1 Gradient Drift Case

$$\mathrm{d}\boldsymbol{x} = -\nabla\psi(\boldsymbol{x})\mathrm{d}t + \sqrt{2\beta^{-1}}dw \tag{1}$$

$$\frac{\partial \rho}{\partial t} = \nabla \cdot (\rho \nabla \psi) + \beta^{-1}\Delta\rho \tag{2}$$

We claim that

$$\begin{aligned}
F(\rho) &= \mathbb{E}_\rho[\psi + \beta^{-1}\log\rho] \\
&= \int \psi(\boldsymbol{x})\rho(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + \beta^{-1}\int \rho(\boldsymbol{x})\log\rho(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\
&= \beta^{-1}\int \rho\log\rho - \rho\log(\exp(-\beta\psi))\,\mathrm{d}\boldsymbol{x} \\
&= \beta^{-1}\int \rho\log\left(\frac{\rho}{\exp(-\beta\psi)}\right)\,\mathrm{d}\boldsymbol{x} \\
&= \beta^{-1}\int \rho\log\left(\frac{\rho}{\frac{1}{Z}\exp(-\beta\psi)}\right)\,\mathrm{d}\boldsymbol{x} + C
\end{aligned}$$

is a Lyapunov functional along the trajectories of (**??**) and $Z = \int \exp(-\beta\psi)\mathrm{d}\boldsymbol{x}$ is a normalization constant. If you have

$$\dot{\boldsymbol{x}} = \boldsymbol{\phi}(\boldsymbol{x}) \tag{3}$$

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \langle \nabla V, \dot{x}\rangle = \langle \nabla V, \boldsymbol{\phi}(\boldsymbol{x})\rangle \leq 0 \tag{4}$$

$$\begin{aligned}
\frac{\mathrm{d}F}{\mathrm{d}t} &= \langle \frac{\delta F}{\delta\rho}, \frac{\partial\rho}{\partial t}\rangle \\
&= \langle \frac{\delta F}{\delta\rho}, \nabla\cdot(\rho\nabla\psi) + \beta^{-1}\Delta\rho\rangle \\
&= \int \frac{\delta F}{\delta\rho}(\nabla\cdot(\rho\nabla\psi) + \beta^{-1}\Delta\rho)\,\mathrm{d}\boldsymbol{x} \\
&= \int (\psi(\boldsymbol{x}) + \beta^{-1}(\log\rho(\boldsymbol{x}) + 1))(\nabla\cdot(\rho\nabla\psi) + \beta^{-1}\Delta\rho)\,\mathrm{d}\boldsymbol{x} \\
&= \int (\psi(\boldsymbol{x}) + \beta^{-1}(\log\rho(\boldsymbol{x}) + 1))(\nabla\cdot(\rho(\nabla\psi + \beta^{-1}\nabla\log\rho)))\,\mathrm{d}\boldsymbol{x} \\
&= -\int \langle \nabla(\psi(\boldsymbol{x}) + \beta^{-1}\log\rho(\boldsymbol{x})), \nabla\psi(\boldsymbol{x}) + \beta^{-1}\nabla\log\rho\rangle\rho\,\mathrm{d}\boldsymbol{x} \\
&= -\int \|\nabla(\psi(\boldsymbol{x}) + \beta^{-1}\log\rho)\|^2\rho(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\
&= -\mathbb{E}_\rho[\|\nabla\zeta\|^2] \leq 0
\end{aligned} \tag{5}$$

where $\zeta = \psi + \log \rho$.
We consider the system of 2nd-order SDEs given by

$$m_i\ddot{\theta}_i + \gamma_i\dot{\theta}_i = P_i - \sum_{j=1}^{n} K_{ij}\sin(\theta_i - \theta_j) + \sigma_i \times \text{ stochastic forcing } i = 1, \ldots, n.$$

We can transform the system into

$$\begin{aligned} \mathrm{d}\boldsymbol{\theta} &= \boldsymbol{\omega}\,\mathrm{d}t \\ \boldsymbol{M}\mathrm{d}\boldsymbol{\omega} &= (-\Gamma\boldsymbol{\omega} - \nabla_{\boldsymbol{\theta}}V(\boldsymbol{\theta}))\mathrm{d}t + \boldsymbol{\Sigma}\,\mathrm{d}\boldsymbol{w} \end{aligned} \tag{6}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n), \boldsymbol{\omega} = (\dot{\theta}_1, \dot{\theta}_2, \ldots, \dot{\theta}_n), \boldsymbol{M} = \mathrm{diag}(m_1, m_2, \ldots, m_n), \boldsymbol{\Gamma} = \mathrm{diag}(\gamma_1, \gamma_2, \ldots, \gamma_n), \boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$ and the potential function is given by

$$V(\boldsymbol{\theta}) := \sum_{i=1}^{n} P_i\theta_i + \sum_{(i,j)\in\mathcal{E}} k_{ij}(1 - \cos(\theta_i - \theta_j)) \tag{7}$$

we can rewrite as

$$\mathrm{d}\underbrace{\begin{bmatrix}\boldsymbol{\theta}\\ \boldsymbol{\omega}\end{bmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{bmatrix}\boldsymbol{\omega}\\ -\boldsymbol{M}^{-1}\boldsymbol{\Gamma}\boldsymbol{\omega} - \boldsymbol{M}^{-1}\nabla_{\boldsymbol{\theta}}V(\boldsymbol{\theta})\end{bmatrix}}_{\boldsymbol{f}}\mathrm{d}t + \underbrace{\begin{bmatrix}\mathbf{0}_{n\times n} & \mathbf{0}_{n\times n}\\ \mathbf{0}_{n\times n} & \boldsymbol{M}^{-1}\boldsymbol{\Sigma}\end{bmatrix}}_{\boldsymbol{g}}\mathrm{d}\boldsymbol{w} \tag{8}$$

1.1 Fokker Planck

$$\begin{aligned} \frac{\partial\rho}{\partial t} &= -\nabla_{\theta,\omega}\cdot\left(\rho\begin{bmatrix}\boldsymbol{\omega}\\ -\boldsymbol{M}^{-1}\boldsymbol{\Gamma}\boldsymbol{\omega} - \boldsymbol{M}^{-1}\nabla_{\boldsymbol{\theta}}V(\boldsymbol{\theta})\end{bmatrix}\right) + \frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2}{\partial\omega_i\omega_j}(\rho\boldsymbol{M}^{-1}\Sigma\Sigma^{\top}\boldsymbol{M}^{-1})_{i,j} \\ &= -\nabla_{\theta}\cdot(\rho\boldsymbol{\omega}) + \nabla_{\boldsymbol{\omega}}\cdot\left(\rho(\boldsymbol{M}^{-1}\boldsymbol{\Gamma}\boldsymbol{\omega} + \boldsymbol{M}^{-1}\nabla_{\boldsymbol{\theta}}V(\boldsymbol{\theta}))\right) + \nabla_{\boldsymbol{\omega}}\cdot(\rho D\nabla_{\boldsymbol{\omega}}\log\rho) \\ &= -\langle\boldsymbol{\omega}, \nabla_{\boldsymbol{\theta}}\rho\rangle + \nabla_{\boldsymbol{\omega}}\cdot\left(\rho\left(\boldsymbol{M}^{-1}\boldsymbol{\Gamma}\boldsymbol{\omega} + \boldsymbol{M}^{-1}\nabla_{\boldsymbol{\theta}}V(\boldsymbol{\theta}) + \boldsymbol{D}\nabla_{\boldsymbol{\omega}}\log\rho\right)\right) \end{aligned} \tag{9}$$

where $\boldsymbol{D} = \frac{1}{2}\boldsymbol{M}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\top}\boldsymbol{M}^{-1}$.
Theorem 1. *(Ito's Lemma for Multi-Dimensional Processes) Let*

$$\mathrm{d}\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t, t)\,\mathrm{d}t + \boldsymbol{g}(\boldsymbol{X}_t, t)\mathrm{d}\boldsymbol{w}_t \tag{10}$$

with $\boldsymbol{X}_t, \boldsymbol{w}_t, \boldsymbol{f}(\cdot,) \in \mathbb{R}^n$ and $\boldsymbol{g}(\cdot) \in \mathbb{R}^{n\times n}$. Then for a given transformation

$$\boldsymbol{Z}_t = \boldsymbol{\psi}(\boldsymbol{X}_t, t) = [\psi_1(\boldsymbol{X}_t, t), \ldots, \psi_n(\boldsymbol{X}_t, t)] \tag{11}$$

where $\boldsymbol{\psi}$ is a function from $\mathbb{R}^n \times [0, \infty]$ into $\mathbb{R}^n$, then $\boldsymbol{Z}_t$ is again an Ito process given by

$$\begin{aligned}
\mathrm{d}Z_{k,t} &= \frac{\partial \psi_k}{\partial t}(\boldsymbol{X}_t, t)\,\mathrm{d}t + \sum_{i=1}^{n} \frac{\partial \psi_k}{\partial x_i}(\boldsymbol{X}_t, t)\mathrm{d}X_{i,t} \\
&\quad + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 \psi_k}{\partial x_i \partial x_j}(\boldsymbol{X}_t, t)\mathrm{d}X_{j,t}\mathrm{d}X_{i,t}
\end{aligned} \tag{12}$$

Theorem 2. *Let $n = 2d$ and consider the transformation defined by*

$$\boldsymbol{Z}_t = \boldsymbol{\psi}(\boldsymbol{X}_t) = \begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\eta} \end{bmatrix} := \begin{bmatrix} \mathrm{diag}(\boldsymbol{m} \oslash \boldsymbol{\sigma}) & 0 \\ 0 & \mathrm{diag}(\boldsymbol{m} \oslash \boldsymbol{\sigma}) \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\omega} \end{bmatrix} \tag{13}$$

i.e.

$$\begin{aligned}
\boldsymbol{Z}_t &= [\psi_1(X_{1,t}), \ldots, \psi_d(X_{d,t}), \psi_{d+1}(X_{d+1,t}), \ldots, \psi_{2d}(X_{2d,t})]^\top \\
&= \left[\frac{m_1}{\sigma_1}\theta_{1,t},\ \ldots\ , \frac{m_d}{\sigma_d}\theta_{d,t}, \ldots, \frac{m_1}{\sigma_1}\omega_{1,t}, \ldots, \frac{m_d}{\sigma_d}\omega_{d,t}\right]^\top
\end{aligned} \tag{14}$$

Proof. Since the transformation (**??**) is a linear transformation and doesn't depend on t, the first and third term in (**??**) vanish so we are left with

$$\begin{aligned}
\mathrm{d}Z_{k,t} &= \sum_{i=1}^{2d} \frac{\partial \psi_k}{\partial x_i}(\boldsymbol{X}_t)\mathrm{d}X_{i,t} \\
&= \sum_{i=1}^{2d} \frac{\partial \psi_k}{\partial x_i}(X_{k,t})\mathrm{d}X_{i,t} \\
&= \frac{\partial \psi_k}{\partial x_k}\mathrm{d}X_{k,t} \\
&= \begin{cases} \frac{m_k}{\sigma_k}\mathrm{d}\theta_{k,t} & 1 \le k \le d \\ \frac{m_k}{\sigma_k}\mathrm{d}\omega_{k,t} & n+1 \le k \le 2d \end{cases} \\
&= \begin{cases} \frac{m_k}{\sigma_k}\omega_{k,t}\mathrm{d}t & 1 \le k \le d \\ \frac{m_k}{\sigma_k}\left(-\frac{\gamma_k}{m_k}\omega_{k,t} - \frac{P_k}{\sigma_k} - \sum_{j=1}^{d}\frac{K_{k,j}}{m_k}\sin(\theta_{k,t} - \theta_{j,t}) + \frac{\sigma_k}{m_k}\mathrm{d}w_{k,t}\right) & n+1 \le k \le 2d \end{cases} \\
&= \begin{cases} \eta_{k,t}\mathrm{d}t & 1 \le k \le d \\ -\frac{\gamma_k}{m_k}\eta_{k,t} - \frac{P_k}{\sigma_k} - \sum_{j=1}^{d}\frac{K_{k,j}}{\sigma_k}\sin\left(\frac{\sigma_k}{m_k}\xi_{k,t} - \frac{\sigma_j}{m_j}\xi_{j,t}\right) + \mathrm{d}w_{k,t} & n+1 \le k \le 2d \end{cases}
\end{aligned}$$

Writing in vector form, we get

$$\begin{bmatrix} \mathrm{d}\boldsymbol{\xi} \\ \mathrm{d}\boldsymbol{\eta} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\eta} \\ -\nabla_{\boldsymbol{\eta}}F(\boldsymbol{\eta}) - \nabla_{\boldsymbol{\xi}}U(\boldsymbol{\xi}) \end{bmatrix}\mathrm{d}t + \begin{bmatrix} \mathbf{0}_{d\times d} & \mathbf{0}_{d\times d} \\ \mathbf{0}_{d\times d} & \boldsymbol{I}_{d\times d} \end{bmatrix}\mathrm{d}\boldsymbol{w} \tag{15}$$

where $F = \frac{1}{2}\langle \boldsymbol{\eta}, \mathrm{diag}(\boldsymbol{\gamma} \oslash \boldsymbol{m})\boldsymbol{\eta}\rangle$ and the potential function in $\boldsymbol{\xi}$ is given by

$$U(\boldsymbol{\xi}) = \sum_{i=1}^{2d} \frac{P_i}{\sigma_i}\xi_i + \sum_{(i,j)\in\mathcal{E}} K_{i,j}\frac{m_i}{\sigma_i^2}\left(1 - \cos\left(\frac{\sigma_i}{m_i}\xi_i - \frac{\sigma_j}{m_j}\xi_j\right)\right) \tag{16}$$

$\square$

The Fokker-Planck equation of this SDE is

$$\frac{\partial \rho}{\partial t} = -\langle \boldsymbol{\eta}, \nabla_{\boldsymbol{\xi}}\rho\rangle + \nabla_{\boldsymbol{\eta}} \cdot \left(\left(\nabla_{\boldsymbol{\eta}}F(\boldsymbol{\eta}) + \nabla_{\boldsymbol{\xi}}U(\boldsymbol{\xi})\right)\rho\right) + \frac{1}{2}\Delta_{\boldsymbol{\eta}}\rho \tag{17}$$

Remark 1. *Given a random vector X, who joint pdf is given by $f_X(x)$ and if $H : \mathbb{R}^n \mapsto \mathbb{R}^n$ is a $1-1$ differentiable function, then the random vector $Y = H(X)$ also has a density function given by*

$$f_Y(y) = \frac{f_X(x)}{|\mathrm{det}\nabla H(x)|}, \quad x \in H^{-1}(y) \tag{18}$$