On Autoencoders and Score Matching for Energy Based Models Supplementary Material

Kevin Swersky* KSWERSKY@CS.UBC.CA
Marc'Aurelio Ranzato† RANZATO@CS.TORONTO.EDU
David Buchman* DAVIDBUC@CS.UBC.CA
Benjamin M. Marlin* BMARLIN@CS.UBC.CA
Nando de Freitas* NANDO@CS.UBC.CA
*Department of Computer Science, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
†Department of Computer Science, University of Toronto, Toronto, ON M5S 2G4, Canada

As a reminder, the score matching objective can be written in two equivalent forms:

$$J(\theta) = \mathbb{E}_{\widetilde{p}(v)}\left[\sum_{i=1}^{n_v}\left(\psi_i(\widetilde{p}(v)) - \psi_i(p_\theta(v))\right)^2\right] = \mathbb{E}_{\widetilde{p}(v)}\left[\sum_{i=1}^{n_v}\frac{1}{2}\left(\psi_i(p_\theta(v))\right)^2 + \frac{\partial \psi_i(p_\theta(v))}{\partial v_i}\right]. \quad (1)$$

Theorem 1 *The score matching objective, Equation (1), for a latent energy based model can be expressed succinctly in terms of either the free energy or expectations of the energy with respect to the conditional distribution $p(h|v)$. Specifically,*

$$J(\theta) = \mathbb{E}_{\widetilde{p}(v)}\left[\sum_{i=1}^{n_v}\frac{1}{2}\left(\psi_i(p_\theta(v))\right)^2 + \frac{\partial \psi_i(p_\theta(v))}{\partial v_i}\right] = \mathbb{E}_{\widetilde{p}(v)}\left[\sum_{i=1}^{n_v}\frac{1}{2}\left(\mathbb{E}_{p_\theta(h|v)}\left[\frac{\partial E_\theta(v,h)}{\partial v_i}\right]\right)^2 + var_{p_\theta(h|v)}\left[\frac{\partial E_\theta(v,h)}{\partial v_i}\right] - \mathbb{E}_{p_\theta(h|v)}\left[\frac{\partial^2 E_\theta(v,h)}{\partial v_i^2}\right]\right].$$

Proof First we calculate $\psi_i(p_\theta(v)$ and $\frac{\partial \psi_i(p_\theta(v))}{\partial v_i}$:

$$\psi_i(p_\theta(v) = \frac{\partial \log(p(v))}{\partial v_i} = \frac{\partial \log\left(\int_h -E_\theta(v,h)dh\right)}{\partial v_i} - \frac{\partial \log(Z(\theta))}{\partial v_i} = \frac{\partial \log\left(\int_h -E_\theta(v,h)dh\right)}{\partial v_i} = \frac{\int_h \frac{\partial -E_\theta(v,h)}{\partial v_i}\exp(-E_\theta(v,h))dh}{\int_{h'} -E_\theta(v,h')dh'} \equiv \int_h \frac{\partial - E_\theta(v,h)}{\partial v_i}p_\theta(h|v) \equiv \mathbb{E}_{p_\theta(h|v)}\left[\frac{\partial - E_\theta(v,h)}{\partial v_i}\right], \quad (2)$$

$$\frac{\partial \psi_i(p_\theta(v))}{\partial v_i} = \int_h \frac{\partial^2 - E_\theta(v,h)}{\partial v_i^2}p_\theta(h|v)dh + \int_h \frac{\partial - E_\theta(v,h)}{\partial v_i}\frac{\partial p_\theta(h|v)}{\partial v_i}dh. \quad (3)$$

Making use of the fact that

$$\frac{\partial p_\theta(h|v)}{\partial v_i} = p_\theta(h|v)\frac{\partial \log(p_\theta(h|v))}{\partial v_i}: \quad (4)$$

$$\frac{\partial \log(p_\theta(h|v))}{\partial v_i} = \frac{\partial}{\partial v_i}\log\left(\frac{\exp(-E_\theta(v,h))}{\int_{h'}\exp(-E_\theta(v,h'))dh'}\right) = \frac{\partial - E_\theta(v,h)}{\partial v_i} - \frac{\partial \log\left(\int_{h'}\exp(-E_\theta(v,h'))dh'\right)}{\partial v_i} = \frac{\partial - E_\theta(v,h)}{\partial v_i} - \mathbb{E}_{p_\theta(h'|v)}\left[\frac{\partial - E_\theta(v,h')}{\partial v_i}\right]. \quad (5)$$

Where (5) comes from the definition given by (2). Using (4) with (5) and putting them together with (3):

$$\begin{aligned}\frac{\partial \psi_i(p_\theta(v))}{\partial v_i} =& \int_h \left(\frac{\partial - E_\theta(v,h)}{\partial v_i}\right)^2 p_\theta(h|v)dh \\ & - \left(\int_h \frac{\partial - E_\theta(v,h)}{\partial v_i} p_\theta(h|v)dh\right)^2 \\ & + \mathbb{E}_{p_\theta(h|v)}\left[\frac{\partial^2 - E_\theta(v,h)}{\partial v_i^2}\right]\end{aligned}$$

$$\begin{aligned}=&\mathbb{E}_{p_\theta(h|v)}\left[\frac{\partial E_\theta(v,h)}{\partial v_i}^2\right] - \mathbb{E}_{p_\theta(h|v)}\left[\frac{\partial E_\theta(v,h)}{\partial v_i}\right]^2 \\ &-\mathbb{E}_{p_\theta(h|v)}\left[\frac{\partial^2 E_\theta(v,h)}{\partial v_i^2}\right] \\ =& var_{p_\theta(h|v)}\left[\frac{\partial E_\theta(v,h)}{\partial v_i}\right] - \mathbb{E}_{p_\theta(h|v)}\left[\frac{\partial^2 E_\theta(v,h)}{\partial v_i^2}\right].\end{aligned}$$

We can therefore express the full objective (1) as:

$$\begin{aligned}J(\theta) =& \mathbb{E}_{\widetilde{p}(v)}\left[\sum_{i=1}^{n_v} \frac{1}{2}\left(\mathbb{E}_{p_\theta(h|v)}\left[\frac{\partial E_\theta(v,h)}{\partial v_i}\right]\right)^2\right. \\ & \left. + var_{p_\theta(h|v)}\left[\frac{\partial E_\theta(v,h)}{\partial v_i}\right] - \mathbb{E}_{p_\theta(h|v)}\left[\frac{\partial^2 E_\theta(v,h)}{\partial v_i^2}\right]\right].\end{aligned}$$

□

Corollary 1 *If the energy function of a latent EBM $E_\theta(v,h)$ takes the following form:*

$$E_\theta(v,h) = \frac{1}{2}(v-\mu(h))^T \Omega(h)(v-\mu(h)) + g(h),$$

where $\mu(h)$ is an arbitrary vector-valued function of length n_v, $g(h)$ is an arbitrary scalar function, and $\Omega(h)$ is an $n_v \times n_v$ positive-definite matrix-valued function, then the vector-valued score function $\psi(p_\theta(v))$ will take the following form: $\mathbb{E}_{p_\theta(h|v)}[\Omega(h)(v-\mu(h))]$. As a result, the score matching objective can be expressed as:

$$\begin{aligned}J(\theta) =& \mathbb{E}_{\widetilde{p}(v)}\left[\sum_{i=1}^{n_v} \frac{1}{2}\left(\mathbb{E}_{p_\theta(h|v)}[\Omega(h)(v-\mu(h))]_i\right)^2\right. \\ & + var_{p_\theta(h|v)}[\Omega(h)(v-\mu(h))]_i \\ & \left. - \mathbb{E}_{p_\theta(h|v)}[\Omega(h)]_{ii}\right].\end{aligned}$$

Proof To prove this we simply need to derive $\psi_i(p_\theta(v))$ and $\frac{\partial \psi_i(p_\theta(v))}{\partial v_i}$, then invoke Theorem 1:

$$\begin{aligned}\psi_i(p_\theta(v)) =& \mathbb{E}_{p_\theta(h|v)}\left[\frac{\partial E_\theta(v,h)}{\partial v_i}\right], \\ =& \mathbb{E}_{p_\theta(h|v)}[\Omega(h)(v-\mu(h))]_i \\ \frac{\partial \psi_i(p_\theta(v))}{\partial v_i} =& var_{p_\theta(h|v)}\left[\frac{\partial E_\theta(v,h)}{\partial v_i}\right] \\ & - \mathbb{E}_{p_\theta(h|v)}\left[\frac{\partial^2 E_\theta(v,h)}{\partial v_i^2}\right] \\ =& var_{p_\theta(h|v)}[\Omega(h)(v-\mu(h))]_i \\ & - \mathbb{E}_{p_\theta(h|v)}[\Omega(h)]_{ii}. \end{aligned} \tag{6}$$

□