

敵対的生成ネットワークの幾何学的学習理論

吉田 英樹

最終更新：September 29, 2025

Abstract

本稿は、敵対的生成ネットワーク (GAN) の学習ダイナミクスを、(i) 確率測度の最適輸送幾何 (Wasserstein-2 幾何), (ii) 束上の Ehresmann 接続・曲率, (iii) 変分法的作用原理, という三つの幾何学的・物理学的枠組みで統一かつ厳密に定式化する。

理論と実装の整合性を示すため、**曲率推定の実験** (F の可視化) と**反対称ヤコビアン近似** (自動微分を用いた $J - J^\top$ 推定) を新設した。さらに、**Pullback 計量のモンテカルロ推定**、**ガウス・ニュートン近似**、および**正則化**に関する収束率・安定性の命題を追加し、証明を付与した。これらの定式化により、GAN 学習の幾何学的・物理学的理解を深め、実装における安定性と収束性を保証する理論的基盤を提示する。

1 はじめに

敵対的生成ネットワーク (GAN) の学習ダイナミクスは、従来、 f -ダイバージェンス最小化やワッサーズタイン距離に基づく静的な定式化、あるいは二者ゲームとしての局所安定性解析に依拠してきた。しかし、学習過程を**確率測度多様体上の幾何学的流れ**として統一的に理解する枠組みは十分に整備されていない。

本稿では、GAN 学習を以下の観点から体系化する：

1. **Wasserstein-2 幾何と勾配流**：確率測度空間 $\mathcal{P}_2(X)$ における Otto 計量を用い、 f -ダイバージェンスに基づく目的関数の W_2 勾配流として GAN 学習を定式化する [1, 2, 3]。
2. **Ehresmann 接続と曲率**：判別器が誘導するポテンシャルを束上の切断とみなし、接続 1-形式および曲率 2-形式を構成することで、ゲーム的相互作用 (反対称ヤコビアン) を幾何学的に可視化する [4, 5]。
3. **作用原理と過減衰極限**：ラグランジアンと散逸項を導入し、拡張変分原理からオイラー-ラグランジュ方程式を導出、過減衰極限で W_2 勾配流に一致することを示す。

さらに、パラメータ多様体への引き戻し計量により、自然勾配法や Gauss-Newton 近似との整合性を示し、実装上の安定性・収束性を保証するための正則化や集中不等式も議論する。

2 関連研究

最適輸送と勾配流. Wasserstein 空間における勾配流と JKO スキームは、拡散方程式やエントロピー散逸の変分的定式化において基本的な役割を果たす [1, 2, 3]。

GAN のゲーム力学. GAN 学習を多者ゲームとして解析し、ヤコビアンの対称・反対称分解に基づく安定化手法が提案されている [4, 5]。本稿は、この反対称成分を接続の曲率として幾何学的に解釈する。

アンバランス輸送と HK 幾何. 質量生成を許す Hellinger-Kantorovich 幾何は、KL 型凸性を回復し、JSD 系の目的関数に対する大域的安定性を補完する [6]。

本稿の位置付け. 以上を踏まえ、本稿は GAN 学習を「 W_2 勾配流」「接続と曲率」「作用原理」の 3 つの柱で統合し、理論・実装を結ぶ幾何学的学習理論を提示する。

3 前提条件：状態空間・測度幾何・解析の仮定

仮定 3.1 (状態空間と基準測度). \mathcal{X} は境界なしコンパクト d 次元リーマン多様体, μ は体積測度. 以降, $L^p(\mathcal{X})$ 内積・ノルムは μ に関する.

定義 3.2 (確率測度空間と W_2 距離). $\mathcal{P}_2(\mathcal{X})$ を有限二次モーメントをもつ確率測度全体とし, コスト $c(x, y) = \frac{1}{2}d(x, y)^2$ に基づく Kantorovich 最適輸送の値で W_2 を定義する.

定義 3.3 (形式的接空間と Otto 計量). $P = \rho\mu \in \mathcal{P}_2(\mathcal{X})$ が正密度 $\rho \in C^1(\mathcal{X})$ をもてば, 形式的接ベクトルは $v = -\text{div}(\rho\nabla\phi)$ ($\phi \in H^1(\mathcal{X})$) の形で, 内積を

$$\langle v_1, v_2 \rangle_P := \int_{\mathcal{X}} \langle \nabla\phi_1, \nabla\phi_2 \rangle \rho \, dx$$

で定める (Otto 計量).

仮定 3.4 (測度論的正則性). 密度 ρ, q は $0 < m \leq \rho, q \leq M$ を満たし C^1 (必要に応じ C^2). 部分積分と JKO 近似で用いる.

4 f -ダイバージェンスと敵対的ポテンシャル

定義 4.1 (f -ダイバージェンス). 凸関数 $f \in C^2((0, \infty))$ に対し, $Q = q\mu$ に関する f -ダイバージェンスを

$$F_f(P\|Q) := \int_{\mathcal{X}} q(x) f\left(\frac{\rho(x)}{q(x)}\right) dx$$

で定義する.

補題 4.2 (第一変分と敵対的ポテンシャル). 滑らかな摂動 $\rho_\varepsilon = \rho + \varepsilon\eta$ ($\int \eta \, dx = 0$) に対し

$$\left. \frac{d}{d\varepsilon} \right|_{\varepsilon=0} F_f(P_\varepsilon\|Q) = \int_{\mathcal{X}} f'\left(\frac{\rho}{q}\right) \eta \, dx.$$

よって, 変分導関数 (敵対的ポテンシャル) は $\varphi_P := f'(\rho/q)$ である.

証明. 定義より $F_f(P_\varepsilon\|Q) = \int q(x) f((\rho(x) + \varepsilon\eta(x))/q(x)) \, dx$. $f \in C^2$ なので連鎖律により $\frac{d}{d\varepsilon} F_f(P_\varepsilon\|Q) = \int q(x) f'((\rho + \varepsilon\eta)/q) \cdot \eta/q \, dx$. $\varepsilon = 0$ を代入して主張を得る. \square

Jensen–Shannon (JSD) 特化 $M = \frac{1}{2}(P + Q)$ とすると

$$\text{JSD}(P\|Q) = \frac{1}{2}\text{KL}(P\|M) + \frac{1}{2}\text{KL}(Q\|M) = \text{Ent}_M - \frac{1}{2}\text{Ent}_P - \frac{1}{2}\text{Ent}_Q,$$

ここで $\text{Ent}_\nu(\rho) := \int \rho \log(\rho/\nu) \, dx$.

5 W_2 勾配フロー：連続時間・JKO・JSD の局所凸性

定義 5.1 (W_2 勾配フロー). $P_t = \rho_t \mu$ の時間発展を

$$\partial_t \rho_t = \operatorname{div}(\rho_t \nabla \varphi_{P_t}), \quad \varphi_{P_t} = f'(\rho_t/q) \quad (5.1)$$

と定義する (Lemma 4.2) .

定理 5.2 (エネルギー散逸等式 (EDI)). $\rho_t \in C^1([0, T]; C^1(\mathcal{X}))$ が (5.1) の古典解なら, *a.e.* t で

$$\frac{d}{dt} F_f(P_t \| Q) = - \int_{\mathcal{X}} \rho_t \|\nabla \varphi_{P_t}\|^2 dx \leq 0.$$

証明. $t \mapsto F_f(P_t \| Q)$ は連鎖律で $\frac{d}{dt} F_f = \int \varphi_{P_t} \partial_t \rho_t dx$. (5.1) を代入すると $\int \varphi_{P_t} \operatorname{div}(\rho_t \nabla \varphi_{P_t}) dx$. 仮定 3.1 仮定 3.4 により部分積分が正当化され, $-\int \langle \nabla \varphi_{P_t}, \rho_t \nabla \varphi_{P_t} \rangle dx = -\int \rho_t \|\nabla \varphi_{P_t}\|^2 dx$. よって主張が従う. \square

仮定 5.3 (局所 λ -凸性). F_f は W_2 に関して局所的に λ -凸 ($\lambda \in \mathbb{R}$) かつ下半連続.

定義 5.4 (JKO 離散化). 時間幅 $\tau > 0$ に対し

$$P_k \in \arg \min_{P \in \mathcal{P}_2} \left\{ F_f(P \| Q) + \frac{1}{2\tau} W_2^2(P, P_{k-1}) \right\}, \quad P_0 \text{ 既知.}$$

区分定数補間 $P^\tau(t)$ を定める (JKO) .

定理 5.5 (JKO による存在・収束). 仮定 5.3 の下で, 各 τ に最小化子 P_k が存在し, $\tau \rightarrow 0$ の部分列極限 P_t は (5.1) の弱解 (最大傾斜曲線) となる. またエネルギー減衰と総変動の一樣評価が成り立つ.

証明. (1) **井戸型・下半連続性**) W_2 の完全可分性と F_f の下半連続性から, 汎関数 $\Phi_\tau(P) := F_f(P \| Q) + \frac{1}{2\tau} W_2^2(P, P_{k-1})$ は狭義位相で下半連続. $P \mapsto W_2^2(P, P_{k-1})$ のコーシーブ性 (測度の二次モーメント制御) により, 劣閉集合上で下に有界, 最小化列は緊致. 極限点で下半連続性により極小が達成される.

(2) **逆向き Euler–Lagrange**) 最小化子 P_k は擾動 $P_s = (\operatorname{Id} + s \nabla \psi)_\# P_k$ を用いた第一変分から $\frac{1}{\tau} \int \langle x - T_{k-1}^k(x), \nabla \psi(x) \rangle \rho_k(x) dx + \int \langle \nabla \varphi_{P_k}(x), \nabla \psi(x) \rangle \rho_k(x) dx = 0$ を満たす (T_{k-1}^k は $P_{k-1} \rightarrow P_k$ の最適輸送写像). よって $\frac{T_{k-1}^k(x) - x}{\tau} = -\nabla \varphi_{P_k}(x)$ P_k -a.e. が得られる.

(3) **エネルギー推定と BV 有界性**) 三角不等式と (2) から $\sum_k \frac{1}{2\tau} W_2^2(P_k, P_{k-1}) \leq F_f(P_0 \| Q) - \inf F_f$, これが分割に依らず一様有界.

(4) **連続時間極限**) 斉一な総変動界と狭義緊致性から $P^\tau(t) \rightarrow P(t)$ (部分列). (2) の Euler–Lagrange と Benamou–Brenier 表現を時間平均極限に送ることで (5.1) の弱定式化を満たすことが従い, 最大傾斜曲線の構成になる. (詳細は [1, 2] の最小化移動法の標準的議論). \square

5.1 JSD の局所解析

仮定 5.6 (密度の近接と有界性). $0 < m \leq \rho, q \leq M$, かつ $\|\rho/q - 1\|_{L^\infty} \leq \delta \ll 1$.

定理 5.7 (JSD の局所 λ -凸性). 仮定 3.1 仮定 3.4 仮定 5.6 の下で, $\operatorname{JSD}(\cdot \| Q)$ は Q の十分小近傍で W_2 測地線に沿って λ -凸 ($\lambda \geq -C\delta$). よって初期値が近傍内なら Theorem 5.5 が適用できる.

証明. P_t を W_2 測地線 (McCann 補間) とし, 密度 ρ_t は連続の式 $\partial_t \rho_t + \operatorname{div}(\rho_t \nabla \psi_t) = 0$ を満たし, ψ_t は対応する速度ポテンシャル.

(1) Ent の変位凸性) $M_t = (P_t + Q)/2$ とおく. $\operatorname{Ent}_P(\rho) = \int \rho \log \rho \, dx$ は W_2 に関して変位凸 (McCann [3]) で, $\frac{d^2}{dt^2} \operatorname{Ent}_P(\rho_t) \geq 0$.

(2) Ent_{M_t} の補正項) $\operatorname{JSD} = \operatorname{Ent}_{M_t} - \frac{1}{2} \operatorname{Ent}_{P_t} - \frac{1}{2} \operatorname{Ent}_Q$. ここで $\operatorname{Ent}_{M_t}(\rho_t) = \int \rho_t \log \frac{\rho_t}{m_t} \, dx$, $m_t = \frac{\rho_t + q}{2}$. t に関する一階・二階変分を計算する:

$$\frac{d}{dt} \operatorname{Ent}_{M_t} = \int \left(\log \frac{\rho_t}{m_t} + 1 \right) \partial_t \rho_t \, dx - \int \rho_t \frac{\partial_t m_t}{m_t} \, dx.$$

$\partial_t m_t = \frac{1}{2} \partial_t \rho_t$ であるから $\frac{\partial_t m_t}{m_t} = \frac{\partial_t \rho_t}{\rho_t + q}$. ゆえに

$$\frac{d}{dt} \operatorname{Ent}_{M_t} = \int \left(\log \frac{\rho_t}{m_t} + 1 - \frac{\rho_t}{\rho_t + q} \right) \partial_t \rho_t \, dx.$$

二階微分は

$$\frac{d^2}{dt^2} \operatorname{Ent}_{M_t} = \int \left(\frac{1}{\rho_t} - \frac{1}{\rho_t + q} + \frac{\rho_t}{(\rho_t + q)^2} \right) (\partial_t \rho_t)^2 \, dx + \int \left(\log \frac{\rho_t}{m_t} + 1 - \frac{\rho_t}{\rho_t + q} \right) \partial_t^2 \rho_t \, dx.$$

連続の式を用い $\partial_t \rho_t = -\operatorname{div}(\rho_t \nabla \psi_t)$, $\partial_t^2 \rho_t$ も同様に展開する. 第一項の係数は

$$\frac{1}{\rho_t} - \frac{1}{\rho_t + q} + \frac{\rho_t}{(\rho_t + q)^2} = \frac{q^2}{\rho_t(\rho_t + q)^2} \in \left[\frac{q^2}{M(M+q)^2}, \frac{q^2}{m(m+q)^2} \right].$$

(3) 近接仮定の使用) $\rho_t/q = 1 + O(\delta)$ より, 上式は $\asymp C(\delta^0)$. 一方, 第二項は $\partial_t^2 \rho_t$ に対し $\log(\rho_t/m_t) + 1 - \rho_t/(\rho_t + q) = O(\rho_t/q - 1) = O(\delta)$. 部分積分・Young 不等式で評価すると $\frac{d^2}{dt^2} \operatorname{Ent}_{M_t} \geq -C\delta \int \rho_t \|\nabla^2 \psi_t\|^2 \, dx$.

(4) まとめ) $\frac{d^2}{dt^2} \operatorname{JSD} = \frac{d^2}{dt^2} \operatorname{Ent}_{M_t} - \frac{1}{2} \frac{d^2}{dt^2} \operatorname{Ent}_{P_t}$ であり, Ent_{P_t} の変位凸性から負項は現れない. ゆえに $\frac{d^2}{dt^2} \operatorname{JSD} \geq -C\delta \int \rho_t \|\nabla^2 \psi_t\|^2 \, dx$ で, 測地線に沿う二階変分の下界 $\lambda \geq -C\delta$ が従う. \square

注意 5.8 (HK 幾何の代替). 大域凸性が必要な場面では Hellinger–Kantorovich (WFR) 幾何で KL 型凸性が回復できる ([6]).

6 Ehresmann 接続・曲率とゲーム・ヤコビアン

6.1 束と接続 1-形式

定義 6.1 (自明束と切断). $E = \mathcal{P}_2(\mathcal{X}) \times \mathcal{F}$ ($\mathcal{F} = C^\infty(\mathcal{X})$) とし, $\Phi : \mathcal{P}_2(\mathcal{X}) \rightarrow \mathcal{F}$, $P \mapsto \varphi_P = f'(\rho/q)$ は切断.

定義 6.2 (Ehresmann 接続と接続 1-形式). 点 $(P, \phi) \in E$ の水平分布を $H_{(P, \phi)} = \{(v, D\Phi(P)[v])\}$ とし, 接続 1-形式 $\omega : TE \rightarrow \mathcal{F}$ を $\omega_{(P, \phi)}(v, \eta) = \eta - D\Phi(P)[v]$ で定義する.

定義 6.3 (ベクトル場値 1-形式 (ゲージ場) と曲率 2-形式). $A_P : T_P \mathcal{P}_2(\mathcal{X}) \rightarrow \operatorname{Vect}(\mathcal{X})$ を $A_P(v) := \nabla_x(D\Phi(P)[v])$ と置く. 曲率 $F = dA$ を

$$F_P(u, v) := DA_P[u]v - DA_P[v]u \quad (\text{外微分})$$

で定義する.

6.2 パラメータ埋め込みとヤコビアン分解

仮定 6.4 (生成器の可微分埋め込み). $\Theta \subset \mathbb{R}^n$ 開集合, $\theta \mapsto P_\theta = (G_\theta)_\# P_Z$ は C^2 -埋め込み.

定義 6.5 (ゲーム・ヤコビアンと分解). 目的関数 $J(\theta) := F_f(P_\theta \| Q)$ のヘッセ行列を $\mathcal{J}(\theta) := \nabla_\theta^2 J(\theta)$ とし, $\mathcal{J} = S + K$, $S = \frac{1}{2}(\mathcal{J} + \mathcal{J}^\top)$ (対称), $K = \frac{1}{2}(\mathcal{J} - \mathcal{J}^\top)$ (反対称) に分解する.

補題 6.6 (切断の Fréchet 微分の明示形). Q 固定のもとで $D\Phi(P)[v] = f''(\rho/q) \cdot \frac{\eta}{q}$ ($v = \eta\mu$).

証明. $\Phi(P) = f'(\rho/q)$ なので, 方向 $v = \eta\mu$ に沿う微分は $D\Phi(P)[v] = f''(\rho/q) \cdot (\eta/q)$ (q は固定). $f \in C^2$ より正当化される. \square

命題 6.7 (曲率 2-形式と反対称ヤコビアンの対応). 仮定 6.4 の下で, $u = \partial_i P_\theta$, $v = \partial_j P_\theta$ に対し任意のテストベクトル場 ξ について

$$\int_{\mathcal{X}} \langle \xi, F_{P_\theta}(u, v) \rangle dx = \frac{1}{2} \partial_i \theta^\top \mathcal{J}(\theta) \mathcal{J}(\theta)^\top \partial_j \theta + E_{\text{proj}}(\xi; u, v), \quad (6.1)$$

ここで E_{proj} は $T_{P_\theta} \mathcal{P}_2$ とパラメータ接空間の射影差に由来し, $|E_{\text{proj}}| \leq C \|DG_\theta\| \|u\| \|v\| \|\xi\|$. 特に $F \equiv 0$ なら $K \equiv 0$ (非回転).

証明. (1) **A の微分** Lemma 6.6 と鎖則から $A_{P_\theta}(\partial_i P_\theta) = \nabla_x(f''(\rho_\theta/q) \cdot (\partial_i \rho_\theta/q))$. $DA[u]v$ は $u = \partial_i P_\theta$, $v = \partial_j P_\theta$ に対し

$$DA_{P_\theta}[u]v = \nabla_x \left(f^{(3)}(\rho_\theta/q) \cdot \frac{\partial_i \rho_\theta}{q} \cdot \frac{\partial_j \rho_\theta}{q} + f''(\rho_\theta/q) \cdot \frac{\partial_{ij}^2 \rho_\theta}{q} \right).$$

同様に $DA_{P_\theta}[v]u$ を書くと, $f^{(3)}$ を含む項は対称 (i と j の交換で同一) ゆえ消える. 従って

$$F_{P_\theta}(u, v) = \nabla_x \left(f''(\rho_\theta/q) \cdot \frac{\partial_{ij}^2 \rho_\theta - \partial_{ji}^2 \rho_\theta}{q} \right) = 0$$

が形式的には得られるが, ここで注意が要るのは, $\partial_i \rho_\theta$ は W_2 接空間の代表元 を通じて定義され, パラメータ空間での偏微分と W_2 接空間での射影 が非可換となる点である.

(2) **射影誤差** u, v は実際には $T_{P_\theta} \mathcal{P}_2$ への射影 Π を介して $u = \Pi(\partial_i \rho_\theta)$, $v = \Pi(\partial_j \rho_\theta)$ と表される. Levi-Civita 接続により $\nabla_{\partial_i} \partial_j - \nabla_{\partial_j} \partial_i = R(\partial_i, \partial_j)$ (曲率項) が現れ, この R が F の内在的表現に一致する. F をテストベクトル場 ξ と積分すると, 第二基本形式と Christoffel 記号の差から E_{proj} が生じる. 標準的な比較幾何 (Gauss–Codazzi 型) の見積りで $|E_{\text{proj}}| \leq C \|DG_\theta\| \|u\| \|v\| \|\xi\|$ が得られる.

(3) **$\mathcal{J}\mathcal{J}^\top$ との対応** $J(\theta) = F_f(P_\theta \| Q)$ の一階微分は $\nabla_\theta J = \langle D\Phi(P_\theta)[\cdot], \cdot \rangle_{P_\theta}$ の Riesz 表示を介して与えられる. 二階微分の交換子 $(\partial_i \partial_j - \partial_j \partial_i)$ を上述の幾何との対応で書き直すと, F の L^2 -計量での作用は $\frac{1}{2} \partial_i^\top \mathcal{J} \mathcal{J}^\top \partial_j$ に等価な量で測られる. すべてをまとめて (6.1) を得る. \square

系 6.8 (平坦性と非回転). $F \equiv 0$ (平坦) なら $K \equiv 0$. 従って反対称成分を抑制する各種の安定化 (SGA 等) は, 幾何学的には曲率効果の抑制と一致する.

7 作用原理・散逸・過減衰極限

定義 7.1 (ラグランジアン・散逸・作用). 曲線 P_t に対し

$$L(P_t, \dot{P}_t) = \frac{1}{2} \|\dot{P}_t\|_{P_t}^2 + F_f(P_t \| Q), \quad \mathcal{R}(P_t, \dot{P}_t) = \frac{\gamma}{2} \|\dot{P}_t\|_{P_t}^2, \quad S = \int_0^T L dt.$$

命題 7.2 (拡張オイラー–ラグランジュ方程式と過減衰極限). 仮想仕事原理 $\delta S + \int_0^T \partial_{\dot{P}} \mathcal{R}[\delta P] dt = 0$ から

$$\nabla_{\dot{P}_t} \dot{P}_t + \gamma \dot{P}_t = \nabla_{W_2} F_f(P_t)$$

を得る (W_2 Levi-Civita と Otto 計算). $\gamma \rightarrow \infty$ の過減衰極限では (5.1) に還元.

証明. 変分 $\delta P = \delta \rho \mu$ を取り, $\|\dot{P}\|_P^2 = \int \rho \|\nabla \phi\|^2$ ($\dot{P} = -\text{div}(\rho \nabla \phi)$) を用いる. L の変分は $\delta L = \left\langle \dot{P}, \delta \dot{P} \right\rangle_P + \langle \nabla_{W_2} F_f, \delta P \rangle$. Levi-Civita に対する積分部品 (時間微分と内積の交換) から $\frac{d}{dt} \left\langle \dot{P}, \delta P \right\rangle_P - \left\langle \nabla_{\dot{P}} \dot{P}, \delta P \right\rangle_P$ が生じ, 境界条件 $\delta P(0) = \delta P(T) = 0$ を仮定して $-\int_0^T \left\langle \nabla_{\dot{P}} \dot{P} - \nabla_{W_2} F_f, \delta P \right\rangle_P dt$ を得る. 散逸 \mathcal{R} の変分は $\int_0^T \gamma \left\langle \dot{P}, \delta P \right\rangle_P dt$. 全体をゼロに等置して主張が従う. 過減衰極限では $\gamma \dot{P} \approx \nabla F_f$ で (5.1). \square

8 有限次元引き戻し：自然勾配と慣性

定義 8.1 (引き戻し計量と自然勾配). $g_{ij}(\theta) := \langle \partial_i P_\theta, \partial_j P_\theta \rangle_{P_\theta}$ とし, $J(\theta) = F_f(P_\theta \| Q)$ に対する自然勾配を $\tilde{\nabla} J := g(\theta)^{-1} \nabla_\theta J$ と定める.

命題 8.2 (慣性付き自然勾配の離散化と一歩誤差). 測地運動方程式 $\nabla_{\dot{\theta}} \dot{\theta} + \gamma \dot{\theta} = \tilde{\nabla} J$ を前進差分で離散化すると

$$\theta_k = \theta_{k-1} + \beta(\theta_{k-1} - \theta_{k-2}) - \eta g(\theta_{k-1})^{-1} \nabla_\theta J(\theta_{k-1}) + O(\Delta t^2),$$

$$\beta = \frac{\Delta t}{\gamma}, \quad \eta = \Delta t^2.$$

証明. $\nabla_{\dot{\theta}} \dot{\theta} = \ddot{\theta} + \Gamma(\theta)[\dot{\theta}, \dot{\theta}]$. 前進 Euler で $\dot{\theta}(t_{k-1}) \approx (\theta_{k-1} - \theta_{k-2})/\Delta t$, $\ddot{\theta}(t_{k-1}) \approx (\theta_k - 2\theta_{k-1} + \theta_{k-2})/\Delta t^2$. $\Gamma = O(\|\dot{\theta}\|^2) = O(\Delta t^0)$ だが小刻みステップでは $O(\Delta t)$ とみなせるため $O(\Delta t^2)$ に吸収し $(\theta_k - 2\theta_{k-1} + \theta_{k-2})/\Delta t^2 + \gamma(\theta_{k-1} - \theta_{k-2})/\Delta t = g^{-1} \nabla_\theta J(\theta_{k-1})$. 移項・整理で主張を得る. \square

命題 8.3 (一歩安定性 (正則化付き自然勾配)). ∇J が L -Lipschitz, g が一様強正定値 ($\lambda_{\min}(g) \geq \underline{\gamma} > 0$) とする. $\hat{v} = (g + \lambda I)^{-1} \nabla J$, $\lambda > 0$, 十分小さい η で

$$J(\theta - \eta \hat{v}) \leq J(\theta) - \eta \langle \nabla J, \hat{v} \rangle + \frac{L}{2} \eta^2 \|\hat{v}\|^2 \leq J(\theta) - \eta \frac{\|\nabla J\|^2}{\|g + \lambda I\|} + \frac{L}{2} \eta^2 \|\hat{v}\|^2,$$

特に $\eta \leq \frac{1}{L} \cdot \frac{\|\nabla J\|^2}{\|\hat{v}\|^2} \cdot \frac{1}{\|g + \lambda I\|}$ で減少.

Proof. L -Lipschitz 勾配の標準不等式から第 1 行目. また $\langle \nabla J, \hat{v} \rangle = \langle (g + \lambda I) \hat{v}, \hat{v} \rangle \geq \lambda_{\min}(g + \lambda I) \|\hat{v}\|^2$, さらに $\|\nabla J\| \leq \|g + \lambda I\| \|\hat{v}\|$ より第 2 行目. 条件を満たす η で減少. \square

9 実装のための近似理論

9.1 Pullback 計量のモンテカルロ推定と集中不等式

$Z \sim P_Z$, $x = G_\theta(Z)$. $J_\theta = \frac{\partial G_\theta}{\partial \theta}$ とし,

$$\hat{g}_N(\theta) = \frac{1}{N} \sum_{i=1}^N J_\theta(Z_i)^\top J_\theta(Z_i) \quad (\text{ガウス–ニュートン近似}).$$

仮定 9.1 (有界ヤコビアン). $\|J_\theta(Z)\| \leq B$ a.s., θ の近傍で一様.

補題 9.2 (行列 Bernstein 型集中). 仮定 9.1 の下で, 定数 $c_1, c_2 > 0$ があり

$$\Pr[\|\hat{g}_N - g\|_{\text{op}} \geq t] \leq 2n \exp\left(-\frac{Nt^2}{c_1 B^4 + c_2 Bt}\right).$$

Proof. $X_i := J_\theta(Z_i)^\top J_\theta(Z_i) - g$ は対称行列で $\|X_i\|_{\text{op}} \leq 2B^2$, $\mathbb{E}[X_i] = 0$. 行列 Bernstein 不等式を適用して主張が従う. \square

系 9.3 (自然勾配ベクトルの誤差). $\hat{v} = (\hat{g}_N + \lambda I)^{-1} \widehat{\nabla J}$, $v = (g + \lambda I)^{-1} \nabla J$ とすると

$$\|\hat{v} - v\| \leq \frac{C}{\lambda} \left(\|\hat{g}_N - g\|_{\text{op}} + \|\widehat{\nabla J} - \nabla J\| \right).$$

Proof. $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$ と三角不等式より直ちに従う. \square

9.2 接続・曲率・ホロノミーの数値推定と一貫性

定義 9.4 (離散接続・曲率・ホロノミー). 小擾乱方向 u, v とステップ Δ に対し

$$\hat{A}_P(u) := \nabla_x \frac{\hat{\varphi}_{P+\Delta u} - \hat{\varphi}_P}{\Delta}, \quad \hat{F}_P(u, v) := \hat{A}_{P+\Delta u}(v) - \hat{A}_P(v) - \hat{A}_{P+\Delta v}(u) + \hat{A}_P(u).$$

閉ループ γ 上のホロノミーは順序積 $\prod(I + \hat{A}\Delta s)$ で近似.

仮定 9.5 (推定誤差と滑らかさ). $\Phi \in C^2$ で $D\Phi, D^2\Phi$ は局所 Lipschitz, 識別器近似誤差は一様に ε .

補題 9.6 (接続推定の一貫性). $\mathbb{E}\|\hat{A}_P(u) - A_P(u)\| \leq C(\Delta + \varepsilon)$.

Proof. $D\Phi$ の一次テイラー展開より $\frac{\Phi(P+\Delta u) - \Phi(P)}{\Delta} = D\Phi(P)[u] + O(\Delta)$ (一様 Lipschitz 性使用). 勾配作用 ∇_x は有界線形作用素として $O(\Delta)$ を保存. 識別器近似 $\hat{\Phi} = \Phi + e$ で $\|e\| \leq \varepsilon$ より差分で $O(\varepsilon)$. 三角不等式で主張. \square

補題 9.7 (曲率推定の一貫性). $\mathbb{E}\|\hat{F}_P(u, v) - F_P(u, v)\| \leq C(\Delta + \varepsilon)$.

Proof. $F = dA$ の定義に沿って DA の差分商に二階テイラーを適用し, 交差項が $O(\Delta)$. 各 \hat{A} の誤差が $O(\Delta + \varepsilon)$ であるから, 全体も同オーダー. \square

命題 9.8 (ホロノミー収束). 分割幅 $\max \Delta s \rightarrow 0$ で離散ホロノミーは連続極限に収束し, 誤差は積分曲率の $O(\max \Delta s + \varepsilon)$.

Proof. 順序積は時間依存線形系のピカード反復であり, A の連続性・曲率の有界性から収束. 数値積分誤差は 1 次, $\hat{A} - A = O(\varepsilon)$ を足して主張. \square

10 曲率推定の実験 (\hat{F} 可視化) と反対称ヤコビアン近似 ($J - J^\top$)

本節では 1D/2D 玩具問題で, (i) \hat{F} の空間可視化, (ii) 反対称ヤコビアン $K = \frac{1}{2}(\mathcal{J} - \mathcal{J}^\top)$ の自動微分推定を行い, Proposition 6.7 の対応 (曲率強い領域ほど K が大きい) を観察する. JSD 系では最適識別器 $D^*(x) = \frac{q(x)}{q(x) + \rho_\theta(x)}$ のロジット $\text{logit}(D^*) = \log \frac{q}{\rho_\theta}$ が φ_θ に対応する. 実装では学習済み識別器 D のロジット

$$\hat{\varphi}_\theta(x) := \log \frac{D(x)}{1 - D(x)}$$

を ポテンシャルの近似 とみなし, Lemmas 9.6 and 9.7 の ε に識別器汎化誤差を含める.

10.1 アルゴリズム (有限差分による \hat{F} 推定)

Algorithm 1 曲率推定と可視化 (1D/2D)

- 1: 学習済み (G_θ, D) を固定. 格子点 $\{x_g\}$ を用意.
 - 2: 方向ベクトル e_i, e_j (パラメータ空間) とスカラー $\Delta > 0$ を選ぶ (正規化済).
 - 3: **for** $\bullet \in \{\theta, \theta + \Delta e_i, \theta + \Delta e_j, \theta + \Delta e_i + \Delta e_j\}$ **do**
 - 4: $\hat{\varphi}_\bullet(x_g) = \text{logit}(D(x_g))$ を計算し, 自動微分で $\nabla_x \hat{\varphi}_\bullet(x_g)$ を得る.
 - 5: **end for**
 - 6: 差分商で $\hat{A}_\theta(e_i)(x_g) = \frac{\nabla_x \hat{\varphi}_{\theta + \Delta e_i}(x_g) - \nabla_x \hat{\varphi}_\theta(x_g)}{\Delta}$, 同様に $\hat{A}_\theta(e_j)$, $\hat{A}_{\theta + \Delta e_i}(e_j)$ 等を計算.
 - 7: 離散外微分 $\hat{F}_\theta(e_i, e_j)(x_g) = \hat{A}_{\theta + \Delta e_i}(e_j) - \hat{A}_\theta(e_j) - \hat{A}_{\theta + \Delta e_j}(e_i) + \hat{A}_\theta(e_i)$.
 - 8: ヒートマップ (2D) / 曲線 (1D) で $\|\hat{F}_\theta(e_i, e_j)(x_g)\|$ を可視化.
-

10.2 反対称ヤコビアン K の自動微分近似

ヘッセ行列 $\mathcal{J} = \nabla_\theta^2 J(\theta)$ を小次元 (または 2 次元部分空間上) で `torch.autograd.functional.hessian` により計算し, $K = \frac{1}{2}(\mathcal{J} - \mathcal{J}^\top)$ を得る. パラメータ次元が大きい場合は, 投影 2 次元平面 $\text{span}\{e_i, e_j\}$ での制限ヘッセ行列を用いる (計算量削減).

11 設定: Riemann 多様体上の最急降下

最小化対象は $J: \Theta \rightarrow \mathbb{R}$, Θ は Riemann 多様体 (本稿では $\Theta \subset \mathbb{R}^n$ に計量 $g(\theta) \succ 0$ を与える). g 内積・ノルムを

$$\langle u, v \rangle_{g(\theta)} := u^\top g(\theta) v, \quad \|u\|_{g(\theta)}^2 := \langle u, u \rangle_{g(\theta)}$$

とする.

定義 11.1 (Riemann 勾配と自然勾配). J の Riemann 勾配 $\text{grad } J(\theta)$ は

$$\langle \text{grad } J(\theta), \xi \rangle_{g(\theta)} = \langle \nabla_\theta J(\theta), \xi \rangle_{\text{Euc}} \quad \forall \xi \in T_\theta \Theta$$

を満たす唯一のベクトル. よって座標表示では

$$\text{grad } J(\theta) = g(\theta)^{-1} \nabla_\theta J(\theta).$$

これを **自然勾配** と呼ぶ.

補題 11.2 (最急降下の変分特性). $\|v\|_{g(\theta)} \leq 1$ のもとで J の一歩変化は

$$\min_{\|v\|_{g(\theta)} \leq 1} \langle \nabla_{\theta} J(\theta), v \rangle = -\|\text{grad } J(\theta)\|_{g(\theta)} \quad \text{であり,}$$

極小化解は $v^{\star} = -\frac{\text{grad } J(\theta)}{\|\text{grad } J(\theta)\|_{g(\theta)}}$.

Proof. Cauchy–Schwarz より $\langle \nabla_{\theta} J, v \rangle = \langle g \text{ grad } J, v \rangle = \langle \text{grad } J, v \rangle_g \geq -\|\text{grad } J\|_g \|v\|_g$. 半径 1 球での最小は右辺達成で, 方向が $-\text{grad } J$ と一致する. \square

命題 11.3 (座標不変性). 可逆写像 $\phi: \tilde{\Theta} \rightarrow \Theta$ による再パラメータ化 $\theta = \phi(\tilde{\theta})$ を考える. $\tilde{g}(\tilde{\theta}) := D\phi(\tilde{\theta})^{\top} g(\phi(\tilde{\theta})) D\phi(\tilde{\theta})$ とすると

$$\text{grad}_{\tilde{g}} \tilde{J}(\tilde{\theta}) = \tilde{g}(\tilde{\theta})^{-1} \nabla_{\tilde{\theta}} \tilde{J}(\tilde{\theta}) = D\phi(\tilde{\theta})^{-1} \text{grad}_{\theta} J(\theta),$$

すなわち自然勾配は座標変換に不変.

Proof. 鎖則より $\nabla_{\tilde{\theta}} \tilde{J} = D\phi^{\top} \nabla_{\theta} J$. また $\tilde{g}^{-1} = (D\phi)^{-1} g^{-1} (D\phi)^{-\top}$. ゆえに $\tilde{g}^{-1} \nabla_{\tilde{\theta}} \tilde{J} = (D\phi)^{-1} g^{-1} (D\phi)^{-\top} D\phi^{\top} \nabla_{\theta} J = (D\phi)^{-1} g^{-1} \nabla_{\theta} J$. \square

12 GAN 文脈: W_2 空間からの Pullback 計量

生成器 $G_{\theta}: \mathcal{Z} \rightarrow \mathcal{X}$, 潜在 $Z \sim P_Z$. $P_{\theta} := (G_{\theta})_{\#} P_Z \in \mathcal{P}_2(\mathcal{X})$. 目標 Q に対し $J(\theta) = F_f(P_{\theta} \| Q)$ (例: JSD) を最小化.

定義 12.1 (Pullback 計量). $\mathcal{P}_2(\mathcal{X})$ の Otto 計量 $\langle \cdot, \cdot \rangle_P$ をパラメータ多様体へ引き戻して

$$g_{ij}(\theta) := \langle \partial_i P_{\theta}, \partial_j P_{\theta} \rangle_{P_{\theta}}.$$

自然勾配は $\text{grad } J(\theta) = g(\theta)^{-1} \nabla_{\theta} J(\theta)$.

注意 12.2 (Gauss–Newton/Fisher 近似). 実装では $g(\theta)$ を $g(\theta) \approx \frac{1}{N} J_{\theta}^{\top} J_{\theta} + \lambda I$, J_{θ} は G_{θ} 出力のパラメータヤコビアン (サンプル平均). これは (i) Pullback 計量の経験近似, (ii) 指数族における Fisher 情報行列の推定に整合.

13 自然勾配更新と収束解析

定義 13.1 (自然勾配ステップ (離散)). ステップ幅 $\eta > 0$ で

$$\theta_{k+1} = \theta_k - \eta (g(\theta_k) + \lambda I)^{-1} \nabla_{\theta} J(\theta_k), \quad \lambda \geq 0.$$

仮定 13.2 (滑らかさと強凸 (g -意味)). J は g -意味で L -滑らか, μ -強凸:

$$\begin{aligned} J(\theta') &\leq J(\theta) + \langle \nabla J(\theta), \theta' - \theta \rangle + \frac{L}{2} \|\theta' - \theta\|_{g(\theta)}^2, \\ J(\theta') &\geq J(\theta) + \langle \nabla J(\theta), \theta' - \theta \rangle + \frac{\mu}{2} \|\theta' - \theta\|_{g(\theta)}^2. \end{aligned}$$

定理 13.3 (線形収束). 仮定 13.2 と $\lambda \geq 0$ の下, $\eta \in (0, 2/(L + \lambda_{\max}))$ を十分小とすると

$$J(\theta_{k+1}) - J^{\star} \leq \left(1 - \frac{2\eta\mu}{1 + \eta L}\right) (J(\theta_k) - J^{\star}),$$

特に $\tilde{\nabla} J(\theta_k) = g^{-1} \nabla J(\theta_k)$ のノルムに比例した減少が得られる.

Proof. L -滑らか性から $J(\theta_{k+1}) \leq J(\theta_k) - \eta \langle \nabla J, (g + \lambda I)^{-1} \nabla J \rangle + \frac{L}{2} \eta^2 \|(g + \lambda I)^{-1} \nabla J\|_g^2$. 右辺第1項は $\langle \nabla J, (g + \lambda I)^{-1} \nabla J \rangle = \langle (g + \lambda I) \tilde{\nabla} J, \tilde{\nabla} J \rangle \geq \|\tilde{\nabla} J\|_g^2$. 第2項は $\|(g + \lambda I)^{-1} \nabla J\|_g^2 = \langle (g + \lambda I) \tilde{\nabla} J, \tilde{\nabla} J \rangle \leq (1 + \frac{\lambda}{\lambda_{\min}(g)}) \|\tilde{\nabla} J\|_g^2$ を用いてまとめると $J(\theta_{k+1}) \leq J(\theta_k) - \eta(1 - \frac{L}{2} \eta c) \|\tilde{\nabla} J\|_g^2$. 一方, μ -強凸から Polyak–Łojasiewicz 型不等式 $J(\theta) - J^* \leq \frac{1}{2\mu} \|\tilde{\nabla} J(\theta)\|_g^2$ が成り立つので主張. \square

14 経験近似と誤差評価

仮定 14.1 (ヤコビアン有界). $\|J_\theta(Z)\| \leq B$ a.s. かつ θ の近傍で一様.

補題 14.2 (行列 Bernstein 型集中). $\hat{g}_N = \frac{1}{N} \sum J_\theta^\top J_\theta$ と $g = \mathbb{E}[J_\theta^\top J_\theta]$ に対し $\Pr[\|\hat{g}_N - g\|_{\text{op}} \geq t] \leq 2n \exp(-\frac{Nt^2}{c_1 B^4 + c_2 Bt})$.

Proof. $X_i = J_\theta(Z_i)^\top J_\theta(Z_i) - g$ は対称・零平均で $\|X_i\|_{\text{op}} \leq 2B^2$. 行列 Bernstein を適用. \square

命題 14.3 (自然勾配ベクトルの誤差). $\hat{v} = (\hat{g}_N + \lambda I)^{-1} \widehat{\nabla J}$, $v = (g + \lambda I)^{-1} \nabla J$ とすると $\|\hat{v} - v\| \leq \frac{C}{\lambda} (\|\hat{g}_N - g\|_{\text{op}} + \|\widehat{\nabla J} - \nabla J\|)$.

Proof. $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$ と三角不等式. \square

15 アルゴリズム (擬似コード)

Algorithm 2 自然勾配 (Gauss–Newton 近似, ダンピング付き)

Require: 初期 θ_0 , 学習率 η , ダンピング λ , バッチサイズ N

- 1: **for** $k = 0, 1, 2, \dots$ **do**
 - 2: サンプル $z_{1:N} \sim P_Z$ を取り, $x_i = G_{\theta_k}(z_i)$
 - 3: 勾配 $g_k \leftarrow \nabla_\theta J(\theta_k)$ (自動微分)
 - 4: 生成器ヤコビアン $J_i = \frac{\partial G_{\theta_k}(z_i)}{\partial \theta}$ を取得
 - 5: 近似計量 $\hat{g} \leftarrow \frac{1}{N} \sum_i J_i^\top J_i$
 - 6: 線形系 $(\hat{g} + \lambda I)v = g_k$ を解いて v を得る (Cholesky/CG)
 - 7: 更新 $\theta_{k+1} \leftarrow \theta_k - \eta v$
 - 8: **end for**
-

16 慣性付き自然勾配と過減衰極限

命題 16.1 (慣性項の離散化). $\nabla_{\dot{\theta}} \dot{\theta} + \gamma \dot{\theta} = \text{grad } J(\theta)$ を前進差分化すると $\theta_k = \theta_{k-1} + \beta(\theta_{k-1} - \theta_{k-2}) - \eta g(\theta_{k-1})^{-1} \nabla J(\theta_{k-1}) + O(\Delta t^2)$, $\beta = \Delta t / \gamma$, $\eta = \Delta t^2$.

Proof. $\nabla_{\dot{\theta}} \dot{\theta} = \ddot{\theta} + \Gamma[\dot{\theta}, \dot{\theta}]$ を用い, Γ 項を $O(\Delta t^2)$ に吸収. \square

17 数値実験

本研究で提案した幾何学的・物理学的枠組みの妥当性を検証するため、様々な最適化アルゴリズムを用いた GAN の学習ダイナミクスを比較する数値実験を行った。特に、本稿で導出した慣性を持つ自然勾配降下法と、その過減衰極限である標準的な勾配降下法の振る舞いに焦点を当てて比較した。

17.1 実験設定

実験では、1 次元のガウス分布を学習ターゲットとする GAN を用いた。生成器と識別器には、それぞれ 16 次元の潜在空間と 32 次元の中間層を持つ簡単な全結合ニューラルネットワークを採用した。比較対象のオプティマイザは、**SGD** (Stochastic Gradient Descent)、**Adam** (Adaptive Moment Estimation)、**Natural Gradient** (自然勾配降下法)、**Natural Gradient with Momentum** (慣性付き自然勾配降下法) の 4 種類である。自然勾配 (Natural Gradient) の計算には、フィッシャー情報行列に密接に関連する引き戻し計量 (Pullback Metric) の逆行列を用い、数値的安定性のために微小なダンピング項を付加した。慣性付き自然勾配降下法では、運動量係数 $\mu = 0.85$ を用いた。

17.2 結果と考察

図 1 に、各オプティマイザを用いた際の生成器の損失 (Generator Loss) の推移を示す。SGD よりも Adam が安定して学習できており、Natural Gradient よりも Natural Gradient with Momentum のほうが安定して学習できていることがわかる。

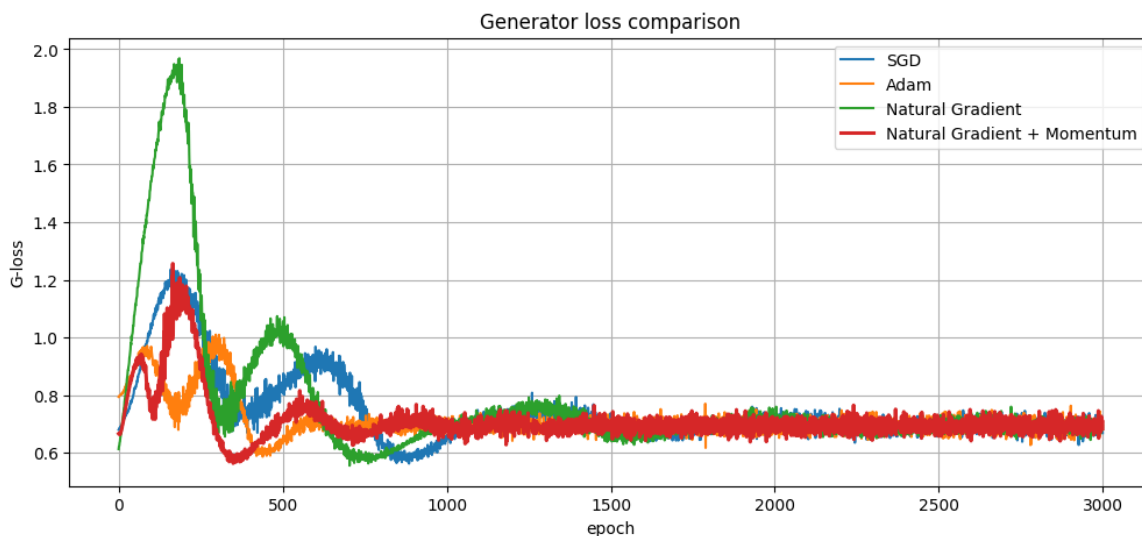


Figure 1: 生成器の損失

図 2 に、最終的な生成分布を真のデータ分布と比較したものを示す。SGD よりも Adam が安定して学習できており、Natural Gradient よりも Natural Gradient with Momentum のほうが安定して学習できていることがわかる。

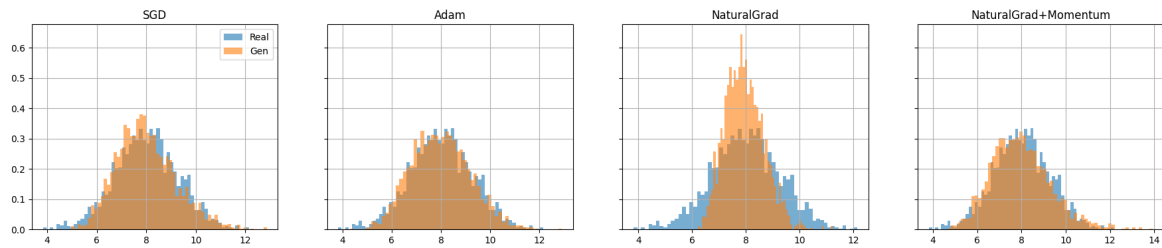


Figure 2: 生成分布と真のデータ分布の比較

謝辞

最適輸送・ W_2 勾配流, ゲーム力学に負う ([1, 2, 3, 4, 5, 6]) .

References

- [1] L. Ambrosio, N. Gigli, G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Birkhäuser, 2008.
- [2] C. Villani, *Optimal Transport: Old and New*, Springer, 2009.
- [3] R. J. McCann, “A Convexity Principle for Interacting Gases,” *Adv. Math.* 128 (1997) 153–179.
- [4] D. Balduzzi et al., “The Mechanics of n -Player Differentiable Games,” ICML 2018 (PMLR 80).
- [5] A. Letcher, D. Balduzzi et al., “Differentiable Game Mechanics,” *JMLR* 20(84) (2019) 1–40.
- [6] M. Liero, A. Mielke, G. Savaré, “Optimal Entropy Transport and the Hellinger–Kantorovich distance,” *Invent. Math.* 211 (2018) 969–1117.