Metropolis Hastings within Partially Collapsed Gibbs
with Application in High-Energy Astrophysics

David A. van Dyk, Department of Statistics, University of California, Irvine, USA, dvd@ics.uci.edu
Taeyoung Park, Department of Statistics, University of Pittsburgh, USA, taeyoung.t.park@gmail.com

Abstract: The recently proposed Partially Collapsed Gibbs (PCG) sampler offers a new strategy for improving the convergence of a Gibbs sampler (van Dyk and Park, 2008, Park and van Dyk, 2009). PCG achieves faster convergence by reducing the conditioning in some of the draws of its parent Gibbs sampler. Although this can significantly improve convergence, care must be taken to ensure the stationary distribution is preserved. The conditional distributions sampled in a PCG sampler may be functionally incompatible and permuting them may upset the stationary distribution. Extra care must be taken when Metropolis steps are used. Reducing the conditioning in an MH within Gibbs sampler can change the stationary distribution, even when the PCG sampler would work perfectly if all the conditional updates were available. We illustrate the challenges that may arise when using an MH within PCG sampler and develop a general strategy for using such updates while maintaining the stationary distribution.

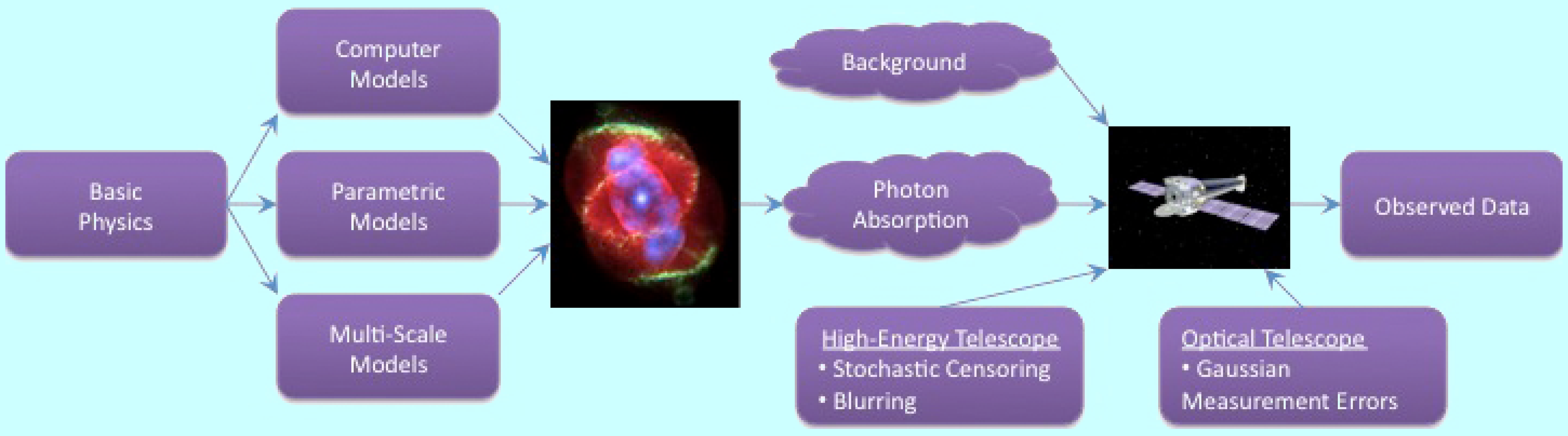

Fig. 2: Overview of Computer-Model Embedded Highly-Structured Models. Basic physics informs the construction of computer, parametric, and multi-scale models for astronomical sources. Photons are partially absorbed, background contaminated and degenerated by instrumental effective area, blurring, and errors. Sophisticated statistical models account for all of these processes.

Two Simple Illustrations:

Illustration I: Consider a two-step MH within Gibbs sampler:

STEP 1: $\psi_1 \sim \mathcal{K}(\psi_1|\psi_2)$ via MH with limiting distribution $p(\psi_1|\psi_2)$
STEP 2: $\psi_2 \sim p(\psi_2|\psi_1)$ (Sampler I)

Sampler I can be slow if ψ_1 and ψ_2 are highly correlated and could be improved if we use a kernel with $p(\psi_1)$ as its limiting distribution in STEP 1 instead of $p(\psi_1|\psi_2)$:

STEP 1: $\psi_1 \sim \mathcal{K}(\psi_1|\psi_2)$ via MH with limiting distribution $p(\psi_1)$
STEP 2: $\psi_2 \sim p(\psi_2|\psi_1)$ (Sampler II)

- To do this, we need only evaluate $p(\psi_1) = p(\psi_1, \psi_2)/p(\psi_2|\psi_1)$ using the properly normalized conditional from STEP 2.
- If we were able to sample $p(\psi_1)$ directly without resorting to MH in STEP 1 we would obtain i.i.d. draws from $p(\psi_1, \psi_2)$.
- This would be a simple special case of Partially Collapsed Gibbs. For this reason, we call Sampler II an MH within PCG sampler.

Illustration II: Consider another two-step MH within Gibbs sampler, similar to Sampler I, but with a MH step is STEP 2:

STEP 1: $\psi_1 \sim p(\psi_1|\psi_2)$ (Sampler III)
STEP 2: $\psi_2 \sim \mathcal{K}(\psi_2|\psi_1)$ via MH with limiting distribution $p(\psi_2|\psi_1)$

As in Illustration I, we may try to improve convergence by replacing STEP 1 with a draw from $p(\psi_1)$:

STEP 1: $\psi_1 \sim p(\psi_1)$ (Sampler IV)
STEP 2: $\psi_2 \sim \mathcal{K}(\psi_2|\psi_1)$ via MH with limiting distribution $p(\psi_2|\psi_1)$

Surprisingly, Sampler IV does not have the correct stationary distribution. To see this suppose (ψ_1, ψ_2) are bivariate normal with zero means, unit variances, and correlation $\rho = 0.9$. Figure 1 shows the results of Samplers III and IV using a Gaussian jumping rule with mean $\psi_2^\star$ and variance τ^2 in STEP 2. (We use a star in the superscript to represent the output from the previous iteration.)

The problem with Sampler IV is that the distribution used to update ψ_2 in STEP 2 depends on both ψ_1 and $\psi_2^\star$:

- The limiting distribution of the MH step depends on ψ_1
- If the proposal is rejected, ψ_2 is set to $\psi_2^\star$.

Because ψ_1 is drawn from its marginal distribution, it is necessarily independent of $\psi_2^\star$ *at every iteration* and STEP 2 will never produce samples from $p(\psi_2|\psi_1)$.

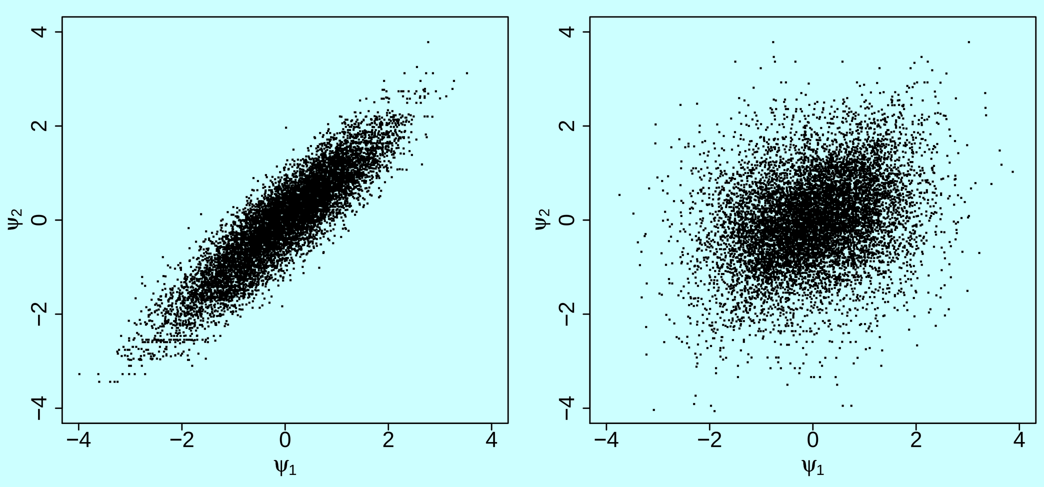

Fig. 1: Monte Carlo samples generated with Sampler III (left) and Sampler IV (right).

The General Strategy

Starting with an MH within Gibbs sampler we can derive a correct MH within PCG sampler by following three simple steps:

1. *Marginalize:* Move parameter components from being conditioned upon to being sampled in one or more step of the sampler.
2. *Permute:* Reorder of sampler steps to facilitate trimming (step 3).
3. *Trim:* Eliminate sampled parameter components that are neither inputs of subsequent steps nor outputs of the PCG iteration.

Example 1: Image Analysis

Image Analysis in High-Energy Astrophysics: The small number of photons in an X-ray or γ-ray image are best modeled with a Poisson process, see schematic in Figure 2:

1. We observe a photon count in each of a large grid of spatial pixels.
2. The counts are blurred; they may be recorded in the wrong pixel.
3. A known point spread function characterizes the blurring.
4. The detector sensitivity may vary across the detector.

NGC 6240, a galaxy that is the product of the collision of two smaller galaxies, is illustrated in Figure 3. The pixelated X-ray image shows far lower resolution than the Hubble image, but highlights the two bright black holes near the galaxy's center.

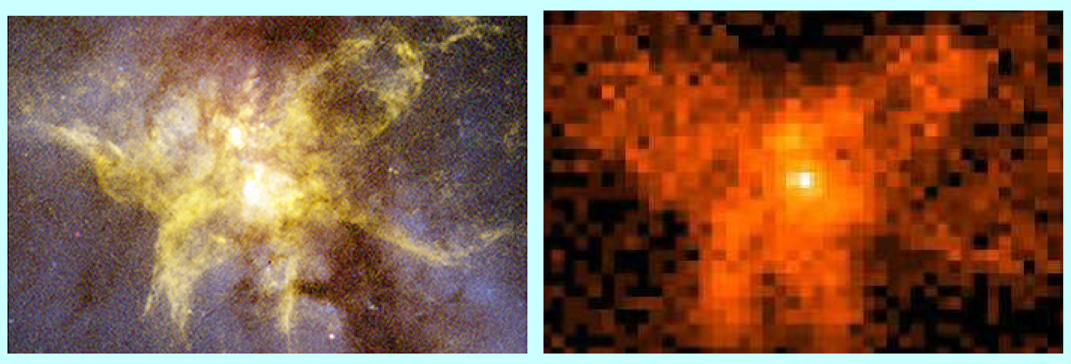

Fig. 3: Optical (left) and x-ray (right) images of NGC 6240.

Model based Analysis: We frame image analysis as a statistical inference problem to facilitate the quantification of uncertainty and the evaluation of evidence for scientific hypotheses. (E.g., the existence of particular structures in the image's source.) Our multi-level statistical model involves:

1. μ: The matrix of underlying poisson intensities;
2. α: The smoothing parameter for μ; and
3. Z: The ideal image of de-blurred photon counts.

Our Multi-Scale smoothing uses a sequential division of the image into quadrants and models the quadrant counts as multinomials with Dirichlet priors. The Dirichlet parameters may differ at different scales, are tabulated in the parameter α, and are fit to the data.

Statistical Computing: A three-step MH within Gibbs sampler can be used to sequentially update μ, α, and Z, using a Metropolis-Hasings step for the Dirichlet parameter, α:

STEP 1: Sample Z given μ, α, and Y
STEP 2: Sample μ given Z, α, and Y.
STEP 3: Sample α given Z, μ, and Y using an MH step.

The result, however, is unsatisfactory as is illustrated by the time-series plot of one of the components of α in the top panel of Figure 4.

By eliminating the conditioning on μ when updating α and permuting the steps, however, we can dramatically improve convergence:

STEP 1: Sample Z given μ, α, and Y
STEP 2: Sample (α, μ) given Z and Y using an MH step.
STEP 3: Sample μ given Z, α, and Y.

The μ sampled in STEP 2 is not used and can be eliminated.

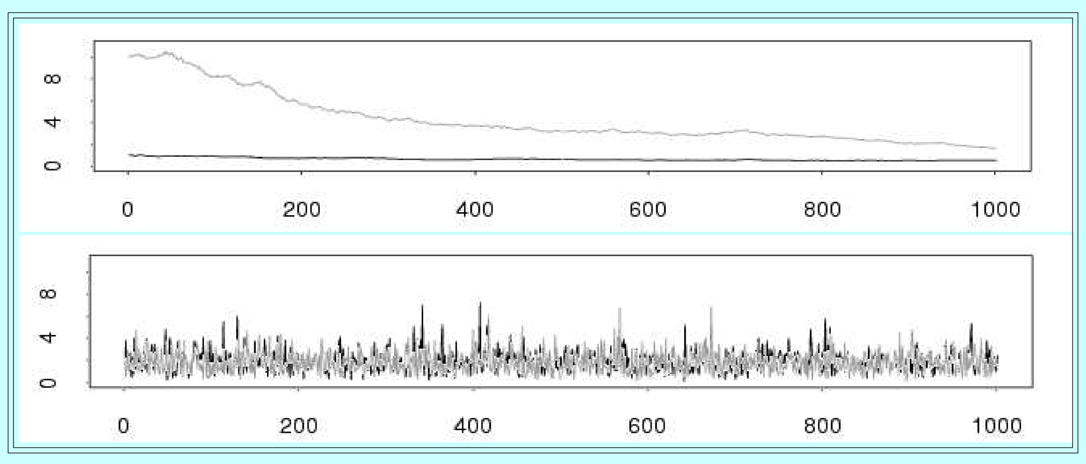

Fig. 4: Time series plot for α using an MH within Gibbs Sampler (top) and an MH within PCG Sampler (bottom).

Example 2: Spectral Calibration

Calibration Uncertainty: Instrumental characteristics such as the point spread function are subject to uncertainty. Since inference depends on these quantities, it should account for calibration uncertainty. Here we discuss the *effective area* of the instrument, which quantifies its varying sensitivity as a function of photon energy.

Uncertainty in the *effective area* can be described by a *calibration sample*, a set of curves provided by calibration scientists that aims to capture the possible range of curves. Figure 5 illustrates a calibration sample. From a Bayesian perspective, we can view the calibration sample as a sample from the *prior of the effective area curve*.

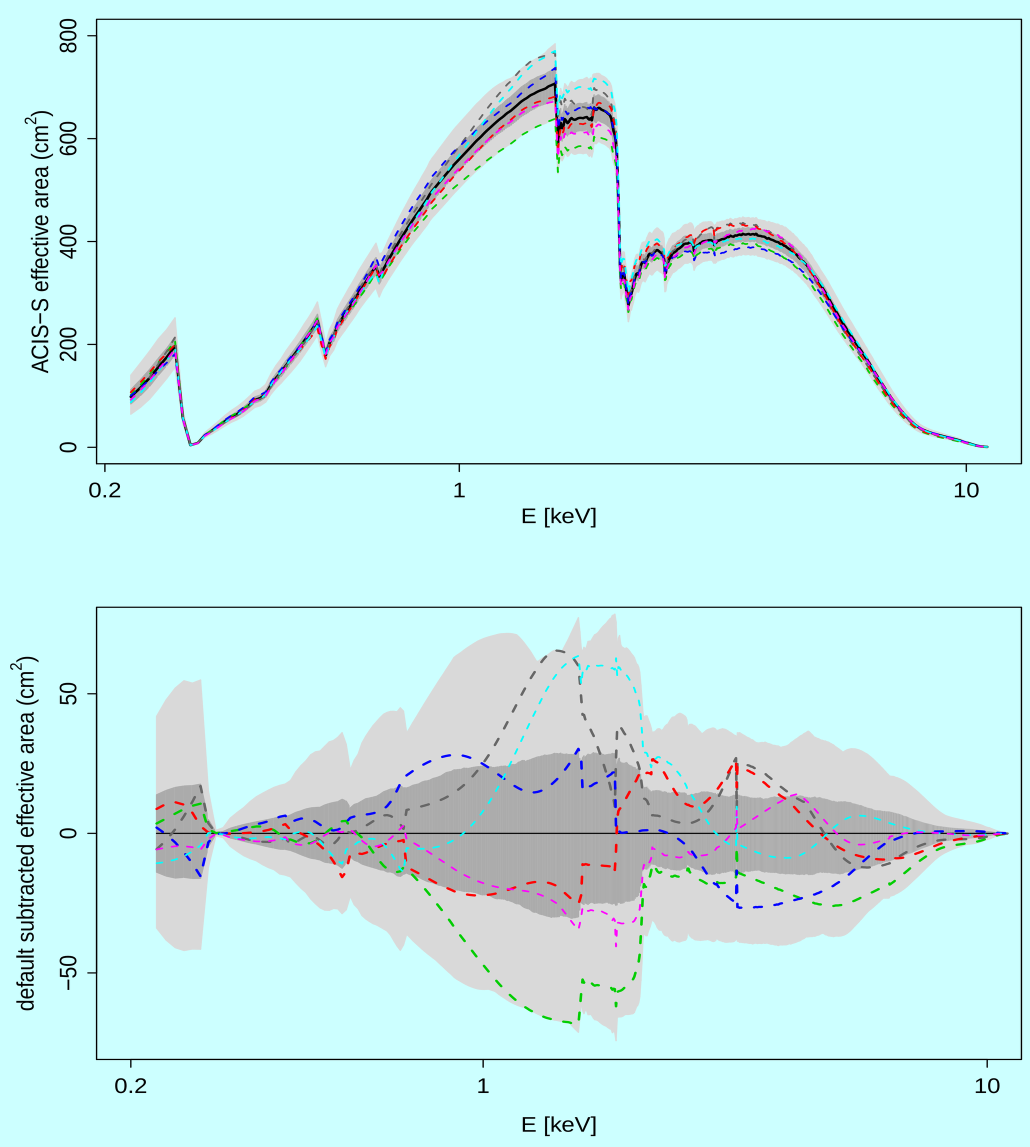

Fig. 5: The point wise range of an effective area calibration sample (top) and the mean-subtracted sample (bottom). The darker region contains 68% of the curves. Five sample curves are highlighted.

Statistical Computing: We use an MH within Gibbs sampler. Letting A represent the effective area curve and θ the parameters:

STEP 1: Sample A given θ and Y.
STEP 2: Sample θ given A and Y using an MH step.

STEP 2 is the standard algorithm with known A. We consider:

PRAGMATIC BAYES TARGET: $\pi_0(A, \theta) = p(A)p(\theta|A, Y)$.
FULLY BAYESIAN TARGET: $\pi(A, \theta) = p(A|Y)p(\theta|A, Y)$.

Sampling $\pi_0(A|\theta, Y)$ is complicated, but sampling $\pi_0(A) = p(A)$ is trivial. This leads to an MH within PCG sampler with the wrong stationary distribution, see Illustration II. To avoid this, we iterate STEP 2, so that it does not depend on $\theta^\star$. The advantages of accounting for calibration uncertainty in general, and the Fully Bayesian solution in particular are illustrated in Figure 6.

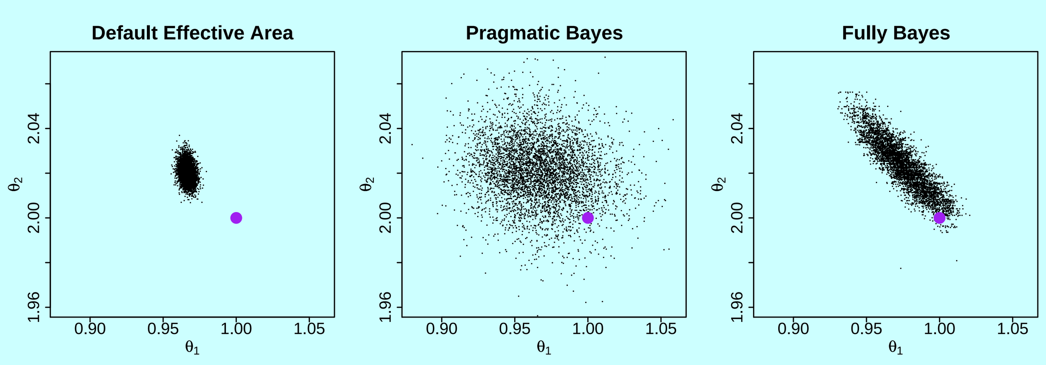

Fig. 6: The posterior distribution of θ when fixing A at a default (left), using the pragmatic Bayes target (middle), and using the fully Bayesian target (right). The purple circle marks the true value of θ.

References:

Lee, H., Kashyap, V. L., van Dyk, D. A., Connors, A., Drake, J. J., Izem, R., Min, S., Park, T., Ratzlaff, P., Siemiginowska, A., and Zezas, A. (2010). Accounting for Calibration Uncertainties in X-ray Analysis: E?ective Areas in Spectral Fitting. *Astrophysical Journal*, submitted.

Park, T. and van Dyk, D. A. (2009). Partially Collapsed Gibbs Samplers: Illustrations and Applications. *Journal of Computational and Graphical Statistics*, **18**, 283-305.

van Dyk, D. A. and Park, T. (2008). Partially Collapsed Gibbs Samplers: Theory and Methods. *Journal of the American Statistical Association*, **103**, 790–796.

van Dyk, D. A., Connors, A., Esch, D. N., Freeman, P., Kang, H., Karovska, M., Kashyap, V., Siemiginowska, A., and Zezas, A. (2006). Deconvolution in High Energy Astrophysics: Science, Instrumentation, and Methods (with discussion). *Bayesian Analysis*, **1**, 189–236.