Prof. Dr. Matthes
Optimaler Transport - Geometrie und Anwendungen

Manuela Lambacher
09.12.2015

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. Stellen wir uns vor, $X(t)$ ist die Position von gegebenen 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 "wohldefiniertes" Energie-Funktional E in einem abstraktem metrischen Raum mit einer metric-denoted Distanz sei gegeben durch:*
Sei die zeitliche Schrittweite $\tau > 0$ gegeben, $n \in \mathbb{N}_0$, dann ist die Folge $(X_\tau^n)_{n\geq 0}$:
$X_\tau^0 := X_0$,
X_τ^{n+1} *ist ein Minimierer von dem Minimerungsproblem*

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

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

Bemerkung 1.2.

1. *Wir setzen zunächst keine Eindeutigkeit des Minimums voraus;*
2. *Existenz des Minimierers muss durch geeignete Eigenschaften von $E(X)$ gesichert werden, z.B. Coercivität, Weakly lower semicontinuity und eine untere Schranke bzw. Wachstumsbedingung.*
3. *Verdeutlichen wir uns, warum dies unser anfängliches Problem approximieren soll:*
   Betrachtet man den euklidischen Abstand, so besitzt das Problem (1.2) einen kritischen Punkt (der für $E(X)$ konvex ein globales Minimum ist) bei X_τ^{n+1} gegeben durch
   $$\frac{X_\tau^{n+1} - X_\tau^n}{\tau} = -\operatorname{grad} E(X_\tau^{n+1}),$$
   wobei die linke Seite offensichtlich eine Approximation der zeitlichen Ableitung von $X(t)$ ist.
4. *Verbindung zum Subdifferential:* $\dfrac{X_\tau^n - X_\tau^n}{\tau} \in \partial E(X_\tau^{n+1})$

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, wollen wir zunächst einige Eigenschaften untersuchen.

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 $X \mapsto E(X) + \dfrac{\operatorname{dist}(X_\tau^n, X)^2}{2\tau}$. Damit ist offensichtlich das Funktional an der Stelle X_τ^{n+1} kleiner gleich dessen Wert an der Stelle X_τ^n (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 steigendes 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), Teleskopsumme, Infimum da $E(X_\tau^n)$ monoton fallend. □

Lemma 1.5 (Hölder 1/2-Abschätzung für X_τ)**.**
Für $s < t$ gilt:

$$\operatorname{dist}(X_\tau(s), X_\tau(t))^2 \overset{1)}{\leq} \left[\frac{t-s}{\tau} + 1\right] \sum_{\frac{s}{\tau}\leq n \leq \frac{t}{\tau}} \operatorname{dist}(X_\tau^n, X_\tau^{n+1})^2 \overset{2)}{\leq} 2[E(X^0) - \inf E[(t-s) + \tau., \tag{1.6}$$

Beweis. 2) folgt wiederum aus Aufsummation von (1.3) und Monotonie.

1) Nach Definition: $X_\tau(s) = X_\tau^n$ für $s \in [n\tau, (n+1)\tau)$. Damit gilt für n: $n = \lfloor \frac{s}{\tau} \rfloor$. Analog sei $X_\tau(t) = X_\tau^m$, $m = \lfloor \frac{t}{\tau} \rfloor$.

$$\begin{aligned}
\operatorname{dist}(X_\tau(s), X_\tau(t)) &= \operatorname{dist}(X_\tau^2, X_\tau^m) \\
&\overset{\triangle-Ungl.}{\leq} \operatorname{dist}(X_\tau^n, X_\tau^{n+1}) + \operatorname{dist}(X_\tau^{n+1}, X_\tau^{n+2}) + ... + \operatorname{dist}(X_\tau^{m-1}, X_\tau^m) \\
&= \sum_{n\leq k\leq m-1} \operatorname{dist}(X_\tau^k, X_\tau^{k+1}) \leq \sum_{n\leq k\leq m} \operatorname{dist}(X_\tau^k, X_\tau^{k+1})
\end{aligned}$$

Anm: warum wird im Buch die höhere Grenze für k genommen?

$$\begin{aligned}
\operatorname{dist}(X_\tau(s) - X_\tau(t))^2 &\leq \left(\sum_{\frac{s}{\tau}\leq k\leq \frac{t}{\tau}} \operatorname{dist}(X_\tau^k, X_\tau^{k+1}) \right)^2 \\
&\overset{CS}{\leq} \left(\sum_{\frac{s}{\tau}\leq k\leq \frac{t}{\tau}} \operatorname{dist}(X_\tau^k, X_\tau^{k+1})^2 \right) \underbrace{\left(\sum_{\frac{s}{\tau}\leq k\leq \frac{t}{\tau}} 1 \right)}_{\frac{t}{\tau} - (\frac{s}{\tau} - 1)}
\end{aligned}$$

□

Lemma 1.6 (Energy gradient estimate)**.** *Sei eine zugrundeliegende Riemannstruktur (Differenzierbare Mannigfaltigkeit, Metrik, Tangentenräume) 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 stetigen Falle für E_τ (nach dessen Definition):

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

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

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

(Also $\lim_{\varepsilon\to 0} \frac{\tilde{X}_\varepsilon - \tilde{X}_0}{\tau} = \omega$)
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 mit Taylor:

$$\begin{aligned} E(\tilde{X}_\varepsilon) &= E(X_0) + \langle \operatorname{grad} E(X_0), \underbrace{X_\varepsilon - X_0}_{\sim\omega\varepsilon}\rangle + O((X_\varepsilon - X_0)^2) \\ &= E(X_\tau^{n+1}) + \varepsilon\langle \operatorname{grad} E(X_\tau^{n+1}), \omega\rangle + O(\varepsilon^2) \end{aligned} \tag{1.11}$$

O oder o?
Außerdem gilt:

$$\begin{aligned} &\operatorname{dist}(X_\tau^n, \tilde{X}_\varepsilon)^2 - \operatorname{dist}(X_\tau^n, X_\tau^{n+1})^2 = \\ &\overset{binom.}{=} \Big(\operatorname{dist}(X_\tau^n, \tilde{X}_\varepsilon) + \operatorname{dist}(X_\tau^n, X_\tau^{n+1})\Big)\Big(\operatorname{dist}(X_\tau^n, \tilde{X}_\varepsilon) - \operatorname{dist}(X_\tau^n, X_\tau^{n+1})\Big) \\ &\overset{\triangle-Ungl.}{\leq} \Big(\operatorname{dist}(X_\tau^n, \tilde{X}_\varepsilon) + \operatorname{dist}(X_\tau^n, X_\tau^{n+1})\Big)\operatorname{dist}(X_\tau^{n+1}, \tilde{X}_\varepsilon) \\ &\overset{\triangle-Ungl.}{\leq} \Big(2\operatorname{dist}(X_\tau^n, X_\tau^{n+1} + \operatorname{dist}(X_\tau^{n+1}, \tilde{X}_\varepsilon)\Big)\operatorname{dist}(X_\tau^{n+1}, \tilde{X}_\varepsilon) \end{aligned}$$

Beobachtung:

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

. 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 (1.10),(1.11) und (1.12): Dazu setzen wir zunächst (**??** in (1.10) ein.

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

Nun wählen wir ω explizit: $\omega = -\operatorname{grad} E(X_\tau^{n+1})$ (im Tangentialraum als Ableitung eines Funktionals auf dem Vektorraum?)

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

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

$$2\tau(E(X^0) - \inf E) \overset{1.5}{\geq} \sum_{n\geq 0} \mathrm{dist}(X_\tau^n, X_\tau^{n+1})^2 \geq \tau^2 \sum_{n\geq 1} \| \operatorname{grad} E(X_\tau^n)\|^2 \tag{1.13}$$

TODO: warum steht im Buch Summe über $n \geq 0$? Ist auf dist-Seite nicht einmal definiert. □

Bemerkung 1.7. *Die Aussage des Lemmas für X_τ gilt sogar in einer nicht-riemannschen Situation mit einer passenden Definition der Norm des Gradienten.*

Diese drei Abschätzungen sollten die relative Kompaktheit von (X_τ) für $\tau \to 0$ sicherstellen, und damit die Existenz einer konvergenten Teilfolge. (Wie folgt relative Kompaktheit für X_τ aus Abschätzungen für das Energiefunktional??)
Um abschließend 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.

2 Die lineare Fokker-Plank-Gleichung, Monge-Kantorovich und approximierter Gradientenfluss

Die hergeleitete Strategie wurde zum ersten Mal am Beispiel der linearen Fokker-Plank-Gleichung angewandt. Dies liefert interessante EInblicke in die Methode, die auch in numerischer Impelementierung Anwendung findet.

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:

$$\begin{aligned} E(\tilde{\rho}_\varepsilon) = & \qquad \textstyle\int \tilde{\rho}_\varepsilon \log \tilde{\rho}_\varepsilon + \int \tilde{\rho}_\epsilon V & (2.1) \\ = & \textstyle\int \rho_\tau^{n+1} \log \frac{\rho_\tau^{n+1}}{\det(I_k + \varepsilon \nabla \xi)} + \int \rho_\tau^{n+1}(x) V(x + \varepsilon \xi(x)) dx. & (2.2) \end{aligned}$$

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

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

Damit kommen wir auf

$$\begin{aligned} & \frac{W_2(\rho_\tau^n, \tilde{\rho}_\varepsilon)^2}{2\tau} + E(\tilde{\rho}_\varepsilon) - \frac{W_2(\rho_\tau^n, \rho_\tau^{n+1})^2}{2\tau} - E(\rho_\tau^{n+1}) & (2.3) \\ & \le \int \rho_\tau^n(x) \left(\frac{|x - \nabla\phi(x) - \varepsilon\xi \circ \nabla\phi(x)|^2 - |x - \nabla\phi(x)|^2}{2\tau} \right) dx & (2.4) \\ & + \int \rho_\tau^{n+1}(x) [V(x + \varepsilon\xi(x)) - V(x)] dx - \int \rho_\tau^{n+1}(x) \log \det(\mathrm{Id} + \varepsilon \nabla \xi(x)) dx \ge 0 & (2.5) \end{aligned}$$

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:

$$\begin{aligned} 0 \le \frac{1}{\tau} \int \rho_\tau^n(x) \langle \nabla\phi(x) - x, \xi \circ \nabla\phi(x) \rangle \, dx + \int \rho_\tau^{n+1}(x) \langle \nabla V(x), \xi(x) \rangle dx \\ - \int \rho_\tau^{n+1}(x) (\nabla \cdot \xi)(x) dx \end{aligned}$$

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

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

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

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

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

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

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 $n_2 = [t_2/\tau]$:

$$\begin{aligned} & \int \rho_\tau(t_2)\zeta - \int \rho_\tau(t_1)\zeta + O\left(\sum_{n=n_1}^{n_2} W_2(\rho_\tau^n, \rho_\tau^{n+1})^2 \right) & (2.8) \\ & = \int_{t_1}^{t_2} \int \rho_\tau(t) (\nabla\zeta - \nabla V \cdot \nabla\zeta) dt + O8\tau) & (2.9) \end{aligned}$$

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 $t_1,\ t_2$:

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

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