Problem 1. Пусть $\mathbf{X}_0 \sim \mathcal{N}(0,1)$ и $\mathbf{X}_1 \sim \mathcal{N}(\mu,1)$ — независимые величины, а векторное поле $f_t^*(\boldsymbol{x})$ является решением задачи оптимизации

$$\int_0^1 \mathbb{E}\left\|f_t(\mathbf{X}_t) - \frac{\partial}{\partial t} I_t(\mathbf{X}_0, \mathbf{X}_1)\right\|^2 \mathrm{d}t \to \min_f,$$

где $\mathbf{X}_t = I_t(\mathbf{X}_0, \mathbf{X}_1)$. Найдите такой интерполянт $I_t(\boldsymbol{x}_0, \boldsymbol{x}_1)$, что ОДУ

$$\mathrm{d}\mathbf{Y}_t = f_t^*(\mathbf{Y}_t)\mathrm{d}t \tag{1}$$

имеет прямые траектории.

Подсказка: оптимальный интерполянт имеет вид $I_t(\boldsymbol{x}_0, \boldsymbol{x}_1) = a_t\boldsymbol{x}_0 + b_t\boldsymbol{x}_1$. Найдя a_t и b_t, подумайте, почему интерполянт имеет именно такой вид.

С одной стороны, т.к. траектории ODE прямые, а $\mathbf{Y}_0 \sim \mathcal{N}(0,1)$ и $\mathbf{Y}_1 \sim \mathcal{N}(\mu,1)$, то

$$\mathbf{Y}_t = \mathbf{Y}_0 + \mu t \sim \mathcal{N}(\mathbf{Y}_t |\, \mu t, 1)$$

С другой стороны, если интерполянт имеет вид $I_t(\boldsymbol{x}_0, \boldsymbol{x}_1) = a_t\boldsymbol{x}_0 + b_t\boldsymbol{x}_1$, то

$$\mathbf{X}_t = a_t\mathbf{X}_0 + b_t\mathbf{X}_1 \sim \mathcal{N}(\mathbf{X}_t |\, b_t\mu, a_t^2 + b_t^2)$$

Т.к. $p_{\mathbf{X}_t} = p_{\mathbf{Y}_t}$, то надо приравнять полученные распределения и найти a_t и b_t:

$$\mu t = \mu b_t \implies b_t = t$$
$$a_t^2 + b_t^2 = 1 \implies a_t = \sqrt{1 - t^2}$$

Problem 2. Пусть $(\widehat{\mathbf{X}}_0, \widehat{\mathbf{X}}_1) \sim p_{01}$ – некоторое распределение парных данных, $\widehat{\mathbf{X}}_0 \sim p_0$ и $\widehat{\mathbf{X}}_1 \sim p_1$. Определим динамику, интерполирующую (почти) p_0 в (почти) p_1:

$$\mathbf{X}_t = I_t(\widehat{\mathbf{X}}_0, \widehat{\mathbf{X}}_1) + \gamma_t\varepsilon,$$

где I_t — некоторый интерполянт, γ_t — дифференцируемая функция от t, такая, что $\gamma_0, \gamma_1 \approx 0$, а $\varepsilon \sim \mathcal{N}(0, I)$ – независимый с $(\widehat{\mathbf{X}}_0, \widehat{\mathbf{X}}_1)$ шум. Таким образом, мы ввели условную динамику

$$p_{t|01}(\boldsymbol{x}_t|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1) = \mathcal{N}(\boldsymbol{x}_t|I_t(\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1), \gamma_t^2 I)$$

и соответствующую ей безусловную динамику

$$p_t(\boldsymbol{x}_t) := p_{\mathbf{X}_t}(\boldsymbol{x}_t) = \int p_{t|01}(\boldsymbol{x}_t|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1)p_{01}(\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1)\mathrm{d}\hat{\boldsymbol{x}}_0\mathrm{d}\hat{\boldsymbol{x}}_1.$$

1. **(0.5 балла)** Найдите такое векторное поле $f_t(\boldsymbol{x}|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1)$, что соответствующее ОДУ порождает условную динамику;
2. **(0.5 балла)** Покажите, что если v решает задачу оптимизации

$$\int_0^1 \mathbb{E}\|v_t(\mathbf{X}_t) - f_t(\mathbf{X}_t|\widehat{\mathbf{X}}_0, \widehat{\mathbf{X}}_1) - \frac{g^2(t)}{2}\nabla \log p_{t|01}(\mathbf{X}_t|\widehat{\mathbf{X}}_0, \widehat{\mathbf{X}}_1)\|^2 \mathrm{d}t \to \min_v, \tag{2}$$

то соответствующее СДУ

$$\mathrm{d}\mathbf{X}_t = v_t(\mathbf{X}_t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{W}_t$$

порождает динамику p_t.

Таким образом, имея задачу с парными данными из доменов p_0 и p_1, можно переводить один домен в другой с помощью СДУ, обучив модель на функционал Задачи 2 и удостоверившись, что $\gamma_0, \gamma_1 \approx 0$, что соответствует распределениям доменов с добавкой очень небольшого шума.

1. Если зафиксировать шум ε, то динамика станет очень простой. Выпишем ODE для неё:

$$\mathrm{d}\mathbf{X}_t = \left(\frac{\partial}{\partial t} I_t(\widehat{\mathbf{X}}_0, \widehat{\mathbf{X}}_1) + \frac{\partial}{\partial t}\gamma_t \varepsilon\right)\mathrm{d}t$$

Получаем, что $f_t(\boldsymbol{x}|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1, \varepsilon) = \frac{\partial}{\partial t} I_t(\widehat{\mathbf{X}}_0, \widehat{\mathbf{X}}_1) + \frac{\partial}{\partial t}\gamma_t\varepsilon$.

Для этой динамики можно выписать уравнение непрерывности. Раз хотим найти векторное поле $f_t(\boldsymbol{x}|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1)$, то попробуем выписать уравнение непрерывности для условной динамики, которое оно порождает:

$$\frac{\partial}{\partial t} p_{t|01}(\boldsymbol{x}_t|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1) = \frac{\partial}{\partial t}\int_0^1 p_{t|01,\varepsilon}(\boldsymbol{x}_t|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1, \varepsilon) p_\varepsilon(\varepsilon)\mathrm{d}\varepsilon = \int_0^1 \frac{\partial}{\partial t} p_{t|01,\varepsilon}(\boldsymbol{x}_t|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1, \varepsilon) p_\varepsilon(\varepsilon)\mathrm{d}\varepsilon =$$

$$= -\int_0^1 \mathrm{div}\left(p_{t|01,\varepsilon}(\boldsymbol{x}_t|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1, \varepsilon) f_t(\boldsymbol{x}|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1, \varepsilon)\right) p_\varepsilon(\varepsilon)\mathrm{d}\varepsilon =$$

$$= -\int_0^1 \mathrm{div}\left(p_{t|01,\varepsilon}(\boldsymbol{x}_t|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1, \varepsilon) f_t(\boldsymbol{x}|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1, \varepsilon) p_\varepsilon(\varepsilon)\right)\mathrm{d}\varepsilon =$$

$$= -\mathrm{div}\left(\int_0^1 p_{t|01,\varepsilon}(\boldsymbol{x}_t|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1, \varepsilon) f_t(\boldsymbol{x}|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1, \varepsilon) p_\varepsilon(\varepsilon)\mathrm{d}\varepsilon\right) =$$

$$= -\mathrm{div}\left(p_{t|01}(\boldsymbol{x}_t|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1)\int_0^1 \frac{p_{t|01,\varepsilon}(\boldsymbol{x}_t|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1, \varepsilon)}{p_{t|01,\varepsilon}(\boldsymbol{x}_t|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1)} f_t(\boldsymbol{x}|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1, \varepsilon) p_\varepsilon(\varepsilon)\mathrm{d}\varepsilon\right) =$$

$$= \{p_\varepsilon = p_{\varepsilon|01}\} = -\mathrm{div}\left(p_{t|01}(\boldsymbol{x}_t|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1)\int_0^1 \frac{p_{t|01,\varepsilon}(\boldsymbol{x}_t|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1, \varepsilon)}{p_{t|01,\varepsilon}(\boldsymbol{x}_t|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1)} f_t(\boldsymbol{x}|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1, \varepsilon) p_{\varepsilon|01}(\varepsilon|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1))\mathrm{d}\varepsilon\right) =$$

$$= \left\{\frac{p_{t|01,\varepsilon}(\boldsymbol{x}_t|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1, \varepsilon)}{p_{t|01,\varepsilon}(\boldsymbol{x}_t|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1)} p_{\varepsilon|01}(\varepsilon|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1)) = p_{\varepsilon|01,t}(\varepsilon|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1, \boldsymbol{x}_t)\right\} =$$

$$= -\mathrm{div}\left(p_{t|01}(\boldsymbol{x}_t|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1)\int_0^1 p_{\varepsilon|01,t}(\varepsilon|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1, \boldsymbol{x}_t) f_t(\boldsymbol{x}|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1, \varepsilon)\mathrm{d}\varepsilon\right)$$

Т.е. получаем, что

$$f_t(\boldsymbol{x}|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1) = \int_0^1 p_{\varepsilon|01,t}(\varepsilon|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1, \boldsymbol{x}_t) f_t(\boldsymbol{x}|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1, \varepsilon)\mathrm{d}\varepsilon =$$

$$= \{\text{при фиксированных } x_t, x_0, x_1\ \varepsilon \text{ востанавливается однозначно}\} =$$

$$= f_t(\boldsymbol{x}|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1, \varepsilon'),$$

где

$$\varepsilon' = \frac{\boldsymbol{x}_t - I_t(\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1)}{\gamma_t}.$$

Итого получаем

$$f_t(\boldsymbol{x}|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1) = \frac{\partial}{\partial t} I_t(\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1) + \frac{\boldsymbol{x}_t - I_t(\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1)}{\gamma_t}\frac{\partial}{\partial t}\gamma_t.$$

2. Если учить $v_t(\mathbf{X}_t)$ на данный лосс, то он выучит

$$v_t(\boldsymbol{x}) = \mathbb{E}\left[f_t(\mathbf{X}_t|\widehat{\mathbf{X}}_0, \widehat{\mathbf{X}}_1) + \frac{g^2(t)}{2}\nabla\log p_{t|01}(\mathbf{X}_t|\widehat{\mathbf{X}}_0, \widehat{\mathbf{X}}_1)\Big|\mathbf{X}_t = \boldsymbol{x}\right] =$$

$$= \mathbb{E}\left[f_t(\mathbf{X}_t|\widehat{\mathbf{X}}_0, \widehat{\mathbf{X}}_1)\Big|\mathbf{X}_t = \boldsymbol{x}\right] + \frac{g^2(t)}{2}\nabla\log p_t(\mathbf{X}_t)$$

Теперь попробуем найти векторное поле, которое порождается p_t:

$$\frac{\partial}{\partial t} p_t(\boldsymbol{x}) = \frac{\partial}{\partial t}\int_0^1 p_{t|01}(\boldsymbol{x}|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1) p_{01}(\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1)\mathrm{d}\hat{\boldsymbol{x}}_0\mathrm{d}\hat{\boldsymbol{x}}_1 =$$

$$= -\int_0^1 \mathrm{div}\left(p_{t|01}(\boldsymbol{x}|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1) f_t(\boldsymbol{x}|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1)\right) p_{01}(\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1)\mathrm{d}\hat{\boldsymbol{x}}_0\mathrm{d}\hat{\boldsymbol{x}}_1 =$$

$$= -\text{div}\left(\int_0^1 p_{t|01}(\boldsymbol{x}|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1) f_t(\boldsymbol{x}|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1) p_{01}(\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1) \mathrm{d}\hat{\boldsymbol{x}}_0 \mathrm{d}\hat{\boldsymbol{x}}_1\right) =$$

$$= -\text{div}\left(p_t(\boldsymbol{x}) \int_0^1 p_{01|t}(\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1|\boldsymbol{x}) f_t(\boldsymbol{x}|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1) \mathrm{d}\hat{\boldsymbol{x}}_0 \mathrm{d}\hat{\boldsymbol{x}}_1\right) = -\text{div}\left(p_t(\boldsymbol{x}) \mathbb{E}\left[f_t(\mathbf{X}_t|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1 \middle| \mathbf{X}_t = \boldsymbol{x}\right]\right) =$$

$$= -\text{div}\left(p_t(\boldsymbol{x}) \left(v_t(\boldsymbol{x}) - \frac{g^2(t)}{2} \nabla \log p_t(\mathbf{X}_t)\right)\right) = -\text{div}\left(p_t(\boldsymbol{x}) v_t(\boldsymbol{x})\right) + \frac{g^2(t)}{2} \Delta p_t(\mathbf{X}_t)$$

Таким образом $p_t(\boldsymbol{x})$ порождает SDE

$$\mathbf{X}_t = v_t(\mathbf{X}_t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{W}_t$$

Problem 3. Пусть $\mathbf{X}_0 \sim p_0$, $\mathbf{X}_1 \sim p_1$, а $I_t(\boldsymbol{x}_0, \boldsymbol{x}_1)$ — интерполянт. Запишем функционал обучения модели Flow Matching:

$$\int_0^1 \mathbb{E}\left\|f_t(\mathbf{X}_t) - \frac{\partial}{\partial t} I_t(\mathbf{X}_0, \mathbf{X}_1)\right\|^2 \mathrm{d}t \to \min_f,$$

где $\mathbf{X}_t = I_t(\mathbf{X}_0, \mathbf{X}_1)$. Раскрыв скобки и избавившись от слагаемого с $\|\partial_t I_t(\mathbf{X}_0, \mathbf{X}_1)\|^2$, не зависящего от f, можно получить эквивалентный функционал

$$\mathcal{L}(f) = \int_0^1 \mathbb{E}\left(\|f_t(\mathbf{X}_t)\|^2 - 2\left\langle f_t(\mathbf{X}_t), \frac{\partial}{\partial t} I_t(\mathbf{X}_0, \mathbf{X}_1)\right\rangle\right) \mathrm{d}t \to \min_f.$$

Пусть $f_t^*(\boldsymbol{x}) = \arg\min_f \mathcal{L}(f)$ — оптимальное векторное поле в данной задаче. Покажите, что

$$\mathcal{L}(f^*) = -\int_0^1 \mathbb{E}\|f_t^*(\mathbf{X}_t)\|^2 \mathrm{d}t.$$

Таким образом, в точке оптимума функционал $\mathcal{L}(f^*)$ с точностью до знака совпадает с функционалом кинетической энергии из задачи динамического оптимального транспорта.

Замечание. Здесь можно считать, что интеграл везде переставляется с матожиданием.

Подставим f_t^* в эквивалентный функционал:

$$\mathcal{L}(f^*) = \int_0^1 \mathbb{E}\left[\|f_t^*(\mathbf{X}_t)\|^2 - 2\left\langle f_t^*(\mathbf{X}_t), \frac{\partial}{\partial t} I_t(\mathbf{X}_0, \mathbf{X}_1)\right\rangle\right] \mathrm{d}t = \int_0^1 \mathbb{E}\left[\|f_t^*(\mathbf{X}_t)\|^2\right] - 2 \cdot \mathbb{E}\left[\left\langle f_t^*(\mathbf{X}_t), \frac{\partial}{\partial t} I_t(\mathbf{X}_0, \mathbf{X}_1)\right\rangle\right] \mathrm{d}t =$$

Отдельно рассмотрим второе слагаемое:

$$\mathbb{E}\left[\left\langle f_t^*(\mathbf{X}_t), \frac{\partial}{\partial t} I_t(\mathbf{X}_0, \mathbf{X}_1)\right\rangle\right] = \{\text{по формуле полного матожидания}\} = \mathbb{E}\left[\mathbb{E}\left[\left\langle f_t^*(\mathbf{X}_t), \frac{\partial}{\partial t} I_t(\mathbf{X}_0, \mathbf{X}_1)\right\rangle \middle| \mathbf{X}_t\right]\right] =$$

$$= \{\mathbb{E}[f(\mathbf{X})\mathbf{Y}|\mathbf{X}] = f(\mathbf{X})\mathbb{E}[\mathbf{Y}|\mathbf{X}]\} = \mathbb{E}\left[\left\langle f_t^*(\mathbf{X}_t), \mathbb{E}\left[\frac{\partial}{\partial t} I_t(\mathbf{X}_0, \mathbf{X}_1) \middle| \mathbf{X}_t\right]\right\rangle\right] =$$

$$= \left\{\mathbb{E}\left[\frac{\partial}{\partial t} I_t(\mathbf{X}_0, \mathbf{X}_1) \middle| \mathbf{X}_t\right] = f_t^*(\mathbf{X}_t)\right\} = \mathbb{E}\left\langle f_t^*(\mathbf{X}_t), f_t^*(\mathbf{X}_t)\right\rangle = \mathbb{E}\|f_t^*(\mathbf{X}_t)\|^2$$

Возращаемся к функционалу:

$$= \int_0^1 \mathbb{E}\|f_t^*(\mathbf{X}_t)\|^2 - 2 \cdot \mathbb{E}\|f_t^*(\mathbf{X}_t)\|^2 \mathrm{d}t = -\int_0^1 \mathbb{E}\|f_t^*(\mathbf{X}_t)\|^2 \mathrm{d}t.$$