

PARALLEL TEMPERING*

BY BŁAŻEJ MIASOJEDOW[¶], MATEUSZ ŁĄCKI[¶] AND WOJCIECH
NIEMIRO^{¶,||}

University of Warsaw[¶] and University of Toruń^{||}

The abstract should summarize the contents of the paper. It should be clear, descriptive, self-explanatory and not longer than 200 words. It should also be suitable for publication in abstracting services. Please avoid using math formulas as much as possible.

This is a sample input file. Comparing it with the output it generates can show you how to produce a simple document of your own.

1. Introduction. Here we shall give a brief introduction to the problem.

2. New Swap Strategies. Here I shall put some info on the swap strategies, like [3, 2].

Now we address the question of how to interlace the independent chains generated with kernel $M_{\Sigma, \beta}$. We want to swap the generated states so that chains with higher temperature influence the regions of lower temperatures. Being more precise, the swapping gives some possibility to the chains with lower temperatures to find themselves in states that would otherwise be very unlikely to happen, the path leading to them being very unlikely to occur - these paths are rendered more probable for the chains with higher temperature. The swaps are made at random so that the probability of swaps takes into account some properties of π evaluated on the last points from all the chains.

Before passing to detailed description of different ways the swaps might be made, which we call Strategies, let us first describe the properties of the swap kernel \mathcal{S}_β common to all Strategies. In all cases, we rely on the following assumption

ASSUMPTION I At one step of the algorithm, there will be only one possible swap between a chosen at random pair of coordinates.

*Special thanks goes to Sir Arthur Conan Doyle for giving us constant thrill and intellectual suspense.

[†]Some comment

[‡]First supporter of the project

[§]Second supporter of the project

The restriction here is not so much in the number of swaps per step. What seems to be more relevant is rather the constraint to a particular subset of the group of all permutations — for one could certainly envisage that any random permutation could take place. However, it simplifies the interpretation of *Strategies* that are to be described.

Let us denote the swap operator by S_{ij} , so that

$$S_{ij}x = (x_1, \dots, x_{i-1}, x_j, x_{i+1}, \dots, x_{j-1}, x_i, x_{j+1}, \dots, x_L).$$

We require $\mathcal{S}_\beta(x, \circ)$ to be a measure concentrated on the set of all possible pairs of swaps between coordinates of $x = (x_1, \dots, x_L)$, i.e. on the set $\mathfrak{S}_x \equiv \{S_{ij}x : i < j\}$. We note that the proposal measure for the swap kernel \mathcal{Q} is of the following form

$$\mathcal{Q}(x, A) \equiv \sum_{i < j} p_{ij}(x) \mathbb{I}_A(S_{ij}x)$$

Then, the swap kernel would be of the following form, for any $x \in \Omega^L$ and A

$$\mathcal{S}_\beta(x, A) = \int_A \alpha_{\text{swap}}(x, y) \mathcal{Q}(x, dy) + r(x) \mathcal{I}(x, A)$$

where $r(x) = 1 - \int_{\Omega^L} \alpha_{\text{swap}}(x, y) \mathcal{Q}(x, dy)$. Plugging \mathcal{Q} permits us to write

$$\begin{aligned} \int_A \alpha_{\text{swap}}(x, y) \mathcal{Q}(x, dy) &= \sum_{i < j} p_{ij}(x) \int \alpha_{\text{swap}}(x, y) \mathbb{I}_A(y) \delta_{S_{ij}x}(dy) \\ &= \sum_{i < j} p_{ij}(x) \alpha_{\text{swap}}(x, S_{ij}x) \mathbb{I}_A(S_{ij}x), \end{aligned}$$

so that finally

$$\mathcal{S}_\beta(x, A) = \sum_{i < j} p_{ij}(x) \alpha_{\text{swap}}(x, S_{ij}x) \mathbb{I}_A(S_{ij}x) + \left(1 - \sum_{i < j} p_{ij}(x) \alpha_{\text{swap}}(x, S_{ij}x)\right) \mathcal{I}(x, A),$$

where

$$\alpha_{\text{swap}}(x, y) = \frac{\pi_\beta(dy) \mathcal{Q}(y, dx)}{\pi_\beta(dx) \mathcal{Q}(x, dy)} \wedge 1$$

is the Radon-Nikodym derivative of two measures. It's calculated with respect to measure $\mu(dx, dy) \equiv \pi(dx)Q(x, dy)$. The measure that is being differentiated is μ composed with the exchange coordinates operation, $R(x, y) = (y, x)$ ¹. So finally $\alpha_{\text{swap}} \equiv \frac{d\mu \circ R^{-1}}{d\mu}$.

Observe that thanks to the proposal's support finiteness we actually get

$$\alpha_{\text{swap}}(x, y) = \frac{\pi_\beta(y)Q(y, x)}{\pi_\beta(x)Q(x, y)} \wedge 1.$$

Now, since $y \in \mathfrak{S}_x$, the tagged π_β is a tensor of measures, $Q(x, S_{ij}x) = p_{ij}(x)$, and $Q(S_{ij}x, x) = p_{ij}(S_{ij}x)$ ², we get finally

$$\alpha_{\text{swap}}(x, S_{ij}x) = \left[\left(\frac{\pi(x_j)}{\pi(x_i)} \right)^{\beta_i - \beta_j} \frac{p_{ij}(S_{ij}x)}{p_{ij}(x)} \right] \wedge 1$$

In our computer simulations several swapping strategies have been tested. For instance,

STRATEGY I $p_{ij}(x) \propto \frac{\pi(x_j)}{\pi(x_i)} \wedge \frac{\pi(x_i)}{\pi(x_j)} = \exp \left(- \left| \log(\pi(x_j)) - \log(\pi(x_i)) \right| \right).$

In this particular the proportionality is in fact equality, because the above expression is transposition symmetric and this considerably simplifies the statistical sum analysis. In general, we want $p_{ij}(x) \propto h(x_i, x_j)$ for some function h . Thus,

$$\frac{p_{kl}(S_{kl}x)}{p_{kl}(x)} = \frac{\frac{h(x_l, x_k)}{A(\tilde{x}_{kl}) + f(x_k) + g(x_l) + h(x_l, x_k)}}{\frac{h(x_k, x_l)}{A(\tilde{x}_{kl}) + f(x_l) + g(x_k) + h(x_k, x_l)}} = \underbrace{\frac{h(x_l, x_k)}{h(x_k, x_l)}}_{\text{'direct effect'}} \times \underbrace{\frac{A(\tilde{x}_{kl}) + f(x_l) + g(x_k) + h(x_k, x_l)}{A(\tilde{x}_{kl}) + f(x_k) + g(x_l) + h(x_l, x_k)}}_{\text{'distribution effect'}}$$

where \tilde{x}_{kl} is x without its k^{th} and l^{th} coordinates, $f(x_l) = \sum_{i < l, i \neq k} h(x_i, x_l)$, and $g(x_k) = \sum_{j > k, j \neq l} h(x_k, x_j)$. As we see it is fairly straightforward to compare the above expression to 1 if the 'distribution effect' equals 1, and difficult had it been the other case. To assure this happens, one can settle for a function h that is transposition invariant, $h(x, y) = h(y, x)$. This is certainly the case of **STRATEGY I**. In this strategy therefore we are promoting exactly swaps between coordinates whose proposals have relatively the same π density values, i.e. $\pi(x_j) \approx \pi(x_i)$.

STRATEGY II $p_{ij}(x) \propto \frac{\pi(x_j)}{\pi(x_i)} \wedge 1 = \exp \left(- \left[\log(\pi(x_j)) - \log(\pi(x_i)) \right] \right) \wedge 1.$

¹Which is obviously measurable in the appropriate sense.

²Here we pass from measure notation to probability function notation, $Q(x, \{S_{ij}x\}) = Q(x, S_{ij}x)$. We can do it because Q is purely atomic given x .

This strategy breaks the symmetry of the previous one rendering the evaluation of the effect on the α_{swap} challenging, if not impossible. The main idea behind this strategy was to

The subsequent strategies are again functions of the first strategy.

$$\text{STRATEGY III } p_{ij}(x) \propto \left(\frac{\pi(x_j)}{\pi(x_i)} \wedge \frac{\pi(x_i)}{\pi(x_j)} \right)^{|\beta_i - \beta_j|} = \exp \left(- |\beta_i - \beta_j| \times |\log(\pi(x_j)) - \log(\pi(x_i))| \right).$$

This strategy permits us to soften a bit the requirement that $\pi(x_j) \approx \pi(x_i)$. This effect is strengthened for coordinates that are similarly tempered, i.e. where $\beta_i - \beta_j \approx 0$. Swaps between adjacent chains will be therefore more probable.

One can also use a strategy that incorporates some distance measure between points, ρ being any quasi-metric. **STRATEGY IV** is a good example of such procedure - the more distant the points, the less probable it is for them to get swapped.

$$\text{STRATEGY IV } p_{ij}(x) \propto \left(\frac{\pi(x_j)}{\pi(x_i)} \wedge \frac{\pi(x_i)}{\pi(x_j)} \right)^{\frac{|\beta_i - \beta_j|}{1 + \rho(x_i, x_j)}} = \exp \left(- \frac{|\beta_i - \beta_j| \times |\log(\pi(x_j)) - \log(\pi(x_i))|}{1 + \rho(x_i, x_j)} \right).$$

Finally, for purposes of comparisons with [1], we include also strategies that are independent of the underlying *State Space*, i.e. the probability of swaps does not depend on the evaluations of the density π in the sample points. Such naïve strategies include

$$\text{STRATEGY V } p_{ij} = \frac{2}{L(L-1)}$$

and

$$\text{STRATEGY VI } p_{ij} = \frac{1}{L-1} \mathbb{I}_{\{|i-j|=1\}}$$

amounting to choosing uniformly among all possible swaps and all neighbouring-in-temperature swap respectively.

References.

- [1] BARAGATTI, M., GRIMAUD, A., AND POMMERET, D. (2013). Parallel tempering with equi-energy moves. *Statistics and Computing* **23**, 3 (may), 323–339.
- [2] DE LA CROIX, D. AND MICHEL, P. (2002). *A Theory of Economic Growth: Dynamics and Policy in Overlapping Generations*. Cambridge University Press.
- [3] ROMER, D. (1997). *Advanced Macroeconomics*. McGraw-Hill/Irwin.