Evolution of Gaussians in the SHK-Boltzmann Gradient Flow

Gradient Flows Face-to-Face, Granada

Oliver Tse

*Joint work with **Jia-Jie Zhu, Matthias Liero,** and **Alexander Mielke***







The Context

Variational Inference: *Given a reference measure* $\pi \in \mathscr{P}(\mathbb{R}^d)$

$$\min_{\mu \in \mathscr{G}} \mathscr{E}(\mu) := \begin{cases} \int \log\left(\frac{d\mu}{d\pi}\right) d\mu & \text{if } \mu \ll \pi \\ +\infty & \text{otherwise} \end{cases}$$

Assumption: $\pi = e^{-V}$, *where* V *is* λ*-convex with* $\lambda > 0$

Possible Approaches ***Otto-Wasserstein flow***

$$\begin{aligned} \partial_t \mu &= \Delta \mu + \mathbf{div}(\mu \nabla V) \\ &= -\mathbb{K}_{\text{Otto}}(\mu) \mathrm{D}\mathscr{E}(\mu) \end{aligned} \qquad \begin{cases} \mathscr{E}(\mu_t) \le O(e^{-2\lambda t}) & \text{if } \lambda > 0, \\ \mathscr{E}(\mu_t) \le O(t^{-1}) & \text{if } \lambda = 0. \end{cases}$$

$$dX_t = \nabla \log \pi \, dt + \sqrt{2}\, dW_t$$ ***Overdamped Langevin***

Jordan-Kinderlehrer-Otto `98* *Ambrosio-Gigli-Savaré `05

The Context

Variational Inference: *Given a reference measure* $\pi \in \mathscr{P}(\mathbb{R}^d)$

$$\min_{\mu \in \mathscr{G}} \mathscr{E}(\mu) := \begin{cases} \int \log\left(\frac{d\mu}{d\pi}\right) d\mu & \text{if } \mu \ll \pi \\ +\infty & \text{otherwise} \end{cases}$$

Assumption: $\pi = e^{-V}$, *where* V *is* λ*-convex with* $\lambda > 0$

Possible Approaches ***Otto-Wasserstein flow***

$$\begin{aligned} \partial_t \mu &= \Delta \mu + \mathbf{div}(\mu \nabla V) \\ &= -\mathbb{K}_{\text{Otto}}(\mu) \mathrm{D}\mathscr{E}(\mu) \end{aligned} \qquad \begin{cases} \mathscr{E}(\mu_t) \le O(e^{-2\lambda t}) & \text{if } \lambda > 0, \\ \mathscr{E}(\mu_t) \le O(t^{-1}) & \text{if } \lambda = 0. \end{cases}$$

Issues: *1) Scales badly with dimension* $d \gg 1$ ⇝ *Parametrized family of measures*

2) Converges slowly for $\lambda \ll 1$ ⇝ *Acceleration methods*

Jordan-Kinderlehrer-Otto `98* *Ambrosio-Gigli-Savaré `05 *E.g. Chen-Li-Wright-Tse `25*

The Context

Variational Inference: *Given a reference measure* $\pi \in \mathscr{P}(\mathbb{R}^d)$

$$\min_{\mu \in \mathscr{G}} \mathscr{E}(\mu) := \begin{cases} \int \log\left(\frac{d\mu}{d\pi}\right) d\mu & \text{if } \mu \ll \pi \\ +\infty & \text{otherwise} \end{cases}$$

Assumption: $\pi = e^{-V}$, *where* V *is* λ*-convex with* $\lambda > 0$

Possible Approaches ***Hamiltonian flow***

$$\partial_t \sigma + \mathbf{div}_x(\sigma v) = \mathbf{div}_v\Big(\sigma\,(v + e^{\beta_t} \nabla \mathrm{D}\mathscr{E}(\mu))\Big)$$

$$\mathscr{E}(\mu_t) \le O(e^{-\beta_t})$$
$$\dot{\beta}_t \le 1$$

$$dX_t = V_t\, dt$$
$$dV_t = -\, V_t\, dt - e^{\beta_t} \nabla \mathrm{D}\mathscr{E}(\mu_t)(X_t)\, dt$$

⇝ *Doubling of dimensions*

⇝ *Well-posedness open!*

Chen-Li-Wright-Tse `25

The Context

Variational Inference: *Given a reference measure* $\pi \in \mathscr{P}(\mathbb{R}^d)$

$$\min_{\mu \in \mathscr{G}} \mathscr{E}(\mu) := \begin{cases} \int \log\left(\frac{d\mu}{d\pi}\right) d\mu & \text{if } \mu \ll \pi \\ +\infty & \text{otherwise} \end{cases}$$

Assumption: $\pi = e^{-V}$, *where* V *is* λ*-convex with* $\lambda > 0$

Possible Approaches ***Spherical Hellinger-Kantorovich flow*** *Wasserstein-Fisher-Rao*

$$\partial_t \mu = \alpha \underbrace{\mathbf{div}(\mu \nabla \mathrm{D}\mathscr{E}(\mu))}_{\text{Kantorovich}} - \beta \underbrace{\mu\left(\mathrm{D}\mathscr{E}(\mu) - \mathscr{E}(\mu)\right)}_{\text{Spherical Hellinger}}$$

$$= -\alpha \mathbb{K}_{\text{Otto}}(\mu)\mathrm{D}\mathscr{E}(\mu) - \beta \mathbb{K}_{\text{SHe}}(\mu)\mathrm{D}\mathscr{E}(\mu)$$

$$= -\mathbb{K}_{\alpha,\beta}(\mu)\mathrm{D}\mathscr{E}(\mu)$$

Liero-Mielke-Savaré `18

The Context

Variational Inference: *Given a reference measure* $\pi \in \mathscr{P}(\mathbb{R}^d)$

$$\min_{\mu \in \mathscr{G}} \mathscr{E}(\mu) := \begin{cases} \int \log\left(\frac{d\mu}{d\pi}\right) d\mu & \text{if } \mu \ll \pi \\ +\infty & \text{otherwise} \end{cases}$$

Assumption: $\pi = e^{-V}$, *where* V *is* λ*-convex with* $\lambda > 0$

Possible Approaches ***Spherical Hellinger-Kantorovich flow***

Wasserstein-Fisher-Rao

$$\partial_t \mu = \alpha\, \mathbf{div}(\mu \nabla \mathrm{D}\mathscr{E}(\mu)) - \beta\, \mu\left(\mathrm{D}\mathscr{E}(\mu) - \mathscr{E}(\mu)\right)$$

Stochastic process?

Feynman-Kac

Birth-death Accelerated Langevin

Convergence rate?

If initial data is sufficiently close, then decay rate is independent of λ

⇝ *Gaussian family*

Le Cavil-Oudjane-Russo '19 *Lu-Lu-Nolen '19* *Mielke-Savaré-Tse ?*

The Agenda

I. Gaussian SHK flow

II. Decay estimates

Geodesic convexity

Gradient dominance

Refined estimates

III. Non-Gaussian targets

IV. Summary + Outlook

The Agenda

I. Gaussian SHK flow

Gaussian SHK Flow

$$\partial_t \mu = -\,\mathbb{K}_{\alpha,\beta}(\mu)\mathrm{D}\mathscr{E}(\mu)$$

Gaussian measures: $\mathscr{G} = \left\{ \mathsf{G}(\Sigma, m) \in \mathscr{P}(\mathbb{R}^d) : \Sigma \text{ s.p.d}, m \in \mathbb{R}^d \right\}$

Let $\mu = \mathsf{G}(\Sigma, m)$ *and* $\pi = \mathsf{G}(\Gamma, n)$. *Then*

(BWFR)

$$\dot{\Sigma} = \alpha\left(2I - \Gamma^{-1}\Sigma - \Sigma\Gamma^{-1}\right) + \beta\left(\Sigma - \Sigma\Gamma^{-1}\Sigma\right)$$

$$\dot{m} = -\,(\alpha + \beta\Sigma)\Gamma^{-1}(m - n)$$

Questions

1) *Is there a gradient structure?*
2) *Does the Hellinger term improve long-time asymptotics?*

Remarks

1) *Equation for* Σ *is standalone!* ⇝ ***Total decay estimate subjected to*** Σ
2) *There is an explicit solution for* Σ*!*

Lambert-Chewi-Bach-Bonnabel-Rigollet '23 ***Chen-Huang-Huand-Reich-Stuart '23***

Gaussian SHK Flow

$$\partial_t \mu = -\,\mathbb{K}_{\alpha,\beta}(\mu)\mathrm{D}\mathscr{E}(\mu)$$

Gaussian measures: $\mathscr{G} = \left\{ \mathsf{G}(\Sigma, m) \in \mathscr{P}(\mathbb{R}^d) : \Sigma \text{ s.p.d}, m \in \mathbb{R}^d \right\}$

Let $\mu = \mathsf{G}(\Sigma, m)$ *and* $\pi = \mathsf{G}(\Gamma, n)$. *Then*

(BWFR)

$$\dot{\Sigma} = \alpha\left(2I - \Gamma^{-1}\Sigma - \Sigma\Gamma^{-1}\right) + \beta\left(\Sigma - \Sigma\Gamma^{-1}\Sigma\right)$$

$$\dot{m} = -\,(\alpha + \beta\Sigma)\Gamma^{-1}(m - n)$$

Questions

1) Is there a gradient structure?

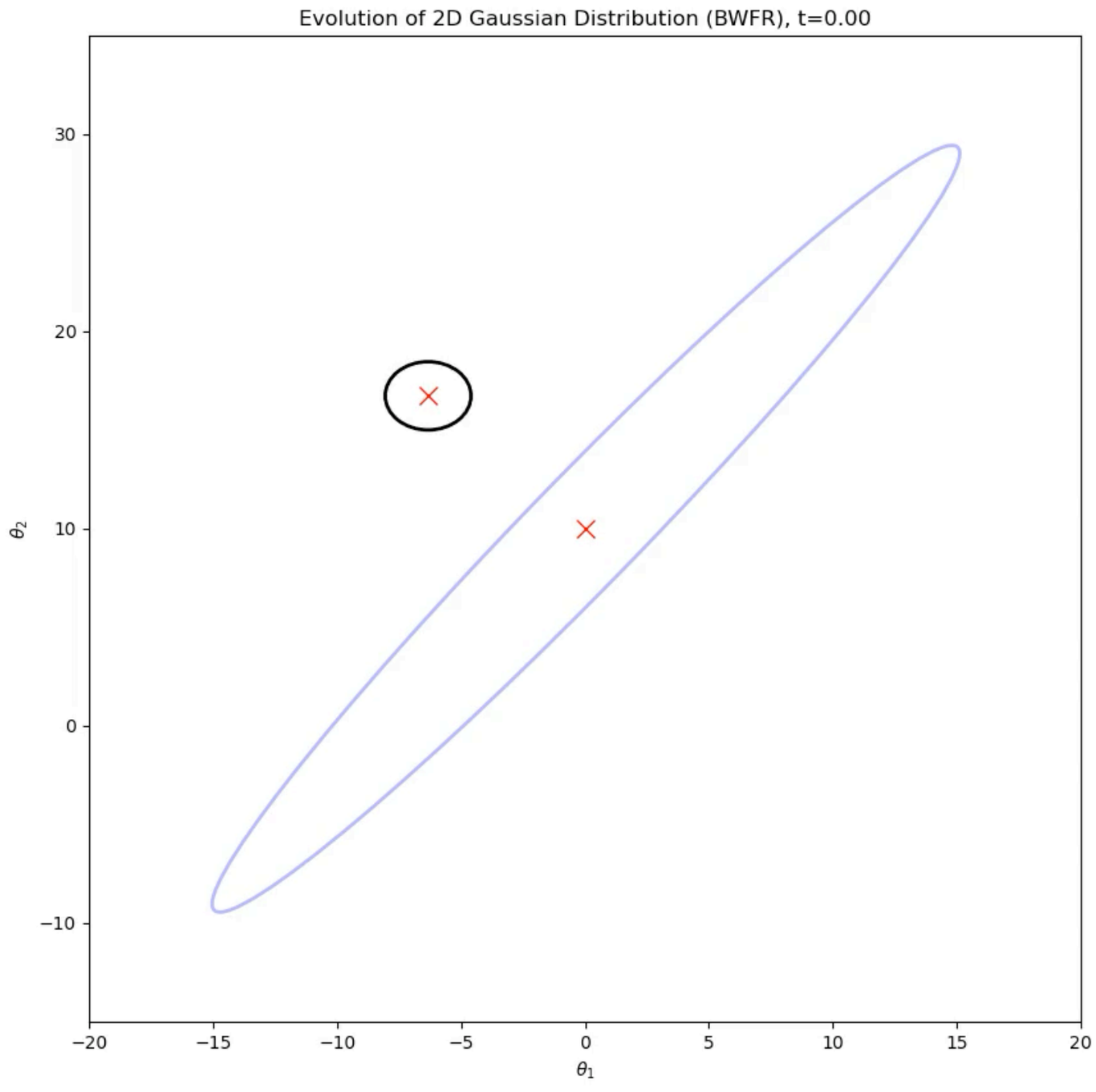
2) Does the Hellinger term improve long-time asymptotics?

Remarks

1) Equation for Σ *is standalone!* ⇝ ***Total decay estimate subjected to*** Σ

2) There is an explicit solution for Σ*!*

Lambert-Chewi-Bach-Bonnabel-Rigollet '23 ***Chen-Huang-Huand-Reich-Stuart '23***


Evolution of 2D Gaussian Distribution (BWFR), t=0.00
30
20
10
0
−10
θ_2
−20
−15
−10
−5
0
5
10
15
20
θ_1

Gaussian SHK Flow

$$\partial_t \mu = -\,\mathbb{K}_{\alpha,\beta}(\mu)\mathrm{D}\mathscr{E}(\mu)$$

Gaussian measures: $\mathscr{G} = \left\{ \mathsf{G}(\Sigma, m) \in \mathscr{P}(\mathbb{R}^d) : \Sigma \text{ s.p.d}, m \in \mathbb{R}^d \right\}$

Proposition (Maas-Mielke '20). *Let $(\mathfrak{X}, \mathscr{E}, \mathbb{K})$ be a gradient system and $\imath : \mathfrak{M} \to \mathfrak{X}$ be an embedding such that $\imath(\mathfrak{M}) \subset \mathfrak{X}$ is a smooth submanifold that is invariant under $\partial_t \mu = -\,\mathbb{K}(\mu)\mathrm{D}\mathscr{E}(\mu)$. Then the restricted flow on $\mathfrak{M}$ has a gradient system $(\mathfrak{M}, \mathscr{E}^{\mathsf{red}}, \mathbb{K}^{\mathsf{red}})$ with reduced energy $\mathscr{E}^{\mathsf{red}} = \mathscr{E} \circ \imath$, and the reduced Onsager operator is given by*

$$\langle \eta, \mathbb{K}^{\mathsf{red}}(\omega)\,\eta \rangle = \inf \left\{ \langle \xi, (\mathbb{K} \circ \imath)(\omega)\,\xi \rangle : \mathrm{D}\imath(\omega)^* \xi = \eta \right\}$$

Reduce SHK gradient structure

$$\mathfrak{M} = \mathsf{Mat}_{+,\mathrm{sym}}(d) \times \mathbb{R}^d, \quad \imath = \mathsf{G} \quad \rightsquigarrow \quad \imath(\mathfrak{M}) = \mathscr{G} \subset \mathscr{P}(\mathbb{R}^d)$$

$$\dot{\omega} = -\,\mathbb{K}^{\mathsf{red}}_{\alpha,\beta}(\omega)\mathrm{D}\mathscr{E}^{\mathsf{red}}(\omega), \qquad \omega = (\Sigma, m)$$

Gaussian SHK Flow

$$\partial_t \mu = -\,\mathbb{K}_{\alpha,\beta}(\mu)\mathrm{D}\mathscr{E}(\mu)$$

Gaussian measures: $\mathscr{G} = \left\{ \mathsf{G}(\Sigma, m) \in \mathscr{P}(\mathbb{R}^d) : \Sigma \text{ s.p.d}, m \in \mathbb{R}^d \right\}$

Reduce SHK gradient structure $\mathfrak{M} = \mathsf{Mat}_{+,\mathrm{sym}}(d) \times \mathbb{R}^d, \quad \iota = \mathsf{G}$

$$\dot{\omega} = -\,\mathbb{K}^{\mathsf{red}}_{\alpha,\beta}(\omega)\mathrm{D}\mathscr{E}^{\mathsf{red}}(\omega), \qquad \omega = (\Sigma, m)$$

$$= -\,\alpha\mathbb{K}^{\mathsf{red}}_{\mathsf{Otto}}(\omega)\mathrm{D}\mathscr{E}^{\mathsf{red}}(\omega) - \beta\mathbb{K}^{\mathsf{red}}_{\mathsf{SHe}}(\omega)\mathrm{D}\mathscr{E}^{\mathsf{red}}(\omega)$$

where

$$\mathscr{E}^{\mathsf{red}}(\omega) = \frac{1}{2}|m-n|^2_{\Gamma^{-1}} + \overbrace{\frac{1}{2}\left(\mathrm{tr}\left[\Gamma^{-1}\Sigma - I\right] - \log\det\left(\Gamma^{-1}\Sigma\right)\right)}^{=:\,\mathscr{F}(\Sigma)}$$

$$\mathbb{K}^{\mathsf{red}}_{\mathsf{Otto}}(\omega)[\Lambda] = \begin{bmatrix} 2(\Lambda_\Sigma\Sigma + \Sigma\Lambda_\Sigma) & 0 \\ 0 & \Lambda_m \end{bmatrix} = \begin{bmatrix} \mathbb{A}(\Sigma)\Lambda_\Sigma & 0 \\ 0 & \Lambda_m \end{bmatrix}$$

$$\mathbb{K}^{\mathsf{red}}_{\mathsf{SHe}}(\omega)[\Lambda] = \begin{bmatrix} 2\Sigma\Lambda_\Sigma\Sigma & 0 \\ 0 & \Sigma\Lambda_m \end{bmatrix} = \begin{bmatrix} \mathbb{B}(\Sigma)\Lambda_\Sigma & 0 \\ 0 & \Sigma\Lambda_m \end{bmatrix}$$

Gaussian SHK Flow

$$\partial_t \mu = -\,\mathbb{K}_{\alpha,\beta}(\mu)\mathrm{D}\mathscr{E}(\mu)$$

Gaussian measures: $\mathscr{G} = \left\{ \mathsf{G}(\Sigma, m) \in \mathscr{P}(\mathbb{R}^d) : \Sigma \text{ s.p.d}, m \in \mathbb{R}^d \right\}$

Reduce SHK gradient structure $\mathfrak{M} = \mathsf{Mat}_{+,\mathrm{sym}}(d) \times \mathbb{R}^d, \quad \iota = \mathsf{G}$

$$\dot{\omega} = -\,\mathbb{K}^{\mathsf{red}}_{\alpha,\beta}(\omega)\mathrm{D}\mathscr{E}^{\mathsf{red}}(\omega), \qquad \omega = (\Sigma, m)$$

$$= -\,\alpha\mathbb{K}^{\mathsf{red}}_{\mathsf{Otto}}(\omega)\mathrm{D}\mathscr{E}^{\mathsf{red}}(\omega) - \beta\mathbb{K}^{\mathsf{red}}_{\mathsf{SHe}}(\omega)\mathrm{D}\mathscr{E}^{\mathsf{red}}(\omega)$$

where

$$\mathscr{E}^{\mathsf{red}}(\omega) = \frac{1}{2}|m-n|^2_{\Gamma^{-1}} + \overbrace{\frac{1}{2}\left(\mathrm{tr}\left[\Gamma^{-1}\Sigma - I\right] - \log\det\left(\Gamma^{-1}\Sigma\right)\right)}^{=:\,\mathscr{F}(\Sigma)}$$

$$\mathbb{K}^{\mathsf{red}}_{\mathsf{Otto}}(\omega)[\Lambda] = \begin{bmatrix} \mathbb{A}(\Sigma)\Lambda_\Sigma & 0 \\ 0 & \Lambda_m \end{bmatrix} \qquad \mathbb{K}^{\mathsf{red}}_{\mathsf{SHe}}(\omega)[\Lambda] = \begin{bmatrix} \mathbb{B}(\Sigma)\Lambda_\Sigma & 0 \\ 0 & \Sigma\Lambda_m \end{bmatrix}$$

$$\dot{\Sigma} = -\,\alpha\mathbb{A}(\Sigma)\mathrm{D}\mathscr{F}(\Sigma) - \beta\mathbb{B}(\Sigma)\mathrm{D}\mathscr{F}(\Sigma), \qquad \mathrm{D}\mathscr{F}(\Sigma) = \frac{1}{2}\left(\Gamma^{-1} - \Sigma^{-1}\right)$$

The Agenda

II. Decay estimates

Geodesic convexity

Gradient dominance

Refined estimates

Decay Estimates

$$\dot{\Sigma} = -\alpha \mathbb{A}(\Sigma) \mathrm{D}\mathscr{F}(\Sigma) - \beta \mathbb{B}(\Sigma) \mathrm{D}\mathscr{F}(\Sigma)$$

$$\mathscr{F}(\Sigma) = \frac{1}{2}\Big(\mathrm{tr}\big[\Gamma^{-1}\Sigma - I\big] - \log\det\big(\Gamma^{-1}\Sigma\big) \Big)$$

Theorem. (Geodesic convexity)

$\mathscr{E}^{\mathrm{red}}$ *is* λ*-convex w.r.t.* $\mathbb{K}^{\mathrm{red}}_{\alpha,\beta}$ *if and only if* $\beta = 0$ *and* $\lambda \leq \alpha \lambda_{\min}(\Gamma^{-1})$

↝ *Hellinger term destroys geodesic convexity!*

Long-time asymptotics

$$\frac{d}{dt}\mathscr{F}(\Sigma) = -\Big\langle \mathrm{D}\mathscr{F}(\Sigma), \big(\alpha \mathbb{A}(\Sigma) + \beta \mathbb{B}(\Sigma)\big) \mathrm{D}\mathscr{F}(\Sigma) \Big\rangle =: -\mathscr{D}_{\alpha,\beta}(\Sigma)$$

Theorem. (Gradient dominance)

Global: $\exists\, \mathsf{c}_{\mathsf{PL}} > 0 : \quad \mathscr{D}_{\alpha,\beta}(\Sigma) \geq \mathsf{c}_{\mathsf{PL}} \mathscr{F}(\Sigma) \quad \Longleftrightarrow \quad \alpha \lambda_{\min}(\Gamma^{-1}) > 0$

Sublevel: $\forall E > 0, \exists\, \mathsf{c}_{\mathsf{PL}}(E) > 0 : \quad \forall \Sigma : \mathscr{F}(\Sigma) \leq E$

$$\mathscr{D}_{\alpha,\beta}(\Sigma) \geq \mathsf{c}_{\mathsf{PL}}(E) \mathscr{F}(\Sigma)$$

Works also when $\alpha \lambda_{\min}(\Gamma^{-1}) = 0$*!*

Decay Estimates

$$\dot\Sigma = -\alpha\mathbb{A}(\Sigma)\mathrm{D}\mathscr{F}(\Sigma) - \beta\mathbb{B}(\Sigma)\mathrm{D}\mathscr{F}(\Sigma)$$
$$\mathscr{F}(\Sigma) = \frac{1}{2}\Big(\operatorname{tr}\big[\Gamma^{-1}\Sigma - I\big] - \log\det\big(\Gamma^{-1}\Sigma\big)\Big)$$

Theorem. (Refined estimates)

For any $\Sigma_0 \in \operatorname{dom}\mathscr{F}$, there exists a constant $\mathsf{c}_\beta = \mathsf{c}_\beta(\Sigma_0)$:

$$\mathscr{F}(\Sigma(t)) \le \mathsf{c}_\beta\, e^{-\nu t}\mathscr{F}(\Sigma_0)$$

with $\nu = 2\alpha\lambda_{\min}(\Gamma^{-1}) + \beta$ ⇝ ***Hellinger term is independent of Γ!***

Idea. *Set $B := \Gamma^{-1/2}\Sigma\Gamma^{-1/2}$ with eigenvalues (b_i). Then*

$$1 \wedge \mathfrak{b}_i^{-1}(t) \le b_i(t) \le 1 \vee \mathfrak{b}_i(t)\,, \qquad \mathfrak{b}_i(t) := 1 + (b_i(0) - 1)\, e^{-\nu t}$$

Noticing that

$$\mathscr{F}(\Sigma) = \frac{1}{2}\sum\nolimits_i \phi(b_i)\,, \qquad \phi(r) = r - 1 - \log r$$

and

$$\frac{d}{dt}\mathscr{F}(\Sigma) \le -\left(2\alpha\lambda_{\min}(\Gamma^{-1}) + \beta b_{\min}\right)\mathscr{F}(\Sigma)\,,$$

a careful application of Gronwall's inequality yields the result.

The Agenda

III. Non-Gaussian targets

Non-Gaussian Target

$$\dot{\omega} = -\,\mathbb{K}^{\mathsf{red}}_{\alpha,\beta}(\omega)\mathrm{D}\mathscr{E}^{\mathsf{red}}(\omega)$$

Let $\mu = \mathsf{G}(\Sigma, m)$ and $\pi = e^{-V}$, where V is λ-convex with $\lambda > 0$. Then

(BWFR)

$$\dot{\Sigma} = \alpha\left(2I - \Gamma^{-1}\Sigma - \Sigma\Gamma^{-1}\right) + \beta\left(\Sigma - \Sigma\Gamma^{-1}\Sigma\right)$$

$$\dot{m} = -\,(\alpha + \beta\Sigma)\int \nabla V(x)\,\mathsf{G}(\Sigma, m)(dx)$$

with $\Gamma^{-1} = \displaystyle\int \nabla^2 V(x)\,\mathsf{G}(\Sigma, m)(dx)$

Theorem. (Existence of minimizer) $\mathscr{E}^{\mathsf{red}}$ *attains a unique minimizer in* $\mathfrak{M}$.

Theorem. (Decay estimate)

Let V be λ-convex with $\lambda > 0$. Then

$$\mathscr{E}^{\mathsf{red}}(\omega(t)) \le e^{-\gamma t}\,\mathscr{E}^{\mathsf{red}}(\omega_0) + (1 - e^{-\gamma t}) \min_{\eta\in\mathfrak{M}} \mathscr{E}^{\mathsf{red}}(\eta)$$

with $\gamma \propto \lambda$

Non-Gaussian Target

$$\dot{\omega} = -\,\mathbb{K}^{\mathsf{red}}_{\alpha,\beta}(\omega)\mathrm{D}\mathscr{E}^{\mathsf{red}}(\omega)$$

Theorem. (Decay estimate)

Let V be λ-convex with $\lambda > 0$. Then

$$\mathscr{E}^{\mathsf{red}}(\omega(t)) \leq e^{-\gamma t}\,\mathscr{E}^{\mathsf{red}}(\omega_0) + (1 - e^{-\gamma t}) \min_{\eta\in\mathfrak{M}} \mathscr{E}^{\mathsf{red}}(\eta)$$

with $\gamma \propto \lambda$

Idea. *Since V is λ-convex, $\mathscr{E}$ is λ-convex w.r.t. $\mathbb{W}_2$. Hence, for any $\omega \in \mathfrak{M}$:*

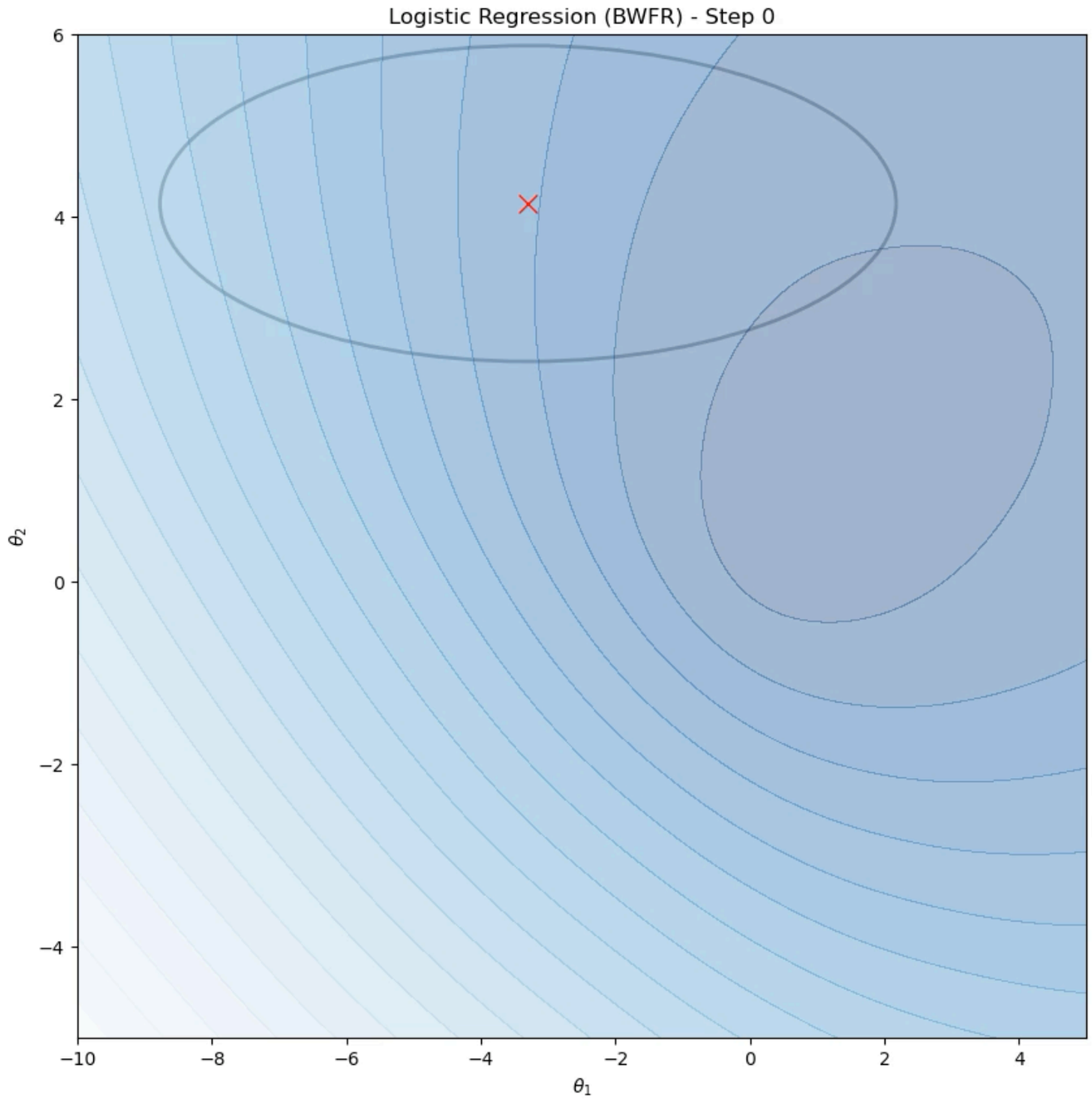
$$\mathscr{E}^{\mathsf{red}}(\omega) - \inf_{\eta\in\mathfrak{M}} \mathscr{E}^{\mathsf{red}}(\eta) \leq \frac{1}{2\lambda}\,|\,\partial_{\mathsf{Otto}}\mathscr{E}^{\mathsf{red}}\,|^2(\omega)$$

Additionally,

$$|\,\partial_{\mathsf{Otto}}\mathscr{E}^{\mathsf{red}}\,|(\omega) \leq \lambda_{\min}(\Sigma)\,|\,\partial_{\mathsf{SHe}}\mathscr{E}^{\mathsf{red}}\,|(\omega)$$

Putting these together yields

$$\begin{aligned}\frac{d}{dt}\mathscr{E}^{\mathsf{red}}(\omega(t)) &= -\,\alpha\,|\,\partial_{\mathsf{Otto}}\mathscr{E}^{\mathsf{red}}\,|^2(\omega) - \beta\,|\,\partial_{\mathsf{SHe}}\mathscr{E}^{\mathsf{red}}\,|^2(\omega)\\ &\leq -\,\big(\alpha + \beta\lambda_{\min}(\Sigma)\big)\,|\,\partial_{\mathsf{Otto}}\mathscr{E}^{\mathsf{red}}\,|^2(\omega)\end{aligned}$$


Logistic Regression (BWFR) - Step 0
6
4
2
0
−2
−4
θ_2
−10
−8
−6
−4
−2
0
2
4
θ_1

The Agenda

IV. Summary + Outlook

Summary

SHK flow

$$\partial_t \mu = -\,\mathbb{K}_{\alpha,\beta}(\mu)\mathrm{D}\mathscr{E}(\mu)$$

⇝

Gaussian SHK flow

$$\dot{\omega} = -\,\mathbb{K}^{\mathsf{red}}_{\alpha,\beta}(\omega)\mathrm{D}\mathscr{E}^{\mathsf{red}}(\omega)$$

Decay estimates for Gaussian target

Geodesic convexity ✗

Gradient dominance ✓

Refined estimates ✓

$$\mathscr{E}^{\mathsf{red}}(\omega(t)) \le \mathsf{c}_\beta\, e^{-\nu t}\, \mathscr{E}^{\mathsf{red}}(\omega_0)$$

$$\nu = 2\alpha\lambda_{\min}(\Gamma^{-1}) + \beta$$

Outlook

SHK flow

$$\partial_t \mu = -\,\mathbb{K}_{\alpha,\beta}(\mu)\mathrm{D}\mathscr{E}(\mu)$$

⇝

Gaussian SHK flow

$$\dot{\omega} = -\,\mathbb{K}^{\mathsf{red}}_{\alpha,\beta}(\omega)\mathrm{D}\mathscr{E}^{\mathsf{red}}(\omega)$$

Decay estimates for Gaussian target

- *Geodesic convexity* ✗
- *Gradient dominance* ✓
- *Refined estimates* ✓

$$\mathscr{E}^{\mathsf{red}}(\omega(t)) \le \mathsf{c}_\beta\, e^{-\nu t}\, \mathscr{E}^{\mathsf{red}}(\omega_0)$$

$$\nu = 2\alpha\lambda_{\min}(\Gamma^{-1}) + \beta$$

Decay estimates for non-Gaussian target

Justification of numerical scheme

Thank you!