OPTIMAL SAMPLING OF CONVOLUTION KERNELS IN W-PROJECTION ALGORITHM

Sarod Yatawatta

ASTRON, The Netherlands Institute for Radio Astronomy,
The Netherlands.
Email: yatawatta@astron.nl

ABSTRACT

Widefield imaging using radio interferometers has a fundamental problem of all the baselines not being on a plane during a long synthesis observation. The most effective way to solve this is by using the w-projection algorithm, which projects the data onto a common plane during convolutional gridding. The exact calculation of the convolution kernel for all values of the w coordinate is expensive, thus it is only calculated for a pre-determined set of values and later interpolated. Determining the best set of values of w to pre-calculate the convolution kernels is an open problem. In this paper, we derive an error bound for convolutional gridding using an interpolated version of convolution kernels calculated at a subset of sampled values in w. Using this bound, we select improved sampling schemes in w to minimize the error in convolutional gridding.

Index Terms— Instrumentation: interferometers; Methods: numerical; Techniques: interferometric

1. INTRODUCTION

Widefield imaging is a crucial stage of radio interferometric data processing, and during this stage, the observed data in the Fourier space are converted to an equivalent real space image. Even when the scientific interest lies on a small area (or a narrow field) in the sky, widefield imaging is still essential to get an accurate sky model for the rest of the sky, thereby improving the quality of calibration of the data [1, 2]. However, imaging a wide field of view can only be achieved with increased computational cost. This is mainly because of the inherent three dimensional sampling of Fourier space which becomes significant in widefield imaging [3]. As a consequence, straightforward use of the two dimensional fast Fourier transform (FFT) is not possible. Therefore, the so called *w-projection* and assorted algorithms [4, 5] are widely used to provide an efficient and accurate solution to this problem.

The w-projection algorithm incorporates a minor modification to the convolution kernels during gridding of Fourier space data such that the gridded data appear to be sampled on a two dimensional plane. Prolate spheroidal wave functions (PSWF) are the de-facto standard used in convolutional gridding [6]. The w-kernel is convolved with the PSWF kernel to get the convolution kernel used in w-projection. We emphasize that the final convolution kernel cannot be calculated in closed form. Therefore, for a pre-determined values of the w-coordinate, this kernel is pre-calculated and afterwards, for any given value of w, an interpolated kernel is used for convolutional gridding. Once the data is gridded onto a two dimensional plane, the FFT can be used for fast image generation.

The selection of a subset of w-coordinates to pre-calculate the convolution kernel can be done in many ways. For instance [4] have proposed square-root spacing in w to pre-calculate the convolution kernel. On the other hand, recent work by [7] have proposed linear spacing. The lack of a formal mathematical framework to evaluate the performance of various sampling schemes of the w coordinate hinders comparisons between such different schemes. This paper focuses exclusively on solving this. We derive error bounds for convolutional gridding that uses interpolated convolution kernels, as opposed to using convolution kernels calculated exactly for any given w coordinate. We also note that we restrict our analysis to piece-wise linear interpolation, which has the least computational cost. Our ultimate objective is to find the optimal sampling in w-coordinate, for a fixed number of samples. In order to do this, we solve a constrained optimization problem that minimizes the cumulative error in gridding, subject to the number of convolution kernels that are pre-calculated being fixed.

Only a few similar problems have been solved in existing research. For instance, [8] consider image based rendering of 3 dimensional scenes in computer vision, using 2 dimensional images captured by a set of cameras. The authors use position of cameras (w coordinate in our case) and the spacing (spacing between two w planes) to define bounds on calculating rendering error. This error is defined as position-interval error (PIE) function. No direct optimization problem is solved but by using the PIE function, they generate sampling points to satisfy various requirements in computer vision applications. On the other hand, in [9], pre-destorter (compander)

This work is supported by the European Research Council project LOFARCORE, grant #339743.

design in communication systems is considered. In this case, by using direct optimization of residual distortion, they manage to design optimal pre-distorter functions. We should also mention that as a general rule the spacing in sampling of the w-coordinate should be inversely proportional to the derivative of the function that is being interpolated. But this is too general and therefore in this paper, we follow a more rigorous approach. We emphasize that each aforementioned problems are unique and therefore, we revert to fundamentals to solve our problem in this paper.

The rest of the paper is organized as follows: In section 2 we give an overview of convolutional gridding and w-projection algorithm. Next, in section 3, we derive bounds for the convolutional gridding error using piecewice linear interpolated convolution kernels. Using this bound, in section 4, we formulate a constrained optimization problem that finds the optimum sampling points in w coordinate to pre-calculate convolution kernels. We demonstrate the results derived in preceding sections in section 5 with various simulations. Finally, we draw our conclusions in section 6.

Notation: Convolution is denoted by $\otimes$. The complex error function is given by $\mathrm{erf}(.)$.

2. CONVOLUTIONAL GRIDDING AND W-PROJECTION

Interferometric data are sampled at irregularly spaced points in the Fourier space. Therefore, to apply the fast Fourier transform, the data is first gridded onto a regular grid. This is termed as convolutional gridding [10, 11] where the data is first convolved with a kernel before resampling onto a regular grid. Prolate spheroidal wave functions [12] are normally used as the convolution kernel.

For data with non zero values of w, the convolution kernel is modified such that the gridded data is projected onto a plane [4] (normally $w = 0$ is used). In spite of the fact that the sampling points in the Fourier space are three dimensional, w-projection enables us to use the two dimensional FFT. This enables a significant computational cost reduction, especially in making widefield images.

We denote the coordinates in the Fourier space as (u, v, w) and on the image plane as (l, m). Given the sky signal $s(l, m)$, the corresponding data observed at (u, v, w) in the Fourier space $S(u, v, w)$ is

$$\begin{aligned} &S(u,v,w) \quad (1)\\ &= \int\int \frac{s(l,m)e^{-2\pi\jmath(ul+vm)}}{\sqrt{1-l^2-m^2}} e^{-2\pi\jmath w(\sqrt{1-l^2-m^2}-1)} dl dm \end{aligned}$$

where

$$f(l,m,w) \stackrel{\triangle}{=} \exp\left(-2\pi\jmath w(\sqrt{1-l^2-m^2}-1)\right) \quad (2)$$

is the term that violates the simple 2 dimensional Fourier relation for $w > 0$. For small l, m we can approximate this as

$$f(l,m,w) \approx \exp\left(-2\pi\jmath w(-l^2/2 - m^2/2)\right) \quad (3)$$

with the Fourier transform of the approximation being

$$F(u,v,w) \approx \frac{\jmath}{w} \exp \frac{\pi\jmath(u^2+v^2)}{w}. \quad (4)$$

Note that we use the exact value given by (2) in actual gridding of the data, however, only in our analysis, we use the approximation (3) throughout the paper.

Given the Prolate spheroidal convolution kernel $c(l, m)$ (and its Fourier equivalent $C(u, v)$), for $w > 0$, the kernel used for gridding is $F(u, v, w) \otimes C(u, v)$. We cannot calculate this in closed form and therefore, this is calculated by taking the Fourier transform of the product $f(l, m, w)c(l, m)$, with appropriate padding and oversampling. We see that computing this for every value of w in the data is computationally not feasible. Therefore, in practice, we pre-calculate $F(u, v, w) \otimes C(u, v)$ at a selected subset of w-coordinate values that cover the full range of possible values of w. Afterwards, during gridding of data, we use (linear) interpolation of the pre-calculated kernels to get an interpolated kernel.

Given any observed data point $S(u, v, w)$ in Fourier space, convolution with the kernel (say projected onto $w = 0$ plane) gives us $\widetilde{S}(u, v, w = 0)$

$$\widetilde{S}(u,v,w=0) = S(u,v,w) \otimes (F(u,v,w) \otimes C(u,v)), \quad (5)$$

which is sampled on a regular u, v grid before taking the 2 dimensional FFT. Thereafter, we extract the true image of the sky $s(l, m)$ by apodization correction [13].

An important property that we need for the rest of the paper is the support of the convolution kernel. This is essentially the bandwidth of $f(l, m, w)$ in (2) plus the bandwidth of $c(l, m)$. Generally the bandwidth of the PSWF is much less compared to the bandwidth of $f(l, m, w)$ (especially for large w), therefore we ignore the bandwidth of $c(l, m)$ in our analysis. There is no direct way of calculating the bandwidth of $f(l, m, w)$, but we can use Carson's rule [14] that is used to (approximately) calculate the bandwidth of frequency modulated communication signals. Using this rule, we see that (2) is a signal with zero carrier frequency and with maximum frequency deviation of $2\pi w(1 - \sqrt{1 - \overline{l}^2 - \overline{m}^2})$ where $\overline{l}$ and $\overline{m}$ are the maximum possible values of l and m. If the field of view is $l \in [-L, L]$,$m \in [-L, L]$, and for $L \ll 1$ we get the approximate upper limit for the bandwidth of $f(l, m, w)$ as $2\pi wL^2$. Therefore, ignoring the bandwidth of PSWF $c(l, m)$, we get the total support of the convolution kernel as $2\pi wL^2$.

3. INTERPOLATION ERROR

To reduce computational cost, for any data point at u, v, w, we do not calculate $F(u, v, w) \otimes C(u, v)$ exactly. Instead we first find w_0 and w_1 such that $w_0 \leq w \leq w_1$ where

w_0 and w_1 already have pre-calculated convolution kernels, i.e. $F(u,v,w_0) \otimes C(u,v)$ and $F(u,v,w_1) \otimes C(u,v)$. In our analysis, we consider that these values are linearly interpolated to find the convolution kernel at w. Linear interpolation is especially suitable for hardware accelerated convolutional gridding for instance by using texture memory in a graphics processing unit (GPU).

Our objective in this section is to find an error bound for the gridded data due to this interpolation. In order to do this, we use Young's inequality for convolution. To elaborate this, let $h(x)$ and $g(x)$ be two functions that are Fourier transformable. Applying Young's inequality [15] to the convolution $h(x) \otimes g(x)$, we get

$$\|h \otimes g\|_r \leq \|h\|_p \|g\|_q, \ \frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1, \ 1 \leq p,q,r \leq \infty \quad (6)$$

where $\|h\|_p$ is the p-norm,

$$\|h\|_p = \left(\int_\Omega |h(x)|^p dx \right)^{\frac{1}{p}} \quad (7)$$

where Ω defines the convolution support.

Applying (6) to (5), we get

$$\|\tilde{S}\|_t \leq \|S\|_p \|F\|_q \|C\|_r \quad (8)$$

where t,p,q,r are appropriately chosen to satisfy (6). Our interest lies in studying $\|F\|_q$. In order to do this, we define a function

$$f(w) \triangleq \int_{-kw}^{kw} \int_{-kw}^{kw} F(u,v,w) du dv \quad (9)$$

with $k = 2\pi L^2$ and let $\widetilde{f}(w)$ be the same function obtained by linear interpolation of $F(u,v,w_0)$ and $F(u,v,w_1)$. We get the error bound for linear interpolation as

$$|f(w) - \widetilde{f}(w)| \leq \frac{1}{8} |w_0 - w_1|^2 \max_{w \in [w_0, w_1]} |f''(w)|. \quad (10)$$

In order to find $f''(w)$ in (10), we apply Leibniz's integral rule

$$\begin{aligned} &\frac{d}{dw} \int_{a(w)}^{b(w)} g(v,w) dv \quad (11) \\ &= \int_{a(w)}^{b(w)} \frac{\partial}{\partial w} g(v,w) dv \\ &+ g(b(w),w) b'(w) - g(a(w),w) a'(w) \end{aligned}$$

where $g(v,w)$ is any differentiable function and $a(w)$ and $b(w)$ are limits of integration [16]. Note that we have to apply (11) twice to (9) and we also exploit the symmetry $F(u,v,w) = F(-u,v,w) = F(u,-v,w) = F(-u,-v,w)$.

After some algebra, we get

$$\begin{aligned} &f''(w) \quad (12) \\ &= -\frac{k}{4} \frac{\exp(\jmath k^2 \pi w)}{w^{1.5}\sqrt{-\jmath}} \Bigg(2(2k^2\pi w + \jmath) \mathrm{erf}(k\sqrt{-\jmath w \pi}) \\ &+ \sqrt{-2\jmath}(2\pi k^2 w(\jmath + 1) + \jmath - 1) \mathrm{erf}\left(\frac{(1-\jmath)}{2} k\sqrt{2\pi w} \right) \Bigg) \\ &+ 2\jmath \frac{k^2}{w} \exp(2\jmath k^2 \pi w). \end{aligned}$$

Given any value of w, we first find the lower and upper planes (bracketing terms) w_0 and w_1 used in the interpolation to calculate the convolution kernel. Afterwards, using (10) and (12), we can estimate an upper bound for the interpolation error. We take the highest possible error, at $(w_0 + w_1)/2$ to evaluate $|f''(w)|$.

4. OPTIMAL SAMPLING

Let the maximum (absolute) value of the w coordinate be W, and we are allowed to only pre-calculate $N+1$ convolution kernels in the w range $[0, W]$. The practical reason for limiting the maximum number of pre-calculated convolution kernels is the memory limit of a compute device such as a GPU. Therefore, we need to find $N+1$ values of w within this range that minimizes the error due to interpolation, and yet covers the complete range $[0, W]$. We reformulate this problem as follows. Let $\mathbf{x} = [x_1, x_2, \ldots, x_N]^T$ be a vector of size $N \times 1$ such that, the i-th value of w coordinate is calculated as

$$w_i = x_1 + x_2 + \ldots \ldots x_i, \ i \in [1, N], w_0 = 0 \quad (13)$$

and given w_i, the $(i+1)$-th value is calculated as $w_i + x_{i+1}$. We define the i-th cost function, or the i-th interpolation error using (10) as

$$J_i(\mathbf{x}) \triangleq \alpha |x_i|^4 \left| f''\left(\frac{(w_{i-1} + w_i)}{2} \right) \right|^2 \quad (14)$$

where α is a constant that we can choose to enlarge the error for numerical algorithms to work (typically 10^3 is sufficient).

Hence, we can restate our problem as

$$\begin{aligned} \mathbf{x} = \arg\min_{\mathbf{x}} \ &\{J_1(\mathbf{x}), J_2(\mathbf{x}), \ldots, J_N(\mathbf{x})\} \quad (15) \\ \text{subject to } &\sum_i x_i = W, \quad x_i > 0 \quad i \in [1, N] \end{aligned}$$

that gives $\mathbf{x}$ as the spacings in w-coordinate. Once we have found $\mathbf{x}$, we use (13) to calculate the values of w at which, we need to pre-calculate convolution kernels.

Note that (15) does not have a single cost function, in fact, it is a multi-objective optimization problem with more than one cost function that needs to be minimized. There are various ways of solving this problem and in this paper, we use goal attainment [17, 18] as our optimization scheme. We omit the details of this method here and refer the reader to some recent examples of its application [19, 20].

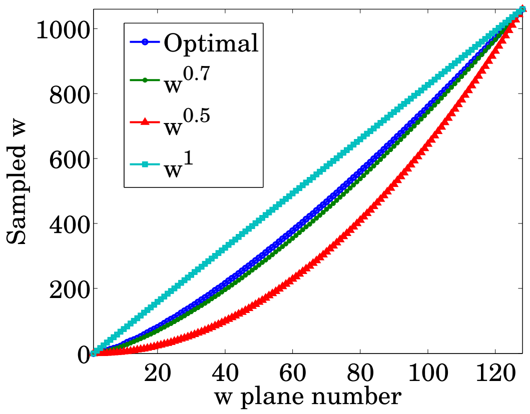

Fig. 1. Four different schemes of sampling the w axis, linear w^1, square root $w^{0.5}$, $w^{0.7}$ and optimal.

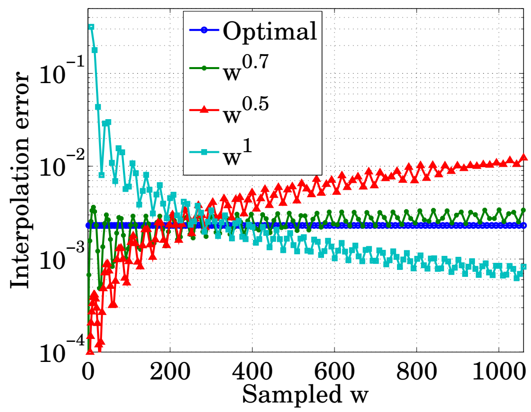

Fig. 2. Interpolation error for the four different schemes of sampling the w axis.

5. SIMULATIONS

In this section, we give simulation results to compare the existing approaches of sampling the w axis to the one proposed in this paper. We consider sampling the w axis in the range $[0, 1060]$ wavelengths ($W = 1060$). We are only allowed to pre-compute 128 convolution kernels and hence, we can only sample 128 values in w, i.e., $N = 127$. We always use the starting value $w_0 = 0$, in other words, the first convolution kernel is always evaluated at $w = 0$. We use these convolution kernels to make images of a field of view of about 30×30 degrees, covered using 4000×4000 pixels with pixel size $30''$.

In Fig. 1, we show the different approaches of sampling the w axis. The commonly used schemes are linear sampling with w^1 spacing and square root sampling with $w^{0.5}$ spacing. We also use the scheme proposed in section 4 to find the optimal sampling, using goal attainment algorithm in MATLAB. In the goal attainment algorithm, we use the intersection of hypographs of the error bounds obtained for w^1 spacing and $w^{0.5}$ spacing as our goal. The solution thus obtained is very similar to a spacing of $w^{0.7}$.

We show the interpolation error bounds obtained for various sampling schemes in Fig. 2.We see that w^1 sampling scheme has higher error at lower values of w while lower error at higher values of w. On the other hand, the $w^{0.5}$ scheme has the exact opposite behavior. The optimal sampling scheme proposed in this paper has evenly distributed error for all values of w, which is closely matched by the $w^{0.7}$ sampling scheme.

Using the pre-computed convolution kernels (of size 2304×2304 pixels) at the w values determined by the aforementioned sampling schemes, we make images of a simulated sky [21]. The simulated sky consists of point sources of unit magnitude flux at various locations in the sky. We compare the recovered peak magnitude flux and find the flux loss. The flux loss is mainly due to the error in calculating the convolution kernels by interpolation of pre-computed kernels. We show the results in Fig. 3. We clearly see that the optimal sampling scheme gives the lowest flux loss, which is slightly better than the $w^{0.7}$ sampling scheme. In contrast, the existing approaches of linear and square root sampling have much higher flux loss. We note here that solving (15) needs to be done only once, and the computations required for this is negligible compared to the cost of convolutional gridding. Therefore, without expending any additional computations, we can improve the accuracy of widefield imaging by selecting the optimal sampling scheme proposed in this paper. Even when solving (15) is not feasible, by choosing the $w^{0.7}$ sampling, we still can get a better result than existing schemes. Further improvements to minimize flux loss can be made by fine tuning the size of the convolution kernels and increasing the number of the pre-calculated kernels.

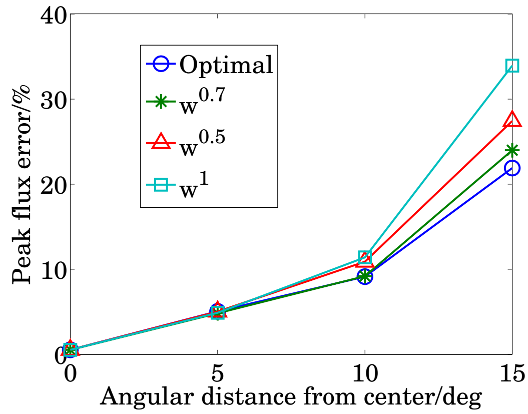

Fig. 3. Error in the recovered flux (flux loss), with the distance from the center of the image.

6. CONCLUSIONS

We have analyzed the error due to interpolation in calculating convolution kernels in widefield radio interferometric imaging using the w-projection algorithm. Using this analysis, we proposed an optimal sampling scheme for the w coordinate to pre-compute convolution kernels. Simulations show improvement in imaging accuracy by using this scheme, which requires very little additional computations. However, we still get substantial flux loss for sources far away from the center. In addition to optimal w sampling, we also need to fine tune the sizes of the convolution kernels to improve this, which is left as future work.

7. ACKNOWLEDGMENTS

We thank Ian Avruch and Wim Brouw for valuable discussions.

8. REFERENCES

[1] J. Bregman, "System design and widefield imaging aspects of synthesis arrays with phased array stations," *PhD Thesis, Univ. Groningen*, Dec. 2012.

[2] S. Yatawatta, A. G. de Bruyn, M. A. Brentjens, P. Labropoulos et al., "Initial deep LOFAR observations of epoch of reionization windows. I. The north celestial pole," *Astronomy & Astrophysics*, vol. 550, pp. A136, Feb. 2013.

[3] T. J. Cornwell and R. A. Perley, "Radio-interferometric imaging of very large fields - The problem of non-coplanar arrays," *Astronomy and Astrophysics*, vol. 261, pp. 353–364, July 1992.

[4] T.J. Cornwell, K. Golap, and S. Bhatnagar, "The non-coplanar baselines effect in radio interferometry: The w-projection algorithm," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 2, no. 5, pp. 647–657, Oct 2008.

[5] S. Bhatnagar, T. J. Cornwell, K. Golap, and J. M. Uson, "Correcting direction-dependent gains in the deconvolution of radio interferometric images," *Astronomy and Astrophysics*, vol. 487, pp. 419–429, Aug. 2008.

[6] F. R. Schwab, "Suppression of Aliasing by Convolutional Gridding Schemes," *VLA Scientific Memorandum 129*, 1978.

[7] A. R. Offringa et al., "WSCLEAN: an implementation of a fast, generic wide-field imager for radio astronomy," *Mon. Not. R. Astron. Soc.*, vol. 444, pp. 606–619, Oct. 2014.

[8] Cha Zhang and Tsuhan Chen, "Non-uniform sampling of image-based rendering data with the position-interval error (PIE) function," in *in Proc. VCIP*, 2003.

[9] S. N. Ba, K. Waheed, and G. T. Zhou, "Optimal spacing of a linearly interpolated complex-gain LUT predistorter," *Vehicular Technology, IEEE Transactions on*, vol. 59, no. 2, pp. 673–681, Feb 2010.

[10] W. N. Brouw, "Aperture synthesis," *in Methods in Computational Physics*, vol. 14, pp. 131–175, 1975.

[11] F. R. Schwab, "Optimal gridding," *VLA Scientific Memorandum 132*, pp. 1–18, 1980.

[12] J. Xiao, V. Rokhlin, and N. Yarvin, "Prolate spheroidal wavefunctions, quadrature and interpolation," *Inverse Problems*, vol. 17, pp. 805–838, Dec. 2001.

[13] S. Yatawatta, "Adaptive weighting in radio interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 444, pp. 790–796, Oct. 2014.

[14] John R. Carson, "Notes on the theory of modulation," *Proceedings of the IEEE*, vol. 51, no. 6, pp. 893–896, June 1963.

[15] J.F. Fournier, "Sharpness in Young's inequality for convolution," *Pacific J. Math.*, vol. 72, no. 2, pp. 383–397, 1977.

[16] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions : With Formulas, Graphs, and Mathematical Tables.*, Dover Publications, New York, 1970.

[17] F.W. Gembicki, *Vector Optimization for Control with Performance and Parameter Sensitivity Indices*, PhD Thesis, Case Western Reserve Univ., 1974.

[18] F. Gembicki and Y.Y. Haimes, "Approach to performance and sensitivity multiobjective optimization: The goal attainment method," *Automatic Control, IEEE Transactions on*, vol. 20, no. 6, pp. 769–771, Dec 1975.

[19] J. Dobes and J. Michal, "An implementation of the circuit multiobjective optimization with the weighted sum strategy and goal attainment method," in *Circuits and Systems (ISCAS), 2011 IEEE International Symposium on*, May 2011, pp. 1728–1731.

[20] J. Dobes, J. Michal, F. Vejrazka, J. Popp, and V. Panko, "Multiobjective optimization of input low noise amplifier for common GPS/Galileo/GLONASS/Compass satellite navigation system receiver," in *SoC Design Conference (ISOCC), 2014 International*, Nov 2014, pp. 88–89.

[21] S. Yatawatta, "Excon: Making billion pixel radio interferometric images," *in proc. URSI GASS 2014, Beijing, China*, Aug. 2014.