Adaptively scaling the Metropolis algorithm using expected squared jumped distance

Cristian Pasarica and Andrew Gelman

J. P. Morgan and Department of Statistics, Columbia University, New York, NY

Supplementary material

This note illustrates the convergence of the algorithm used in Section 4 for two examples: correlated normal target distribution and extreme starting values.

S-1 Correlated normal target distribution

We illustrate adaptive scaling for target distribution with unknown covariance matrix. We follow the adaptive scheme of Haario et al. (1999, 2001) to adapt the covariance matrix, in conjunction with our optimization routine to adapt the scaling parameter parameter. We consider a two-dimensional target distribution with covariance $\Sigma = \begin{pmatrix} 100 & 9 \\ 9 & 1 \end{pmatrix}$. A choice for the covariance matrix of the initial Gaussian proposal is the inverse of the Hessian, $-\bigtriangledown^2 \log(\pi)$, computed at the maximum of π. Unfortunately, numerical optimization methods can perform very badly for high dimensions when the starting point of the algorithm is not close to the maximum. Even for such a simple distribution, starting the optimization algorithm far from the true mode might not find the global maximum, resulting in a bad initial proposal covariance matrix Σ_0. To represent this possibility, we start here with an independent proposal $\Sigma_0 = \begin{pmatrix} 25 & 0 \\ 0 & 1 \end{pmatrix}$. Figure 1 shows the performance of our algorithm; approximate convergence is achieved in 20 steps. We will add this example to the online appendix.

S-2 Extreme starting values

We have also successfully tested the robustness of our method for extreme starting values. For faster convergence we recommend using an optimization procedure (see Brent (1973)).

Figure 2 shows the convergence of the optimized scale parameter

Insert Figure 2 here “Extreme starting points”

The convergence from extreme starting points shows the same pattern of our initial runs: spurious drops for high starting value and high upward jumps for the lowest starting values, but still converge in a few steps to the optimal value.

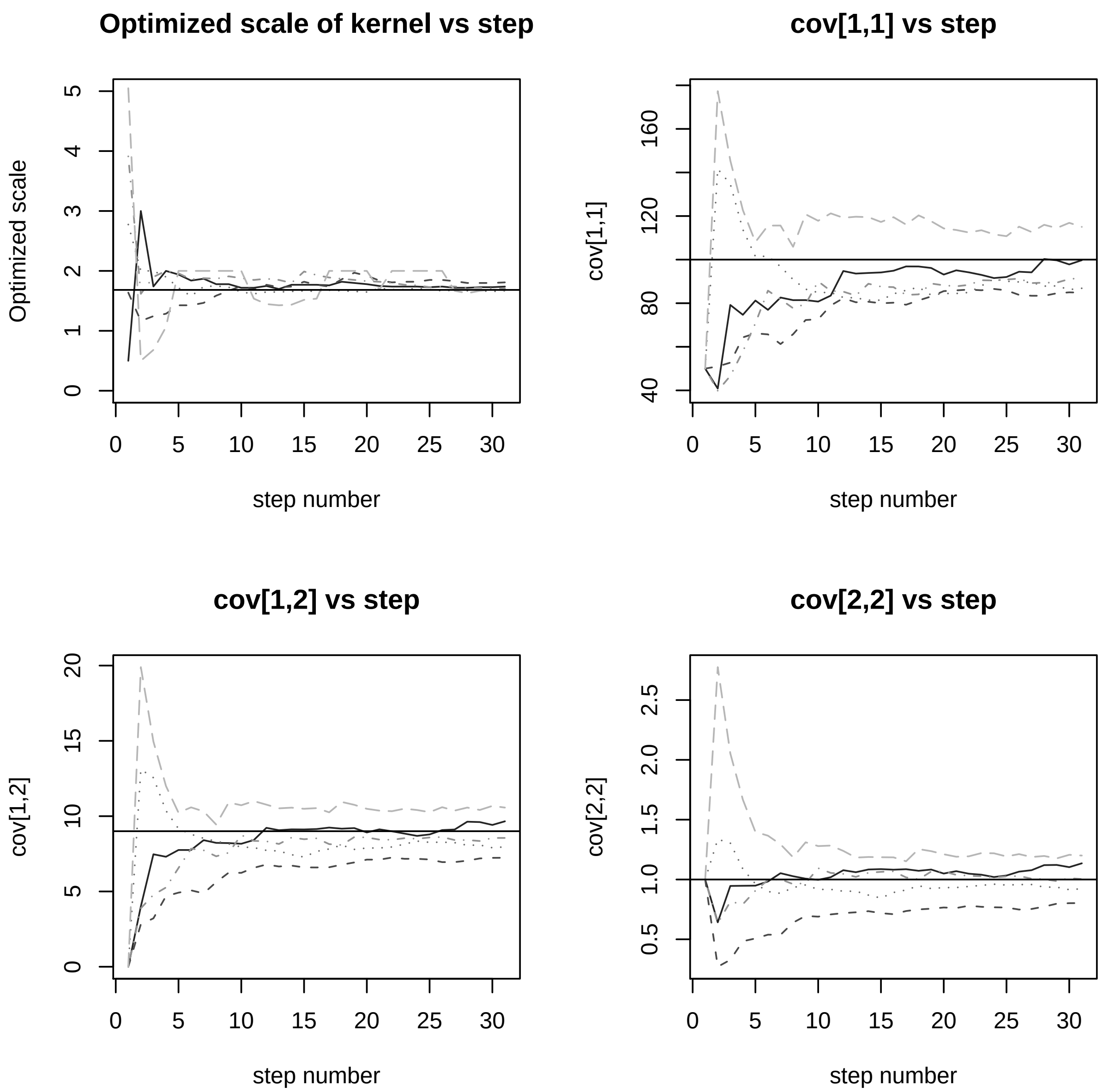

Figure 1: Convergence of the adaptive optimization procedure that maximizes the ESJD by scaling and updating the covariance matrix, starting with independent proposal density
$\Sigma_0 = \begin{pmatrix} 25 & 0 \\ 0 & 1 \end{pmatrix}$. with 50 iterations per step. Convergence of the sample covariance matrix is attained in 20 steps and convergence to optimal scaling in 30 steps.

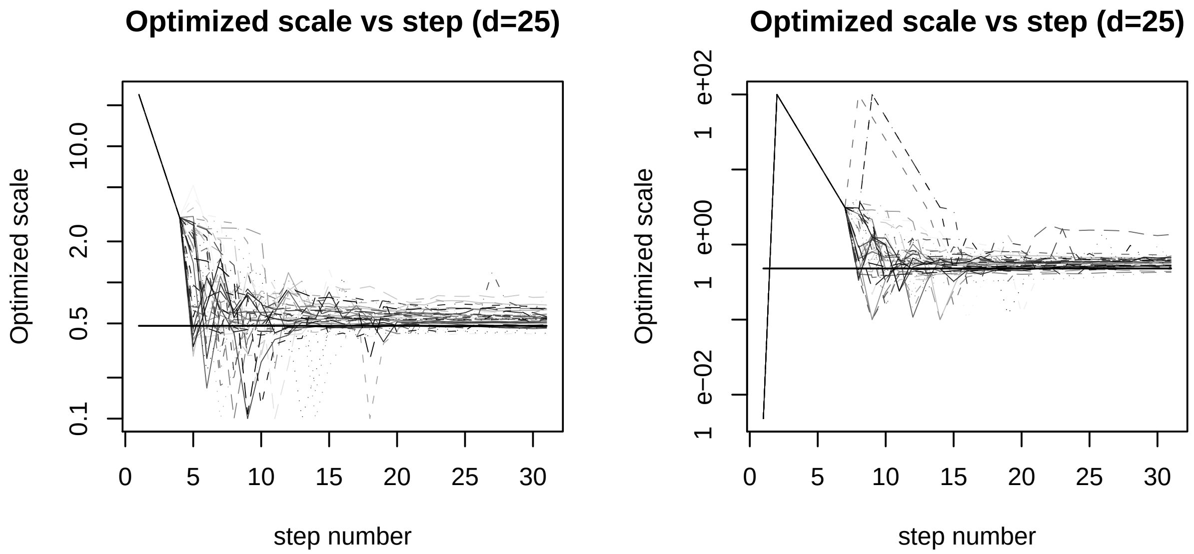

Figure 2: Convergence of the adaptive optimization procedure with extreme starting points $(0.01, 50) \times$ optimal value, for dimension $d = 25$ with multivariate normal target distribution, for 50 independent paths with 50 steps per iteration plotted on log-scale on the y-axis.