Prof. Dr. Matthes
Optimaler Transport - Geometrie und Anwendungen

Manuela Lambacher
xx.12.2015

tolles referat

1 Generalisierte Gradientenflüsse

Wir betrachten folgende partielle Diffgleichung:

$$\frac{dX}{dt} = -\operatorname{grad} E(X) \tag{1.1}$$

E ist hierbei die Energie im Bezug auf den Optimalen Transport, die Diffgl beschreibt z.b. die kinetische Energie von Partikeln. TODO: Räume, was woraus...

Um dieses Problem nicht explizit zu lösen, wollen wir es zeitlich diskretisieren, sodann zum Limes übergehen während unser Zeitschritt gegen Null geht. Dabei können subdifferentiale und tangentenräume schön umgangen werden und der GradientenOperator nicht explizit benutzt werden.

Definition 1.1. *Der approximierte Gradientenfluss für ein Energie-Funktional E in einem abstraktem Metrischen Raum mit einer metric-denoted Distanz sei gegeben durch:*
Sei der timestep $\tau > 0$, dann ist die Folge $(X_\tau^n)_{n\geq 0}$:
$X_\tau^0 := X_0$,
X_τ^{n+1} *ist der Minimierer (oder ein Minimierer, wir verlangen keine Eindeutigkeit TODO: ex eine lösung?) von*

$$\min \left[E(X) + \frac{\operatorname{dist}(X_\tau^n, X)^2}{2\tau} \right] \tag{1.2}$$

Sei X_τ auf $\mathbb{R}_+$ als stückweise konstante Funktion mit Wert $X_\tau(t) = X_\tau^n$ für $t \in [n\tau, (n+1)\tau)$.

Bemerkung 1.2. *Betrachtet man den euklidischen Abstand, ist die Euler-Lagrange-Gleichung zu dem Minimierungsproblem (1.2)*

$$\frac{X_\tau^{n+1} - X_\tau^n}{\tau} = -\operatorname{grad} E(X_\tau^{n+1})$$

Nach Aufstellen der Diskretisierung wollen wir $\tau \to 0$ gehen lassen und betrachten den Grenzwert, den "generalisierten Gradientenfluss". Um zu zeigen, dass dieser Grenzwert exisitert und die Ausgangsgleichung erfüllt, müssen dessen Eigenschaften untersucht werden.

1.1 Drei Ungleichungen

Wir nehmen an, dass E nach unten durch eine absolute Konstante ist, d.h. $E(X) \geq C$ unabh. von X.

X_τ^{n+1} ist ein Minimierer von dem Funktional $X \mapsto E(X) + \dfrac{\operatorname{dist}(X_\tau^n, X)^2}{2\tau}$, und damit ist offensichtlich das Funktional an X_τ^{n+1} kleiner gleich dem Wert dessen an der Stelle X_τ^n (Eigenschaften Metrik: $\operatorname{dist}(X_\tau^n, X_\tau^n) = 0$). Damit haben wir

$$E(X_\tau^{n+1}) + \frac{\operatorname{dist}(X_\tau^n, X_\tau^{n+1})^2}{2\tau} \leq E(X_\tau^n). \tag{1.3}$$

Lemma 1.3 (Energy estimate)**.**

$$\sup_{n\geq 0} E(X_\tau^n) \leq E(X^0) \tag{1.4}$$

Beweis. Folgt sofort aus (1.3), da der zweite Term nichtnegativ und somit die Folge der $(E(X_\tau^n)$ für n monoton abnehmend ist. □

Lemma 1.4 (total square distance estimate)**.**

$$\sum_{n\geq 0} \operatorname{dist}(X_\tau^n, X_\tau^{n+1})^2 \leq 2\tau(E(X^0) - \inf E) \tag{1.5}$$

Beweis. Summation über (1.3) □

Bemerkung 1.5. *Daraus kann auch eine Hölder 1/2-Abschätzung für X_τ abgeleitet werden: für $s < t$ gilt:*

$$\operatorname{dist}(X_\tau(s), X_\tau(t))^2 \leq \left[\frac{t-s}{\tau} + 1\right] \sum_{\frac{s}{\tau}\leq n\leq \frac{t}{\tau}} \operatorname{dist}(X_\tau^n, X_\tau^{n+1})^2 \leq C[(t-s)+\tau], \tag{1.6}$$

wobei $C = 2[E(X^0) - \inf E]$.
Folgt aus Dreiecksunglichung für W_2 und Cauchy-Schwarz.

Lemma 1.6 (Energy gradient estimate)**.** *Sei eine zugrundeliegende Riemannstruktur gegeben und das Energie Funktional hinreichend glatt. Dann gilt*

$$\tau \sum_{n\geq 0} \|\operatorname{grad} E(X_\tau^n)\|^2 \leq 2[E(X^0) - \inf E] \tag{1.7}$$

bzw. im kontinuierlichen Falle:

$$\int_0^\infty \|\operatorname{grad} E(X_\tau^n)\|^2 dt \leq 2[E(X^0) - \inf E] \tag{1.8}$$

Beweis. Sei ω ein beliebiger Tangentenvektor auf X_τ^{n+1}. Definiere für ein kleines $\varepsilon > 0$ einen "Pfad" $\tilde{X}_\varepsilon$ folgendermaßen:

$$\tilde{X}_0 = X_\tau^{n+1}, \quad \left.\frac{d\tilde{X}_\varepsilon}{d\varepsilon}\right|_{\varepsilon=0} = \omega \tag{1.9}$$

Da X_τ^{n+1} Minimierer ist, gilt offensichtlich

$$E(X_\tau^{n+1}) + \frac{\operatorname{dist}(X_\tau^n, X_\tau^{n+1})^2}{2\tau} \leq E(\tilde{X}_\varepsilon) + \frac{\operatorname{dist}(X_\tau^n, \tilde{X}_\varepsilon)^2}{2\tau} \quad \forall \varepsilon > 0 \tag{1.10}$$

Unter der Annahme, dass das Energiefunktional glatt ist, gilt:

$$E(\tilde{X}_\varepsilon) = E(X_\tau^{n+1}) + \varepsilon \langle \operatorname{grad} E(X_\tau^{n+1}), \omega\rangle + O(\varepsilon^2) \tag{1.11}$$

Außerdem gilt:

$$\begin{aligned}
&\operatorname{dist}(X_\tau^n, \tilde{X}_\varepsilon)^2 - \operatorname{dist}(X_\tau^n, X_\tau^{n+1})^2 \\
= &\left(\operatorname{dist}(X_\tau^n, \tilde{X}_\varepsilon) + \operatorname{dist}(X_\tau^n, X_\tau^{n+1})\right)\left(\operatorname{dist}(X_\tau^n, \tilde{X}_\varepsilon) - \operatorname{dist}(X_\tau^n, X_\tau^{n+1})\right) \\
&\leq \left(\operatorname{dist}(X_\tau^n, \tilde{X}_\varepsilon) + \operatorname{dist}(X_\tau^n, X_\tau^{n+1})\right) \operatorname{dist}(X_\tau^{n+1}, \tilde{X}_\varepsilon) \\
&\leq \left(2 \operatorname{dist}(X_\tau^n, X_\tau^{n+1} + \operatorname{dist}(X_\tau^{n+1}, \tilde{X}_\varepsilon)\right) \operatorname{dist}(X_\tau^{n+1}, \tilde{X}_\varepsilon)
\end{aligned}$$

Wir wissen, dass

$$\operatorname{dist}(X_\tau^{n+1}, \tilde{X}_\varepsilon) = \varepsilon\|\omega\| + o(\varepsilon)$$

. Damit wird aus obiger Gleichung

$$\frac{\operatorname{dist}(X_\tau^n, \tilde{X}_\varepsilon)^2}{2\tau} \leq \frac{\operatorname{dist}(X_\tau^n, X_\tau^{n+1})^2}{2\tau} + \varepsilon \operatorname{dist}(X_\tau^n, X_\tau^{n+1})\frac{\|\omega\|}{\tau} + o(\varepsilon) \tag{1.12}$$

Jetzt kombinieren wir die drei Aussagen 1.101.111.12: Dazu setzen wir zunächst 1.11in 1.10ein.

$$
\begin{aligned}
E(X_\tau^{n+1}) + \frac{\operatorname{dist}(X_\tau^n, X_\tau^{n+1})^2}{2\tau} &\leq E(X_\tau^{n+1}) + \varepsilon \langle \operatorname{grad} E(X_\tau^{n+1}), \omega\rangle + O(\varepsilon^2) + \frac{\operatorname{dist}(X_\tau^n, \tilde{X}_\varepsilon)^2}{2\tau} \\
\Rightarrow \frac{\operatorname{dist}(X_\tau^n, X_\tau^{n+1})^2}{2\tau} &\overset{1.12}{\leq} \varepsilon \langle \operatorname{grad} E(X_\tau^{n+1}), \omega\rangle + O(\varepsilon^2) + \frac{\operatorname{dist}(X_\tau^n, X_\tau^{n+1})^2}{2\tau} + \varepsilon \operatorname{dist}(X_\tau^n, X_\tau^{n+1}) \frac{\|\omega\|}{\tau} + o(\varepsilon) \\
&\Rightarrow 0 \leq \varepsilon \left(\langle \operatorname{grad} E(X_\tau^{n+1}), \omega\rangle) + \operatorname{dist}(X_\tau^n, X_\tau^{n+1}) \frac{\|\omega\|}{\tau} + \underbrace{\frac{o(\varepsilon)}{\varepsilon} + \frac{O(\varepsilon^2)}{\varepsilon}}_{\to 0 \text{ für } \varepsilon \to 0} \right)
\end{aligned}
$$

Nun wählen wir das explizite $\omega = -\operatorname{grad} E(X_\tau^{n+1})$ (warum ist das im geforderten raum?)

$$
\begin{aligned}
0 \leq \underbrace{\langle \operatorname{grad} E(X_\tau^{n+1}), -\operatorname{grad} E(X_\tau^{n+1})\rangle}_{-\|\operatorname{grad} E(X_\tau^{n+1})\|^2} + \frac{\operatorname{dist}(X_\tau^n, X_\tau^{n+1}) \|\operatorname{grad} E(X_\tau^{n+1})\|}{\tau} \\
\Rightarrow \frac{\operatorname{dist}(X_\tau^n, X_\tau^{n+1})^2}{\tau^2} \geq \|\operatorname{grad} E(X_\tau^{n+1})\|^2
\end{aligned}
$$

In Verbindung mit der square distance estimate (1.5) kommen wir schließlich zu der geforderten Gleichung:

$$
2\tau(E(X^0) - \inf E) \geq \sum_{n \geq 0} \operatorname{dist}(X_\tau^n, X_\tau^{n+1})^2 \geq \tau^2 \sum_{n \geq 1} \|\operatorname{grad} E(X_\tau^n)\|^2 \tag{1.13}
$$

TODO: warum kann ich den term für n=0 auch mitnehmen? also summe über $n \geq 0$? □

Bemerkung 1.7. *Die kontinueriliche Funktion gilt sogar in einer nicht-Riemanschen Situation mit einer passenden Definition der Norm des Gradienten. (todo: riemann setting häh?)*

Diese drei Abschätzungen sollten die relative Kompaktheit von (X_τ) für $\tau \to 0$ sicherstellen.
Um aber zum Limes überzugehen, müssen wir ein dem letzten Beweis ähnliches Prozedere unterlaufen: Kleine perturbationen, ω beliebig lassen. Eine approximierte euler-lagrange-gl finden. siehe nächstes kapteil. außerdem: ????

2 Anwendung auf das Monge-Kantorovich Problem

Die oben erwähnte Strategie wurde zum ersten mal am Beispiel der linearen Fokker-Plank-Gleichung angewandt. Dies liefert interessante EInblicke in die Methode. Die übrigens auch implementiert werden kann.

setting:
Betrachte das Energie-Funktional

$$
E(\rho) = \int \rho \log \rho + \int \rho V
$$

auf $P(\mathbb{R}^k)$, wobei $V(x)$ ein glattes Potential (??) ist, das folgende Wachstumsbedingung erfüllt: $V(x) = O(|x|^2)$ für $x \to \infty$.
Sei ρ_0 eine Wahrscheinlichkeitsdichte als Anfangswert mit endlichem $E(\rho)$.
Sei weiterhin $\tau > 0$ und definiere sodann die Folge (ρ_τ^n) analog zu Punkt 1 des Vortrages:

$$
\rho_\tau^0 = \rho^0;
$$

gegeben ρ_τ^n definiere den nächsten Schritt als Minimierer von

$$
\rho \mapsto E(\rho) + \frac{W_2(\rho_\tau^n, \rho)^2}{2\tau}
$$

Da E strikt konvex im üblichen Sinne ist (displacement konvex nur wenn V konvex ist), ebenso wie die Wassersteinmetrik, und damit ist unser Minimierer eindeutig.
Wie im vorherigen Abschnitt kann man eine uniform control über $\sup_{t\geq 0} E(\rho_\tau(t))$ erlangen, die einem die tightness der familie $(rho_\tau)_{\tau>0}$ in der x-varibale sowie schwache L^1-kompaktheit garantiert.
Die approximate Hölder 1/2 continuity estimate in der quadraticshen wasserstein distanz sichert die approximative uniforme equikontinuität in zeit von $(\rho_\tau)_\tau$. All dies ist hinreichend um (betrachte evtl subsequence $(\tau_k)_k$) $(\rho_\tau)_\tau$ zu einer Funktion $\rho : \mathbb{R}_+ \mapsto P_{ac}(\mathbb{R}^n)$ konvergiert, stetig falls $P_{ac}(\mathbb{R}^n)$ mit der schwachen L^1-Topologie verknüft ist.
Um nun den Grenzewrt zu betrachten, brauche nwir eine geeignete Variation von ρ_τ^{n+1}.
Sei ξ ein glattes Vektorfeld mit kompaktem Träger, und T_ε die Familie von Trajektoren verknüpft mit dem Vektorfeld $\mathrm{Id} + \varepsilon\xi$. Dann definiere

$$\tilde{\rho}_\varepsilon = T_\varepsilon \# \rho_\tau^{n+1}$$

Für ε klein genug ist T_ε ein C^1-diffeomorphismus (fixed point theoren und $\det(\nabla T_\varepsilon) > 0$.
Nach Definition gilt:

$$E(\tilde{\rho}_\varepsilon) = \int \tilde{\rho}_\varepsilon \log \tilde{\rho}_\varepsilon + \int \tilde{\rho}_\epsilon V \tag{2.1}$$

$$= \int \rho_\tau^{n+1} \log \frac{\rho_\tau^{n+1}}{\det(I_k + \varepsilon \nabla \xi)} + \int \rho_\tau^{n+1}(x) V(x + \varepsilon\xi(x)) dx. \tag{2.2}$$

Aus der energy estimate folgt andererseits, dass $\rho_\tau^n, \rho_\tau^{n+1}$ absolut stetig sind. Damit existert eine optimale Abbildung $\nabla\phi$ s.t. $\nabla\phi \# \rho_\tau^n = \rho_\tau^{n+1}$ und es gilt

$$W_2(\rho_\tau^n, \tilde{\rho}_\varepsilon)^2 \leq \int \rho_\tau^n(x) |x - \nabla\phi(x) - \varepsilon\xi \circ \nabla\phi(x)|^2 dx$$

Damit kommen wir auf

$$\frac{W_2(\rho_\tau^n, \tilde{\rho}_\varepsilon)^2}{2\tau} + E(\tilde{\rho}_\varepsilon) - \frac{W_2(\rho_\tau^n, \rho_\tau^{n+1})^2}{2\tau} - E(\rho_\tau^{n+1}) \tag{2.3}$$

$$\leq \int \rho_\tau^n(x) \left(\frac{|x - \nabla\phi(x) - \varepsilon\xi \circ \nabla\phi(x)|^2 - |x - \nabla\phi(x)|^2}{2\tau} \right) dx \tag{2.4}$$

$$+ \int \rho_\tau^{n+1}(x)[V(x + \varepsilon\xi(x)) - V(x)]dx - \int \rho_\tau^{n+1}(x) \log\det(\mathrm{Id} + \varepsilon\nabla\xi(x))dx \geq 0 \tag{2.5}$$

wobei die nichtnegativität der linken Seite aus der Minimierungseigenschaft von ρ_τ^{n+1} folgt. Dividiere durch ε und lasse dieses gegen 0^+ gehen, so bekommen wir:

$$0 \leq \frac{1}{\tau} \int \rho_\tau^n(x) \left\langle \nabla\phi(x) - x, \xi \circ \nabla\phi(x) \right\rangle dx + \int \rho_\tau^{n+1}(x) \langle \nabla V(x), \xi(x) \rangle dx$$
$$- \int \rho_\tau^{n+1}(x)(\nabla \cdot \xi)(x) dx$$

Da dies genauso für $-\xi$ gilt, bekommen wir Gleichheit. Umforuliert:

$$\frac{1}{\tau} \int \rho_\tau^n(x) \left\langle \nabla\phi(x) - x, \xi \circ \nabla\phi(x) \right\rangle dx = \int \rho_\tau^{n+1}(x) \left[(\nabla \cdot \xi)(x) - \langle \nabla V(x), \xi(x) \rangle \right] dx \tag{2.6}$$

Jetzt sei ξ derart, dass für ein $\zeta \in \mathcal{D}(\mathbb{R}^n)$ gilt $\xi = \nabla\zeta$. Es gilt (erweiterung):

$$\zeta(\nabla\phi(x)) - \zeta(x) = \langle \nabla\phi(x) - x, \nabla\zeta \circ \nabla\phi(x) \rangle + O(|x - \nabla\phi(x)|^2 \tag{2.7}$$

und damit kann die linke Seite von (2.6)ersetzt werden durch

$$\frac{1}{\tau} \left(\int \rho_\tau^n(x) \zeta \circ \nabla\phi(x) dx - \int \rho_\tau^n(x) \zeta(x) dx \right) + O\left(\frac{1}{\tau} \int \rho_\tau^n(x) |x - \nabla\phi(x)|^2 dx \right)$$
$$= \frac{1}{\tau} \left(\int \rho_\tau^{n+1} \zeta - \int \rho_\tau^n \zeta \right) + O\left(\frac{W_2(\rho_\tau^n, \rho_\tau^{n+1})^2}{\tau} \right)$$

Setzt dies in(2.6) ein und nun wird aufsummiert für t_1, t_2 beliebige Zeiten von $n_1 = [t_1/\tau]$ bis zu $n_2 = [t_2/\tau]$:

$$\int \rho_\tau(t_2)\zeta - \int \rho_\tau(t_1)\zeta + O\left(\sum_{n=n_1}^{n_2} W_2(\rho_\tau^n, \rho_\tau^{n+1})^2\right) \tag{2.8}$$

$$= \int_{t_1}^{t_2} \int \rho_\tau(t)(\nabla\zeta - \nabla V \cdot \nabla\zeta)dt + O8\tau) \tag{2.9}$$

Wobei wir benutzt haben, dass $\nabla\zeta$ und $\nabla V \cdot \nabla\zeta$ beschränkte Funktionien sind und $\rho_\tau(t) = \rho_\tau(t, \cdot)$ Wahrscheinlichkeitsmaß auf $\mathbb{R}^n$.

Jetzt benutzen wir die total square estimate wrt W_2 und deduzieren für alle fixed $t_1,\ t_2$:

$$\int \rho_\tau(t_2)\zeta - \int \rho_\tau(t_1)\zeta + O(\tau) = \int_{t_1}^{t_2} \int \rho_\tau(t)(\nabla\zeta - \nabla V \cdot \nabla\zeta)dt + O(\tau) \tag{2.10}$$

Grenzwert $\tau \to 0$:

$$\int \rho(t_2)\zeta - \int \rho(t_1)\zeta = \int_{t_1}^{t_2} \int \rho(t)(\nabla\zeta - \nabla V \cdot \nabla\zeta)dt \tag{2.11}$$

die schwache Formulierng der linearen Fokker-Planck-Gleichung mit Potential V, d.h.

$$\frac{\partial\rho}{\partial t} = \nabla\rho + \nabla \cdot (\rho\nabla V) \tag{2.12}$$

Diese Gleichung ist erfüllt bei jedem “weak cluster point’"der Folge (ρ_τ); da es eine eindeutige Lösung annimmt, muss diese Lösung der Grenzwert für $\tau \to 0$ sein.
Damit ist bewiesen:

Theorem 2.1. *Für den Zeitschritt $\tau \to 0$ konvergiert die Lösung des Zeit-diskretisierten Gradienten gegen die Lösung der linearen Fokker-Planck-Gleichung mit Anfangswert ρ_0.*
Damit gilt: Die Fokker-Planck-Gleichung ist der Graidentenfluss für das freie EnergieFunktional mit der Wassersteindistanz.

Bemerkung 2.2.

3 stuff