Draft version April 27, 2021
Typeset using LaTeX **twocolumn** style in AASTeX631

Perturbative Approach to Solve the Map-Making Equation

Bai-Qiang Qiang[1] and Kevin M. Huffenberger[1]

[1] *Florida State University*

ABSTRACT

Some abstract

Keywords: Some Keywords

1. INTRODUCTION

Map-making is an intermeidate process between data collection and estimate various cosmological parameters. As the next generation CMB observations will have much higher resolution and generate more data, we need an efficient way to process the data. There are many map-making methods introduced in Tegmark (1997a). Currently, one the most commonly used method is COBE method.

Recently Elsner & Wandelt (2013) introduced a new method called messenger field to solve Wiener filter, and then this technique was being applied to map-making by Huffenberger & Næss (2018), and messenger field could outperform traditional conjugate gradient mehtod, with proper cooling technique. It has been shown by Papež et al. (2018) this messenger field method is equivalent to applying a preconditioner to the original problem and introducing an extra cooling parameter λ, but whether this cooling parameter will boost performance compare to the (traditional) conjugate gradient method is still controversial . Here I give a detailed analysis of this parameter and show that it may improve performance under some circumstances, if we properly choose its values.

The map making procedure could be summarized in equation

$$\mathbf{d} = P\mathbf{m} + \mathbf{n} \tag{1}$$

where $\mathbf{d}$, P, $\mathbf{m}$, $\mathbf{n}$ are time-ordered data (TOD), pointing matrix, CMB map, and noise. Here we assume that the noise has zero mean $\langle \mathbf{n} \rangle = \mathbf{0}$, and noise covariance matrix could be written as $N = \langle \mathbf{n}\mathbf{n}^\dagger \rangle$.

As we can see the map making model Eq.(1) mathematically is a standard linear regression problem, with *design matrix* being pointing matrix P, and *regression coefficients* are $\mathbf{m}$. For COBE method, we estimate linear regression coefficients $\mathbf{m}$ with *generalized least square* (GLS) technique, since the noise $\mathbf{n}$ is *heteroscedastic*. The GLS minimize

$$\chi^2(\mathbf{m}) \equiv (\mathbf{d} - P\mathbf{m})^\dagger N^{-1} (\mathbf{d} - P\mathbf{m}). \tag{2}$$

and the estimated map $\hat{\mathbf{m}}$ is given by

$$\hat{\mathbf{m}} = \arg\min_{\mathbf{m}} \chi^2(\mathbf{m}) = \left(P^\dagger N P\right)^{-1} P^\dagger N^{-1} \mathbf{d} \tag{3}$$

Or rewrite it as

$$\left(P^\dagger N^{-1} P\right) \hat{\mathbf{m}} = P^\dagger N^{-1} \mathbf{d} \tag{4}$$

This is the map-making equation we need to solve. However, based on current computation power, it is impossible to solve $\hat{\mathbf{m}}$ by calculating $\left(P^\dagger N^{-1} P\right)^{-1} P^\dagger N^{-1} \mathbf{d}$ directly, since the noise covariance matrix N is sparse in frequency domain, and pointing matrix P is sparse in (time by pixel) domain. In experiments currently under design, there may be $\sim 10^{16}$ time samples and $\sim 10^9$ pixels, so these matrix inversions are intractable. Therefore we use Conjugate Gradient method, which is an iterative algorithm, to solve this map-making equation. For simplicity we fix the preconditioner being $M = P^\dagger P$ for all of calculations.

The structure of this paper is organized as follows.

2. MESSENGER FIELD METHOD

Messenger field method separate noise covariance matrix $N = \bar{N} + T$, with $T = \tau I$ and τ being the minimum eigenvalue of N. Then there is a cooling parameter λ such that $N(\lambda) = \bar{N} + \lambda T$, with initial λ being a very large number and final λ being 1.

After apply preconditioner $P^\dagger T^{-1} P$ to the map making equation Eq.(4), we would get:

$$\begin{aligned}\hat{\mathbf{m}} = &\left(P^\dagger T^{-1} P\right)^{-1} P^\dagger T^{-1} \left(T^{-1} + \bar{N}^{-1}\right)^{-1} \\ &\times \left[T^{-1} P \hat{\mathbf{m}} + \bar{N}^{-1} \mathbf{d}\right]\end{aligned} \tag{5}$$

To add cooling parameter λ, we need to change T to λT and N to $N(\lambda)$. Then we could rewrite it as a fixed point iteration form

$$\begin{cases} \mathbf{t}_i = \left((\lambda T)^{-1} + \bar{N}^{-1}\right)^{-1}\left[(\lambda T)^{-1}P\hat{\mathbf{m}}_i + \bar{N}^{-1}\mathbf{d}\right] \\ \hat{\mathbf{m}}_{i+1} = \left(P^\dagger(\lambda T)^{-1}P\right)^{-1}P^\dagger(\lambda T)^{-1}\mathbf{t}_i \end{cases} \tag{6}$$

This is fixed point iteration form of messenger field method. It's equivalent to solving map-making equation Eq.(4) with preconditioner $P^\dagger(\lambda T)^{-1}P = \tau^{-1}P^\dagger P$. This preconditioner is equivalent to preconditioner $M = P^\dagger P$, since multiply a constant won't change condition number. Therefore messenger field is solving modified map making equation

$$P^\dagger N(\lambda)^{-1}P\ \hat{\mathbf{m}} = P^\dagger N(\lambda)^{-1}\mathbf{d} \tag{7}$$

with preconditioner $M = P^\dagger P$. More detailed analysis could be found in Papež et al. (2018).

3. PARAMETERIZED CONJUGATE GRADIENT METHOD

3.1. *Introduce the Idea*

The messenger field method introduced an extra cooling parameter λ to map-making equation, so we want to know how to choose this parameter. In Kodi Ramanah et al. (2017), they showed that for Wiener filter the cooling parameter should be chosen as a geometric series. In this work, we are going to give an alternative interpretation of the parameterizing process then show that for map-making equation the optimal choice for λ would also be a geometric series.

Based on previous analysis, we know that what messenger field method really does is parameterizing the map-making equation. Here to avoid confusion, we introduce another parameter η, such that the parameterized map-making equation is

$$P^\dagger N(\eta)^{-1}P\ \hat{\mathbf{m}} = P^\dagger N(\eta)^{-1}\mathbf{d} \tag{8}$$

The idea is that map-making equation Eq.(4) is hard to solve due to noise covariance matrix is sandwiched between $P^\dagger P$. But if noise covariance matrix N is proportional to identity matrix I, then its solution is given by simple binned map $\mathbf{m}_0 = \left(P^\dagger P\right)^{-1}P^\dagger\mathbf{d}$, which could be solved directly. We can parameterize the noise covariance matrix N with a parameter η, such that initially $\eta = \eta_i$ we have $N(\eta_i) \propto I$. In the end $\eta = \eta_f$ and $N(\eta_f) \propto N$, such that the final solution is what we want. We expect that the parameterized noise covariance matrix $N(\eta)$ would connect our initial guess $\hat{\mathbf{m}}_0$ and final solution $\hat{\mathbf{m}}$ as we change η from η_i to η_f, such that it would help improve convergence speed.

Now question is how to find $N(\eta)$ such that $N(\eta_i) \propto I$ and $N(\eta_f) \propto N$? Since the non-white noise part of N is the difficult portion, we could think of it as a perturbation term, which adds upon the white noise. Initially there is only white noise and solution is given by $\hat{\mathbf{m}}_0$, then we gradually add extra noise into this equation by changing η from 0 to 1. At the end when $\eta = 1$ we are solving equation Eq.(4).

Therefore we separate noise covariance matrix into two parts $N = \tau I + \bar{N}$ where τ is the minimum eigenvalue of N. Then we define $N(\eta) = \tau I + \eta\bar{N}$, with perturbation parameter η which satisfies $\eta_i = 0$ and $\eta_f = 1$.

Eq.(8) then becomes

$$\left(P^\dagger(\tau I + \eta\bar{N})^{-1}P\right)\hat{\mathbf{m}}(\eta) = P^\dagger\left(\tau I + \eta\bar{N}\right)^{-1}\mathbf{d} \tag{9}$$

We require the perturbation parameter η being monotonically increase series $0 = \eta_0 < \eta_1 < \cdots < \eta_n = 1$. For some specific η_m, we use conjugate gradient method to solve equation $\left(P^\dagger N(\eta_m)^{-1}P\right)\hat{\mathbf{m}}(\eta_m) = P^\dagger N(\eta_m)^{-1}\mathbf{d}$ with simple preconditioner $P^\dagger P$, and using $\hat{\mathbf{m}}(\eta_{m-1})$ as the initial value. The initial guess is $\hat{\mathbf{m}}(\eta_0) = \mathbf{m}_0 = \left(P^\dagger P\right)^{-1}P^\dagger\mathbf{d}$.

As you can see, η is the reciprocal of λ. The reason I switch to η in stead of keeping λ is that it would be easier for further derivations, and it's a different interpretation.

3.2. *Choosing perturbation parameters η*

The next question is how we choose these monotonically increasing parameters η. If we choose these parameters inappropriately, it would only makes it converge slower. Also we want to determine $\eta_1, \cdots, \eta_{n-1}$ before starting conjugate gradient iteration. That's because time ordered data $\mathbf{d}$ is very large, and we don't want to keep it in the system RAM during calculation. If $\eta_1, \cdots, \eta_{n-1}$ could be determined before the iterations, then we can first calculate $P^\dagger N(\eta)^{-1}\mathbf{d}$ for each η_m and store these map-sized objects in RAM, instead of the entire time-ordered data $\mathbf{d}$.

First let us try to find out our starting point η_1. What would be good value for η_1?

Here to simplify notation, I will use N_η to denote $N(\eta)$. The parameterized estimated map $\hat{\mathbf{m}}(\eta) = \left(P^\dagger N_\eta^{-1}P\right)^{-1}P^\dagger N_\eta^{-1}\mathbf{d}$ minimizes the parameterized

$$\chi^2(\mathbf{m}, \eta) = (\mathbf{d} - P\mathbf{m})^\dagger N_\eta^{-1}(\mathbf{d} - P\mathbf{m}). \tag{10}$$

For some specific η value, the minimum χ^2 value is given by

$$\chi^2(\hat{\mathbf{m}}(\eta), \eta) = \left(\mathbf{d} - P\hat{\mathbf{m}}(\eta)\right)^\dagger N_\eta^{-1}\left(\mathbf{d} - P\hat{\mathbf{m}}(\eta)\right) \tag{11}$$

To further simplify the analysis, let's assume that the noise covariance matrix $N = \left\langle\mathbf{n}\mathbf{n}^\dagger\right\rangle$ is diagonal in the frequency domain.

Let's first consider $\eta_1 = \eta_0 + \delta\eta = \delta\eta$ such that $\eta_1 = \delta\eta$ is very small quantity. Then the relative decrease of $\chi^2(\hat{\mathbf{m}}(0), 0)$ from $\eta_0 = 0$ to $\eta_1 = \delta\eta$ is

$$-\frac{\delta\chi^2(\hat{\mathbf{m}}(0),0)}{\chi^2(\hat{\mathbf{m}}(0),0)} = \delta\eta\frac{1}{\tau}\frac{(\mathbf{d}-P\hat{\mathbf{m}}(0))^\dagger\bar{N}(\mathbf{d}-P\hat{\mathbf{m}}(0))}{\left(\mathbf{d}-P\hat{\mathbf{m}}(0)\right)^\dagger\left(\mathbf{d}-P\hat{\mathbf{m}}(0)\right)} \tag{12}$$

Here we put a minus sign in front of this expression such that it's non-negative.

Ideally, we want $\delta\chi^2(\hat{\mathbf{m}}(0), 0) = \chi^2(\hat{\mathbf{m}}(1), 1) - \chi^2(\hat{\mathbf{m}}(0), 0)$, such that it would get close to the final χ^2 at next iteration. Here if we assume that initial χ^2 value $\chi^2(\hat{\mathbf{m}}(0), 0)$ is much larger than final value $\chi^2(\hat{\mathbf{m}}(1), 1)$, then we would expect $\left|\delta\chi^2(\hat{\mathbf{m}}(0), 0)/\chi^2(\hat{\mathbf{m}}(0), 0)\right| \approx 1^-$, strictly smaller than 1. To make sure it will not start too fast, we could set its upper bound equal to 1, $\delta\eta \max(\bar{N}_f)/\tau = 1$. This gives

$$\eta_1 = \frac{\tau}{\max(\bar{N}_f)} = \frac{\min(N_f)}{\max(N_f) - \min(N_f)} \tag{13}$$

Here N_f and $\bar{N}_f$ are the eigenvalues of N and $\bar{N}$ under frequency domain. If the condition number of noise covariance matrix $\kappa(N) = \max(N_f)/\min(N_f) \gg 1$, then $\eta_1 \approx \kappa^{-1}(N)$.

What about the other parameters η_m with $m > 1$? We could use a similar analysis, let $\eta_{m+1} = \eta_m + \delta\eta_m$ with a small $\delta\eta_m$, and set the upper bound of relative decrease equal to 1. And we would get

$$\delta\eta_m = \min\left(\frac{\tau + \eta_m\bar{N}_f}{\bar{N}_f}\right) = \eta_m + \frac{\tau}{\max(\bar{N}_f)}. \tag{14}$$

Therefore

$$\eta_{m+1} = \eta_m + \delta\eta_m = 2\eta_m + \frac{\tau}{\max(\bar{N}_f)} \tag{15}$$

As we can see, $\eta_1, \cdots, \eta_n$ increase like a geometric series.

$$\eta_i = \min\left\{1,\ \frac{\tau}{\max(\bar{N}_f)}\left(2^i - 1\right)\right\} \tag{16}$$

Here we need to truncate the series when $\eta_i > 1$.

This is the main result. Eq.(16) tells us not only how to choose parameters η_i, but also when we should stop the perturbation, and set $\eta = 1$. For example, if noise covariance matrix N is almost white noise, then $\bar{N} = N - \tau I \approx 0$, and we would have $\frac{\tau}{\max(\bar{N}_f)} \gg 1$. This tell us that we don't need to use parameterized method at all, because $\eta_1 = 1$. Note that the vanilla conjugate gradient method with simple binned map as initial guess corresponds to choosing $\eta_0 = 0$ and $\eta_1 = \eta_2 = \cdots = 1$.

3.3. *Intuitive Interpretation of η*

In this section, let me introduce another way to understand the role of η. Our ultimate goal is to find $\hat{\mathbf{m}}(\eta = 1)$ which minimizes $\chi^2(\mathbf{m}) = (\mathbf{d} - P\mathbf{m})^\dagger N^{-1}(\mathbf{d} - P\mathbf{m})$. Here we also assumed that N is diagonal in frequency space. With this condition χ^2 could be written as a sum of all frequency mode $|(\mathbf{d} - P\mathbf{m})_f|^2$ with weight N_f^{-1}, such as $\chi^2(\mathbf{m}) = \sum_f |(\mathbf{d} - P\mathbf{m})_f|^2 N_f^{-1}$. N_f^{-1} is large when there is little noise at that frequency, and vice versa. Which means $\chi^2(\mathbf{m})$ would favor the low noise frequency mode over high noise ones, because low noise part has higher weight. In other words the optimal map $\hat{\mathbf{m}}$ focusing on minimize the error $\mathbf{r} \equiv \mathbf{d} - P\mathbf{m}$ in the low-noise part.

After introducing η, we minimize $\chi^2(\mathbf{m}, \eta) = (\mathbf{d} - P\mathbf{m})^\dagger N_\eta^{-1}(\mathbf{d} - P\mathbf{m})$ for each η value as it increase from 0 to 1. For $\eta = 0$, $N_{\eta=0}^{-1} \propto I$ and the estimated map $\hat{\mathbf{m}}(\eta = 0)$ does not prioritize any frequency mode when minimizing the error. As we slowly increase η, we decrease the weight for the frequency modes which have large noise, and focusing minimizing error for low noise part. If we start with $\eta_1 = 1$ directly, which corresponds to the vanilla conjugate gradient method, then the entire conjugate gradient solver will only focusing on minimizing low noise part, such that χ^2 would converge very fast at low noise region, but relative slow on high noise part. However by introducing η parameter, we let the solver first treat every frequency equally. Then as η slowly increases, it gradually shifts focus to low noise part. If we write the difference between final and initial χ^2 value as $\chi^2(\hat{\mathbf{m}}(1), 1) - \chi^2(\hat{\mathbf{m}}(0), 0) = \int_0^1 \mathrm{d}\eta\, \frac{\mathrm{d}}{\mathrm{d}\eta}\chi^2(\hat{\mathbf{m}}(\eta), \eta)$, and use Eq.(**??**). We note that when η is very small, the $\frac{\mathrm{d}}{\mathrm{d}\eta}\chi^2(\hat{\mathbf{m}}(\eta), \eta)$ would have relatively large contribution from medium to large noise region, comparing to large η. So introducing η might improve the convergence of χ^2 at these regions, because the vanilla conjugate gradient method only focuses on the low noise part and it may have difficulty at these regions.

3.4. *Computational Cost*

To properly compare the performance cost of this method with respect to vanilla conjugate gradient method with simple preconditioner, we need to compare their computational cost at each iteration. The right hand side of parameterized map making equation Eq.(**??**) could be computed before iterations, so it won't introduce extra computational cost during iterations. The most demanding part of conjugate gradient method is calculating $P^\dagger N^{-1} P\hat{\mathbf{m}}$, because it contains a Fourier transform of $P\hat{\mathbf{m}}$ from time domain to frequency domain and an inverse Fourier transform of $N^{-1}P\hat{\mathbf{m}}$ from frequency domain back to time domain, which is order

$\mathcal{O}(n \log n)$ with n being the length of time ordered data. If we change N^{-1} to $N^{-1}(\eta)$, it won't add extra cost. Therefore the computational cost it the same for one step.

However in previous analysis, our choice of parameters η_i is based on $\delta\chi^2(\hat{\mathbf{m}}(\eta_i), \eta_i)$ which is evaluated at $\mathbf{m} = \hat{\mathbf{m}}(\eta_i)$ the estimated map at η_i. We update η_i to η_{i+1} when $\mathbf{m} \approx \hat{\mathbf{m}}(\eta_i)$. How do we know current map $\mathbf{m}$ is close to $\hat{\mathbf{m}}(\eta_i)$? Because for each new η_i value, we are solving a new set of linear equations $A(\eta_i)\hat{\mathbf{m}} = \mathbf{b}(\eta_i)$ with $A(\eta_i) = P^\dagger N(\eta_i)^{-1} P$ and $\mathbf{b}(\eta_i) = P^\dagger N(\eta_i)^{-1}\mathbf{d}$, and we could stop calculation and moving to next value η_{i+1} when the norm of residual $||\mathbf{r}(\eta_i)|| = ||\mathbf{b}(\eta_i) - A(\eta_i)\mathbf{m}||$ smaller than some specific value. Since when doing conjugate gradient algorithm we calculate $\mathbf{r}$ and stop the iteration when $||\mathbf{r}||$ is small enough, now after introducing parameter η, we move to next parameter η_{i+1} when $||\mathbf{r}(\eta_i)||$ is small enough. Again, this won't add extra cost compare to vanilla conjugate gradient method.

Therefore we find that the only significant cost after adding perturbation parameter η, is to to find out $\mathbf{b}(\eta_i)$ for each $\eta_i \neq 1$ before starting the iterations. And this is one time calculation, it's negligible compare to remaining calculations.

4. NUMERICAL SIMULATIONS

To compare these algorithms, we need to do some simple simulation of scanning process, and generate time ordered data from random sky signal. Our sky is a small rectangular area, with two orthogonal directions x and y, both with range from $-1°$ to $+1°$. The electromagnetic signal is described as four stokes parameters $(S_0, S_1, S_2, S_3) = (I, Q, U, V)$. We model the overall electromagnetic signal is created by some normal distributed sources in the sky, with intensity $I_i(x, y) = A_i \exp\left(-\frac{1}{2}\frac{(x-x_i)^2+(y-y_i)^2}{\sigma_i^2}\right)$, for each source centered at (x_i, y_i). In our simulation, $A_i \sim \mathrm{Unif}(-100, 100)$, $\sigma_i \sim \mathrm{Unif}(0.05°, 0.2°)$ and the center of each source $x_i, y_i \sim \mathrm{Unif}(-1°, +1°)$. Every source has its degree of polarization $p_i \sim \mathrm{Unif}(0, 1)$ and polarization angle $\psi_i \sim \mathrm{Unif}(0, \pi)$. Here we ignored angle χ_i, because our detectors won't be sensitive to circular polarization. Finally, the stokes parameters over sky is given by $S_0(x, y) = \sum_i I_i(x, y)$, $S_1(x, y) = \sum_i I_i(x, y) p_i \cos(2\psi_i)$, $S_2(x, y) = \sum_i I_i(x, y) p_i \sin(2\psi_i)$. Again, we ignored S_3, because it describes circular polarization.

For the scanning process, our single telescope contains nine detectors, each has different sensitivity to polarization S_1 and S_2. It scans the sky with a raster scanning pattern and scanning frequency $f_{\mathrm{scan}} = 0.1$ Hz sampling frequency $f_{\mathrm{sample}} = 100$ Hz. The telescope scans the sky horizontally and then vertically, and then digitizes the position (x, y) into 512×512 pixel. This gives noiseless signal $\mathbf{s}$.

The noise power spectrum is given by

$$P(f) = \sigma^2 \left(1 + \frac{f_{\mathrm{knee}}^\alpha + f_{\mathrm{apo}}^\alpha}{f^\alpha + f_{\mathrm{apo}}^\alpha}\right) \tag{17}$$

Here we fixed $\sigma^2 = 10\ \mu\mathrm{K}^2$, $\alpha = 2$ and $f_{\mathrm{knee}} = 10$ Hz, and change f_{apo} to compare the performance under different noise models. Note that as $f_{\mathrm{apo}} \to 0$, $P(f) \to \sigma^2\big(1 + (f/f_{\mathrm{knee}})^{-1}\big)$, it becomes a $1/f$ noise model. The noise covariance matrix

$$N_{ff'} = P(f)\frac{\delta_{ff'}}{\Delta_f} \tag{18}$$

is a diagonal matrix in frequency space, where Δ_f is equal to reciprocal of total scanning time T.

Finally, we get the simulated time ordered data $\mathbf{d} = \mathbf{s} + \mathbf{n}$ by adding up signal and noise.

5. RESULTS

First let's compare the results with vanilla conjugate gradient method with simple preconditioner $P^\dagger P$. The results are showed in Figure (1) for different kinds of noise power spectra. Here note that χ^2 in Figure (**??**) is calculated based on Eq.(2)

$$\chi^2(\mathbf{m}) = (\mathbf{d} - P\mathbf{m})^\dagger N^{-1}(\mathbf{d} - P\mathbf{m}) \tag{2}$$

not $\chi^2(\mathbf{m}, \eta)$ in Eq.(10). The $\chi^2_{\min}$ is calculated from perturbative conjugate gradient method with more intermediate η values, and more iterations after $\eta = 1$.

As we can see in Figure(**??**), if the condition number of noise covariance matrix $\kappa(N)$ is small, and the noise is almost white noise, the performance between different these two methods is small. The vanilla conjugate gradient method converge faster, because its perturbation parameter $\eta_i = \{0, 1, 1, \cdots\}$, however for the perturbation method its η value will slowly reach 1 in first few iterations as we can see in Figure(**??**).

Notice that as we increase $\kappa(N)$, or equivalently decrease f_{apo}, the perturbation parameter η starts showing its benefits, as showed in Figure(2) and Figure(3). It outperforms the vanilla conjugate gradient method when $f_{\mathrm{apo}} \approx 0$ and the noise power spectrum becomes the $1/f$ noise model, which usually is the intrinsic noise of instrumentsTegmark (1997b).

In the conjugate gradient method with messenger cooling parameter λ, the number of cooling parameters we need is an extra free parameter. After the number of λ is determined, we construct a geometric series with

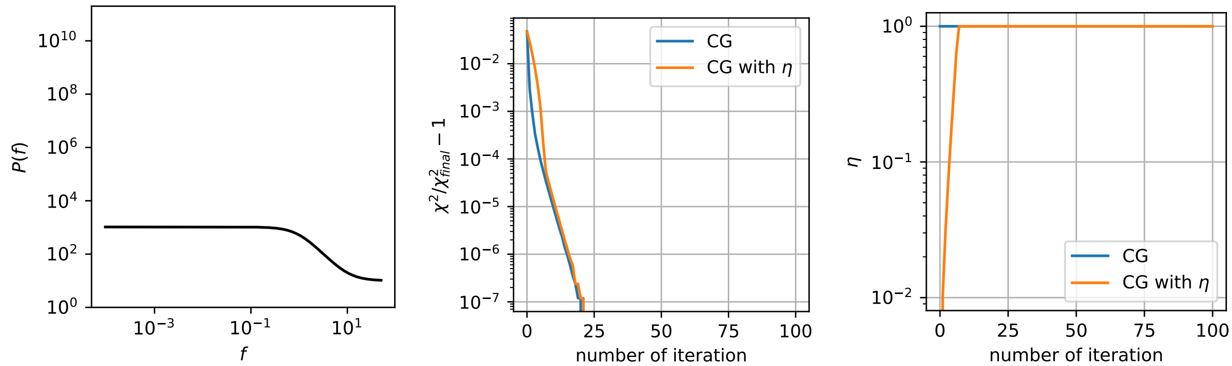

Figure 1. The left graph shows the noise power spectrum Eq.(17) with $f_{\rm apo} \approx 0.99$ and $\kappa(N) = 10^2$. The center one shows the $\chi^2(\mathbf{m})/\chi^2_{\rm final} - 1$, with $\chi^2(\mathbf{m})$ calculated based on Eq.(2). The right one shows the η value for each iteration. For vanilla conjugate gradient method η always equal to 1, so it's a horizontal line at $\eta = 1$.

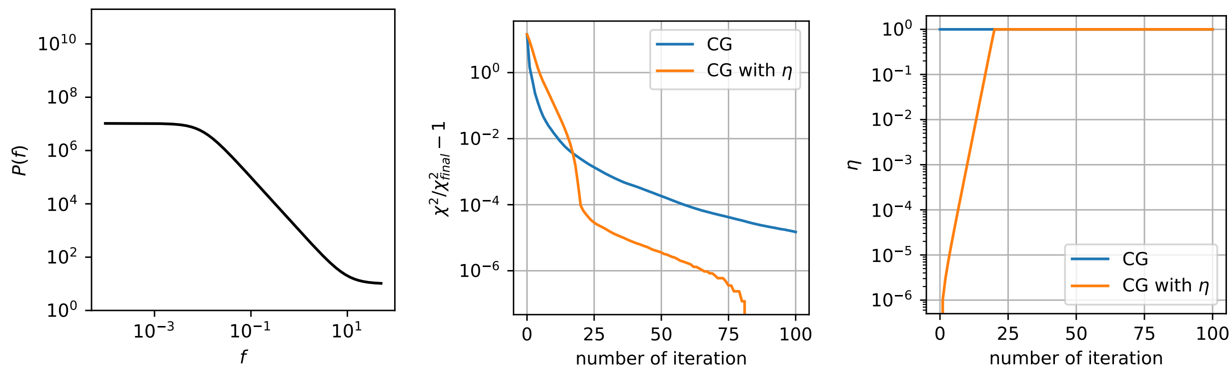

Figure 2. The figure shows results for $f_{\rm apo} \approx 9.8 \times 10^{-3}$ and $\kappa(N) = 10^6$.

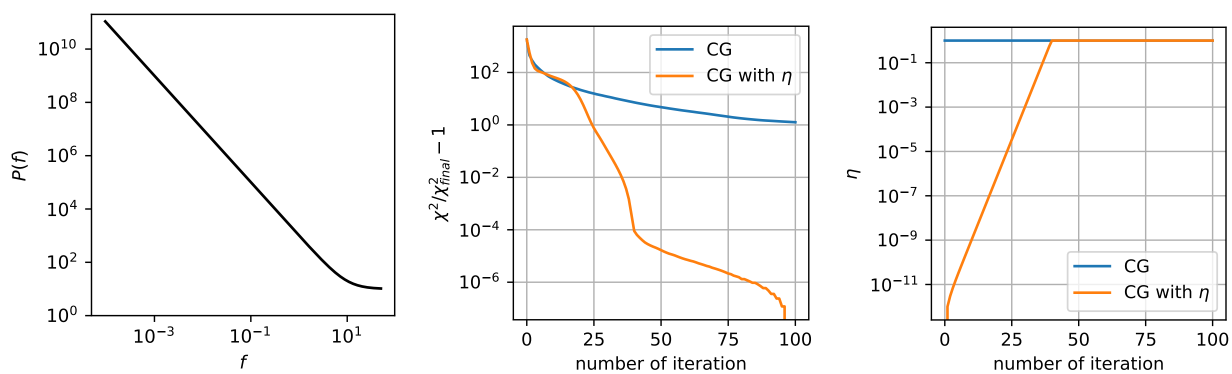

Figure 3. The figure shows results for $f_{\rm apo} \approx 9.8 \times 10^{-6}$ and $\kappa(N) = 10^{12}$.

fixed initial and final value, which uses `logspace` function in `numpy`. Since I show in Eq.(**??**) that the messenger field cooling parameter λ is equivalent to $1/\eta$, I use η for further analysis.

Now let us compare the performance difference between choosing η parameters based on Eq.(16) and fixing number of η parameters n_η manually. Here we choose the η_i values using function `numpy.logspace(start=`$\ln(\eta_1)$`, stop=0, num=`n_η`, base=`e`)`. The results are showed in Figure(4), (5), and (6).

When $\kappa(N)$ is small, and Eq.(16) tells us that only a few η parameters are good enough, see Figure(**??**). If unfortunately we choose n_η being large value, like 15 or 30, then it will ends up converge slowly, because it needs at least 15 or 30 iterations to reach $\eta = 1$, at least one iteration per η level.

On the other hand if $\kappa(N)$ is very large and the power spectrum is $1/f$ noise, we need more η parameters. If n_η is too small, for example $n_\eta = 5$ in Figure(**??**), it may be better than the vanilla conjugate gradient method, but it is still far from optimal.

6. POSSIBLE IMPROVEMENTS

As you may have noticed in Figure(5) and Figure(6), the perturbation parameter based on Eq.(16) is more than needed, especially for $1/f$ noise case. From Figure(**??**) we know that Eq.(16) gives us $n_\eta \approx 40$, however based on χ^2 result in Figure(**??**), we notice that $n_\eta \approx 30$ or even $n_\eta \approx 15$ is good enough. Also, for the nearly-white-noise case, we could certainly choose $n_\eta = 1$ such that $\eta_1 = 1$ which corresponds to vanilla conjugate gradient method, based on χ^2 result in Figure(**??**). However Eq.(16) gives us $n_\eta \approx 6$, see Figure(**??**), even though it does not make the final χ^2 result much different at the end.

Is it possible to further improve the analysis, such that it produces smaller n_η? Let's examine how we get η_i series. Remember that we determine $\delta\eta$ value based on the upper bound of $-\delta\chi^2(\hat{\mathbf{m}}(\eta),\eta)/\chi^2(\hat{\mathbf{m}}(\eta),\eta)$, in Eq.(**??**). Here I rewrite it in a simplified form

$$\begin{aligned} -\frac{\delta\chi^2(\hat{\mathbf{m}}(\eta),\eta)}{\chi^2(\hat{\mathbf{m}}(\eta),\eta)} &= \delta\eta \frac{\hat{\mathbf{r}}_\eta^\dagger N_\eta^{-1}\bar{N}N_\eta^{-1}\hat{\mathbf{r}}_\eta}{\hat{\mathbf{r}}_\eta^\dagger N_\eta^{-1}\hat{\mathbf{r}}_\eta} \\ &\le \frac{\delta\eta}{\eta + \frac{\tau}{\max(\bar{N}_f)-\tau}} \end{aligned} \tag{19}$$

with $\mathbf{r}_\eta = \mathbf{d} - P\hat{\mathbf{m}}(\eta) = \left[1 - P\left(P^\dagger N_\eta^{-1}P\right)^{-1}P^\dagger N_\eta^{-1}\right]\mathbf{d} \equiv \mathcal{P}_\eta\mathbf{d}$. We treated $\mathbf{r}_\eta$ as an arbitrary vector in frequency domain, since we don't know how to calculate $\mathcal{P}_\eta$ for $\eta \neq 0$, and it's hard to analyze the projection matrix $\mathcal{P}_\eta$ in frequency space, as it contains $\left(P^\dagger N_\eta^{-1}P\right)^{-1}$. Note that we have to determine all of η value before calculation, because we don't want to keep the time ordered data in system RAM, so we need to somehow analytically analyze $\mathcal{P}_\eta$, and its behavior in frequency space.

Unless $\mathbf{r}_\eta$ almost only has large noise modes, $\frac{\mathrm{d}}{\mathrm{d}\eta}\chi^2(\hat{\mathbf{m}}(\eta),\eta)/\chi^2(\hat{\mathbf{m}}(\eta),\eta)$ won't get close to the upper bound $1/\left(\eta + \frac{\tau}{\max(\bar{N}_f)-\tau}\right)$. Based on the analysis in Section(3.3), for small η the estimated map $\hat{\mathbf{m}}(\eta)$ does not only focusing on minimizing error $\mathbf{r}_\eta$ at low noise region. So we would expect that there would be a fair amount of low noise modes contribution in $\mathbf{r}_\eta$ especially for the first few η values. Which means if we could somehow know the frequency distribution of $\mathbf{r}_\eta$, we could tighten the boundary of $\frac{\mathrm{d}}{\mathrm{d}\eta}\chi^2(\hat{\mathbf{m}}(\eta),\eta)/\chi^2(\hat{\mathbf{m}}(\eta),\eta)$, and get larger $\delta\eta$ value. This should make η goes to 1 faster, and yields the fewer η parameters we need.

Also notice that the η values determined from Eq.(16)

$$\eta_i = \min\left\{1,\ \frac{\tau}{\max(\bar{N}_f)}\left(2^i - 1\right)\right\} \tag{16}$$

are not dependent on any scanning information, it only depends on noise power spectrum $P(f)$, or noise covariance matrix N. Figure(7) and Figure(8) show two examples with same parameters as in Figure(6) except scanning frequency f_{scan}, in Figure(7) it scans very slow and in Figure(8) it's very fast. In these two cases our η values based on Eq.(16) are better than manually selected values. Based on these two results we know, the η values should somehow depends on scanning scheme. Again that's because when we determine the upper bound of $\frac{\mathrm{d}}{\mathrm{d}\eta}\chi^2(\hat{\mathbf{m}}(\eta),\eta)$ we treat $\mathbf{r}_\eta = \mathbf{d} - P\hat{\mathbf{m}} = \mathcal{P}_\eta\mathbf{d}$ as an arbitrary vector, such that we lose all information related to scanning scheme in the pointing matrix P.

7. CONCLUSION

Here we discussed a method to solve map making equation Eq.(**??**)

$$\hat{\mathbf{m}} = \left(P^\dagger N P\right)^{-1} P^\dagger N^{-1}\mathbf{d} \tag{??}$$

by separating noise covariance matrix N into two parts, white noise part τI and the remaining noise $\bar{N}$. Then we could think $\bar{N}$ as a perturbation added to white noise, by introducing a parameter η, as η change from 0 to 1, we gradually add this non white noise in to system.

The η values can be predetermined analytically. This property is very important, because we don't want to keep entire time ordered data in system RAM. If these η values can be determined before calculation, then we only need to keep several map sized object, which is much smaller than timed ordered data. Also we showed that this method has same computational cost as vanilla conjugate gradient method but performs better when the condition number of noise covariance matrix $\kappa(N)$ is large, especially in $1/f$ noise case. The only extra free parameter added is to determine whether the error at current step $\mathbf{r}(\eta_i) = ||\mathbf{b}(\eta_i) - A(\eta_i)\mathbf{m}||$ is small enough such that we change advance to next value η_{i+1}.

The perturbation parameter η get from Eq.(16) are not perfect. Since it only takes in to account the noise information in N, but ignored all scanning information contained in pointing matrix P, because we are unable to analyze the structure of $\mathbf{r}_\eta = \mathbf{d} - P\hat{\mathbf{m}}(\eta) = \mathcal{P}_\eta\mathbf{d}$ in frequency space.

The analysis of η value also explains why cooling parameters $\lambda = 1/\eta$ in messenger field are chosen to be geometric series or `logspace` Huffenberger & Næss (2018).

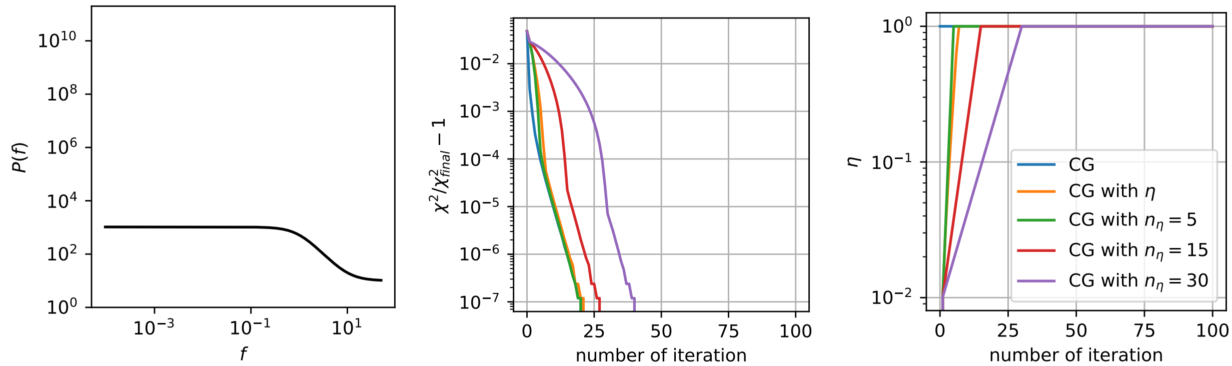

Figure 4. Same as Figure(1) with extra manually chosen n_η results.

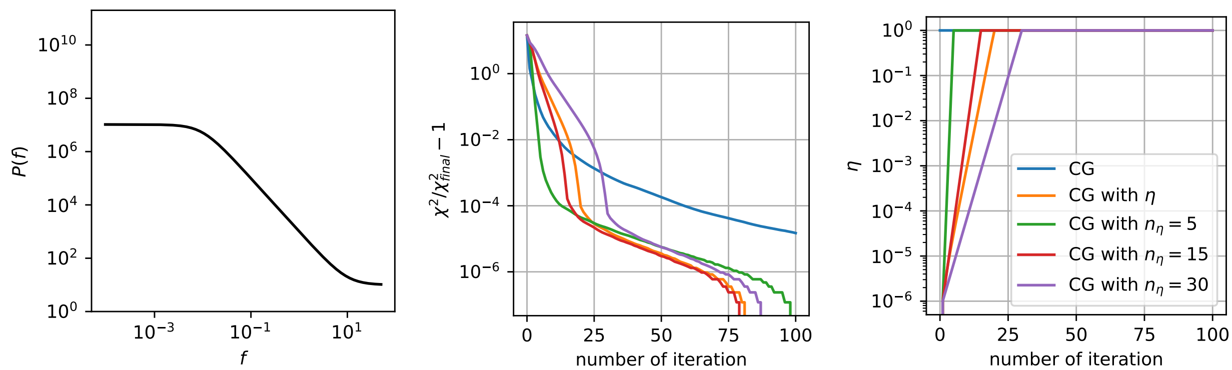

Figure 5. Same as Figure(2) with extra manually chosen n_η results.

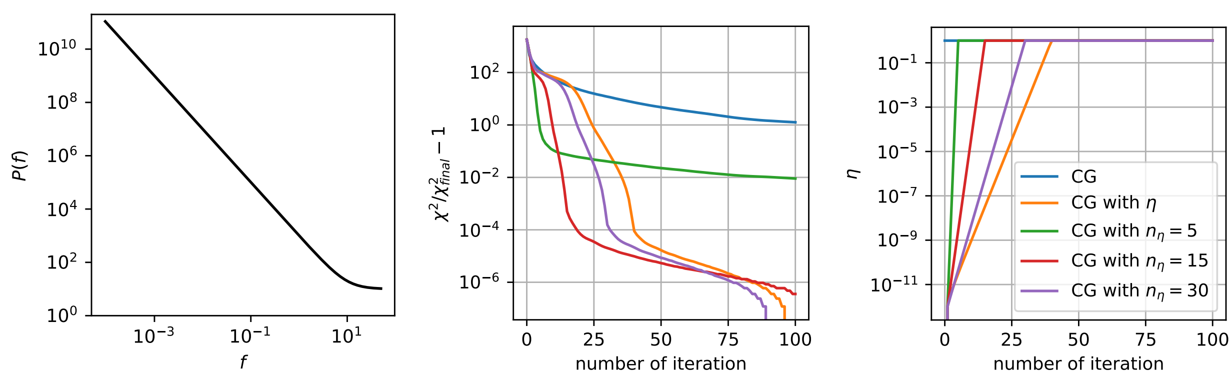

Figure 6. Same as Figure(3) with extra manually chosen n_η results.

All of the calculation are using simple preconditioner $P^\dagger P$, but the entire analysis is independent of preconditioner. Using better preconditioners, it would also have improvements.

APPENDIX

REFERENCES

Elsner, F., & Wandelt, B. D. 2013, A&A, 549, A111, doi: 10.1051/0004-6361/201220586

Huffenberger, K. M., & Næss, S. K. 2018, The Astrophysical Journal, 852, 92, doi: 10.3847/1538-4357/aa9c7d

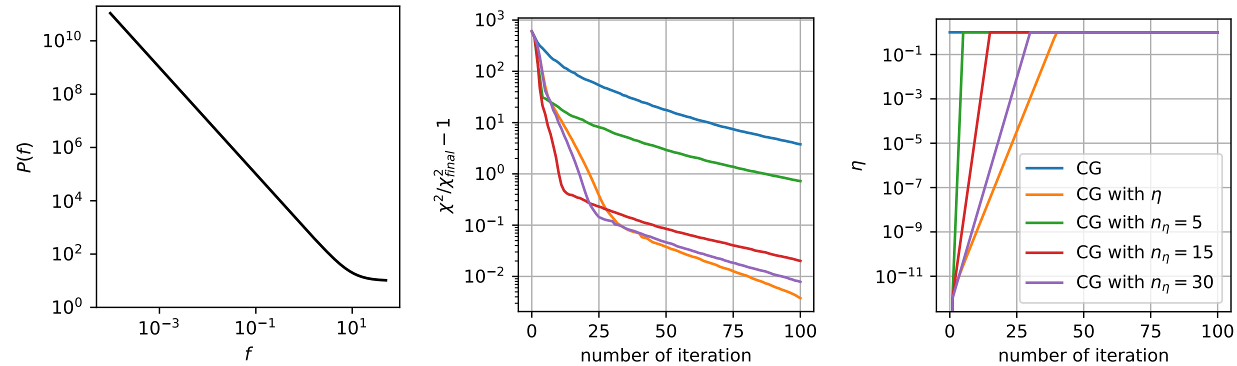

Figure 7. In this case all frequencies are the same as Figure(6) except $f_{\rm scan} = 0.001$.

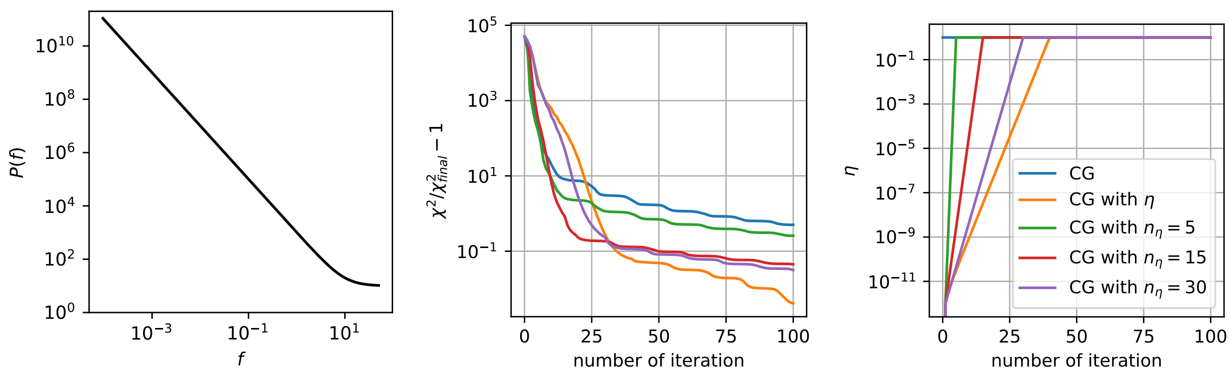

Figure 8. In this case all frequencies are the same as Figure(6) except $f_{\rm scan} = 10$.

Kodi Ramanah, D., Lavaux, G., & Wandelt, B. D. 2017, MNRAS, 468, 1782, doi: 10.1093/mnras/stx527

Papež, J., Grigori, L., & Stompor, R. 2018, A&A, 620, A59, doi: 10.1051/0004-6361/201832987

Tegmark, M. 1997a, ApJL, 480, L87, doi: 10.1086/310631

—. 1997b, PhRvD, 56, 4514, doi: 10.1103/PhysRevD.56.4514