Gradient Flows — Exercises

Exercise 1 (φ-divergences and Wasserstein gradient flow)**.** Let ν be a reference probability measure on $\mathbb{R}^d$ and let $\varphi : \mathbb{R}_+ \to \mathbb{R}$ be a convex $\mathcal{C}^1$ function with $\varphi(1) = 0$. For a probability measure α such that $\alpha \ll \nu$ with density $\rho := \frac{d\alpha}{d\nu}$, the φ-divergence of α with respect to ν is defined by

$$D_\varphi(\alpha \mid \nu) := \int_{\mathbb{R}^d} \varphi\left(\frac{d\alpha}{d\nu}(x)\right) d\nu(x) = \int_{\mathbb{R}^d} \varphi(\rho(x))\, d\nu(x).$$

(1) Show that $D_\varphi(\alpha \mid \nu) \geq 0$.

(2) Consider the functional $f(\alpha) := D_\varphi(\alpha \mid \nu)$ on $\mathsf{W}_2(\mathbb{R}^d)$. Assume that both ν and α admit smooth strictly positive densities with respect to Lebesgue measure.

- **(a)** Compute the first variation $\frac{\delta f}{\delta \alpha}(\alpha)(x)$ and deduce the Wasserstein gradient $\nabla_{\mathsf{W}} f(\alpha)(x)$.
- **(b)** Write the corresponding Wasserstein gradient flow PDE.
- **(c)** Specialize your formulas to the case $\varphi(s) = s \log s$.

Exercise 2 (Gradient flows of Dirichlet energy and Fisher information)**.** We denote α_t be a family of probability measures on $\mathbb{R}^d$ with smooth positive density $\alpha_t(dx) = \rho_t(x)\, dx$.

(1) Consider the Dirichlet energy $\mathcal{E}(\alpha) := \int_{\mathbb{R}^d} |\nabla \rho(x)|^2\, dx, \qquad \rho = \frac{d\alpha}{dx}$.

- **(a)** Compute first variation $\delta f(\alpha)$ and write down the "$L^2(dx)$" gradient flow PDE, which is $\partial_t \rho_t = \delta f(\alpha_t)$ ie. the steepest descent for $\mathcal{E}$ with respect to the $L^2(dx)$ metric.
- **(b)** Compute the W_2-gradient of $\mathcal{E}$ and write the associated Wasserstein gradient flow PDE.

(2) Consider now the Fisher information $\mathcal{I}(\alpha) := \int_{\mathbb{R}^d} |\nabla \log \rho(x)|^2\, dx, \qquad \rho = \frac{d\alpha}{dx}$. Same questions.

Exercise 3 (Gaussian evolution under linear dynamics)**.** Let $(X_t)_{t\in\mathbb{N}}$ be a sequence of random variables in $\mathbb{R}^d$ defined by the linear recursion

$$X_{t+1} = X_t - \tau A_t X_t, \qquad t = 0, 1, 2, \ldots,$$

where $\tau > 0$ is a fixed step size and, for each t, $A_t \in \mathbb{R}^{d\times d}$ is a deterministic matrix. Denote by α_t the law of X_t.

(1) Assume that the initial condition is Gaussian $X_0 \sim \mathcal{N}(0, \Sigma_0)$, for some symmetric positive semidefinite covariance matrix $\Sigma_0 \in \mathbb{R}^{d\times d}$. Show that each X_t is also Gaussian, and describe the evolution of the law $\alpha_t = \mathcal{L}(X_t)$ by giving an explicit recursion for the covariance matrices Σ_t.

(2) Consider now the continuous time ODE $\dot{X}_t = -A_t X_t, \qquad t \geq 0$, with $A_t \in \mathbb{R}^{d\times d}$ a (say, continuous) time dependent matrix and random initial condition $X_0 \sim \mathcal{N}(0, \Sigma_0)$. Explain why X_t is Gaussian for every $t \geq 0$, say $X_t \sim \mathcal{N}(0, \Sigma_t)$, and, by considering the limit $\tau \to 0$ in the discrete time recursion from part **(1)**, show that $(\Sigma_t)_{t\geq 0}$ solves a matrix ODE that you should identify.

Exercise 4 (Wasserstein flow for quadratic kernels)**.** Let $\alpha \in \mathcal{P}_2(\mathbb{R}^d)$ and denote its mean and covariance by

$$m(\alpha) := \int_{\mathbb{R}^d} x\, d\alpha(x), \qquad \Sigma(\alpha) := \int_{\mathbb{R}^d} (x - m(\alpha))(x - m(\alpha))^\top d\alpha(x).$$

Consider the functionals

$$f(\alpha) := \iint_{\mathbb{R}^d\times\mathbb{R}^d} k(x, y)\, d\alpha(x)\, d\alpha(y)$$

on the Wasserstein space $\mathsf{W}_2(\mathbb{R}^d)$, where the kernels k are given below.

(1) Let $k(x,y) = \langle Ax, y\rangle, \qquad A \in \mathbb{R}^{d\times d}$.

- **(a)** Express $f(\alpha)$ in terms of $m(\alpha)$ and A.
- **(b)** Compute the Wasserstein gradient $\nabla_{\mathsf{W}} f(\alpha)$.
- **(c)** Write the Wasserstein gradient flow PDE, and, for a general initial condition $\alpha_{t=0} = \alpha_0$, give an explicit expression of α_t in terms of α_0.

(2) Let $k(x,y) = \langle Ax, y\rangle \langle Bx, y\rangle, \qquad A, B \in \mathbb{R}^{d\times d}$.

- **(a)** Compute the Wasserstein gradient $\nabla_{\mathsf{W}} f(\alpha)$.
- **(b)** Using the previous exercise on linear ODEs with Gaussian initial data, show that if $\alpha_{t=0} = \mathcal{N}(0, \Sigma_0)$, then $\alpha_t = \mathcal{N}(0, \Sigma_t)$ for all $t \geq 0$, and derive the matrix ODE satisfied by $(\Sigma_t)_{t\geq 0}$.

Optional – More difficult exercises

Exercise 5 (Mean-field two-layer linear network under Wasserstein flow)**.** Let ρ be a probability measure on $\mathbb{R}^d$ (the data distribution) and $h : \mathbb{R}^d \to \mathbb{R}$ a target function. For a parameter distribution

$$\alpha \in \mathcal{P}_2(\mathbb{R} \times \mathbb{R}^d), \qquad x = (u, v) \in \mathbb{R} \times \mathbb{R}^d,$$

consider the two-layer linear mean-field model

$$g_\alpha(z) := \int u \, \langle z, v \rangle \, d\alpha(u, v), \qquad z \in \mathbb{R}^d,$$

and the quadratic loss

$$f(\alpha) := \int_{\mathbb{R}^d} \left| g_\alpha(z) - h(z) \right|^2 d\rho(z).$$

(1) Show that f can be written as a quadratic functional of α of the form

$$f(\alpha) = \iint k(x, y) \, d\alpha(x) \, d\alpha(y) + \int b(x) \, d\alpha(x) + c,$$

where $x = (u, v)$, $y = (u', v')$, and identify explicit formulas for $k(x, y)$ and $b(x)$ in terms of ρ and h. You may leave the constant c implicit.

(2) Consider the Wasserstein gradient flow of f on $\mathsf{W}_2(\mathbb{R} \times \mathbb{R}^d)$:

$$\partial_t \alpha_t + \nabla \cdot \big(\alpha_t v_t \big) = 0, \qquad v_t(x) = -\nabla_{\mathsf{W}} f(\alpha_t)(x).$$

Using the structure from part **(1)** and the previous exercise on quadratic kernels, show that if

$$\alpha_{t=0} = \mathcal{N}(0, \Sigma_0),$$

then α_t remains Gaussian for all $t \geq 0$.

Solution. **(1) Quadratic representation of the loss.**
Write $x = (u, v)$ and $y = (u', v')$, with $u, u' \in \mathbb{R}$ and $v, v' \in \mathbb{R}^d$. The model output can be written as

$$g_\alpha(z) = \int u \, \langle z, v \rangle \, d\alpha(u, v) = \int \phi_x(z) \, d\alpha(x), \qquad \phi_x(z) := u \, \langle z, v \rangle.$$

Then

$$f(\alpha) = \int \big(g_\alpha(z) - h(z) \big)^2 \, d\rho(z) = \int g_\alpha(z)^2 \, d\rho(z) - 2 \int g_\alpha(z) h(z) \, d\rho(z) + \int h(z)^2 \, d\rho(z).$$

The last term does not depend on α, so it can be absorbed into a constant c.
For the quadratic term,

$$\begin{aligned} \int g_\alpha(z)^2 \, d\rho(z) &= \int \left(\int \phi_x(z) \, d\alpha(x) \right)^2 d\rho(z) \\ &= \int \left(\int \phi_x(z) \, d\alpha(x) \right) \left(\int \phi_y(z) \, d\alpha(y) \right) d\rho(z) \\ &= \iint \left(\int \phi_x(z) \, \phi_y(z) \, d\rho(z) \right) d\alpha(x) \, d\alpha(y). \end{aligned}$$

This suggests the kernel

$$k(x, y) := \int \phi_x(z) \, \phi_y(z) \, d\rho(z) = \int u \, u' \, \langle z, v \rangle \, \langle z, v' \rangle \, d\rho(z).$$

It is convenient to express this in terms of the covariance of ρ:

$$K := \int z z^\top \, d\rho(z) \in \mathbb{R}^{d \times d}.$$

Then

$$\int \langle z, v\rangle \, \langle z, v'\rangle \, d\rho(z) = v^\top \left(\int z z^\top \, d\rho(z) \right) v' = v^\top K v'.$$

Hence

$$k(x, y) = u \, u' \, v^\top K v'.$$

For the linear term,

$$\begin{aligned} -2 \int g_\alpha(z) h(z) \, d\rho(z) &= -2 \int \left(\int \phi_x(z) \, d\alpha(x) \right) h(z) \, d\rho(z) \\ &= -2 \int \left(\int \phi_x(z) h(z) \, d\rho(z) \right) d\alpha(x), \end{aligned}$$

so we can write

$$b(x) := -2 \int \phi_x(z) h(z) \, d\rho(z) = -2u \int \langle z, v\rangle \, h(z) \, d\rho(z).$$

Define

$$c_h := \int z \, h(z) \, d\rho(z) \in \mathbb{R}^d,$$

so that

$$\int \langle z, v\rangle \, h(z) \, d\rho(z) = \langle v, c_h\rangle.$$

Then

$$b(x) = -2u \, \langle v, c_h\rangle.$$

Putting these pieces together,

$$f(\alpha) = \iint k(x, y) \, d\alpha(x) \, d\alpha(y) + \int b(x) \, d\alpha(x) + c,$$

with

$$k(x, y) = u \, u' \, v^\top K v', \qquad b(x) = -2u \, \langle v, c_h\rangle,$$

and $c = \int h(z)^2 \, d\rho(z)$.

(2) Gaussian preservation under Wasserstein flow.
We now consider the Wasserstein gradient flow of f:

$$\partial_t \alpha_t + \nabla \cdot \big(\alpha_t v_t \big) = 0, \qquad v_t(x) = -\nabla_{\mathsf{W}} f(\alpha_t)(x).$$

First variation and gradient structure. From the representation

$$f(\alpha) = \iint k(x, y) \, d\alpha(x) \, d\alpha(y) + \int b(x) \, d\alpha(x) + c,$$

with a symmetric kernel k, the first variation can be written as

$$\frac{\delta f}{\delta \alpha}(\alpha)(x) = 2 \int k(x, y) \, d\alpha(y) + b(x) + \text{const}.$$

Therefore the Wasserstein gradient is

$$\nabla_{\mathsf{W}} f(\alpha)(x) = \nabla_x \left(2 \int k(x, y) \, d\alpha(y) + b(x) \right).$$

Using the explicit formulas for k and b, we see that for fixed α the map

$$x \mapsto 2 \int k(x, y) \, d\alpha(y) + b(x)$$

is a polynomial of degree at most two in the coordinates of $x = (u, v)$. Indeed:

- $k(x, y)$ is bilinear in x and y, so integrating in y leaves a function that is quadratic in x.
- $b(x)$ is bilinear in (u, v), so it is also quadratic in x.

As a consequence, its gradient with respect to x is *linear* in x:

$$\nabla_{\mathsf{W}} f(\alpha)(x) = M(\alpha)\, x + \ell(\alpha),$$

for some matrix $M(\alpha)$ and vector $\ell(\alpha)$ that depend on α only through its low order moments (essentially mean and covariance).

Hence the velocity field in the gradient flow has the affine form

$$v_t(x) = -\nabla_{\mathsf{W}} f(\alpha_t)(x) = -M_t\, x - \ell_t,$$

where we write $M_t := M(\alpha_t)$ and $\ell_t := \ell(\alpha_t)$.

Reduction to linear ODEs. A Wasserstein gradient flow with velocity field of the form

$$v_t(x) = -M_t\, x - \ell_t$$

corresponds to a nonlinear (in law) dynamics for a random parameter $X_t = (U_t, V_t)$:

$$\dot{X}_t = -M_t X_t - \ell_t,$$

in the sense that $\alpha_t = \mathcal{L}(X_t)$ is a weak solution of the continuity equation. Here M_t and ℓ_t are deterministic functions of the current law α_t, but for each fixed t the dynamics in x are linear and homogeneous plus a translation.

Consider an initial condition

$$\alpha_0 = \mathcal{N}(0, \Sigma_0),$$

so that X_0 is Gaussian with mean zero and covariance Σ_0. The previous exercise established that linear ODEs of the form

$$\dot{X}_t = -A_t X_t \quad \text{or more generally} \quad \dot{X}_t = -A_t X_t - b_t,$$

with deterministic A_t and b_t, preserve Gaussianity of the law: if X_0 is Gaussian, then X_t is Gaussian for all t, and the mean and covariance solve closed ODEs.

In our setting, at each time t the right-hand side is linear in X_t plus a constant vector:

$$\dot{X}_t = -M_t X_t - \ell_t.$$

Conditionally on the path $\big(M_s, \ell_s\big)_{s \le t}$, this is a linear deterministic ODE in X_t. Therefore, starting from a Gaussian X_0, the solution X_t remains Gaussian for every $t \ge 0$.

Equivalently, the law $\alpha_t = \mathcal{L}(X_t)$ stays in the family of Gaussian measures:

$$\alpha_t = \mathcal{N}(m_t, \Sigma_t),$$

with (m_t, Σ_t) following a closed system of ODEs derived by applying the linear ODE covariance computation from the previous exercise. In particular, if $m_0 = 0$, then m_t stays zero and only the covariance evolves.

Conclusion. Since the Wasserstein gradient of f is affine linear in the parameter x, the associated Wasserstein gradient flow acts on parameters through a linear ODE with time dependent coefficients. By the result on linear dynamics with Gaussian initial data, this implies that if

$$\alpha_{t=0} = \mathcal{N}(0, \Sigma_0),$$

then α_t is Gaussian for every $t \ge 0$.

Exercise 6 (Wasserstein flow of W_2^2)**.** Fix $\beta \in \mathcal{P}_2(\mathbb{R}^d)$. For $\alpha \in \mathcal{P}_2(\mathbb{R}^d)$ define $f(\alpha) := W_2(\alpha, \beta)^2$. Assume α has a smooth strictly positive density and consider the quadratic cost $c(x, y) = \frac{1}{2}|x - y|^2$.

(1) Using the Kantorovich dual problem, and using the enveloppe theorem (that was used the same way in the previous exercise sheet for discrete measure), compute the first variation $\delta f(\alpha)$.

(2) Assume α is absolutely continuous. Leveraging Brenier's theorem connecting the transport map T to the dual potential, compute the Wasserstein gradient $\nabla_W f(\alpha)$.

(3) Relate the Wasserstein flow $(\alpha_t)_{t \ge 0}$ to the McCann interpolation (you should use Benamou-Brenier dynamical formulation which details a valid field advecting the interpolant).

Solution. Throughout we work with the quadratic cost $c(x,y) = \frac{1}{2}|x-y|^2$ so that $F(\alpha) := \frac{1}{2}W_2(\alpha,\beta)^2$ has a clean Kantorovich dual representation.

(1) Dual formulation and envelope theorem.
The Kantorovich dual problem for $F(\alpha) = \frac{1}{2}W_2(\alpha,\beta)^2$ reads

$$F(\alpha) = \sup_{\varphi,\psi}\left\{\int \varphi\, d\alpha + \int \psi\, d\beta :\ \varphi(x)+\psi(y) \le \frac{1}{2}|x-y|^2\ \forall x,y\right\}.$$

For the moment fix β and view F as a functional of α. For each admissible pair (φ,ψ), define

$$\mathcal{J}(\alpha;\varphi,\psi) := \int \varphi\, d\alpha + \int \psi\, d\beta.$$

Then

$$F(\alpha) = \sup_{(\varphi,\psi)\in\mathcal{A}} \mathcal{J}(\alpha;\varphi,\psi),$$

where $\mathcal{A}$ is the set of admissible potentials. Suppose that for the given α there exists an optimal pair $(\varphi_\alpha,\psi_\alpha) \in \mathcal{A}$ such that

$$F(\alpha) = \mathcal{J}(\alpha;\varphi_\alpha,\psi_\alpha),$$

and that this pair is unique up to an additive constant (this holds under mild conditions).

Formally, one can apply an infinite-dimensional variant of the envelope theorem: the derivative of a supremum with respect to a parameter is obtained by differentiating the objective at an optimal argument, ignoring the derivative of the optimizer itself. Concretely, let σ be a signed measure with zero total mass and consider a perturbation $\alpha_\varepsilon = \alpha + \varepsilon\sigma$ (for small ε and in a formal sense). Then

$$F(\alpha_\varepsilon) = \sup_{(\varphi,\psi)\in\mathcal{A}} \mathcal{J}(\alpha_\varepsilon;\varphi,\psi) \ge \mathcal{J}(\alpha_\varepsilon;\varphi_\alpha,\psi_\alpha) = \int \varphi_\alpha\, d\alpha_\varepsilon + \int \psi_\alpha\, d\beta.$$

On the other hand, optimality of $(\varphi_\alpha,\psi_\alpha)$ at α ensures that

$$F(\alpha) = \mathcal{J}(\alpha;\varphi_\alpha,\psi_\alpha) \ge \mathcal{J}(\alpha;\varphi,\psi) \quad \text{for all } (\varphi,\psi)\in\mathcal{A}.$$

Assuming enough regularity and differentiability so that the envelope theorem applies, the directional derivative of F at α in direction σ is obtained by differentiating $\mathcal{J}$ at the optimal pair:

$$\frac{d}{d\varepsilon}F(\alpha_\varepsilon)\Big|_{\varepsilon=0} = \frac{d}{d\varepsilon}\mathcal{J}(\alpha_\varepsilon;\varphi_\alpha,\psi_\alpha)\Big|_{\varepsilon=0} = \int \varphi_\alpha\, d\sigma.$$

By definition of the first variation, this means

$$\frac{\delta F}{\delta \alpha}(\alpha)(x) = \varphi_\alpha(x) + \text{const}.$$

Since $f(\alpha) = W_2(\alpha,\beta)^2 = 2F(\alpha)$, we obtain

$$\frac{\delta f}{\delta \alpha}(\alpha)(x) = 2\,\varphi_\alpha(x) + \text{const}.$$

The additive constant is irrelevant for the Wasserstein gradient.

(2) Link to Brenier map and identification of $\nabla_{\mathsf{W}} f$.
Assume now that α has a density with respect to Lebesgue measure. By Brenier's theorem, there exists a convex function $u_\alpha : \mathbb{R}^d \to \mathbb{R}$ such that

$$T_\alpha(x) := \nabla u_\alpha(x)$$

is the unique optimal transport map from α to β for the cost $c(x,y) = \frac{1}{2}|x-y|^2$, and

$$(T_\alpha)_\#\alpha = \beta, \qquad \frac{1}{2}W_2(\alpha,\beta)^2 = \int \frac{1}{2}|x - T_\alpha(x)|^2\, d\alpha(x).$$

For the quadratic cost, any optimal potential φ_α can be written as

$$\varphi_\alpha(x) = \frac{1}{2}|x|^2 - u_\alpha(x) + c_\alpha,$$

for some constant c_α. This comes from the c-transform representation: the c-transform of a convex function u with cost $\frac{1}{2}|x-y|^2$ is

$$\varphi(x) = \inf_y \left\{ \frac{1}{2}|x-y|^2 - \psi(y) \right\} = \frac{1}{2}|x|^2 - u(x),$$

with u convex. For an optimal pair, u coincides with the Brenier potential u_α.

Differentiating the expression for φ_α gives

$$\nabla\varphi_\alpha(x) = x - \nabla u_\alpha(x) = x - T_\alpha(x).$$

From the first variation above,

$$\frac{\delta f}{\delta \alpha}(\alpha)(x) = 2\varphi_\alpha(x) + \text{const},$$

so the Wasserstein gradient (in Otto's formal Riemannian calculus) is the spatial gradient of the first variation:

$$\nabla_{\mathsf{W}} f(\alpha)(x) = \nabla_x \left(\frac{\delta f}{\delta \alpha}(\alpha)(x) \right) = 2\,\nabla\varphi_\alpha(x) = 2\big(x - T_\alpha(x)\big).$$

This recovers, in a more systematic way, the informal formula used earlier.

(3) Wasserstein flow and McCann interpolation.

The Wasserstein gradient flow of f is defined by

$$\partial_t \alpha_t + \nabla \cdot \big(\alpha_t v_t\big) = 0, \qquad v_t(x) = -\nabla_{\mathsf{W}} f(\alpha_t)(x).$$

Using the expression from the previous item, if T_t denotes the optimal transport map from α_t to β, we have

$$v_t(x) = -2\big(x - T_t(x)\big) = 2\big(T_t(x) - x\big).$$

So the PDE for the density ρ_t of α_t is

$$\partial_t \rho_t + \nabla \cdot \big(\rho_t\, 2(T_t - \mathrm{Id})\big) = 0, \qquad (T_t)_{\#}\alpha_t = \beta.$$

Now let us connect this with the McCann interpolation between α_0 and β. Let T be the optimal transport map from α_0 to β (for the same cost). McCann's displacement interpolation defines a constant-speed geodesic $(\mu_s)_{s\in[0,1]}$ in $(\mathcal{P}_2, W_2)$ by

$$\mu_s = \big((1-s)\mathrm{Id} + sT\big)_{\#}\alpha_0.$$

One can check that

$$W_2(\mu_s, \beta) = (1-s)\, W_2(\alpha_0, \beta).$$

Formally, Wasserstein space behaves like a Riemannian manifold, and the tangent vector to the geodesic (μ_s) at time s is represented by the velocity field

$$w_s(x) = T_s(x) - x,$$

where T_s is the optimal map from μ_s to β. Along this geodesic, the gradient of $f(\cdot) = W_2(\cdot, \beta)^2$ at μ_s points in the direction of $x - T_s(x)$, that is opposite to the direction w_s. More precisely, from the computation above

$$\nabla_{\mathsf{W}} f(\mu_s)(x) = 2\big(x - T_s(x)\big) = -2\, w_s(x),$$

and the negative gradient (the direction of steepest decrease) is

$$-\nabla_{\mathsf{W}} f(\mu_s)(x) = 2\, w_s(x),$$

which is exactly the geodesic velocity scaled by a factor 2.

This is completely analogous to the finite dimensional situation on a Riemannian manifold with metric distance d: the gradient of the function $p \mapsto \frac{1}{2}d(p,q)^2$ at a point on the geodesic from p_0 to q is a multiple of the geodesic velocity vector. As a consequence, the gradient flow of $d(\cdot, q)^2$ follows the geodesic to q, but with a nonlinear time parametrization (exponential in time rather than linear).

Translating this heuristic picture to Wasserstein space:

- Fix the geodesic $(\mu_s)_{s\in[0,1]}$ from α_0 to β given by McCann's interpolation.
- Its initial velocity at $s = 0$ is $w_0(x) = T(x) - x$.
- The Wasserstein gradient at α_0 is $\nabla_{\mathsf{W}} f(\alpha_0)(x) = 2(x - T(x)) = -2w_0(x)$, so the gradient flow starting at α_0 moves in the same direction as the McCann geodesic, only faster by a factor 2 in this tangent sense.

With a bit more formal Otto calculus, one can show that the trajectory $(\alpha_t)_{t\geq 0}$ of the gradient flow can be written as

$$\alpha_t = \mu_{s(t)}, \qquad s(0) = 0,$$

where $s(t) \in [0, 1)$ solves a scalar ODE of the type

$$\dot{s}(t) = 2\big(1 - s(t)\big),$$

so that $s(t) = 1 - e^{-2t}$. In other words, the gradient flow follows the McCann interpolation, but reparametrized in time so that it reaches β only in the infinite time limit and with an exponential rate of convergence in W_2:

$$W_2(\alpha_t, \beta) = (1 - s(t))\, W_2(\alpha_0, \beta) = e^{-2t} W_2(\alpha_0, \beta).$$