Lemma 1 (**Botlzmann Equilibrium**). *The Boltzmann equilibrium distribution $p(x) \propto \exp(-2\phi(x)/\sigma^2)$ minimizes the free energy,*

$$F(p) = -\int p(x)\phi(x)dx - \frac{\sigma^2}{2}\int p(x)\log p(x)dx. \tag{88}$$

□

Lemma 3 (**Score Matching**). *Let p be the pdf of the rv X and q a pdf, such that*

$$\lim_{u\to-\infty} E\Big[\frac{\partial \log q(X)}{\partial x_i} \mid X_i = u\Big] = \lim_{u\to\infty} E\Big[\frac{\partial \log q(X)}{\partial x_i} \mid X_i = u\Big] = 0 \tag{93}$$

□

for $i = 1\cdots n$. Then

$$\begin{aligned} \frac{1}{2}E[\|\nabla_x \log p(X) - \nabla_x \log q(X)\|^2] &= \frac{1}{2}E[\|\nabla_x \log q(X)\|^2] \\ &+ E[\nabla_x^2 \cdot \log q(X)] + \frac{1}{2}E[\|\nabla_x \log p(X)\|^2]. \end{aligned} \tag{94}$$

Corollary 1.

$$\begin{aligned} \frac{1}{2}E[\|\nabla_x \log p^\theta_\infty(X^\theta_t) - \nabla_x \log p^\theta_t(X^\theta_t)\|^2] &= \frac{1}{2}E[\|\nabla_x \log p^\theta_t(X^\theta_t)\|^2] \\ &+ \frac{1}{2}E[\|\nabla_x \frac{2}{\sigma^2}\phi(X^\theta_t,\theta)\|^2] + \frac{2}{\sigma^2}E[\nabla_x^2 \cdot \phi(X^\theta_t,\theta)]. \end{aligned} \tag{101}$$

Corollary 2 (**Probability Velocity**).

$$\begin{aligned} \frac{1}{2}E[\|V^\theta_t(X^\theta_t)\|^2 &= \frac{1}{2}E[\|\frac{\sigma^2}{2}\nabla_x \log p^\theta_t(X^\theta_t)\|^2] \\ &+ \frac{1}{2}E[\|\nabla_x \phi(X^\theta_t,\theta)\|^2] + \frac{\sigma^2}{2}E[\nabla_x^2 \cdot \phi(X^\theta_t,\theta)]. \end{aligned} \tag{102}$$

where

$$V^\theta_t(x) \stackrel{def}{=} \nabla_x\Big(\phi(x,\theta) - \frac{\sigma^2}{2}\log p^\theta_t(x)\Big) \tag{103}$$

is the probability velocity at state x, and time t, on a network with parameter θ.

Theorem 1 (**Velocity of the Average Potential).**

$$\frac{dE[\phi(X_t^\theta,\theta)]}{dt} = \frac{1}{2}\Big(E[\|V_t^\theta(X_t^\theta)\|^2] + E[\|\nabla_x\phi(X_t^\theta,\theta)\|^2] - E[\|\nabla_x\frac{\sigma^2}{2}\log p_t^\theta(X_t^\theta)\|^2]\Big).$$

Theorem 2 (**Velocity of the Entropy).** *Let $\phi, {p_t}^\theta$ such that*

$$\lim_{t\to-\infty}\phi(x,\theta)\frac{\partial}{\partial x_i}p_t^\theta(x) = \lim_{t\to\infty}\phi(x,\theta)\frac{\partial}{\partial x_i}p_t^\theta(x) = 0$$

for $i = 1,\cdots,n$. Then

$$\frac{dH(X_t^\theta)}{dt} = E[\nabla_x^2\cdot\phi(X_t^\theta,\theta)] - \frac{\sigma^2}{2}E[\nabla_x^2\cdot\log p_t^\theta(X_t^\theta)],$$

where $H({X_t}^\theta)$ is the entropy of ${X_t}^\theta$. □

Theorem 3 (**Velocity of the Free Energy).**

$$\frac{dF_t^\theta(X_t^\theta)}{dt} = -E[\|V_t^\theta(X_t^\theta)\|^2],$$

where F_t^θ is the free energy of X_t^θ.

Theorem 4 (**Exact Contrastive Divergence).** *Exact Contrastive Divergence is a Minimum Velocity Algorithm.*

Theorem 5 (**Standard Contrastive Divergence).** *As $t \to 0$ standard CD converges to exact CD.*

Given a fixed network parameter θ'

$$\nabla_\theta\left.\frac{dE[\phi(X_t^{\theta'},\theta)]}{dt}\right|_{t=0,\theta=\theta'} = \nabla_\theta\frac{1}{2}E[\|V_0^\theta(\xi)\|^2]$$

□