Supplementary Material to
Uncertainty Quantification in Dynamic Image Reconstruction with Applications to Cryo-EM

Tze Leung Lai[1], Shao-Hsuan Wang[2], Szu-Chi Chung[3], Wei-hau Chang[4], and I-Ping Tu*[5]

[1]*Department of Statistics, Stanford University,*

[2] *Graduate Institute of Statistics, National Central University,*

[3]*Department of Applied Mathematics, National Sun Yat-sen University,*

[4]*The Institute of Statistical Science, Academia Sinica*

[5]*Institute of Chemistry, Academia Sinica*

S1. "Einstein from noise" and model bias in image alignment

The study of Wang et al. (2021, Section 2) treats a white-noise image as a uniformly distributed random vector on the $(p-1)$-dimensional sphere and an image of p pixels as a vector in the p-dimensional Euclidean space. A widely used similarity measure in image processing is the cross correlation (CC) of two images. The simulation study considers $n = 2\times 10^6$ while-noise images with pixel number $p = 120\times 120 = 1.44\times 10^4$. Among the n white-noise images, the largest cross correlation with Einstein's facial image (the reference) was 0.039 but increased to 0.65 after averaging the $m = 800$ images that are most highly cross-correlated with the reference image. Similar results were obtained for other reference images in the simulation study. Let $\rho_{n,p,m}$ denote the cross correlation of the reference image and the average of the m white-noise images that are most highly cross-correlated with the reference image. In particular, $\rho_{n,p,1}$ is the largest cross correlation value with the reference image among the n white-noise images. The distribution of the bias $\rho_{n,p,m}$ depends on n, p, and m in a convoluted manner. Section 3 of Wang et al. (2021) uses extreme value theory and multivariate analysis to derive approximations to the distribution of $\rho_{n,p,1}$ when n and p are both large and to the distribution of $\rho_{n,p,m}$ when $p = p_n$ and $m = m_n$ both grow with n such that $(\log n)^2 = o(p_n)$ and $m_n = o(n)$.

Wang et al. (2021) were able to derive these explicit expressions to explain the "Einstein from noise"phenomenon by assuming i.i.d. standard normal components of n white-noise images. In practice, a cryo-EM image consists of regions which contain only noise whose variance may vary over the regions and need not equal 1, regions with noisy 2D projections, and regions with other contaminants. The bias of particle picking for reference matching becomes much more difficult to quantify, and the work of the

Validation Task Force is a community attempt to validate cryo-EM image reconstruction.

S2. Markov Chain Monte Carlo with Sequential Substitutions for Joint Parameter and State Estimation in HMMs

Lai (2021) summarizes major breakthroughs in this very important area in time series analysis, image and signal processing, robotics, automatic navigation, bioengineering, and control systems, these methodological advances was in martingale representation of the particle filter in a hidden Markov model (HMM) given by Chan and Lai (2013, Lemmas 1 and 4). The particle filter (or sequential Monte Carlo method) was introduced by Gordon et al. (1993). The assumption of a single, fully specified, HMM is too restrictive in applications since the model parameters are usually unknown and need to be estimated sequentially from the observed data, hence joint state and parameter estimation is the usual estimation task in the aforementioned applications of hidden Markov models. In the Bayesian framework, the unknown parameter θ has a prior distribution so that the posterior joint distribution of θ and the states is the target distribution to be estimated.

The Metropolis–Hastings (MH) Markov chain Monte Carlo (MCMC) scheme provides a simulation-based method to estimate a target distribution. Andrieu et al. (2010) and Chopin et al. (2013) used this approach in their "particle MCMC" and "SMC2" procedures, respectively, but encountered difficulties in approximating the joint target and state distribution. Lai, together with Hock Peng Chan who received his Ph.D. under Lai in 1998 and his Ph.D. students Huangzhong Xu and Michael Hongyu Zhu, circumvented this difficulty by developing a new MCMC scheme called MCMC with sequential substitutions (MCMC-SS); see Lai (2021, Theorem 2) that provides (i) a comprehensive asymptotic theory of MCMC-SS showing its asymptotic optimality with respect to computational and statistical criteria, and (ii) consistent estimators of standard errors of the Monte Carlo state and parameter estimates. MCMC-SS uses B disjoint blocks $\Theta_{b,k}$ of ν atoms within each block and uses the following sequential substitution procedure $SS(\Theta_{b,k}, \mathbf{w}_k^b)$ at stage k to update the atom set in $\Theta_{b,k}$, $b = 1, \ldots, B$:

(a) Let $\{q(\cdot|\gamma) : \gamma \in \Gamma\}$ be a family of proposal densities with respect to some measure m, and sample $\widetilde{\theta}$ from $q(\cdot|\gamma_{b,k-1})$ as the candidate atom.

(b) Let $\theta^b_{\nu+1,k-1} = \widetilde{\theta}$ and compute $\lambda^b_{i,k} = q\left(\theta^b_{i,k-1}|\gamma_{b,k-1}\right)/f\left(\theta^b_{i,k-1}\right)$, $i = 1, \ldots, \nu+1$, in which f is a given function that is proportional to the target density.

(c) Sample J from $\{1, \ldots, \nu+1\}$ with probability $\pi_{i,k} = \lambda^b_{i,k}/\left(\sum_{j=1}^{\nu+1} \lambda^b_{j,k}\right)$ for i.

(d) If $J = \nu + 1$, let $\Theta_{b,k} = \Theta_{b,k-1}$. Otherwise let $\Theta_{b,k} = \left(\Theta_{b,k-1} \cup \{\widetilde{\theta}\}\right) \setminus \left\{\theta^b_{\widetilde{J},k-1}\right\}$.

(e) Let $w^b_{i,k} = 1/\pi^b_{i,k}$ for $i = 1, \ldots, \nu$, and $\mathbf{w}^b_k = \left(w^b_{1,k}, \ldots, w^b_{\nu,k}\right)$.

In many applications, the parameter γ in the proposal density $q(\cdot|\gamma)$ is a function $\gamma : \mathcal{P} \to \Gamma$, where $\mathcal{P}$ is the space of probability measures on Θ. Assuming this framework, Lai (2021) describes the choice of $\gamma_{b,k-1}$ in the updating algorithm of MCMC-SS. Let κ represent an initial burn-in period that is asymptotically negligible in comparison with the total number K of iterations in the asymptotic theory with $\kappa \to \infty$ such that $\kappa = o(K)$. For $k \leq \kappa$, let $\gamma_{b,k-1} = \nu^{-1} \sum_{\theta \in \Theta_{b,k-1}} \gamma(\theta)$, which is the mean of the empirical measure of the atoms in the bth block at the end of stage $k-1$. On the other hand, for $k > \kappa$, pool across blocks by letting $\widetilde{\gamma}_{k-1} = B^{-1} \sum_{b=1}^{B} \gamma_{b,k-1}$, which we use as the modified $\gamma_{b,k-1}$ for all blocks. Therefore, after the burn-in period, carry out the update $SS(\theta_{b,k})$ in the order $b = 1, \ldots, B$, so that if the candidate atom in $SS(\theta_{b,k})$ is not used for block b, it can serve as candidate atom for block $b+1 (\leq B)$, which then does not need to generate another random variable from $q(\cdot|\widetilde{\gamma}_{k-1})$, an obvious advantage for high-dimensional complicated states. MCMC-SS estimates $\mu = E_p \psi(\theta)$ for which $E_p\psi^2(\theta) < \infty$ by

$$\widehat{\psi} = \frac{1}{B(K-\kappa)} \sum_{b=1}^{B} \sum_{k=\kappa+1}^{K} \widehat{\psi}_{b,k}, \quad \text{with } \widehat{\psi}_{b,k} = \frac{\sum_{i=1}^{\nu} w_{i,k}^{b} \psi(\theta_{i,k}^{b})}{\sum_{i=1}^{\nu} w_{i,k}^{b}}, \tag{S.1}$$

and $\sigma^2 := \mathrm{Var}_p\left(\psi(\theta)\right)$ by

$$\hat{\sigma}^2 = \frac{1}{B(K-\kappa)} \sum_{b=1}^{B} \sum_{k=\kappa}^{K} \frac{1}{\nu-1} \sum_{\theta \in \Theta_{b,k}} (\psi(\theta) - \widehat{\psi}_{b,k})^2. \tag{S.2}$$

Moreover, $\hat{\sigma}^2$ is a consistent estimate of σ^2 and with probability approaching 1 for large k, the candidate atom $\widetilde{\theta}$ indeed substitutes some existing atom in $\Theta_{b,k-1}$. Hence, similar to the case of known target density p, each random variable generated in the MCMC-SS scheme asymptotically contributes weight $(B(K-\kappa))^{-1}$, to (a) the estimate $\widehat{\psi}$ of μ and (b) the asymptotic variance of $\widehat{\psi}$. Wu et al. (2021) describe applications of MCMC-SS to image reconstruction and latent variable analysis with uncertainty quantification in infants' brain network development.

Additional References

Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B* **72,** 269–342.

Chan, H. P. and Lai, T. L. (2013). A general theory of particle filters in hidden Markov models and some applications. *The Annals of Statistics* **41,** 2877–2904.

Chopin, N., Jacob, P. E., and Papaspiliopoulos, O. (2013). SMC2: an efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society: Series B* **75,** 397–426.

Gordon, N. J., Salmond, D. J., and Smith, A. F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *In IEE Proceedings F-radar and Signal Processing* **140,** 107–113.

Wang, S. H., Yao, Y. C., Chang, W. H., and Tu, I P. (2021). Quantification of model bias underlying the phenomenon of "Einstein from noise". *Statistica Sinica,* in press (doi: 10.5705/ss.202020.0334).

Wu, H. T., Lai, T. L., Haddad, G. G., and Muotri, A. (2021). Oscillatory biomedical signals: Frontiers in mathematical models and statistical analysis. *Frontiers in Applied Mathematics and Statistics* (doi: fams.2021.689991).