

Supplement to "Atmospheric inverse modeling with known physical bounds: an example from trace gas emissions"

Scot M. Miller, Anna M. Michalak, and Patricia J. Levi

This supplement describes in greater detail the multiple-try Metropolis-Hastings algorithm and the Gibbs sampler implementation.

1 Existing Metropolis-Hastings implementations

Metropolis-Hastings algorithms can be used in a number of ways to enforce inequality constraints, and existing literature in hydrology (Michalak and Kitanidis, 2004; Wang and Zabarar, 2006; Zanini and Kitanidis, 2009) and atmospheric sciences (Rigby et al., 2011; Burrows et al., 2013) implement different approaches. The algorithms implemented in the cited hydrology papers use Lagrange multipliers to enforce constraints on individual realizations. Such an approach is computationally attractive for larger-scale inverse problems, with hundreds or thousands of elements of the state space for which constraints must be enforced. The cited studies generally use the following approach:

1. Add noise to the observations and the prior. The vectors of random noise are generated using the covariance matrices (\mathbf{R} and \mathbf{Q} , respectively) and the last accepted realization.
2. Create a conditional, constrained realization of the emissions ($\mathbf{s}_{cc,j}$) by minimizing the inversion cost function subject to the randomly-generated inputs above and enforcing constraints on the state vector using the method of Lagrange multipliers.
3. Calculate the likelihood of the candidate realization ($\mathbf{s}_{cc,j}$) relative to the previous, accepted realization ($\mathbf{s}_{cc,j-1}$). Based upon this likelihood, either accept or reject the realization, and begin again with step 1.

These steps are described in more detail in section 2.

The cited existing atmospheric studies use a different approach (Rigby et al., 2011; Burrows et al., 2013). In these studies, a new realization ($\mathbf{s}_{cc,j}$) is generated by adding a random quantity directly to the previous candidate realization ($\mathbf{s}_{cc,j-1}$). If the candidate realization ($\mathbf{s}_{cc,j}$) does not obey the inequality constraints, it is discarded. If the constraints are met in the realization, it is again accepted or rejected based on its relative probability compared to the last accepted realization. This implementation omits steps 1 and 2 above. In other words, the modeler is not required to minimize the cost function and can instead skip directly to step 3. This feature affords greater flexibility in the inversion setup (e.g., Rigby et al., 2011); the modeler can use a complex probability density function (pdf) for the inversion even if calculating its maximum is computationally-intractable. A larger number of multivariate distributions could be used for the prior pdf, not just a multivariate Gaussian distribution (implemented with Lagrange multipliers in the hydrology studies above). The fraction of realizations that are rejected, however, can be very high for two reasons. First, for large state vectors, a very high fraction of candidate realizations will not obey all bounds and will be rejected. Second, there are no formal steps to

ensure that each proposed realization reproduces the atmospheric observations. Therefore, the fraction of realizations that are rejected will be high, even among those realizations that obey the inequality constraints.

The implementation used in the cited hydrology studies is therefore better for larger problems because it samples the posterior probability space more strategically. The implementation generates candidate realizations that explicitly sample the covariance matrices (\mathbf{R} and \mathbf{Q}) and honor the observations (\mathbf{z}). Refer to Chib and Greenberg (1995) or Bolstad (2012) for a general discussion on creating candidate realizations. In general, larger problems require strategic generation of candidate realizations; the efficiency and/or acceptance probability of Metropolis-Hastings decreases as the number of unknowns increase for problems with comparable implementation and covariance matrices (Gelman, 2004, ch. 11). This fact makes the approach used in the cited hydrology literature more suitable for larger problems like the methane case study in the main manuscript. However, this approach also restricts the inversion setup to pdfs with computationally-tractable maxima.

2 The multiple-try Metropolis-Hastings

The following section describes the multiple-try Metropolis-Hastings algorithm modified to accommodate inequality constraints. This algorithm is more computationally tractable for large problems than many traditional Metropolis-Hastings implementations.

This algorithm (Liu et al., 2000) first requires the generation of an unconstrained unconditional realization, denoted \mathbf{s}_{uu} . The realization for step j , denoted $\mathbf{s}_{uu,j}$, is created by applying a modification to $\mathbf{s}_{uu,j-1}$. The modification to the previous realization is provided by what is known as the jumping distribution $T()$. This distribution should create new realizations that are sufficiently different from the previous one such that the algorithm effectively samples the entire probability space. However, the jumping distribution should avoid creating subsequent realizations that are so different such that $\mathbf{s}_{uu,j}$ gets rejected by the algorithm (e.g., Chib and Greenberg, 1995).

The jumping distribution used here requires taking the Cholesky decomposition of \mathbf{Q} :

$$\mathbf{Q} = \mathbf{C}\mathbf{C}^T \tag{1}$$

The distribution $T()$ can be chosen in any number of ways (e.g., Chib and Greenberg, 1995), but we generate new unconditional unconstrained realizations as follows (where \mathbf{u} is a random vector with distribution $\mathcal{N}(0, 1)$):

$$\begin{aligned} \mathbf{s}_{uu,0} &= \mathbf{C}\mathbf{u} \\ \mathbf{s}_{uu,j} &= \phi\mathbf{s}_{uu,j-1} + \sqrt{1 - \phi^2}\mathbf{C}\mathbf{u} \end{aligned} \tag{2}$$

In this case, we set $\phi = 0.9$, though any value greater than zero and less than one is acceptable (e.g., Michalak and Kitanidis, 2004). The multiple-try Metropolis-Hasting with inequality constraints has the following steps:

1. Draw k trial proposals for $\mathbf{s}_{uu,j}$ from the jumping distribution described by Eq. 2.
2. Compute a conditional constrained realization ($\mathbf{s}_{cc,j}$) for each of the trial proposals by minimizing the posterior negative log-likelihood via Lagrange multipliers:

$$L_{s,\beta} = \frac{1}{2}(\mathbf{z} + \mathbf{v} - \mathbf{H}\mathbf{s}_{cc,j}^*)^T \mathbf{R}^{-1}(\mathbf{z} + \mathbf{v} - \mathbf{H}\mathbf{s}_{cc,j}^*) + \frac{1}{2}(\mathbf{s}_{cc,j}^* - \mathbf{s}_{uu,j}^*)^T \mathbf{G}(\mathbf{s}_{cc,j}^* - \mathbf{s}_{uu,j}^*) \quad (3)$$

where \mathbf{v} is a random vector with covariance \mathbf{R} . In this case, the asterisk (*) indicates that the candidate is one of k trial proposals for the realization.

3. Compute the weighting function for each trial proposal:

$$w(\mathbf{s}_{cc,j}^* | \mathbf{s}_{cc,j-1}) = \frac{p''(\mathbf{s}_{cc,j}^* | \mathbf{z}, \mathbf{H}, \mathbf{X})}{T(\mathbf{s}_{cc,j}^* | \mathbf{s}_{cc,j-1})} \quad (4)$$

where $p''(\mathbf{s}_{cc,j}^* | \mathbf{z}, \mathbf{H}, \mathbf{X})$ indicates the posterior probability of $\mathbf{s}_{cc,j}^*$, and $T(\mathbf{s}_{cc,j}^* | \mathbf{s}_{cc,j-1})$ is the jumping probability of $\mathbf{s}_{cc,j}^*$ given $\mathbf{s}_{cc,j-1}$. The posterior probability and approximate jumping probability can be calculated as follows (Michalak and Kitanidis, 2004):

$$p''(\mathbf{s}_{cc,j}^* | \mathbf{z}, \mathbf{H}, \mathbf{X}) \propto \exp[-\frac{1}{2}(\mathbf{z} - \mathbf{H}\mathbf{s}_{cc,j}^*)^T \mathbf{R}^{-1}(\mathbf{z} - \mathbf{H}\mathbf{s}_{cc,j}^*) - \frac{1}{2}\mathbf{s}_{cc,j}^{*T} \mathbf{G}\mathbf{s}_{cc,j}^*] \quad (5)$$

$$T(\mathbf{s}_{cc,j}^* | \mathbf{s}_{cc,j-1}) \propto \exp[-\frac{1}{2}(\mathbf{s}_{uu,j}^* - \phi\mathbf{s}_{uu,j-1})^T \frac{\mathbf{Q}^{-1}}{(1-\phi^2)}(\mathbf{s}_{uu,j}^* - \phi\mathbf{s}_{uu,j-1})] \quad (6)$$

4. Select $\mathbf{s}_{cc,j}$ from the trial proposals by individually, randomly drawing each element from $\mathbf{s}_{cc,j}^*$ with probability proportional to the weighting function $w(\mathbf{s}_{cc,j}^* | \mathbf{s}_{cc,j-1})$. Select the corresponding elements of $\mathbf{s}_{uu,j}^*$ to construct $\mathbf{s}_{uu,j}$.
5. Create $(k-1)$ new trial proposals for $\mathbf{s}_{cc,j-1}$. To do this, draw samples from the jumping distribution $T(\mathbf{s}_{uu,j-1}^* | \mathbf{s}_{uu,j})$ (i.e., $\mathbf{s}_{uu,j-1}^* = \phi\mathbf{s}_{uu,j} + \sqrt{1-\phi^2}\mathbf{C}\mathbf{u}$). Calculate the trial conditional constrained realizations $\mathbf{s}_{cc,j-1}^*$ using the procedure outlined in step 2. Set trial proposal k to $\mathbf{s}_{cc,j-1}$. Finally, calculate the weighting function for each trial conditional constrained realization, $w(\mathbf{s}_{cc,j-1}^* | \mathbf{s}_{cc,j})$.
6. Calculate the acceptance/rejection probability (Liu et al., 2000):

$$\xi = \min \left\{ 1, \frac{\sum^k w(\mathbf{s}_{cc,j}^* | \mathbf{s}_{cc,j-1})}{\sum^k w(\mathbf{s}_{cc,j-1}^* | \mathbf{s}_{cc,j})} \right\} \quad (7)$$

Accept $\mathbf{s}_{cc,j}$ if $\xi > \mathcal{U}(0,1)$. Otherwise, set $\mathbf{s}_{cc,j} = \mathbf{s}_{cc,j-1}$.

Repeat steps 1 – 6 until a sufficient number of realizations have been generated to sample across the entire posterior probability space. Note that unlike the Gibbs sampler, this multiple-try Metropolis-Hastings algorithm does not require discarding realizations from an initial spin-up period. For this application, we choose $k = 8$. Larger values for k can lead to greater acceptance rates but higher computational cost. Liu et al. (2000) note that an acceptance rate of 0.4 – 0.5 is ideal for a multiple-try Metropolis-Hastings algorithm.

3 The Gibbs sampler implementation

The Gibbs sampler requires generating the element-wise conditional probability density, the probability of any individual element in \mathbf{s} given an estimate of all other elements in \mathbf{s} . This conditional density is denoted $p(s_i|\mathbf{s}, \mathbf{z})$ where i is one of m elements in \mathbf{s} . The equations for $p(s_i|\mathbf{s}, \mathbf{z})$ can be found in Michalak (2008) for the inversion setup discussed in this paper.

The Gibbs sampler has the following steps:

1. Make an initial guess for \mathbf{s}_1 where the subscript ‘1’ denotes the first realization of \mathbf{s} .
2. Obtain a new realization, \mathbf{s}_j , from the previous realization, \mathbf{s}_{j-1} . To do this, successively generate a conditional probability for each element in \mathbf{s} , and draw a random sample from each one:

$$\begin{aligned} p(s_{1,j}) &= p(s_{1,j}|s_{2,j-1}, \dots, s_{m,j-1}) \\ p(s_{2,j}) &= p(s_{2,j}|s_{1,j}, s_{3,j-1}, \dots, s_{m,j-1}) \\ p(s_{i,j}) &= p(s_{i,j}|s_{1,j}, \dots, s_{i-1,j}, s_{i+1,j-1}, \dots, s_{m,j-1}) \\ p(s_{m,j}) &= p(s_{m,j}|s_{1,j}, \dots, s_{m-1,j}) \end{aligned} \tag{8}$$

3. Update j to $j + 1$ and continue generating realizations.

Create a large number of realizations (in this case 1200) to fully sample across the posterior probability space. The initial realizations are usually discarded as a “spin-up” period (in this case, the first 200).

In this implementation $p(s_i|\mathbf{s}, \mathbf{z})$ is Gaussian. To enforce the inequality constraints, Michalak (2008) draws a random sample from $p(s_i|\mathbf{s}, \mathbf{z})$ until the random draw falls within the bounds. This draw becomes the estimate for $s_{i,j}$. The approach is equivalent to using a truncated Gaussian as a the prior probability density function, but this implementation avoids the computational challenge of directly computing a multivariate, truncated probability distribution.

This study uses a modified approach for the methane case study. If the random sample from $p(s_i|\mathbf{s}, \mathbf{z})$ is positive, it becomes the estimate for $s_{i,j}$. If the random sample is negative, set $s_{i,j} = 0$. This approach is equivalent to sampling from a truncated normal distribution with an added Dirac delta function. The method adapted here increases the probability of estimating zero emissions for a given flux or emissions location.

References

- Bolstad, W.: Understanding Computational Bayesian Statistics, Wiley Series in Computational Statistics, John Wiley & Sons, 2012.
- Burrows, S. M., Rayner, P. J., Butler, T., and Lawrence, M. G.: Estimating bacteria emissions from inversion of atmospheric transport: sensitivity to modelled particle characteristics, Atmospheric Chemistry and Physics, 13, 5473–5488, doi:10.5194/acp-13-5473-2013, 2013.
- Chib, S. and Greenberg, E.: Understanding the Metropolis-Hastings Algorithm, Am. Stat., 49, pp. 327–335, 1995.
- Gelman, A.: Bayesian Data Analysis., Texts in Statistical Science, Chapman & Hall/CRC, Boca Raton, Florida, 2 edn., 2004.

- Liu, J. S., Liang, F., and Wong, W. H.: The Multiple-Try Method and Local Optimization in Metropolis Sampling, *J. Am. Stat. Assoc.*, 95, 121–134, doi:10.1080/01621459.2000.10473908, 2000.
- Michalak, A. M.: A Gibbs sampler for inequality-constrained geostatistical interpolation and inverse modeling, *Water Resour. Res.*, 44, doi:10.1029/2007WR006645, 2008.
- Michalak, A. M. and Kitanidis, P. K.: Application of geostatistical inverse modeling to contaminant source identification at Dover AFB, Delaware, *Journal of Hydraulic Research*, 42, 9–18, doi:10.1080/00221680409500042, 2004.
- Rigby, M., Manning, A. J., and Prinn, R. G.: Inversion of long-lived trace gas emissions using combined Eulerian and Lagrangian chemical transport models, *Atmospheric Chemistry and Physics*, 11, 9887–9898, doi:10.5194/acp-11-9887-2011, 2011.
- Wang, J. and Zabaras, N.: A Markov random field model of contamination source identification in porous media flow, *Int. J. Heat Mass. Tran.*, 49, 939–950, doi:10.1016/j.ijheatmasstransfer.2005.09.016, 2006.
- Zanini, A. and Kitanidis, P. K.: Geostatistical inversing for large-contrast transmissivity fields, *Stoch. Env. Res. Risk. A.*, 23, 565–577, doi:10.1007/s00477-008-0241-7, 2009.