978-1-4615-1229-5_14

[5] O. Christensen *et al.*, *An introduction to frames and Riesz bases*, 2nd ed. Springer, 2016.

[6] H. G. Feichtinger, K. Grochenig, and T. Strohmer, "Efficient numerical methods in non-uniformsampling theory," *Numerische Mathematik*, vol. 69, no. 4, pp. 423–440, Feb 1995. [Online]. Available: https://doi.org/10.1007/s002110050101

[7] D. J. Wingham, "The reconstruction of a band-limited function and its fourier transform from a finite number of samples at arbitrary locations by singular value decomposition," *IEEE Transactions on Signal Processing*, vol. 40, no. 3, pp. 559–570, March 1992.

[8] A. M. Davie and A. J. Stothers, "Improved bound for complexity of matrix multiplication," *Proceedings of the Royal Society of Edinburgh: Section A Mathematics*, vol. 143, no. 2, p. 351–369, 2013.

[9] D. Coppersmith and S. Winograd, "Matrix multiplication via arithmetic progressions," *Journal of Symbolic Computation*, vol. 9, no. 3, pp. 251 – 280, 1990, computational algebraic complexity editorial. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0747717108800132

[10] F. Le Gall, "Faster algorithms for rectangular matrix multiplication," in *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, Oct 2012, pp. 514–523.

[11] L. Greengard and V. Rokhlin, "A fast algorithm for particle simulations," *Journal of Computational Physics*, vol. 73, no. 2, pp. 325 – 348, 1987. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0021999187901409

[12] J. Carrier, L. Greengard, and V. Rokhlin, "A fast adaptive multipole algorithm for particle simulations," *SIAM Journal on Scientific and Statistical Computing*, vol. 9, no. 4, pp. 669–686, 1988. [Online]. Available: https://doi.org/10.1137/0909044

[13] A. Dutt, M. Gu, and V. Rokhlin, "Fast algorithms for polynomial interpolation, integration, and differentiation," *SIAM Journal on Numerical Analysis*, vol. 33, no. 5, pp. 1689–1711, 1996. [Online]. Available: http://www.jstor.org/stable/2158326

[14] J.-P. Cardinal, "On a property of cauchy-like matrices," *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, vol. 328, no. 11, pp. 1089 – 1093, 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0764444299803309

[15] V. Y. Pan and A. Zheng, "Superfast algorithms for Cauchy-like matrix computations and extensions," *Linear Algebra and its Applications*, vol. 310, no. 1, pp. 83 – 108, 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0024379500000410

[16] D. Slepian and H. Pollak, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty. I," *Bell Systems Tech. J.*, vol. 40, no. 1, pp. 43–64, 1961.

[17] H. Landau and H. Pollak, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty. II," *Bell Systems Tech. J.*, vol. 40, no. 1, pp. 65–84, 1961.

[18] ——, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty. III," *Bell Systems Tech. J.*, vol. 41, no. 4, pp. 1295–1336, 1962.

[19] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty. IV," *Bell Systems Tech. J.*, vol. 43, no. 6, pp. 3009–3058, 1964.

[20] ——, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty. V – The discrete case," *Bell Systems Tech. J.*, vol. 57, no. 5, pp. 1371–1430, 1978.

[21] J. R. Shewchuk, "An introduction to the conjugate gradient method without the agonizing pain," 1994.

[22] J. A. Tropp *et al.*, "An introduction to matrix concentration inequalities," *Foundations and Trends in Machine Learning*, vol. 8, no. 1-2, pp. 1–230, 2015.

[23] A. Osipov, "Certain upper bounds on the eigenvalues associated with prolate spheroidal wave functions," *Applied and Computational Harmonic Analysis*, vol. 35, no. 2, pp. 309 – 340, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1063520313000237