

In this Chapter we motivate the use of the `PARALLEL TEMPERING` making use of an example widely-cited in the literature, see ? and ?.

As pointed out in the Introduction, the reason for using the `PARALLEL TEMPERING` is the multimodality of the distribution of interest,  $\pi$ . In our simulations we assume to know the density function of  $\pi$  to be the following mixture of normal densities:

$$f(x) = \sum_{i=1}^{20} \frac{\omega_i}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{(x - \mu_i)^t(x - \mu_i)}{2\sigma_i^2}\right), \quad (1)$$

where  $\sigma_1 = \dots = \sigma_{20} = 0.1$ ,  $\omega_1 = \dots = \omega_{20} = 0.05$ , and where the means  $\mu_i$  are given by

$\mu_1$	$\mu_2$	$\mu_3$	$\mu_4$	$\mu_5$	$\mu_6$	$\mu_7$	$\mu_8$	$\mu_9$	$\mu_{10}$
2.18	8.67	4.24	8.41	3.93	3.25	1.70	4.59	6.91	6.87
5.76	9.59	8.48	1.68	8.82	3.47	0.50	5.60	5.81	5.40

and

$\mu_{11}$	$\mu_{12}$	$\mu_{13}$	$\mu_{14}$	$\mu_{15}$	$\mu_{16}$	$\mu_{17}$	$\mu_{18}$	$\mu_{19}$	$\mu_{20}$
5.41	2.70	4.98	1.14	8.33	4.93	1.83	2.26	5.54	1.69
2.65	7.88	3.70	2.39	9.50	1.50	0.09	0.31	6.86	8.11

Figure 1 inspects visually the density. One notices easily that different peaks get intermingled, as only 15 peaks are visible out of 20 peaks are clearly visible. The peaks are usually separated by at least an area of unit diameter of lowered probability mass.

The `METROPOLIS-HASTINGS` is known to be a local algorithm and the reasons for it will be exposed in Chapter ?. To motivate the use of the `PARALLEL TEMPERING` it is enough to say for now that sample points generated by the usual `METROPOLIS-HASTINGS`<sup>1</sup> will be more often drawn from the vicinity of modes. This in turn might give the impression that most of the probability is concentrated somewhere next to the starting points<sup>2</sup>. We shall refer to the percentages of iterations in the vicinity of a particular mode its sojourn time, postponing a more thorough definition till Chapter ?.

By the results that stem from the Ergodic Theory, the sojourn times for each component of the mixture given by Eq. 1 should converge to the probabilities of the modes' vicinity areas,  $\omega_i$ , and so be roughly equal in the above-mentioned example. However, in practical terms, this convergence might be prohibitively long, way longer than what a typical user would consider fortunate to wait for. This is clearly a big problem, since in practical applications, as mentioned in the ??, one considers the approximation of the needed integral to be simply the mean evaluation of a function on the generated sample point and these approximations cannot possibly be good if we cut the procedures too quickly.

To investigate the `METROPOLIS-HASTINGS`'s localness, we have run the `METROPOLIS` routine. Figures 2 and 3 show that the localness of the algorithm is persistent

<sup>1</sup>For introduction to the theory behind the Metropolis-Hastings algorithm consider ?.

<sup>2</sup>We choose the starting points at random

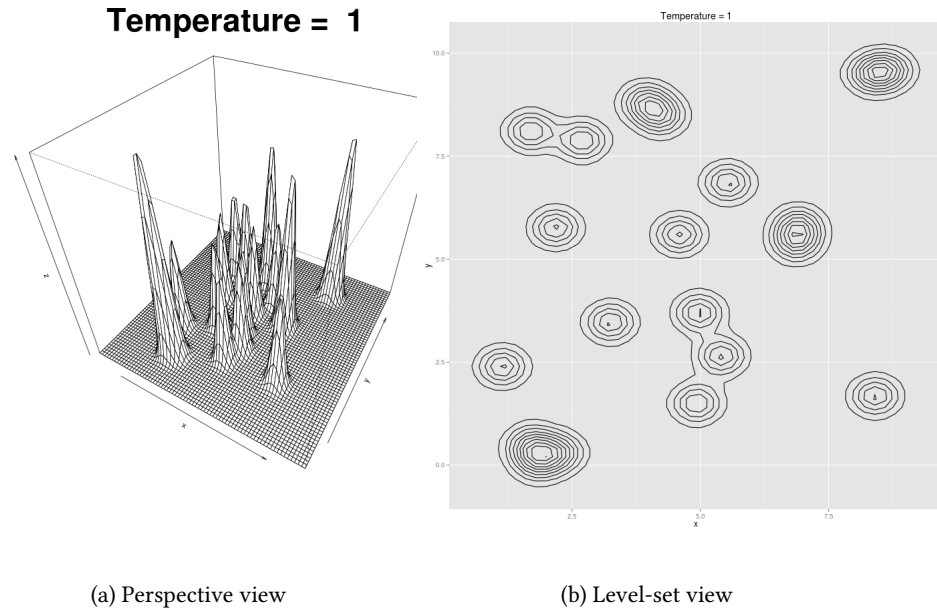
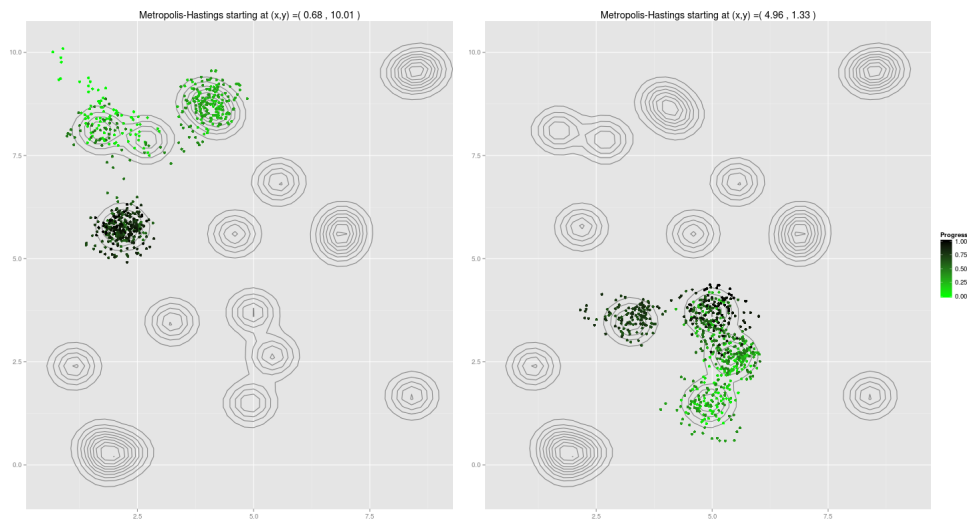


Figure 1: Toy-example density.

even with longer runs of the METROPOLIS-HASTINGS, the reasons for which will be mentioned in Chapter ???. Of course any type of inference based on results based on sample points as in Figure 3 would be totally not true.

Everything changes when one considers the PARALLEL TEMPERING instead, as seen in Figures 4 and 5. In Figure 4a one can visually ascertain himself that most of the STATE SPACE got explored. In Figure 4b one notices that only one mode was never explored. The simulation however was not a very long one. If one sets the number of iterations to be higher there are no modes that are unexplored in the studied example, as shown in Figure 5. In this picture one notices clearly that points accumulate in the regions with higher probability levels.



(a) Thousand steps...

(b) ...that lead nowhere.

Figure 2: Thousand Iterations of METROPOLIS-HASTINGS - about 250 different points generated. Colour intensity increases as points get generated in more advanced stages of the algorithm. One notices easily that short runs of the algorithm do not explore the whole STATE SPACE.

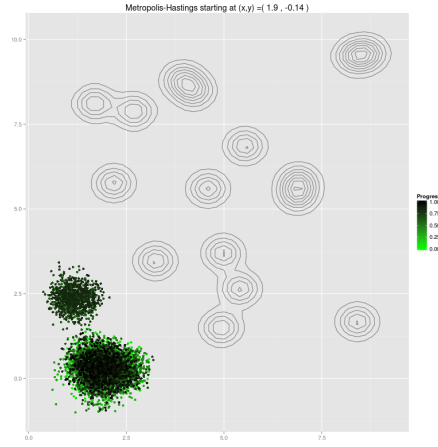
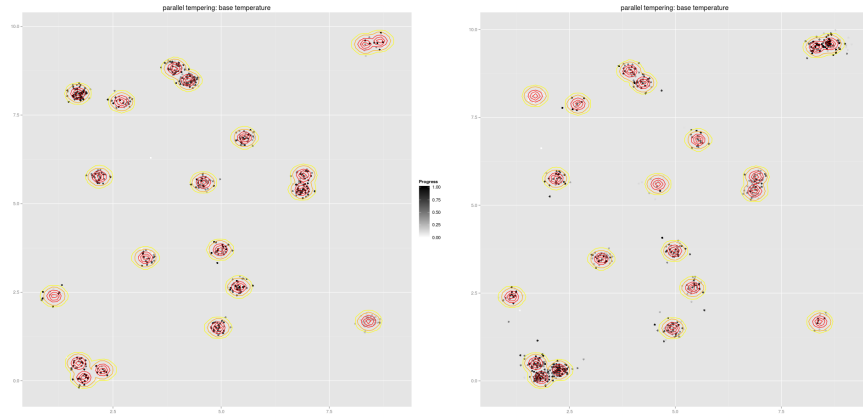


Figure 3: Ten Thousand Iterations of METROPOLIS-HASTINGS - about a quarter of which resulted in different points. Again the METROPOLIS-HASTINGS failed at exploring the whole STATE SPACE.



(a) Two thousand steps under STRATEGY 1.... (b) Two thousand steps under STRATEGY 2.

Figure 4: Results of the PARALLEL TEMPERING when and STRATEGY 2 from Chapter ?? are applied. One notices immediately the qualitative difference between these examples and the METROPOLIS-HASTINGS results, as the STATE SPACE is explored more thoroughly. The results are obtainable under a minute on a modern computer equipped with one core processor unit, 1.5 GHz.

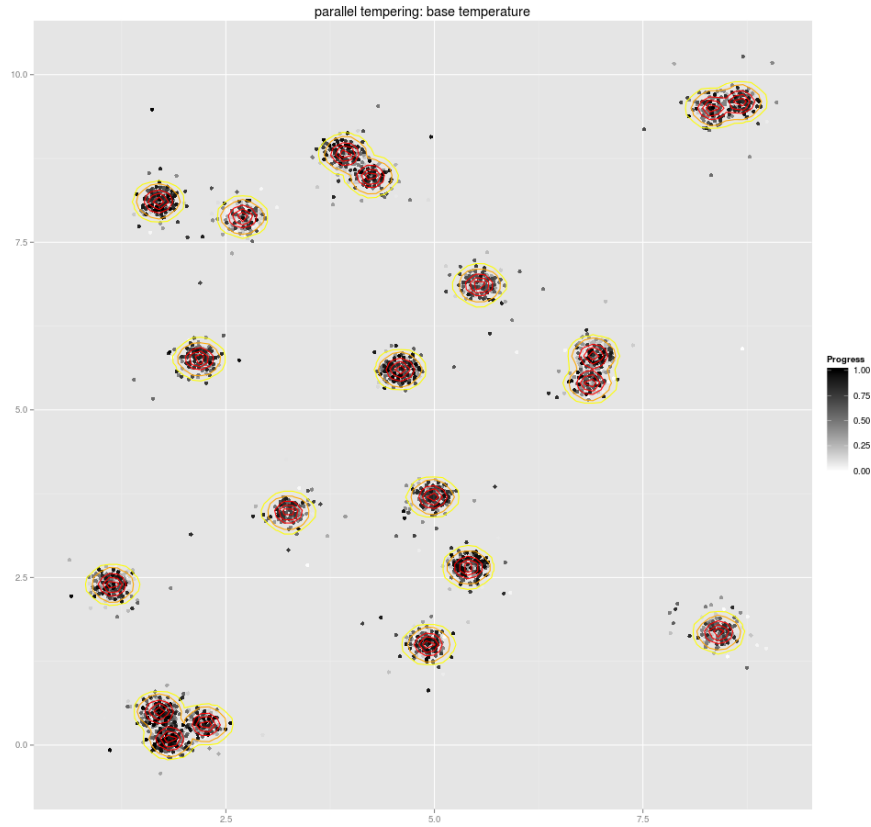


Figure 5: Ten Thousand Iterations of PARALLEL TEMPERING with 2500 iterations of burn-in, under  $??$ . About a quarter of random-walk steps were accepted. The simulation lasted about 140 seconds on a 8 years-old computer equipped with a one core computer. The isoquants encircle regions that contain one percent of overall probability mass (in yellow), five percent (orange), and twenty five, fifty and seventy five (all in red). The shading of sample points varies from white to black, getting darker as the simulation proceeds.