Solution Manifolds

Jiyoung Park

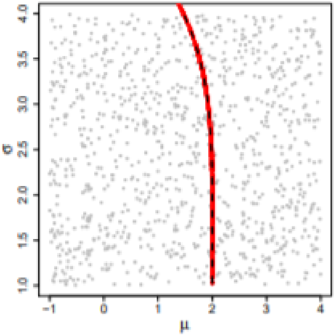
Statistics Ph.D. Student, Texas A&M University, College Station, TX

October 10, 2025

Section 1

Simple Example

- Question: Find parameters (μ, σ^2) for $Y \sim N(\mu, \sigma^2)$ satisfying $P(-5 < Y < 2) = 0.5$
- Total parameter space: $\mathbb{R} \times \mathbb{R}^+$
- Possible parameter space w.r.t. constraint: 1-dimensional curve (figure)
- Idea: Space of solutions forms a manifold (Solution Manifold)



The goal of the paper

- What is the property of solution manifold?
- How to find a Solution Manifold numerically?
- Concept of Solution Manifold applied to actual problems?

Notations

- $\Psi : \mathbb{R}^d \to \mathbb{R}^s$ Target function.
- $M_\Psi := \{x \mid \Psi(x) = 0\}$ Solution Manifold given Ψ. Often denoted just M.
- $\widehat{M} := M_{\widehat{\Psi}}$ An approximation (estimation) of M, if $\widehat{\Psi}$ is an approximation (estimation) of Ψ.
- $||\Psi||_\infty^J = \sup_x \max_i \max_{j_1,\dots j_J} |\frac{\partial^J \Psi_i(x)}{\partial x_{j_1},\dots \partial x j_J}|$, $||\Psi||_{\infty,J}^* = \max_{j=0,\dots J} ||\Psi||_\infty^j$
- $G_\Psi(x) = \nabla\Psi(x)$, $H_\Psi(x) = \nabla^2\Psi(x)$.
- $M \oplus r := \{x \mid d(x, M) \leq r\}$

Flow of the presentation

- **Properties of Solution Manifold**: lemma 1 $\rightarrow$ Assumptions, lemma 2 $\rightarrow$ Smoothness Thm (Thm 3) $\rightarrow$ Stability Thm (Thm 4)
- **Numerically obtaining Solution Manifold**: Conv in grad flow (Thm 6) $\rightarrow$ Conv in GD (Thm 8) $\rightarrow$ MCGD algorithm (Algo 1)
- **Statistical Applications of Solution Manifold**: Theorem 7 $\rightarrow$ pushforward measure on M $\rightarrow$ Bayesian Prior and Posterior sampling from Solution Manifold (Algo 3)

Main results

- Properties of solution manifold
  - **Smoothness theorem**: Smoothness of $\Psi \Rightarrow$ Smoothness(?) of M_Ψ. (Lem 1 and Thm 3).
  - **Stability theorem**: $d(\widehat{\Psi}, \Psi) \to 0 \Rightarrow \widehat{M} \to M$ w.r.t. the Hausdorff distance (Thm 4).
- Numerically finding Solution Manifold
  - **Convergence of a gradient descent algorithm**: With good initialization and proper update step, Gradient Descent algorithm can obtain points in M_Ψ. (Thm 6, Thm 8)
  - **Monte Carlo gradient descent algorithm**: Generating point clouds over M, using only Ψ and its gradient. (Algorithm 1)
- Statistical Applications of Solution Manifold
  - **Local center manifold theorem**: For $z \in M$, $A(z) := \{x \mid x \text{ converges to } z\}$ forms an s-dimensional manifold. (Thm 7).
  - **Approximated manifold posterior algorithm**: A Bayesian procedure that approximates the posterior distribution on a manifold. (Algorithm 3)

Section 2

Assumptions

- **Note 1** $\forall x \in M$, rank$(\nabla\Psi(x)) = s \Rightarrow M$ forms a $(d-s)$ dimensional manifold. ($\because$ Implicit function theorem)
- **Assumptions**
  - **F1** $\exists\, ||\Psi||^*_{\infty,3} < \infty$
  - **F2** $\exists\, \lambda_0, \delta_0, c_0 > 0$ s.t.
    - $\forall x \in M \oplus \delta_0$, $\lambda_{\min}(G_\Psi(x)G_\Psi(x)^T) = \lambda_{\min>0}(G_\Psi(x)^T G_\Psi(x)) > \lambda_0^2$.
    - $\forall x \notin M \oplus \delta_0$, $||\Psi(x)||_{\max} > c_0$.
  - F1 is related with smoothness (in fact, differentiability) of Ψ.
  - F2 is related with curvature information around M. ($\because$ lemma 1)
- **Lemma 1** $\forall x \in M$, Span(Row($G_\Psi(x)$)) = Normal($M(x)$).

Assumptions: Examples

- Are these assumptions reasonable?
- Rephrase of F2 in specific problems
  - mode estimation: $\Psi = \nabla p(x) \Rightarrow$ F2 $\approx$ pdf: Morse function.
  - Solving MLE: $\Psi = \nabla l(\theta) \Rightarrow$ F2 $=$ positive definite fisher information.

Smoothness Theorem

- In general, Smoothness of $\Psi \not\Rightarrow$ smoothness of M,
  but under above assumptions, Smoothness of $\Psi \Rightarrow$ positive reach.
- **Definition** Reach
  $\text{Reach}(M) \ := \ \inf\{r > 0 \,|\, \forall x \in M \oplus r,\ x$ has a unique projection onto M$\}$
- **Smoothness Theorem** Under above assumptions,
  $\text{reach}(M) \geq \min(\frac{\delta_0}{2}, \frac{\lambda_0}{||\Psi||^*_{\infty,2}})$
- **Remark 1** $\frac{\delta_0}{2}$ is related with folding, and $\frac{\lambda_0}{||\Psi||^*_{\infty,2}}$ is related with curvature.
- **Note** Relationship with Smoothness of Manifold?
  - Positive reach $\Rightarrow C^{1,1}$ manifold [A. Lytchak, 2005]
  - $C^{k,\alpha}$ manifold: Transition maps are $C^{k,\alpha}$ Hölder continuous.
- **Note** Why positive reach matters?
  - Positive reach set and convex set shares some properties (e.g. Steiner formula)
    $\Rightarrow$ Positive reach can substitute convexity condition in some cases.
    ([A. Cuevas, 2012])

Stability Theorem

- **Stability Theorem** Ψ satisfies above assumptions. $\widehat{\Psi}$ is bounded 2-differentiable. If $||\widehat{\Psi} - \Psi||^*_{\infty,2}$ is sufficiently small, then
    - $\widehat{\Psi}$ also satisfies F2.
    - $d_{\text{Hausdorff}}(\widehat{M}, M) = O(||\widehat{\Psi} - \Psi||^0_\infty)$
    - reach$(\widehat{M}) \geq \min(\frac{\delta_0}{2}, \frac{\lambda_0}{||\Psi||^*_{\infty,2}}) + O(||\widehat{\Psi} - \Psi||^*_{\infty,2})$
- Meaning: Using $\widehat{\Psi}$ instead of Ψ does not ruin theoretical guarantees.

Section 3

Goal

- How to obtain points in M, when we cannot directly solve $\Psi(x) = 0$?
- Goal: Use gradient descent.
  - Will GD find points in M?
  - If it does, what would be the condition for GD to be well-behave?

Reformulazation of Problem: Gradient flow

- Take any $\Lambda \in \mathbb{R}^{s\times s}$: Positive definite matrix. (e.g. $\Lambda = I$).
  Let $\Lambda_{\max}, \Lambda_{\min}$ largest, smallest eigenvalues of Λ.
- Define $f(x) := \Psi(x)^t \Lambda \Psi(x)$
- Then, $M_\Psi = M_f$
- And, Define a gradient flow $\pi_x(t)$ as below.
  - $\pi_x(0) = x$, $\pi_x'(t) = -\nabla f(\pi_x(t))$
  - Note: This is continuous version of Gradient descent.
  - Note 2: points in M will be stationary points of this flow.
- **Lemma 5** Behavior of $H_f(x)$: Normal to $M(x)$ when $x \in M$, and well-bahave when $x \in$ Reach(M).

Convergence of Gradient flow

- Under assumptions, this gradient flow converges to M.
- **Theorem 6** Under assumptions, $\pi_x(t)$ satisfies the followings.
  - $\forall x \in M \oplus \delta_c$, $\pi_x(\infty) \in M$.
  - $\pi_x(\infty) \in M \Rightarrow \lim_{t\to\infty} \frac{\pi_x'(t)}{||\pi_x'(t)||} \perp M(\pi_x(\infty))$
  - $f(x_n) \to 0 \Rightarrow ||x_n - \pi_{x_n}(\infty)|| = O(\sqrt{f(x_n)})$.
- Notes on Theorem 6
  - 1: Convergence radius. Guarantees 'not too far' points from M converging to M.
  - 2: The arrives to M with the direction of normal vector. (i.e. does not take detour)
  - 3: Initial point and convergence point are not too far. (i.e. it does not converges to far points. It converges to close point in manifold)
  - $\therefore$ Gradient flow with suitable initial points converges to M, with efficient path.

Gradient Descent

- Gradient Flow framework $\Rightarrow$ Gradient Descent framework
  - Gradient Descent: $x_{t+1} = x_t - \gamma \nabla f(x_t)$
  - requires one more hyperparameter: the step size γ
  - γ determines whether it converges, as well as the speed of convergence.
- What should be the condition for γ for convergence and speed of convergence?

Gradient Descent: Convergence

- Condition of γ for GD to work well
- **Theorem 8** Assume above assumptions. Then,
  - Initial point condition: $\forall x \in M \oplus \delta_c$ (Same with Conv of gradient flow)
  - step size condition: $\forall \gamma < \min(\frac{1}{\Lambda_{\max}||\Psi||^*_{\infty,2}}, \delta_c)$

    $\Rightarrow x_\infty \in M$.
  - Speed of conv: In addition, if $\gamma < \frac{\Lambda_{\max}||\Psi||^*_{\infty,2}}{\lambda_0^4 \Lambda_{\min}^2}$

    $\Rightarrow f(x_t) \leq f(x_0)(1-\gamma\frac{\lambda_0^4\Lambda_{\min}^2}{\Lambda_{\max}||\Psi||^*_{\infty,2}})^t$, $d(x_t, M) \leq d(x_0, M)(1-\gamma\lambda_0^2\Lambda_{\min})^{t/2}$
- Meaning: with suitable initial points and step size, GD will converge to M.
- Need $O(log(1/\epsilon))$ iterations for ϵ error.
- first element of min: decreasing objective function.
  2nd element of min: Well-behaving hessian
- Note: GD still converge even if f: non-convex. (Did not required such properties)

Gradient Descent: Implementation

- Algorithm 1
  - hyperparameters: γ: Step size; Λ: P.D. Matrix
  - input: N: the data size, Ψ: the target function
    1. $f(x) := \Psi(x)^t \Lambda \Psi(x)$
    2. randomly choose x_0 from region of interest
    3. iterates $x_{t+1} = x_t - \gamma \nabla f(x_t)$ until converge
    4. if $\Psi(x_\infty) = 0$, keep the point. otherwise discard.
    5. Repeat until reaches N samples.
- Slight modified version of algorithm 1 in paper

Section 4

Simple Example: Univariate Gaussian by Bayesian

- Parameter of interest: $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$ for $N(\mu, \sigma^2)$
- Constraint: Second moment $= s_0^2$ for some s_0.
- $M = \{(\mu, \sigma^2) \mid s_0^2 - \mu^2 - \sigma^2 = 0\}$
- Obstacle: How to construct prior on M?
- Let Q: distribution on total space, Q_M: distribution on M.
- We want to link $Q \sim Q_M$, so that we can generate prior Q_M using Q.

Distribution on Solution Manifold

- **Definition** Given gradient flow $\pi_x(t)$ and $z \in M$, $A(z) := \{x \,|\, \pi_x(\infty) = z\}$ is called basin of attraction for z,
- Meaning: points whose gradient flow converges to z.
- **Theorem 7** Under assumptions, $A(z)$ forms s dimensional manifold for each z.

Distribution on Solution Manifold: Cont'D

- **Outcome of Thm 7**: Define Q_M to be a pushforward measure of Q, with pushforward function defined by the gradient flow.
  $\Rightarrow Q_M$ has a s dimensional Hausdorff density function.
- $\because$ Besicovitch-Marstrand-Preiss Theorem. [C. De Lellis, 2008]
  - **Rectifiability**: Set $E \subset \mathbb{R}^d$ is Rectifiable with k-dimension if $\exists$ countably many C^1 k-dimensional submanifolds of $\mathbb{R}^d$ which cover $\mathcal{H}^k$-almost all E.
  - Theorem 7 shows M is Rectifiable. ($A(z)$ will be submanifold.)
  - **BMP Theorem** If μ is a locally finite Radon measure on $\mathbb{R}^d$, k is an integer, E is a Rectifiable set with k-dimension. Then $\exists f$ s.t.
    $\mu(S) = \int_{S \cap E} f \, d\mathcal{H}^k$ for $\forall S \subset \mathbb{R}^d$
    And such f coincides with Hausdorff density if density exists.
  - let $E = M$, $\mu = Q_M$, $k = s \Rightarrow$ our conditions.
- **Meaning**: If density exists, it matches with the pushforward measure.
- $\therefore$ Sampling from Q and applying GD would result in points on Q_M, and densities of samples matches with Q_M

Bayesian Posterior Algorithm

- Algorithm 3
  - hyperparameters: h: normalizing; K: Smooth function
  - input: N: the data size, Ψ: the target function
    1. Apply Algorithm 1, with modifying step 2 (instead of randomly choose x_0, sample x_0 from π). We get $(Z_1, Z_2, ..., Z_N)$, points from π_M.
    2. Estimate density score of Z_i using $\hat{\rho}_{i,N} = \frac{1}{N}\sum_{j=1}^{N} K(\frac{||Z_i - Z_j||}{h})$
    3. $\hat{\pi}_{i,N} = \pi_(Z_i)\prod_{j=1}^{n} p(X_j|Z_i)$; $\hat{\omega}_{i,N} = \frac{\hat{\pi}_{i,N}}{\hat{\rho}_{i,N}}$
    4. return $(Z_i, \hat{\omega}_{i,N})_{i=1,...,N}$
- $\hat{\omega}_{i,N}$ contains weights of the point.
- This weights can be regarded as distribution on M due to Silde 23.

Summary

- **Assumptions**: Are they reasonable? How strong the assumptions?
- **Properties of solution manifold**: positive reach, stability
- **Finding a solution manifold**: Theoretical guarantees for using GD? How fast it converges?
- **Applications**: How to use bayesian model on solution manifold?

References

Yen-Chi Chen (2020)
Solution Manifold and Its Statistical Applications
Arxiv

Antonio Cuevas (2012)
On statistical properties of sets fulfilling rolling-type conditions
Advances in Applied Probability 44 (2), 311-329

Alex Ander Lytchak (2005)
Almost convex subsets
Geom Dedicata 115, 201–218

C. De Lellis (2008)
"Rectifiable sets, densities and tangent measures" Zurich Lectures in Advanced Mathematics.
European Mathematical Society (EMS), Zürich, 2008.

Thank You!