Structure Preserving Discretizations of Wasserstein Gradient Flows Part II of IV: Time discretization

Daniel Matthes (TU Munich)

"A school on gradient flows"
ICTS Bengaluru 7.–11.9.2026

Remember yesterday. . . (1/2)

The $\mathbb{W}$-gradient flow of $\boldsymbol{E} : \mathcal{P}_2(\mathbb{R}^d) \to [0, +\infty]$ is

$$\partial_t \rho = \operatorname{div}\left(\rho \nabla \frac{\delta \boldsymbol{E}}{\delta \rho}\right).$$

Remember yesterday... (1/2)

The $\mathbb{W}$-gradient flow of $\boldsymbol{E} : \mathcal{P}_2(\mathbb{R}^d) \to [0, +\infty]$ is

$$\partial_t \rho = \operatorname{div}\left(\rho \nabla \frac{\delta \boldsymbol{E}}{\delta \rho}\right).$$

Theorem

Under basic hypotheses on $\boldsymbol{E}$, solutions $(\rho_t)_{t\geq 0}$ satisfy:

- *Energy monotonicity:* $\boldsymbol{E}(\rho_{t_2}) \leq \boldsymbol{E}(\rho_{t_1})$.
- *Dissipation:* $\displaystyle\int_{t_1}^{t_2} |\partial \boldsymbol{E}(\rho_t)|^2 \,\mathrm{d}t = \boldsymbol{E}(\rho_{t_1}) - \boldsymbol{E}(\rho_{t_2})$,
- *Hölder continuity:* $\displaystyle\frac{\mathbb{W}(\rho_{t_1}, \rho_{t_2})^2}{t_2 - t_1} \leq \boldsymbol{E}(\rho_{t_1}) - \boldsymbol{E}(\rho_{t_2})$.

Recall:

$$|\partial \boldsymbol{E}(\rho)|^2 = \int \left|\nabla \frac{\delta \boldsymbol{E}}{\delta \rho}\right|^2 \rho \,\mathrm{d}x.$$

Remember yesterday. . . (2/2)

Theorem

If $\boldsymbol{E}$ is λ-geodesically convex with $\lambda > 0$ then:

1. $\boldsymbol{E}$ *possesses precisely one minimizer* ρ_∞*;*
2. *the speed decays exponentially:*

$$\|\varphi_t\|_{\rho_t}^2 \leq e^{-2\lambda t}\|\varphi_0\|_{\rho_0}^2;$$

and so do $|\partial \boldsymbol{E}(\rho_t)|^2$, $\boldsymbol{E}(\rho_t) - \boldsymbol{E}(\rho_\infty)$ *and* $\mathbb{W}(\rho_t, \rho_\infty)^2$*;*

3. *the* $(\mathrm{EVI})_\lambda$ *holds*

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{W}(\rho_t, \eta)^2 + \frac{\lambda}{2}\mathbb{W}(\rho_t, \eta)^2 \leq \boldsymbol{E}(\eta) - \boldsymbol{E}(\rho_t),$$

and in particular, the flow is λ-contractive:

$$\mathbb{W}(\hat{\rho}_t, \check{\rho}_t)^2 \leq e^{-2\lambda t}\mathbb{W}(\hat{\rho}_0, \check{\rho}_0)^2$$

Where do we need to go today?

Question

Do time discretizations preserve these properties — or not?

Do we need to start with first order? Booooring!

First order time discretization
has mysterious powers.

An ODE analogy for JKO (1/2)

The euclidean gradient flow

$$\dot{x} = -\nabla \boldsymbol{e}(x)$$

has the implicit Euler discretization with time step $\tau > 0$

$$\frac{x_\tau^n - x_\tau^{n-1}}{\tau} = -\nabla \boldsymbol{e}(x_\tau^n).$$

An ODE analogy for JKO (1/2)

The euclidean gradient flow

$$\dot{x} = -\nabla \boldsymbol{e}(x)$$

has the implicit Euler discretization with time step $\tau > 0$

$$\frac{x_\tau^n - x_\tau^{n-1}}{\tau} = -\nabla \boldsymbol{e}(x_\tau^n).$$

This is actually a variational scheme since

$$x_\tau^n \in \operatorname*{argmin}_x \left(\frac{1}{2\tau} \|x - x_\tau^{n-1}\|^2 + \boldsymbol{e}(x) \right).$$

An ODE analogy for JKO (2/2)

Lemma

If $\boldsymbol{e}$ is λ-uniformly convex with $\lambda > 0$, then

- $\boldsymbol{e}$ possesses a unique minimizer x_∞.
- the discrete speed decays exponentially,

$$\left\| \frac{x_\tau^{n+1} - x_\tau^n}{\tau} \right\|^2 \leq (1 + 2\lambda\tau)^{-1} \left\| \frac{x_\tau^n - x_\tau^{n-1}}{\tau} \right\|^2,$$

  and so do $\|\nabla \boldsymbol{e}(x_\tau^n)\|^2$, $\boldsymbol{e}(x_\tau^n) - \boldsymbol{e}(x_\infty)$, $\|x_\tau^n - x_\infty\|^2$.
- the time-discrete $(\mathrm{EVI})_\lambda$ holds:

$$\frac{\|x_\tau^n - y\|^2 - \|x_\tau^{n-1} - y\|^2}{2\tau} + \frac{\lambda}{2}\|x_\tau^n - y\|^2 \leq \boldsymbol{e}(y) - \boldsymbol{e}(x_\tau^n).$$

...see computations on board ...

Recall JKO

In Filippo's lecture you have seen:

JKO scheme

Given an initial datum ρ^0 and a time step $\tau > 0$, define approximations $(\rho_\tau^n)_{n=0,1,\ldots}$ inductively by

1. **Step** 0**:** ρ_τ^0 as approximation of ρ^0.
2. **Step** n**:** given ρ_τ^{n-1}, choose

$$\rho_\tau^n \in \operatorname*{argmin}_\rho \left(\frac{1}{2\tau} \mathbb{W}(\rho, \rho_\tau^{n-1})^2 + \boldsymbol{E}(\rho) \right).$$

Recall JKO

In Filippo's lecture you have seen:

JKO scheme

Given an initial datum ρ^0 and a time step $\tau > 0$, define approximations $(\rho_\tau^n)_{n=0,1,\dots}$ inductively by

1. **Step** 0**:** ρ_τ^0 as approximation of ρ^0.
2. **Step** n**:** given ρ_τ^{n-1}, choose

$$\rho_\tau^n \in \operatorname*{argmin}_\rho \left(\frac{1}{2\tau} \mathbb{W}(\rho, \rho_\tau^{n-1})^2 + \boldsymbol{E}(\rho) \right).$$

Under certain reasonable hypotheses on $\boldsymbol{E}$, the scheme is well-defined, and (ρ_τ^n) converges to a weak solution of the PDE.

JKO has it all (1/2)

Theorem

Under basic hypotheses on $\boldsymbol{E}$, the JKO steps (ρ_τ^n) satisfy:

- *Energy monotonicity:* $\boldsymbol{E}(\rho_\tau^{n_2}) \leq \boldsymbol{E}(\rho_\tau^{n_1})$.
- *Dissipation:* $\displaystyle \frac{1}{2}\,\tau \sum_{n=n_1+1}^{n_2} |\partial \boldsymbol{E}(\rho_\tau^n)|^2 \leq \boldsymbol{E}(\rho_\tau^{n_1}) - \boldsymbol{E}(\rho_\tau^{n_2})$.
- *Hölder continuity:* $\displaystyle \frac{1}{2}\,\frac{\mathbb{W}(\rho_\tau^{n_1},\rho_\tau^{n_2})^2}{\tau(n_2-n_1)} \leq \boldsymbol{E}(\rho_\tau^{n_1}) - \boldsymbol{E}(\rho_\tau^{n_2})$.

...see computations on board ...

JKO has it all (2/2)

Theorem

If $\boldsymbol{E}$ *is* λ*-displacement convex with* $\lambda > 0$, *then*

- *energy and dissipation decrease exponentially,*

$$\left|\partial \boldsymbol{E}(\rho_\tau^n)\right|^2 \leq (1+\lambda\tau)^{-2}\left|\partial \boldsymbol{E}(\rho_\tau^{n-1})\right|^2,$$
$$\boldsymbol{E}(\rho_\tau^n) - \boldsymbol{E}(\rho_\infty) \leq (1+\lambda\tau)^{-2}\big(\boldsymbol{E}(\rho_\tau^{n-1}) - \boldsymbol{E}(\rho_\infty)\big);$$

- *a time-discrete* $(\mathrm{EVI})_\lambda$ *holds:*

$$\frac{1}{2}\frac{\mathbb{W}(\rho_\tau^n,\eta)^2 - \mathbb{W}(\rho_\tau^{n-1},\eta)^2}{\tau} + \frac{\lambda}{2}\mathbb{W}(\rho_\tau^n,\eta)^2 \leq \boldsymbol{E}(\eta) - \boldsymbol{E}(\rho_\tau^n).$$

The curvature property

The key to deriving the discrete $(\mathrm{EVI})_\lambda$

$$\frac{1}{2}\frac{\mathbb{W}(\rho_\tau^n,\eta)^2-\mathbb{W}(\rho_\tau^{n-1},\eta)^2}{\tau}+\frac{\lambda}{2}\mathbb{W}(\rho_\tau^n,\eta)^2\le \boldsymbol{E}(\eta)-\boldsymbol{E}(\rho_\tau^n)$$

is the

First curvature property of $\mathbb{W}$

If $(\rho_s)_{s\in[0,1]}$ is a segment w.r.t. η, then

$$\mathbb{W}(\gamma_s,\eta)^2\le(1-s)\mathbb{W}(\rho_0,\eta)^2+s\mathbb{W}(\rho_1,\eta)^2-s(1-s)\mathbb{W}(\rho_0,\rho_1)^2.$$

...see computations on board ...

And . . . what about DLSS?

Recall: the rescaled DLSS equation

$$\partial_t \rho = -2\,\mathrm{div}\left(\rho\,\nabla \frac{\Delta\sqrt{\rho}}{\sqrt{\rho}}\right) + \mathrm{div}(x\,\rho)$$

is the $\mathbb{W}$-gradient flow of the relative Fisher information

$$\boldsymbol{F}'(\rho) = 2\int |\nabla\sqrt{\rho}|^2\,\mathrm{d}x + \frac{1}{2}\int |x|^2\rho\,\mathrm{d}x,$$

which is **totally not** displacement convex.

And . . . what about DLSS?

Recall: the rescaled DLSS equation

$$\partial_t \rho = -2\,\mathrm{div}\left(\rho\,\nabla \frac{\Delta\sqrt{\rho}}{\sqrt{\rho}}\right) + \mathrm{div}(x\,\rho)$$

is the $\mathbb{W}$-gradient flow of the relative Fisher information

$$\boldsymbol{F}'(\rho) = 2\int |\nabla\sqrt{\rho}|^2\,\mathrm{d}x + \frac{1}{2}\int |x|^2 \rho\,\mathrm{d}x,$$

which is **totally not** displacement convex.

Theorem

The JKO iterates (ρ_τ^n) for $\boldsymbol{F}'$ satisfy:

$$\begin{aligned}\boldsymbol{H}'(\rho_\tau^n) - \boldsymbol{H}'(\rho_\infty) &\le (1+2\tau)^{-1}\big(\boldsymbol{H}'(\rho_\tau^{n-1}) - \boldsymbol{H}'(\rho_\infty)\big),\\ \boldsymbol{F}'(\rho_\tau^n) - \boldsymbol{F}'(\rho_\infty) &\le (1+2\tau)^{-1}\big(\boldsymbol{F}'(\rho_\tau^{n-1}) - \boldsymbol{F}'(\rho_\infty)\big).\end{aligned}$$

And . . . what about DLSS?

Recall: the rescaled DLSS equation

$$\partial_t \rho = -2\,\mathrm{div}\left(\rho\,\nabla \frac{\Delta\sqrt{\rho}}{\sqrt{\rho}}\right) + \mathrm{div}(x\,\rho)$$

is the $\mathbb{W}$-gradient flow of the relative Fisher information

$$\boldsymbol{F}'(\rho) = 2\int |\nabla\sqrt{\rho}|^2\,\mathrm{d}x + \frac{1}{2}\int |x|^2 \rho\,\mathrm{d}x,$$

which is **totally not** displacement convex.

Theorem

The JKO iterates (ρ_τ^n) for $\boldsymbol{F}'$ satisfy:

$$\boldsymbol{H}'(\rho_\tau^n) - \boldsymbol{H}'(\rho_\infty) \le (1+2\tau)^{-1}\big(\boldsymbol{H}'(\rho_\tau^{n-1}) - \boldsymbol{H}'(\rho_\infty)\big),$$
$$\boldsymbol{F}'(\rho_\tau^n) - \boldsymbol{F}'(\rho_\infty) \le (1+2\tau)^{-1}\big(\boldsymbol{F}'(\rho_\tau^{n-1}) - \boldsymbol{F}'(\rho_\infty)\big).$$

Proof: via relation of $\boldsymbol{F}'$ to 1-displacement convex relative entropy $\boldsymbol{H}'$,

$$\boldsymbol{F}'(\rho) - \boldsymbol{F}'(\rho_\infty) = \frac{1}{2}|\partial \boldsymbol{H}'(\rho)|^2.$$

Okay, done?!? Can we do faster stuff, please!

If you go to second order,
some things break, some remain.

An ODE analogy for BDF2 (1/2)

The euclidean gradient flow

$$\dot{x} = -\nabla \boldsymbol{e}(x)$$

has the BDF2 discretization with time step $\tau > 0$

$$2\frac{x_\tau^n - x_\tau^{n-1}}{\tau} - \frac{x_\tau^n - x_\tau^{n-2}}{2\tau} = -\nabla \boldsymbol{e}(x_\tau^n). \qquad \text{(BDF2)}$$

Under reasonable conditions, the iteration $(x_\tau^{n-1}, x_\tau^{n-2}) \mapsto (x_\tau^n, x_\tau^{n-1})$ is well-defined, and approximates the gradient flow to second order.

An ODE analogy for BDF2 (1/2)

The euclidean gradient flow

$$\dot{x} = -\nabla \boldsymbol{e}(x)$$

has the BDF2 discretization with time step $\tau > 0$

$$2\frac{x_\tau^n - x_\tau^{n-1}}{\tau} - \frac{x_\tau^n - x_\tau^{n-2}}{2\tau} = -\nabla \boldsymbol{e}(x_\tau^n). \tag{BDF2}$$

Under reasonable conditions, the iteration $(x_\tau^{n-1}, x_\tau^{n-2}) \mapsto (x_\tau^n, x_\tau^{n-1})$ is well-defined, and approximates the gradient flow to second order.

(BDF2) is actually a variational scheme since

$$x_\tau^n \in \operatorname*{argmin}_x \left(\frac{1}{\tau}\|x - x_\tau^{n-1}\|^2 - \frac{1}{4\tau}\|x - x_\tau^{n-1}\|^2 + \boldsymbol{e}(x) \right).$$

An ODE analogy for BDF2 (2/2)

Example: In $n = 1$, consider $\boldsymbol{e}(x) = x^2/2$ with gradient flow

$$\dot{x} = -x \quad \text{that is} \quad x_t = x_0 e^{-t}.$$

JKO and BDF2 amount to, respectively

$$x_\tau^n = \frac{x_\tau^{n-1}}{1+\tau} \quad \text{and} \quad x_\tau^n = \frac{4x_\tau^{n-1} - x_\tau^{n-2}}{3+2\tau}.$$

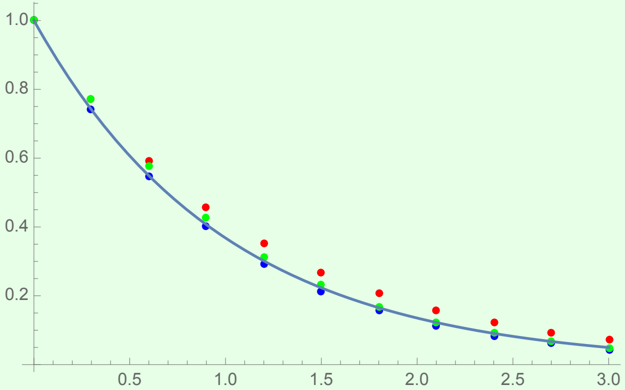
An ODE analogy for BDF2 (2/2)

Example: In $n = 1$, consider $\boldsymbol{e}(x) = x^2/2$ with gradient flow

$$\dot{x} = -x \quad \text{that is} \quad x_t = x_0 e^{-t}.$$

JKO and BDF2 amount to, respectively

$$x_\tau^n = \frac{x_\tau^{n-1}}{1+\tau} \quad \text{and} \quad x_\tau^n = \frac{4x_\tau^{n-1} - x_\tau^{n-2}}{3+2\tau}.$$



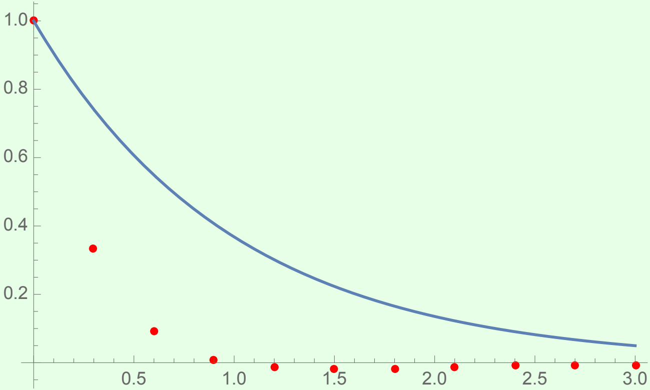
An ODE analogy for BDF2 (2/2)

Example: In $n = 1$, consider $\boldsymbol{e}(x) = x^2/2$ with gradient flow

$$\dot{x} = -x \quad \text{that is} \quad x_t = x_0 e^{-t}.$$

JKO and BDF2 amount to, respectively

$$x_\tau^n = \frac{x_\tau^{n-1}}{1+\tau} \quad \text{and} \quad x_\tau^n = \frac{4x_\tau^{n-1} - x_\tau^{n-2}}{3+2\tau}.$$



BDF2 for Wasserstein

In full analogy to JKO, introduce a BDF2-like two-step time discretization of $\boldsymbol{E}$'s gradient flow in $\mathbb{W}$:

BDF2 scheme

Given an initial datum ρ^0 and a time step $\tau > 0$, define approximations $(\rho_\tau^n)_{n=0,1,\ldots}$ inductively by

1. **Step** 0**:** ρ_τ^0 as approximation of ρ^0.
2. **Step** 1**:** ρ_τ^1 as approximation of genuine solution at time $t = \tau$.
3. **Step** n**:** given ρ_τ^{n-1} and ρ_τ^{n-2}, choose

$$\rho_\tau^n \in \operatorname*{argmin}_\rho \left(\frac{1}{\tau} \mathbb{W}(\rho, \rho_\tau^{n-1})^2 - \frac{1}{4\tau} \mathbb{W}(\rho, \rho_\tau^{n-2})^2 + \boldsymbol{E}(\rho) \right).$$

BDF2 has ... some of it! (1/2)

Theorem

Under basic hypotheses on $\boldsymbol{E}$, the BDF2 steps (ρ_τ^n) satisfy:

- *Almost energy monotonicity:*

$$\boldsymbol{E}(\rho_\tau^{n_2}) \leq \boldsymbol{E}(\rho_\tau^{n_1}) + \frac{1}{4\tau}\mathbb{W}(\rho_\tau^{n_1}, \rho_\tau^{n_1-1})^2;$$

- *Nothing noteworthy concerning dissipation.*
- *Almost Hölder continuity:*

$$\frac{\mathbb{W}\big(\rho_\tau^{n_1}, \rho_\tau^{n_2}\big)^2}{4(n_2 - n_1)\tau} \leq \boldsymbol{E}(\rho_\tau^{n_1}) - \boldsymbol{E}(\rho_\tau^{n_2}) + \frac{1}{4\tau}\mathbb{W}(\rho_\tau^{n_1}, \rho_\tau^{n_1-1})^2,$$

BDF2 has ... some of it! (1/2)

Theorem

Under basic hypotheses on $\boldsymbol{E}$, the BDF2 steps (ρ_τ^n) satisfy:

- *Almost energy monotonicity:*

$$\boldsymbol{E}(\rho_\tau^{n_2}) \leq \boldsymbol{E}(\rho_\tau^{n_1}) + \frac{1}{4\tau}\mathbb{W}(\rho_\tau^{n_1}, \rho_\tau^{n_1-1})^2;$$

- *Nothing noteworthy concerning dissipation.*
- *Almost Hölder continuity:*

$$\frac{\mathbb{W}\big(\rho_\tau^{n_1}, \rho_\tau^{n_2}\big)^2}{4(n_2-n_1)\tau} \leq \boldsymbol{E}(\rho_\tau^{n_1}) - \boldsymbol{E}(\rho_\tau^{n_2}) + \frac{1}{4\tau}\mathbb{W}(\rho_\tau^{n_1}, \rho_\tau^{n_1-1})^2,$$

... see computations on board ...

BDF2 has . . . some of it! (2/2)

Lemma

If $\boldsymbol{E}$ is λ-displacement convex with $\lambda > 0$, then

- *no result on exponential decrease of energy, speed etc.*
- *a discrete variant of* $(\mathrm{EVI})_\lambda$ *holds:*

$$\frac{\mathbb{W}(\rho_\tau^n,\eta)^2 - \mathbb{W}(\rho_\tau^{n-1},\eta)^2}{\tau} - \frac{1}{2}\left(\frac{\mathbb{W}(\rho_\tau^n,\eta)^2 - \mathbb{W}(\rho_\tau^{n-2},\eta)^2}{2\tau}\right)$$
$$+ \frac{\lambda}{2}\mathbb{W}(\rho_\tau^n,\eta)^2$$
$$\leq \boldsymbol{E}(\eta) - \boldsymbol{E}(\rho_\tau^n) - \frac{\mathbb{W}(\rho_\tau^n,\rho_\tau^{n-1})^2}{\tau} + \frac{\mathbb{W}(\rho_\tau^n,\rho_\tau^{n-2})^2}{4\tau}.$$

Another curvature property

The key to deriving the discrete $(\mathrm{EVI})_\lambda$

$$\frac{\mathbb{W}(\rho_\tau^n,\eta)^2-\mathbb{W}(\rho_\tau^{n-1},\eta)^2}{\tau}-\frac{1}{2}\left(\frac{\mathbb{W}(\rho_\tau^n,\eta)^2-\mathbb{W}(\rho_\tau^{n-2},\eta)^2}{2\tau}\right)$$
$$+\frac{\lambda}{2}\mathbb{W}(\rho_\tau^n,\eta)^2$$
$$\leq \boldsymbol{E}(\eta)-\boldsymbol{E}(\rho_\tau^n)-\frac{\mathbb{W}(\rho_\tau^n,\rho_\tau^{n-1})^2}{\tau}+\frac{\mathbb{W}(\rho_\tau^n,\rho_\tau^{n-2})^2}{4\tau}.$$

is the

Second curvature property of $\mathbb{W}$

If ρ_s is a segment w.r.t. η, then for $\alpha>\beta>0$, and any $\nu\in\mathcal{P}_2(\mathbb{R}^d)$,

$$\alpha\,\mathbb{W}(\rho_s,\eta)^2-\beta\,\mathbb{W}(\rho_s,\nu)^2$$
$$\leq(1-s)\big[\alpha\,\mathbb{W}(\rho_0,\eta)^2-\beta\,\mathbb{W}(\rho_0,\nu)^2\big]+s\big[\alpha\,\mathbb{W}(\rho_1,\eta)^2-\beta\,\mathbb{W}(\rho_1,\nu)^2\big]$$
$$-s(1-s)(\alpha-\beta)\mathbb{W}(\rho_0,\rho_1)^2,$$

. . . see computations on board . . .

That's all for today

Thank you, and see you tomorrow!