Notes on SGNHT and a possible relativistic extension

Leonard Hasenclever

February 27, 2016

1 Where does the thermostat come from in physics?

The Nosé -Hoover thermostat was developed to simulate systems at a constant temperature. The idea is to introduce a fictitious variable ξ that 'cools down' the systems if the temperature is higher than the target temperature and 'heats up' the system if the temperature is lower than the target temperature. It does so by acting on the generalised momenta (speed up for higher temperature, slow down for lower temperature). The equations for a deterministic Nosé-Hoover thermostat in classical thermodynamics are:

$$\dot{q}_i = \frac{p_i}{m_i} \tag{1}$$

$$\dot{p}_i = f_i - \xi p_i \tag{2}$$

$$\dot{\xi} = \frac{1}{\mu}\left(\sum_i \frac{p_i^2}{m_i} - dk_b T\right) \tag{3}$$

where $m_1, \ldots, m_d)$ are the particle masses and d is the dimension. The choice of derivative for ξ is motivated by the fact that $\langle \sum_i \frac{p_i^2}{m_i} \rangle = dk_b T$ by the equipartition theorem and $\langle \cdot \rangle$ denotes the average over phase space (or, assuming ergodicity over time). In steady state ($\dot{\xi} = 0$) the kinetic energy is given by $dk_b T$ (equipartition).

2 Is there a relativistic generalisation?

I did not find any work on a relativistic thermostat but I think we can use a similar idea for a relativistic system. The generalised equipartition theorem for the canonical ensemble states that:

$$\langle x_i \frac{\partial H}{\partial x_j} \rangle = k_B T \delta_{ij} \tag{4}$$

So in particular for a separable Hamiltonian $H(q, p) = U(q) + K(p)$, we have $\langle p^T \nabla K(p) \rangle = dk_b T$. This motivate looking at the following generalised thermostat equations:

$$\dot{q} = \nabla K(p) \tag{5}$$

$$\dot{p} = -\nabla U(q) - \xi p \tag{6}$$

$$\dot{\xi} \propto \left(p^T \nabla K(p) - dk_b T\right) \tag{7}$$

Note that these equations would make sense with our relativistic Hamiltonian. In our applications we can always choose $k_b T = 1$.

3 How does the Nosé-Hoover Thermostat work in the framework of Ma, Chen & Fox?

Ma, Chen and Fox derive a general recipe for stochastic gradient MCMC methods, given a Hamiltonian H they consider the SDE

$$dz = f(z)dt + \sqrt{2D(z)}dW_t \tag{8}$$

where $z = (q, p)$, W is a Brownian motion and

$$f(z) = -[D(z) + Q(z)]\nabla H(z) + \Gamma(z) \tag{9}$$

$$\Gamma_i(z) = \sum_{j=1}^{d} \frac{\partial}{\partial z_j}(D_{ij}(z) + Q_{ij}(z)) \tag{10}$$

and Q is a skew-symmetric matrix. This SDE can be shown to have the right stationary distribution.

In their paper they claim that SGNHT without a mass matrix corresponds to the following choices

$$H(q, p, \xi) = U(q) + \frac{1}{2}p^T p + \frac{1}{2d}(\xi - A)^2 \tag{11}$$

$$D(q, p, \xi) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & A \cdot I & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{12}$$

$$Q(q, p, \xi) = \begin{pmatrix} 0 & -I & 0 \\ I & 0 & p/d \\ 0 & -p^T/d & 0 \end{pmatrix} \tag{13}$$

I think there is a small mistake in their work.

$$\nabla H(q, p, z) = \begin{pmatrix} \nabla U \\ p \\ \frac{1}{d}(\xi - A) \end{pmatrix} \tag{14}$$

Next we can calculate Γ and hence f.

$$D + Q = \begin{pmatrix} 0 & -I & 0 \\ I & A \cdot I & p/d \\ 0 & -p^T/d & 0 \end{pmatrix} \tag{15}$$

$$\frac{\partial}{\partial z_j}(D_{ij} + Q_{ij}) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -1/d & 0 \end{pmatrix} \tag{16}$$

$$\Gamma = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} \tag{17}$$

Now we can calculate $f(z)$.

$$f(z) = -\begin{pmatrix} 0 & -I & 0 \\ I & A \cdot I & p/d \\ 0 & -p^T/d & 0 \end{pmatrix} \begin{pmatrix} \nabla U \\ p \\ \frac{1}{d}(\xi - A) \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} p \\ -\nabla U - Ap - \frac{p}{d^2}(\xi - A) \\ \frac{1}{d}p^T p - 1 \end{pmatrix} \tag{18}$$

The two A terms should cancel but in this formulation they don't. Note however that everything works out if we change the ξ term in the Hamiltonian to $\frac{d}{2}(\xi - A)^2$. In that case we obtain

$$f(z) = \begin{pmatrix} p \\ -\nabla U - p\xi \\ \frac{1}{d}p^T p - 1 \end{pmatrix} \tag{19}$$

and the full SDE is given by

$$d\begin{pmatrix} q \\ p \\ \xi \end{pmatrix} = \begin{pmatrix} p \\ -\nabla U - p\xi \\ \frac{1}{d}p^T p - 1 \end{pmatrix} dt + \begin{pmatrix} 0 \\ \sqrt{2AI} \\ 0 \end{pmatrix} dW_t \tag{20}$$

4 Can we introduce masses in the Ma/Chen/Fox framework?

If we introduce a mass matrix M, the Hamiltonian becomes

$$H(q,p,\xi) = U(q) + \frac{1}{2}p^T M^{-1} p + \frac{d}{2}(\xi - A)^2 \tag{21}$$

$$\tag{22}$$

and

$$\nabla H(q,p,z) = \begin{pmatrix} \nabla U \\ M^{-1}p \\ d(\xi - A) \end{pmatrix} \tag{23}$$

To obtain the same cancellation of the A terms we can choose:

$$D(q,p,\xi) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & A \cdot M & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{24}$$

$$Q(q,p,\xi) = \begin{pmatrix} 0 & -I & 0 \\ I & 0 & p/d \\ 0 & -p^T/d & 0 \end{pmatrix} \tag{25}$$

This leads to

$$D + Q = \begin{pmatrix} 0 & -I & 0 \\ I & A \cdot M & p/d \\ 0 & -p^T/d & 0 \end{pmatrix} \tag{26}$$

$$\frac{\partial}{\partial z_j}(D_{ij} + Q_{ij}) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -1/d & 0 \end{pmatrix} \tag{27}$$

$$\Gamma = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} \tag{28}$$

as before.

$$f(z) = -\begin{pmatrix} 0 & -I & 0 \\ I & A \cdot M & p/d \\ 0 & -p^T/d & 0 \end{pmatrix} \begin{pmatrix} \nabla U \\ M^{-1}p \\ d(\xi - A) \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} M^{-1}p \\ -\nabla U - Ap - p(\xi - A) \\ \frac{1}{d}p^T M^{-1} p - 1 \end{pmatrix} \tag{29}$$

and the A terms cancel as required. The whole dynamics become:

$$d\begin{pmatrix} q \\ p \\ \xi \end{pmatrix} = \begin{pmatrix} M^{-1}p \\ -\nabla U - p\xi \\ \frac{1}{d}p^T M^{-1} p - 1 \end{pmatrix} dt + \begin{pmatrix} 0 \\ \sqrt{2A} M^{\frac{1}{2}} \\ 0 \end{pmatrix} dW_t \tag{30}$$

5 Can we build a stochastic gradient relativistic Nosé-Hoover thermostat algorithm?

Note that for a relativistic Hamiltonian $\nabla_{p_i} K(p) = \frac{p_i}{m_i}\left(\frac{p_i^2}{m_i^2 c_i^2} + 1\right)^{-\frac{1}{2}}$ so we can write $\nabla K(p) = \tilde{M}^{-1}(p)p$. Note that $\tilde{M}^{-1}$ is a diagonal matrix that is certainly invertible. This is very similar to the classical mass matrix, except that the mass is now momentum-dependent. As mentioned above, I think the ξ equation should read

$$\dot{\xi} \propto \left(p^T \nabla K(p) - d\right) \tag{31}$$

using the generalised equipartition theorem. The relativistic Hamiltonian is:

$$H(q,p,\xi) = U(q) + \sum_i m_i^2 c_i^2 \left(\frac{p_i^2}{m_i^2 c_i^2} + 1 \right)^{\frac{1}{2}} + \frac{d}{2}(\xi - A)^2 \tag{32}$$

$$\tag{33}$$

and

$$\nabla H(q,p,z) = \begin{pmatrix} \nabla U \\ \tilde{M}^{-1}(p)p \\ d(\xi - A) \end{pmatrix} \tag{34}$$

To obtain the same cancellation of the A terms we can choose:

$$D(q,p,\xi) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & A \cdot \tilde{M}(p) & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{35}$$

$$Q(q,p,\xi) = \begin{pmatrix} 0 & -I & 0 \\ I & 0 & p/d \\ 0 & -p^T/d & 0 \end{pmatrix} \tag{36}$$

This leads to

$$D + Q = \begin{pmatrix} 0 & -I & 0 \\ I & A \cdot \tilde{M}(p) & p/d \\ 0 & -p^T/d & 0 \end{pmatrix} \tag{37}$$

$$\frac{\partial}{\partial z_j}(D_{ij} + Q_{ij}) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \nabla_p \tilde{M}(p) & 0 \\ 0 & -1/d & 0 \end{pmatrix} \tag{38}$$

Note that

$$\frac{\partial}{\partial p_j}\tilde{M}_{ij} = \delta_{ij}\frac{\partial}{\partial p_j}\left(m_i \left(\frac{p_i^2}{m_i^2 c_i^2} + 1 \right)^{\frac{1}{2}} \right) = \delta_{ij}\frac{m_i}{\left(\frac{p_i^2}{m_i^2 c_i^2} + 1 \right)^{\frac{1}{2}}} \frac{1}{2}\frac{2p_i}{m_i^2 c_i^2} = \delta_{ij}\frac{1}{\left(\frac{p_i^2}{m_i^2 c_i^2} + 1 \right)^{\frac{1}{2}}} \frac{p_i}{m_i c_i^2} = \tilde{M}_{ii}^{-1}\frac{p_i}{c_i^2} \tag{39}$$

Hence

$$\Gamma = \begin{pmatrix} 0 \\ \tilde{M}^{-1}Cp \\ -1 \end{pmatrix} \tag{40}$$

where $C_{ij} = \delta_{ij} c_i^{-2}$. The resulting dynamics are

$$d\begin{pmatrix} q \\ p \\ \xi \end{pmatrix} = \begin{pmatrix} \tilde{M}^{-1}p \\ -\nabla U - p\xi + \tilde{M}^{-1}(p)Cp \\ \frac{1}{d}p^T \tilde{M}(p)^{-1}p - 1 \end{pmatrix} dt + \begin{pmatrix} 0 \\ \sqrt{2A}\tilde{M}^{\frac{1}{2}}(p) \\ 0 \end{pmatrix} dW_t \tag{41}$$

Note the additional term $\tilde{M}^{-1}(p)Cp$. Interestingly this terms seems to speed up the system.