Escaping from saddle points on Riemannian manifolds

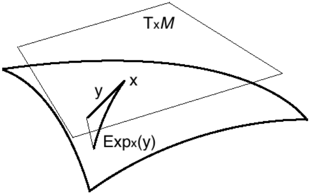
Yue Sun[†], Nicolas Flammarion[‡], Maryam Fazel[†]

[†] Department of Electrical and Computer Engineering, University of Washington, Seattle
[‡] Department of Electrical Engineering and Computer Science, University of California, Berkeley

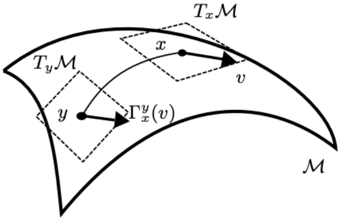
October 12, 2019

Prelim: operations on manifold

Exponential map and parallel transport.



Exponential map is a projection-like operation mapping from tangent space to manifold, where the curve from $x \to \mathrm{Exp}_x(y)$ is a geodesic with initial velocity y.



Parallel transport is an operation that translates a tangent vector from $\mathcal{T}_x\mathcal{M}$ to $\mathcal{T}_y\mathcal{M}$ along a geodesic.

Manifold constrained optimization

We consider the manifold constrained optimization problem

$$\underset{x}{\operatorname{minimize}} \quad f(x), \text{ subject to } \; x \in \mathcal{M}$$

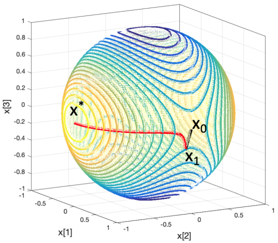
assuming the function and manifold satisfying

1. There is a finite constant β such that
$$\|\mathrm{grad} f(y) - \Gamma_x^y \mathrm{grad} f(x)\| \leq \beta d(x,y) \quad \text{for all } x, y \in \mathcal{M}.$$
2. There is a finite constant ρ such that
$$\|H(y) - \Gamma_x^y H(x) \Gamma_y^x\|_2 \leq \rho d(x,y) \quad \text{for all } x, y \in \mathcal{M}.$$
3. There is a finite constant K such that
$$|K(x)[u,v]| \leq K \quad \text{for all } x \in \mathcal{M} \text{ and } u, v \in \mathcal{T}_x\mathcal{M}^1$$

f may not be convex.

[1] $K(x)[u,v]$ denotes the curvature constant of $\mathcal{M}$ at x in direction u, v.

Algorithm



Hope to escape from saddle point and converge to an approximate local minimum.

1. At iterate x, check the norm of gradient.
2. If large: do $x^+ = \mathrm{Exp}_x(-\eta \mathrm{grad} f(x))$ to decrease function value.
3. If small: near either a saddle point or a local min. Perturb iterate by adding appropriate noise, run a few iterations.

   3.1 if f decreases, iterates escape saddle point (and alg continues).
   3.2 if f doesn't decrease: at approximate local min (alg terminates).

Theorem

Theorem (Jin et al., Eucledean space)

Perturbed GD converges to a $(\epsilon, -\sqrt{\rho\epsilon})$-stationary point of f in

$$O\left(\frac{\beta(f(x_0)-f(x^*))}{\epsilon^2}\log^4\left(\frac{\beta d(f(x_0)-f(x^*))}{\epsilon^2\delta}\right)\right)$$

iterations.

We replace Hessian Lipschitz ρ by $\hat{\rho}$ as a function of ρ and K and we quantify it in the paper.

Theorem (manifold)

Perturbed RGD converges to a $(\epsilon, -\sqrt{\hat{\rho}(\rho, K)\epsilon})$-stationary point of f in

$$O\left(\frac{\beta(f(x_0)-f(x^*))}{\epsilon^2}\log^4\left(\frac{\beta d(f(x_0)-f(x^*))}{\epsilon^2\delta}\right)\right)$$

iterations.

Experiment

Burer-Monteiro facotorization.

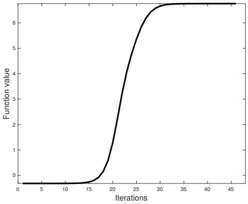
Let $A \in \mathbb{S}^{d\times d}$, the problem

$$\max_{X\in\mathbb{S}^{d\times d}} \operatorname{trace}(AX),$$
$$s.t.\ \operatorname{diag}(X) = 1, X \succeq 0, \operatorname{rank}(X) \leq r.$$

can be factorized as

$$\max_{Y\in\mathbb{R}^{d\times p}} \operatorname{trace}(AYY^T),\ s.t.\ \operatorname{diag}(YY^T) = 1.$$

when $r(r+1)/2 \leq d$, $p(p+1)/2 \geq d$.



Iteration versus function value.