# Supplementary Material B - Random Initial Values

Andrew J. Pohl, Matthew R. Schofield & Reed Ferber

University of Calgary April 21, 2021

Both numerical optimization used for least squares (LS) methods and MCMC used for Bayesian inference require the specification of initial values, the choice of which can influence the output of each respective algorithm. Unlike the main article in which the true values of each simulation were used as initial values for numerical minimization of LS error and the resulting LS estimate used to initialize MCMC sampling; this appendix focus on the performance of each model when initial values are specified randomly. For LS optimisation this equates to selecting initial values with uniform probability from the support of each parameter. For MCMC sampling this equates to selecting initial values at random from the prior distribution for the respective Bayesian model. Arguably this approach is more reflective of real world application given that in practice, the true values of inverse kinematic problem are unknown. All research code to replicate this analysis can be found at: https://github.com/AndyPohlNZ/BayesKin.

## 1 Results

Performance of each model on identifying underling pose parameters when random values are used as initial values for MCMC sampling or numerical optimization are presented for 1, 2 and 3-link chains in Figures 1, 2, 3 respectfully and the bias, variance and RMSE for each model are presented in Table 1 Distinctive banding at non-optimal estimators occurs in all models with the exception of the third set of priors for the Bayesian model. This is reflected in increased bias, variance, and RMSE for P1, P2 and LS models.

## 2 Discussion

The banding of estimators highlighted in Figures 1 2 and 3 suggests that there exists local minima within the LS cost surface or multimodality within the Bayesian Posterior. This is supported by the partial cost surface highlighted Figure 4. We can see that the the numerical optimizer settles close to a small local minimum and that this local minima corresponds to a solution which matches marker plates located on the first two links

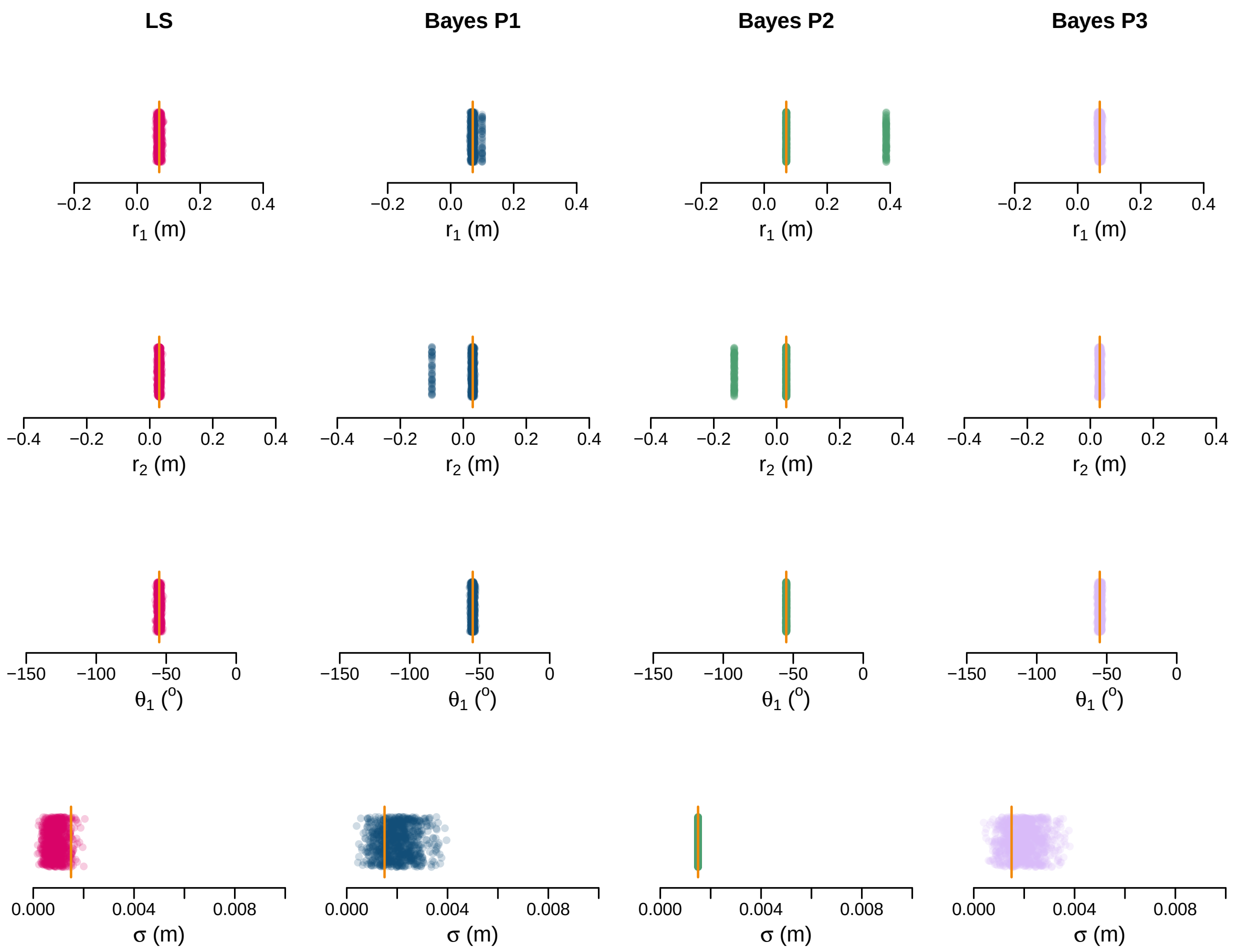



Figure 1: Performance of the estimators from each model (columns) on each parameter (rows) for 1000 single link simulations where initial values were specified using random values. True values for each parameter identified in orange.

| | | LS | | | Bayes P1 | | | Bayes P2 | | | Bayes P3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Bias | Variance | RMSE | Bias | Variance | RMSE | Bias | Variance | RMSE | Bias | Variance | RMSE |
| Single Link | $r_1$ (mm) | $-0.029$ | 18.459 | 4.297 | 1.556 | 56.497 | 7.676 | 44.100 | 12058.962 | 118.337 | $-0.260$ | 13.221 | 3.645 |
| | $r_2$ (mm) | $-0.073$ | 9.486 | 3.081 | $-6.855$ | 833.156 | 29.667 | $-22.956$ | 3266.412 | 61.591 | $-0.300$ | 6.788 | 2.623 |
| | $\theta_1$ (°) | 59.77 | 0.847 | 0.920 | 12.23 | 0.603 | 0.776 | 32.66 | 0.020 | 0.141 | 0.072 | 0.608 | 0.782 |
| | $\sigma$ (mm) | $-0.609$ | 0.092 | 0.680 | 1.152 | 8.751 | 3.175 | 0.000 | $< 0.001$ | $< 0.001$ | 0.461 | 0.375 | 0.766 |
| Double Link | $r_1$ (mm) | $-0.045$ | 6.704 | 2.590 | 5.322 | 122.894 | 12.297 | 68.903 | 21234.964 | 161.191 | 0.065 | 5.330 | 2.310 |
| | $r_2$ (mm) | 0.005 | 4.653 | 2.157 | $-22.887$ | 2234.145 | 52.516 | $-50.044$ | 8325.410 | 104.067 | 0.094 | 3.561 | 1.889 |
| | $\theta_1$ (°) | 79.20 | 0.272 | 0.522 | 25.61 | 0.202 | 0.449 | 33.84 | 0.010 | 0.102 | $-0.020$ | 0.202 | 0.451 |
| | $\theta_2$ (°) | 85.34 | 0.170 | 0.413 | 34.10 | 0.127 | 0.356 | 68.18 | 0.087 | 0.294 | 0.039 | 0.127 | 0.358 |
| | $\sigma$ (mm) | $-0.158$ | 0.073 | 0.314 | 2.721 | 27.815 | 5.935 | 0.000 | $< 0.001$ | $< 0.001$ | 0.150 | 0.104 | 0.355 |
| Triple Link | $r_1$ (mm) | 18.202 | 1862.157 | 46.834 | 8.275 | 175.028 | 15.605 | 93.421 | 27338.532 | 189.911 | 0.013 | 4.976 | 2.231 |
| | $r_2$ (mm) | 10.584 | 612.644 | 26.920 | $-28.887$ | 3165.409 | 63.244 | $-66.770$ | 13884.938 | 135.437 | 0.071 | 3.185 | 1.786 |
| | $\theta_1$ (°) | 75.99 | 0.252 | 0.502 | 32.83 | 0.183 | 0.428 | 35.22 | 0.010 | 0.098 | $-0.008$ | 0.183 | 0.429 |
| | $\theta_2$ (°) | 91.20 | 0.052 | 0.227 | 41.82 | 0.038 | 0.196 | 81.72 | 0.026 | 0.161 | 0.014 | 0.039 | 0.197 |
| | $\theta_3$ (°) | 16.83 | 0.229 | 0.479 | 5.63 | 0.166 | 0.408 | 0.99 | 0.146 | 0.383 | $-0.046$ | 0.166 | 0.408 |
| | $\sigma$ (mm) | 15.512 | 966.393 | 34.716 | 3.944 | 37.030 | 7.252 | 0.000 | $< 0.001$ | $< 0.001$ | 0.082 | 0.065 | 0.267 |

Table 1: Bias, Variance and RMSE of estimators for each model on single, double and triple link simulations when random values are used as initial conditions. The equivalent results using random values for initial conditions can be found in table 1 in the main article.

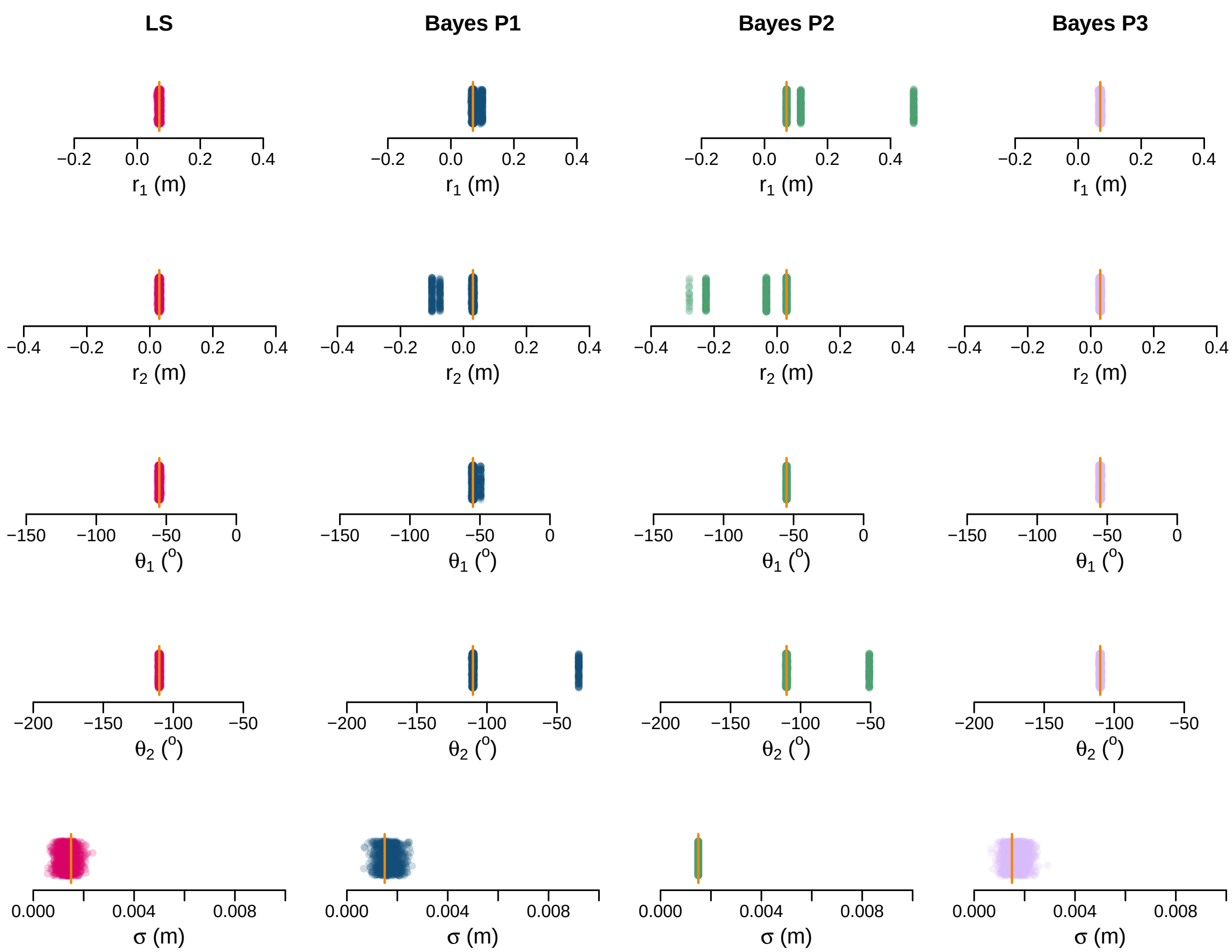



Figure 2: Performance of the estimators from each model (columns) on each parameter (rows) for 1000 double link simulations where initial values were specified using random values. True values for each parameter identified in orange.

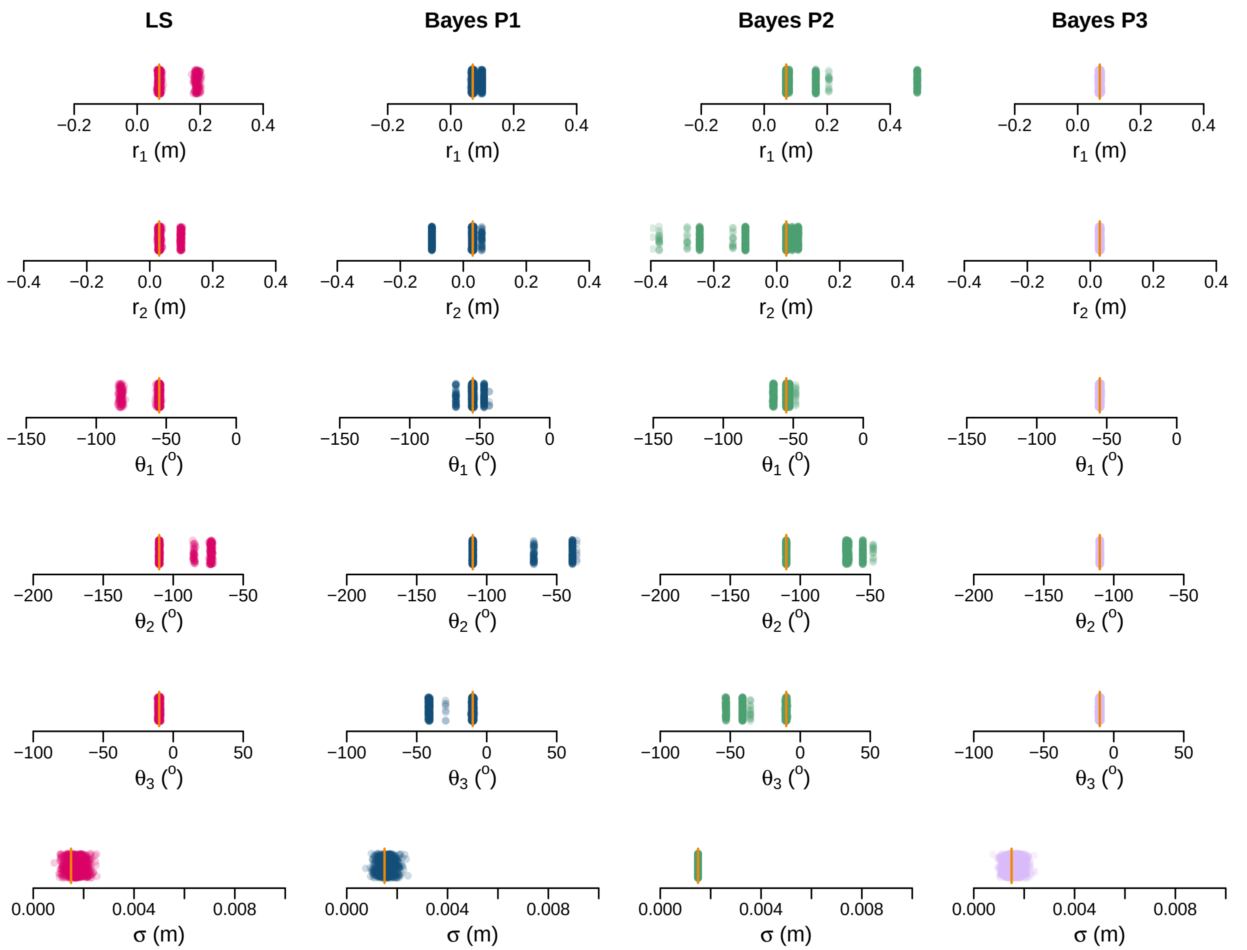
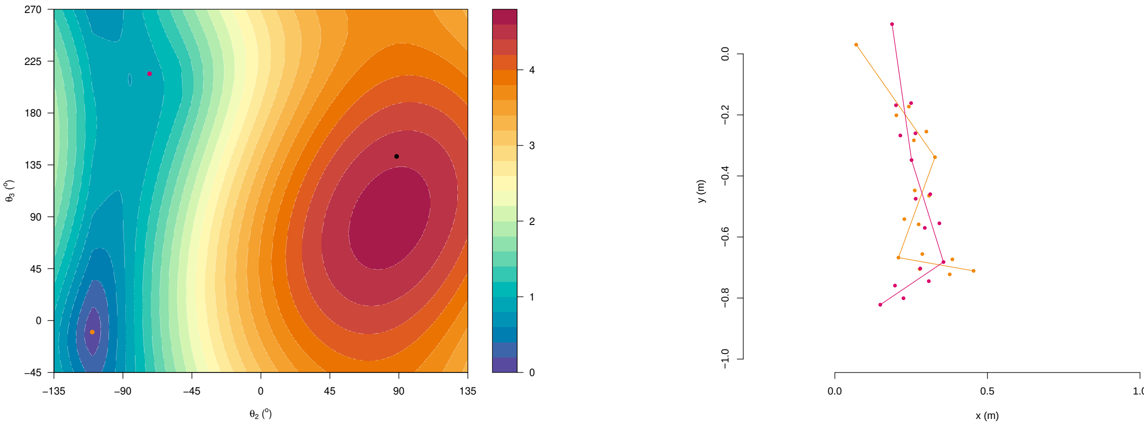



Figure 3: Performance of the estimators from each model (columns) on each parameter (rows) 1000 triple link simulations where initial values were specified using random values. True values for each parameter identified in orange.




(a) (b)

Figure 4: Partial cost surface for a typical example of poor performance of the LS model on a 3-link problem 4a. Orange denotes the true value, pink the result of optimization and black the initial value. The resulting pose estimate (pink) is compared to the true value (orange) in 4b.

despite having $156°$ of error for $\theta_3$ and $0.14\,\mathrm{m}$ of error for $\boldsymbol{r}$. Such a solution could be described as a reflection of links 1 and 2 to approximately align the first two marker plates. A similar phenomenon occurs when examining trace plots of MCMC chains for the same example (Appendix B). It is clear that poor mixing occurs between chains suggestive of a chain becoming 'stuck' at this reflective solution and suggesting multimodality within the posterior.

In general the effect of initial values is well understood for optimisation problems. Iterative methods such as the quasi-Newton BFGS method used in this work can settle on local optima or at saddle points where gradients change sign. Alternative optimization algorithms such as simulated annealing (Bertsimas & Tsitsiklis, 1993), may be effective at preventing numerical optimization from settling on these local minima.

For MCMC based sampling in theory the specification of initial values does not prevent the samples from being obtained from the appropriate posterior distribution. However, these theoretical arguments are made on the basis of asymptotic theory (Gilks et al., 1996). In practice we must terminate sampling at some finite value and as a result we hope that independent MCMC chains show adequate mixing to have confidence in the MCMC samples obtained. When random values were used to initiate MCMC chains we observe considerably higher $\hat{R}$ statistics (Brooks & Gelman, 1998) than when the true values were specified as a result $56\%$ of simulations were excluded on the basis of poor convergence (much higher than the $< 0.1\%$ excluded when true values were specified).

In Figure 5 a typical example of poor convergence is provided where independent Markov Chains become 'stuck' at different modes of the posterior distribution. These modes correspond to the true solution and the ‘reflective’ solution highlighted in 4b. In practice several solutions exist, in an analogous manner to simulated annealing for LS optimisation simulated tempering (Marinari & Parisi, 1992) provides a modification to MCMC sampling which improves sampling efficiency for multi-modal posteriors. Alternatively weakly informative priors (such as that proposed in Bayes P3) are not as effected by this phenomenon. The ‘reflective’ solution obtained in the case of random initial values pertains to a pose with extreme amount of ankle dorsiflexion or hip rotation, well beyond normal anatomical limits. Such solutions have little prior probability if prior distributions are based on previous literature which describes the normal range of anatomically plausible motion as with our third set of priors. A similar approach could be used for LS sampling in that the support of parameter values of which initial values are sampled from is limited to those based on previous literature.

# References


Bertsimas, D., & Tsitsiklis, J. (1993). Simulated Annealing. *Statistical Science*, *8*, 10–15.

Brooks, S. P., & Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, *7*, 434–455.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in practice*. (1st ed.). Springer Science + Business Media.

Marinari, E., & Parisi, G. (1992). Simulated Tempering: A New Monte Carlo Scheme. *Europhysics Letters (EPL)*, *19*, 451–458.

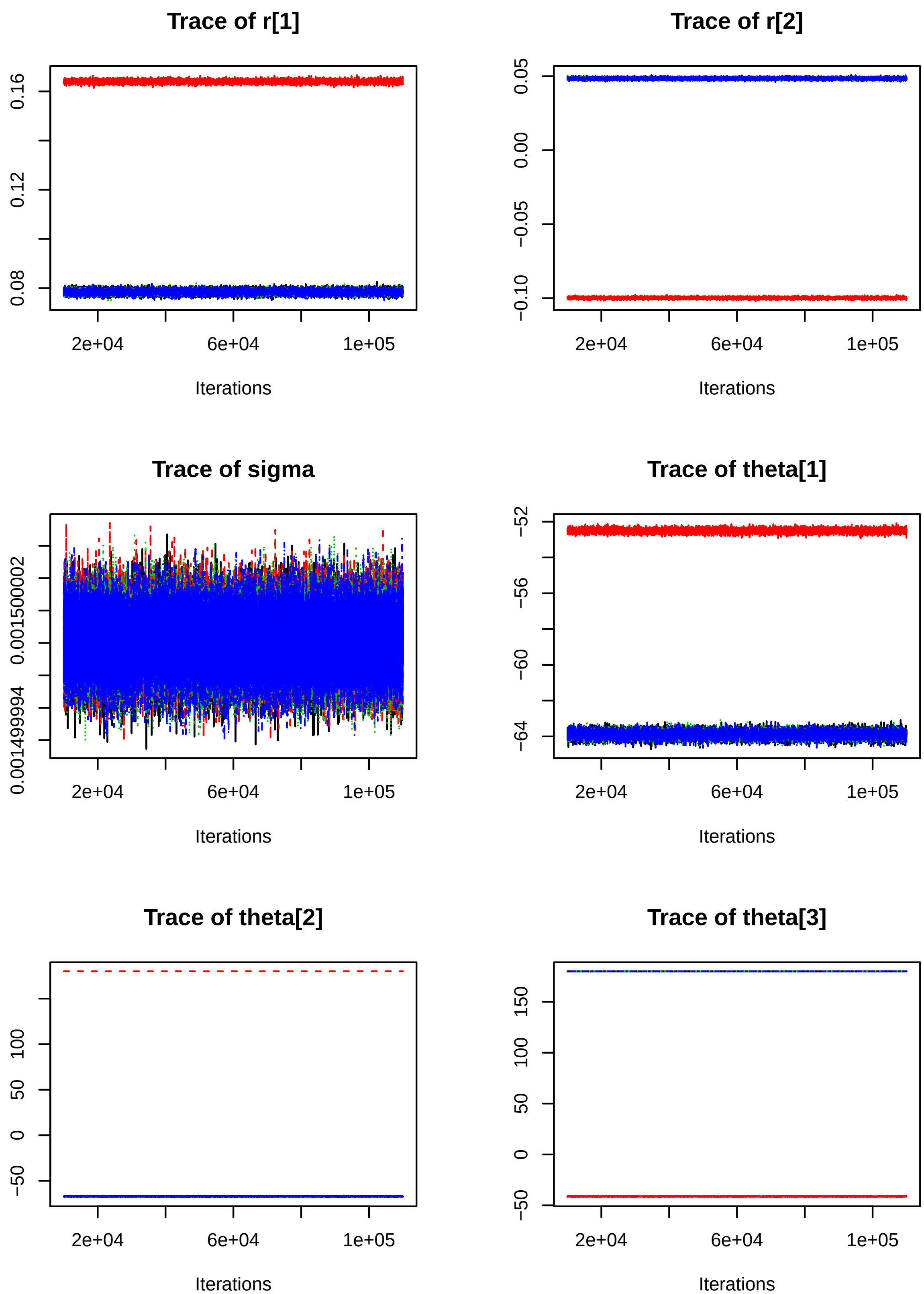



Figure 5: Traceplots demonstrating poor convergence with chains sampling from distinctively different modes within the posterior distribution.