ROBUST ADAPTIVE WIENER FILTERING

Ethan Anderes

Statistics Department
University of California, Davis, CA 95616

ABSTRACT

A recent paper by Anderes and Paul [1] analyze a regression characterization of a new estimator of lensing from cosmic microwave observations, developed by Hu and Okamoto [2, 3, 4]. A key tool used in that paper is the application of the robust generalized shrinkage priors developed 30 years ago in [5, 6, 7] to the problem of adaptive Wiener filtering. The technique requires the user to propose a fiducial model for the spectral density of the unknown signal but the resulting estimator is developed to be robust to misspecification of this model. The role of the fiducial spectral density is to give the estimator superior statistical performance in a "neighborhood of the fiducial model" while controlling the statistical errors when the fiducial spectral density is drastically wrong. One of the main advantages of this adaptive Wiener filter is that one can easily obtain posterior samples of the true signal given the unknown data. These posterior samples are particularly advantageous when studying non-linear functions of the signal, cross correlating with other independent measurements of the same signal and can be used to propagate uncertainty when the filtering is done in a scientific pipeline. In this paper we explore these advantages with simulations and examine the possibility of widespread application in more general image and signal processing problems.

***Index Terms*—** Wiener filtering, shrinkage priors, Bayesian priors, cosmic microwave background, robust filtering

1. INTRODUCTION

The main object of study in this paper is the generic filtering problem where one observes a random field $y(\boldsymbol{x})$ on $\boldsymbol{x} \in \mathbb{R}^d$ which has the following decomposition

$$y(\boldsymbol{x}) = \phi(\boldsymbol{x}) + \epsilon(\boldsymbol{x}).$$

The function ϕ represents the signal, ϵ denotes the noise and both are assumed to be stationary, mean zero, Gaussian random fields. The goal is to then estimate $\phi(\boldsymbol{x})$ from the observations $y(\boldsymbol{x})$. The optimal solution is normally given by Wiener filtering. However, this requires knowledge of the spectral density of ϕ. This paper analyzes a Bayesian adaptive approach to Wiener filtering when there is uncertainty in the spectral density of ϕ.

Under the assumption that ϕ and ϵ are both stationary Gaussian random fields one can characterize the covariance structure of the Fourier transform with spectral densities. In particular, if we let $\phi_{\boldsymbol{l}}$ denote the Fourier transform $\int e^{-i\boldsymbol{x}\cdot\boldsymbol{l}}\phi(\boldsymbol{x})\frac{d\boldsymbol{x}}{(2\pi)^{d/2}}$ (similarly for $\epsilon_{\boldsymbol{l}}$) then the spectral density for ϕ and ϵ, denoted by $C_{\boldsymbol{l}}^{\phi\phi}$ and $C_{\boldsymbol{l}}^{\epsilon\epsilon}$ satisfy

$$\langle \phi_{\boldsymbol{l}}\, \phi_{\boldsymbol{l}'}^* \rangle = \delta_{\boldsymbol{l}-\boldsymbol{l}'} C_{\boldsymbol{l}}^{\phi\phi} \tag{1}$$
$$\langle \epsilon_{\boldsymbol{l}}\, \epsilon_{\boldsymbol{l}'}^* \rangle = \delta_{\boldsymbol{l}-\boldsymbol{l}'} C_{\boldsymbol{l}}^{\epsilon\epsilon} \tag{2}$$

Supported by NSF grant 1007480.

where $\delta_{\boldsymbol{l}} \equiv \int e^{i\boldsymbol{x}\cdot\boldsymbol{l}}\frac{d\boldsymbol{x}}{(2\pi)^{d/2}}$ and $\langle\cdot\rangle$ denotes expectation. The Wiener filter in this context is then given by

$$\hat{\phi}_{\boldsymbol{l}} = \left[1 - \frac{C_{\boldsymbol{l}}^{\epsilon\epsilon}}{C_{\boldsymbol{l}}^{\phi\phi} + C_{\boldsymbol{l}}^{\epsilon\epsilon}}\right] y_{\boldsymbol{l}}. \tag{3}$$

Notice that to compute the Wiener filter, the spectral density $C_{\boldsymbol{l}}^{\phi\phi}$ must be known. Unfortunately, in typical situations $C_{\boldsymbol{l}}^{\phi\phi}$ is not known or at least only partially known. One of the ways account for this uncertainty in $C_{\boldsymbol{l}}^{\phi\phi}$ is to utilize Bayesian techniques, with a prior on $C_{\boldsymbol{l}}^{\phi\phi}$, then using Markov Chain Monte Carlo (MCMC) sampling techniques to obtain approximate samples from the posterior distribution on ϕ given y. The main difficulty with this approach is the elicitation of an appropriate prior and the development of accurate MCMC technique with good diagnostics. In this paper we explore an alternative approach which utilizes robust Bayesian techniques (originally developed in [6, 7] for statistical regression) to generate an adaptive shrinkage adjustment to the nominal Wiener filter. The new estimator is given by

$$\tilde{\phi}_{\boldsymbol{l}} = \left[1 - F_{\boldsymbol{l}}^{y}\frac{C_{\boldsymbol{l}}^{\epsilon\epsilon}}{C_{\boldsymbol{l},\text{fid}}^{\phi\phi} + C_{\boldsymbol{l}}^{\epsilon\epsilon}}\right] y_{\boldsymbol{l}} \tag{4}$$

where $C_{\boldsymbol{l},\text{fid}}^{\phi\phi}$ is a fiducial spectral density model and $F_{\boldsymbol{l}}^{y}$ is an adaptive shrinkage factor defined in Section 2.1. The adaptive shrinkage factor $F_{\boldsymbol{l}}^{y}$ is derived as a posterior expected value in a robust Bayesian procedure and adapts the Wiener filter to account for uncertainty associated with misspecification of $C_{\boldsymbol{l},\text{fid}}^{\phi\phi}$. Indeed, the Bayesian viewpoint is the principal advantage of the estimate: using posterior draws of the lensing potential ϕ, one can construct estimates and quantify the uncertainty for any non-linear function of the gravitational potential, including spectral density estimate.

In this paper, adaptivity and robustness have two slightly different meanings. Robustness describes an estimator which is insensitive to the fiducial spectral density. For example, one of the estimates shown in Section 3 uses a fiducial spectral density which is 100 times smaller than the simulation truth spectral density. However, the results are very similar if one uses a fiducial model 100 times too large (see Figure 2). This is robustness. An adaptive estimate, on the other hand, uses the data to adapt to the scenario that one uses a fiducial model which is seriously wrong and produces an estimate which behaves nearly optimally. Indeed, we show in Section 3 that even with a fiducial model 100 times too small (or too large), our estimate adapts to the data and produces a filter that behaves like the optimal Wiener filter when using the true, but unknown spectral density. This is adaptivity.

In Section 2 we derive our new estimator and show that the robust adaptive Bayesian procedure is easy to simulate without resorting to expensive MCMC. We also discuss an illuminating connection with a generalized James-Stein shrinkage estimator [8], where the role of the fiducial spectral density is essentially to specify a shrinkage direction. We then finish the paper in Section 3 with simulations that illustrate the advantages provided by the robust Bayesian characterization of the adaptive Wiener filter.

2. ADAPTIVE WIENER FILTERING

To derive the adaptive Wiener filter (4) it will be useful to work under the discrete approximation to the continuous Fourier transform that arises when observing $y(\boldsymbol{x})$ on a finite grid in $\mathbb{R}^d$. For example, the function $y_{\boldsymbol{l}}$ will specify the matrix obtained from discrete Fourier transform and $\delta_{\boldsymbol{l}}$ is approximated as $1/\Delta\boldsymbol{l}$ when $\boldsymbol{l}=0$ and zero otherwise, where $\Delta\boldsymbol{l}$ is the area element of the grid in Fourier space. We also tacitly assume $y(\boldsymbol{x})$ is appropriately periodic and sampled at a sufficient resolution to avoid the typical complications when relating the continuous and discrete Fourier transform.

To align our notation with standard regression theory it will be useful to concatenate the values of $y_{\boldsymbol{l}}$, for different frequencies $\boldsymbol{l}$, into one data vector of length n, denoted $\mathbf{y}$. Define $\boldsymbol{\phi}$ similarly as the vector of $\phi_{\boldsymbol{l}}$ values for the matching frequencies $\boldsymbol{l}$ used to construct $\mathbf{y}$. We make the additional assumption that $\boldsymbol{\phi}$ and $\mathbf{y}$ contain only unique elements up to complex conjugation (so there are no distinct coordinates with values w and z such that $z=w$ or $z=w^*$). Working with the vectors $\boldsymbol{\phi}$ and $\mathbf{y}$, instead of the functions ϕ and y, has the additional advantage that one can easily extend to the case where the adaptive shrinkage is done on separate annuli in Fourier space (which is discussed in Remark 1 at the end of Section 2.1).

Translating equation (3) to our vector notation, the Wiener filter is given simply by matrix multiplication

$$\tilde{\boldsymbol{\phi}}_\Lambda \equiv \Lambda(\Lambda+\Sigma)^{-1}\mathbf{y} \tag{5}$$

where the matrices Σ and Λ are defined by

$$\Sigma \equiv \left(\delta_{\boldsymbol{l}_k-\boldsymbol{l}_j} C^{\epsilon\epsilon}_{\boldsymbol{l}_k}\right)_{k,j=1}^n$$
$$\Lambda \equiv \left(\delta_{\boldsymbol{l}_k-\boldsymbol{l}_j} C^{\phi\phi}_{\boldsymbol{l}_k}\right)_{k,j=1}^n.$$

In addition, our notation allows us to clearly write the relationship between $\boldsymbol{\phi}$ and $\mathbf{y}$ in a hierarchal Bayesian setting

$$\mathbf{y}\big|\boldsymbol{\phi} \sim \mathcal{N}(\boldsymbol{\phi},\Sigma) \tag{6}$$
$$\boldsymbol{\phi}\big|\Lambda \sim \mathcal{N}(0,\Lambda) \tag{7}$$

where $\boldsymbol{\phi}\big|\Lambda \sim \mathcal{N}(0,\Lambda)$ means Re$(\boldsymbol{\phi})$ and Im$(\boldsymbol{\phi})$ are independent Gaussian random vectors with individual distributions given by Re$(\boldsymbol{\phi}) \sim \mathcal{N}(0,\Lambda/2)$ and Im$(\boldsymbol{\phi}) \sim \mathcal{N}(0,\Lambda/2)$.

2.1. Robust generalized Bayesian adaptive shrinkage

The Bayesian paradigm is the clearest way to understand how one adapts (5) when Λ is unknown. Indeed, if one is willing to model the uncertainty in Λ (equivalently in $C^{\phi\phi}_{\boldsymbol{l}}$) using a prior probability density $\mathcal{P}(\Lambda)$, then Bayes theorem in conjunction with (6) and (7) gives a posterior density $\mathcal{P}(\boldsymbol{\phi},\Lambda|\mathbf{y}) \propto \mathcal{P}(\mathbf{y}|\boldsymbol{\phi})\mathcal{P}(\boldsymbol{\phi}|\Lambda)\mathcal{P}(\Lambda)$. The posterior density $P(\boldsymbol{\phi},\Lambda|\mathbf{y})$ quantifies the joint uncertainty in the unobserved $\boldsymbol{\phi}$ and Λ when observing the data $\mathbf{y}$. Notice that one can marginalize out Λ to obtain a posterior on $\boldsymbol{\phi}$ exclusively, $\mathcal{P}(\boldsymbol{\phi}|\mathbf{y}) = \int \mathcal{P}(\boldsymbol{\phi},\Lambda|\mathbf{y})d\Lambda = \int \mathcal{P}(\boldsymbol{\phi}|\Lambda,\mathbf{y})\mathcal{P}(\Lambda|\mathbf{y})d\Lambda$ where $d\Lambda$ corresponds

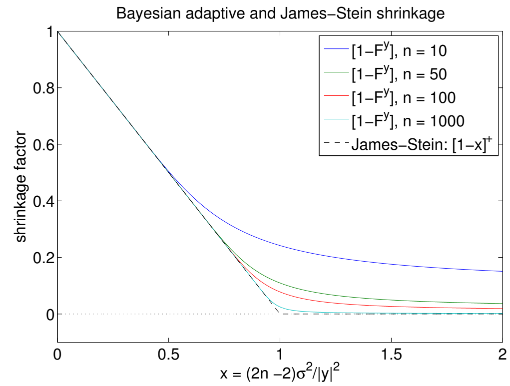

Fig. 1. The colored plots show the Bayesian adaptive shrinkage factor $1-F^y$ as a function of $(2n-2)\sigma^2/\mathbf{y}^\dagger\mathbf{y}^*$, for different values of n. The dashed black line shows the classic James-Stein shrinkage factor $\left[1-\frac{(2n-2)\sigma^2}{\mathbf{y}^\dagger\mathbf{y}^*}\right]^+$.

to coordinate-wise area element. Now the posterior expected value of $\boldsymbol{\phi}$ given the data $\mathbf{y}$ can be computed as

$$\int \boldsymbol{\phi}\mathcal{P}(\boldsymbol{\phi}|\mathbf{y})\,d\boldsymbol{\phi} = \int \underbrace{\left[\int \boldsymbol{\phi}\mathcal{P}(\boldsymbol{\phi}|\Lambda,\mathbf{y})\,d\boldsymbol{\phi}\right]}_{\text{Wiener filter (5)}} \mathcal{P}(\Lambda|\mathbf{y})\,d\Lambda \tag{8}$$
$$= \underbrace{\left[\int \Lambda(\Lambda+\Sigma)^{-1}\mathcal{P}(\Lambda|\mathbf{y})\,d\Lambda\right]}_{\substack{\text{posterior expected}\\ \text{shrinkage factor}}} \mathbf{y}. \tag{9}$$

The advantage of (8) is that it clearly exposes how to handle the situation when Λ is unknown: average the Wiener filter $\tilde{\boldsymbol{\phi}}_\Lambda$ over different possibilities for Λ supported by the data (through the posterior $\mathcal{P}(\Lambda|\mathbf{y})$). In fact, equation (9) shows that to account for uncertainty in Λ when estimating $\boldsymbol{\phi}$ simply replace the shrinkage factor $\Lambda(\Lambda+\Sigma)^{-1}$ in (5) with it's expected value under the posterior distribution on Λ, $\mathcal{P}(\Lambda|\mathbf{y})$.

The non-informative generalized prior on Λ we use to generate our robust adaptive Wiener filter was originally developed in [6, 7]. This prior requires the user to specify a fiducial spectral density model, denoted $C^{\phi\phi}_{\boldsymbol{l},\text{fid}}$, and is designed to be both robust against alternative spectral truths and computationally simple. To specify the prior start by defining the following matrix based on the fiducial spectral density

$$\Lambda_{\text{fid}} \equiv \left(\delta_{\boldsymbol{l}_k-\boldsymbol{l}_j} C^{\phi\phi}_{\boldsymbol{l}_k,\text{fid}}\right)_{k,j=1}^n.$$

The non-informative generalized prior for Λ is given by

$$\Lambda \sim \xi(\Sigma+\Lambda_{\text{fid}})-\Sigma \tag{10}$$
$$\xi \text{ has generalized density} \propto \xi^{-1} \text{ on } (\rho,\infty) \tag{11}$$

where ρ is the largest eigenvalue of $\Sigma(\Sigma+\Lambda_{\text{fid}})^{-1}$ (which ensures that $\xi(\Sigma+\Lambda_{\text{fid}})-\Sigma$ is always positive definite). Since the support of ξ contains 1, the prior includes the fiducial model Λ_{fid} as a possible truth. The generalized density ξ^{-1} is an improper prior (it integrates to infinity) which can be derived as the well known Jefereys' non-informative prior within the class of distributions $\Lambda \sim$

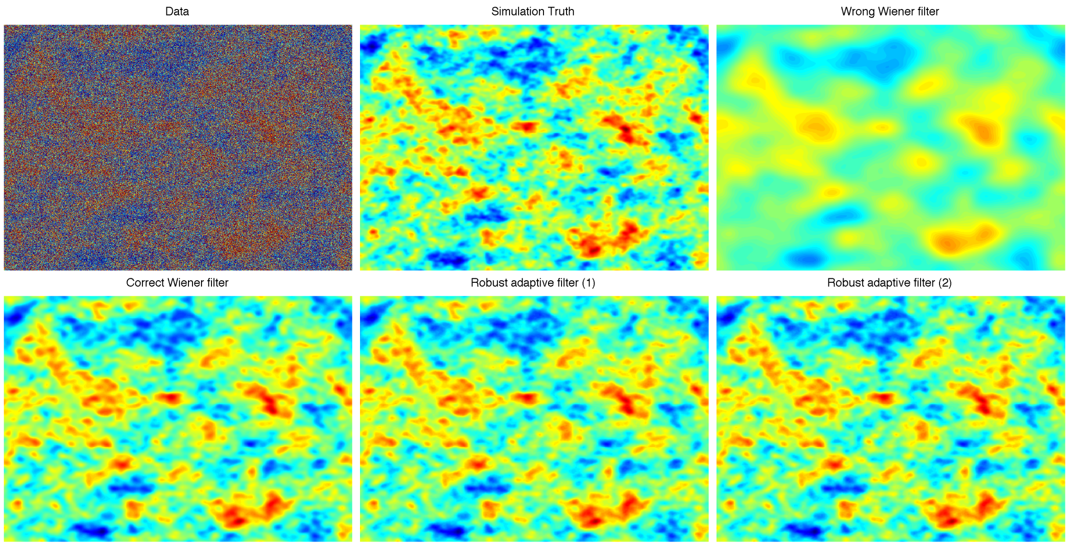

Fig. 2. Illustration of the adaptive Wiener filter. The simulation truth $\phi(\boldsymbol{x})$ is shown in the top-middle image and the data is shown top-left. Two robust Bayes estimates are shown using an input fiducial spectral model $C_{\boldsymbol{l},\text{fid}}^{\phi\phi}$ set 100 times too small (bottom-middle) and set 100 times too large (bottom-right). In contrast, the Wiener filter based on a fiducial spectral density set 100 times too small is shown in the top-right whereas the optimal Wiener filter based on the true but unknown spectral density $C_{\boldsymbol{l}}^{\phi\phi}$ is shown bottom-left.

$\xi(\Sigma + \Lambda_{\text{fid}}) - \Sigma$ for the hierarchal parameter ξ. In [1] the authors derive the posterior distribution for ξ as

$$\mathcal{P}\big(\xi|\mathbf{y}\big) \propto \xi^{-n-1} \exp\big(-\|\mathbf{y}\|^2_{\text{fid}}/\xi\big) \tag{12}$$

on (ρ, ∞) where $\|\mathbf{y}\|^2_{\text{fid}}$ is defined by

$$\|\mathbf{y}\|^2_{\text{fid}} \equiv \Big\|(\Sigma + \Lambda_{\text{fid}})^{-1/2}\text{Re}(\mathbf{y})\Big\|^2 + \Big\|(\Sigma + \Lambda_{\text{fid}})^{-1/2}\text{Im}(\mathbf{y})\Big\|^2.$$

This shows that the posterior $\mathcal{P}(\xi|\mathbf{y})$ has a truncated inverse gamma distribution, which can easily be sampled from. The posterior samples from $\mathcal{P}(\xi|\mathbf{y})$ do not have a direct physical interpretation. However, Algorithm 1 shows how sampling from $\mathcal{P}(\xi|\mathbf{y})$ allows easy sampling from $\mathcal{P}(\boldsymbol{\phi}|\mathbf{y})$. In addition, the posterior expected shrinkage factor in (9) can be computed as follows

$$\int \Lambda(\Sigma + \Lambda)^{-1}\mathcal{P}(\Lambda|\mathbf{y})d\Lambda = I - F^y\, \Sigma(\Sigma + \Lambda_{\text{fid}})^{-1} \tag{13}$$

where $F^y \equiv \int \xi^{-1}\mathcal{P}(\xi|\mathbf{y})$ and I denotes the $n \times n$ identity matrix. Keeping track of the normalization factor in (12), one obtains the following analytic expression for F^y

$$F^y = \frac{n-1}{\|\mathbf{y}\|^2_{\text{fid}}} \frac{P\big(n, \|\mathbf{y}\|^2_{\text{fid}}/\rho\big)}{P\big(n-1, \|\mathbf{y}\|^2_{\text{fid}}/\rho\big)} \tag{14}$$

where $P(a, x)$ is the normalized incomplete gamma function given by $P(a, x) \equiv \frac{1}{\Gamma(a)}\int_0^x t^{a-1}e^{-t}dt$. Combining equations (13) and (9) one obtains the following adaptive shrinkage estimate of $\boldsymbol{\phi}$ given $\mathbf{y}$

$$\tilde{\boldsymbol{\phi}} \equiv \int \boldsymbol{\phi}\mathcal{P}(\boldsymbol{\phi}|\mathbf{y})d\boldsymbol{\phi} = \big[I - F^y\, \Sigma(\Sigma + \Lambda_{\text{fid}})^{-1}\big]\mathbf{y} \tag{15}$$

which is recognized as the matrix form of (4).

Remark 1: In our implementation we partition the Fourier frequencies $\boldsymbol{l}$ into concentric annuli around the origin and construct the posteriors $\mathcal{P}(\xi|\mathbf{y})$ and $\mathcal{P}(\boldsymbol{\phi}|\mathbf{y})$ separately on each annuli. In this way we obtain distinct shrinkage factor adjustments F^y for each annuli, hence the dependence of F^y on $\boldsymbol{l}$, written $F^y_{\boldsymbol{l}}$ in (4). This essentially adds flexibility by allowing independent priors $\mathcal{P}(\xi)$ for each annuli. In fact, if $C_{\boldsymbol{l}}^{\epsilon\epsilon}$ changes drastically within an annuli, one may further partition with the goal of obtaining partitions within which the values of $C_{\boldsymbol{l}}^{\epsilon\epsilon}$ are similar.

Remark 2: Although one has access to an analytic expression for the F^y given in (14), this formula becomes numerically unstable when either $\|\mathbf{y}\|^2_{\text{fid}}$ is small or n large. Therefore we recommend approximating $F^y_{\boldsymbol{l}} = \int \xi^{-1}\mathcal{P}(\xi|\mathbf{y})d\xi$ by averaging samples of ξ^{-1} obtained from step 3 in Algorithm 1.

Algorithm 1 Sample from $\mathcal{P}(\boldsymbol{\phi}|\mathbf{y})$

1: **Simulate** a Gamma random variable ζ with density proportional to $\zeta^{n-1}\exp(-\zeta\|\mathbf{y}\|^2_{\text{fid}})$.
2: **If** $\zeta^{-1}(\Sigma + \Lambda_{\text{fid}}) - \Sigma$ is positive definite **then** go to step 3, **else** go back to step 1.
3: **Set** $\xi \leftarrow \zeta^{-1}$ and $\Lambda \leftarrow \xi(\Sigma + \Lambda_{\text{fid}}) - \Sigma$.
4: **Simulate** $\boldsymbol{\phi} \sim \mathcal{N}\Big(\Lambda(\Sigma+\Lambda)^{-1}\mathbf{y},\, [\Sigma^{-1} + \Lambda^{-1}]^{-1}\Big)$.
5: **Return** $\boldsymbol{\phi}$.

2.2. James-Stein shrinkage

There is an interesting connection with the adaptive Wiener filter (4) and the famous James-Stein shrinkage estimate [8] which sheds light on the nature of the adaptivity factor F^y. The cleanest connection occurs when the fiducial model $C_{\boldsymbol{l},\text{fid}}^{\phi\phi}$ is set to 0 (so that $\Lambda_{\text{fid}} = 0$) which postulates no signal in the observations $y(\boldsymbol{x})$. We also make the assumption that $C_{\boldsymbol{l}}^{\epsilon\epsilon}$ is a constant function of $\boldsymbol{l}$ so that Σ can be written as a constant multiple of the identity, $2\sigma^2 I$. Now the estimate (15) simplifies to $\tilde{\boldsymbol{\phi}} = \big[1 - F^y\big]\mathbf{y}$ and the results found in the Appendix of [1] establish that

$$[1 - F^y] = \left[1 - \frac{(2n-2)\sigma^2}{\mathbf{y}^\dagger \mathbf{y}^*}\right]^+ + o(1) \tag{16}$$

where $o(1) \to 0$ uniformly in $\mathbf{y}$ as $n \to \infty$ and $x^+ \equiv \max\{0, x\}$. The right hand side of (16) is the shrinkage factor used in the famous James-Stein shrinkage estimator [8]. This follows since the real and imaginary parts of $\mathbf{y}$ constitute $2n$ independent Gaussian random variables, each with variance σ^2. In fact, for any fixed number $\epsilon > 0$, the supremum of $|o(1)|$ over the region $\frac{\mathbf{y}^\dagger \mathbf{y}^*}{2n-2} \geq \sigma^2 + \epsilon$ converges to zero exponentially fast as $n \to \infty$. Therefore if one changes $2n-2$ in the numerator of (16) to, say $2n$, this exponential convergence fails to hold. Even more is true: under the same assumptions on Λ_{fid} and Σ, the results of [9] show that $\tilde{\boldsymbol{\phi}}$ is a minimax and admissible estimate of $\boldsymbol{\phi}$ with respect to the quadratic loss when $n \geq 2$. Similar results for (15) and other Bayes estimates have been derived in the statistical literature (see [10, 5] for a review of the literature). For non-asymptotic n, the shrinkage factor is essentially a smoothed out version of the James-Stein shrinkage factor. This can be seen in Figure 1 which compares $1 - F^y$ to the classic James-Stein shrinkage factor.

3. SIMULATION

In Figures 2 and 3 we illustrate the performance of the adaptive Wiener filter, when using a fiducial spectral model is wrong by a factor of 100, to the problem of filtering a noisy observation of the cosmic microwave background, denoted here as ϕ, (under a flat sky approximation) observed on a $8.5^o \times 8.5^o$ patch of sky with 0.5 arcmin pixels. The CAMB code [11] is used to generate the theoretical power spectra of the true signal $C_{\boldsymbol{l}}^{\phi\phi}$ with similar cosmological parameters as used in [1].

The two images in upper left corner of Figure 2 show the data $y(\boldsymbol{x})$ (top-left) and the true signal $\phi(\boldsymbol{x})$ (top-middle). The remaining images in Figure 2 show four different estimates of $\phi(\boldsymbol{x})$: the *non-adaptive* Wiener filter of $y(\boldsymbol{x})$ based on $C_{\boldsymbol{l},\text{fid}}^{\phi\phi}$ which is 100 times too small (top right); the *non-adaptive* Wiener filter of $y(\boldsymbol{x})$ based on the correct $C_{\boldsymbol{l}}^{\phi\phi}$ (bottom-left); the *adaptive* Wiener filter based on $C_{\boldsymbol{l},\text{fid}}^{\phi\phi}$ which is 100 times too small (bottom-middle) and 100 times too large (bottom-right). Figure 3 shows the simulation truth spectrum $|\boldsymbol{l}|^2 C_{\boldsymbol{l}}^{\phi\phi}$ (solid blue line), one of the fiducial spectrums $|\boldsymbol{l}|^2 C_{\boldsymbol{l},\text{fid}}^{\phi\phi}$ which is 100 times too small (solid red line) and the noise spectrum $|\boldsymbol{l}|^2 C_{\boldsymbol{l}}^{\epsilon\epsilon}$ (solid black line). In addition, Figure 3 shows the posterior confidence regions of spectral power using Algorithm 1 (shown in light blue) compared to the non-adaptive Wiener filter technique (shown in light red), where both the adaptive and non-adaptive filters are using the same fiducial model (solid red line).

In Figure 2 there are two things to notice. First, if one uses the Wiener filter based on the wrong fiducial model $C_{\boldsymbol{l},\text{fid}}^{\phi\phi}$ one seriously over shrinks the data (i.e. top-right shrinks too much). Secondly, the new estimate is both robust and adaptive: robust because the bottom-right and bottom-middle are similar (even though the fiducial models are different by a factor of 10^4); adaptive because the bottom-right and bottom-middle behave like the optimal Wiener filter shown bottom-left. In addition, Figure 3 gives further evidence that the posterior samples of spectral power (shown in light blue), from Algorithm 1, are adapting to the observations by producing confidence regions which are scattered around the true spectrum rather than the fiducial model (solid red line). As a comparison, the corresponding posterior samples using the non-adaptive Wiener filter (light red) based on the same fiducial model completely miss the true spectrum.

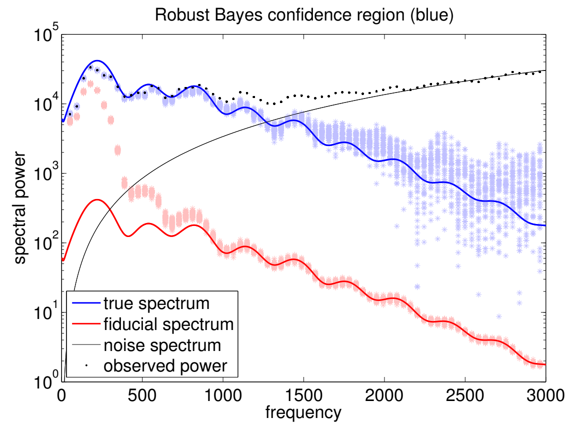

Fig. 3. The light blue marks illustrate the adaptivity of the posterior samples from Algorithm 1 to estimate the true spectrum (solid blue line) even when the fiducial spectrum $C_{\boldsymbol{l},\text{fid}}^{\phi\phi}$ is 100 times too small (solid red line). In contrast, the light red marks illustrate how wrong the corresponding non-adaptive Wiener filter can be when using the same fiducial model . The noise spectrum $|\boldsymbol{l}|^2 C_{\boldsymbol{l}}^{\epsilon\epsilon}$ is shown in black.

4. REFERENCES

[1] E. Anderes and D. Paul, “Shrinking the Quadratic Estimator,” *ArXiv/1110.1694*, Oct. 2011.

[2] W. Hu, “Mapping the Dark Matter through the Cosmic Microwave Background Damping Tail,” *The Astrophysical Journal*, vol. 557, pp. L79–L83, Aug. 2001.

[3] W. Hu and T. Okamoto, “Mass Reconstruction with Cosmic Microwave Background Polarization,” *The Astrophysical Journal*, vol. 574, pp. 566–574, Aug. 2002.

[4] T. Okamoto and W. Hu, “Cosmic microwave background lensing reconstruction on the full sky,” *Physical Review D*, vol. 67, no. 8, pp. 083002, Apr. 2003.

[5] J. O. Berger, *Statistical decision theory and Bayesian analysis*, 1985.

[6] James Berger, “A robust generalized Bayes estimator and confidence region for a multivariate normal mean,” *Ann. Statist.*, vol. 8, no. 4, pp. 716–761, 1980.

[7] William E. Strawderman, “Proper Bayes minimax estimators of the multivariate normal mean,” *Ann. Math. Statist.*, vol. 42, no. 1, pp. 385–388, 1971.

[8] W. James and C. Stein, “Estimation with quadratic loss,” in *Proc. Fourth Berkeley Symp. Math. Statist. Prob*, 1961, vol. 1, pp. 361–379.

[9] James O. Berger, “Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss,” *Ann. Statist.*, vol. 4, no. 1, pp. 223–226, 1976.

[10] EL Lehmann and George Casella, *Theory of Point Estimation (Springer Texts in Statistics)*, Springer, 2nd edition, 1998.

[11] Antony Lewis, Anthony Challinor, and Anthony Lasenby, “Efficient computation of CMB anisotropies in closed FRW models,” *Astrophys. J.*, vol. 538, pp. 473–476, 2000.