Non-convex smooth optimization on weakly convex manifolds: projected gradient and conditional gradient methods

Maxim Balashov, Boris Polyak, Andrey Tremba

Laboratory of Adaptive and Robust Systems
Institute of Control Sciences, Moscow

September 27, 2019

Plan of the talk[1]

Smooth optimization on manifolds:
statement and issues

1. Weakly convex sets
   - Projected gradient method
   - Polyak-Łojasiewicz-like property
2. Almost linear functions
   and Frank-Wolfe method

[1] Balashov, Polyak, Tremba, arXiv:1906.11580

Problem statement and motivation

$$\min_{x \in S} f(x)$$

- Sphere

$$S = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$$

  - direction factoring
  - eigenvector problem $f(x) = (Ax, x)$
  - trust-region methods
- Stiefel manifold $X^T X = I_m$
- Functional constraints

$$S = \{x \in \mathbb{R}^n : g_i(x) = 0,\ i = 1, \ldots, m\}$$

Main issue: non-convexity

- Unit ball is convex, unit sphere is not.
- Target function may be non-convex as well.
- Optimization on Riemannian manifolds included (no need to introduce metric or retractions* though).

Key difference:

Extremal points $\rightarrow$ stationary points: gradient belongs to normal cone(s).

Requirements

- $f \in \mathcal{C}^1$ - continuously differentiable (with Lipschitz constant L_0), its gradient $f'(\cdot)$ is also Lipschitz with constant L_1.
- 1. S - a weakly convex set with constant R.

  2. S - boundary of a strongly convex set B with constant r.

Requirements

- $f \in \mathcal{C}^1$ - continuously differentiable (with Lipschitz constant L_0), its gradient $f'(\cdot)$ is also Lipschitz with constant L_1.
- 1. S - a weakly convex set with constant R. (optionally): Ležansky-Polyak-Łojasievicz condition
  2. S - boundary of a strongly convex set B with constant r.

Weakly convex sets (proximally smooth sets)

A closed set Q is proximally smooth with constant R, if (equivalent definitions)

- Distance function $\rho(x, S) = \inf_{a \in S} \|x - a\|$ is continuously differentiable in tube $U_S(R) = \{x \in \mathbb{R}^n \mid 0 < \rho(x, S) < R\}$

Weakly convex sets (proximally smooth sets)

A closed set Q is proximally smooth with constant R, if (equivalent definitions)

- Metric projection $P_S : U_S(R) \to S$ is a single-valued continuous function,

Weakly convex sets (proximally smooth sets)

A closed set Q is proximally smooth with constant R, if (equivalent definitions)

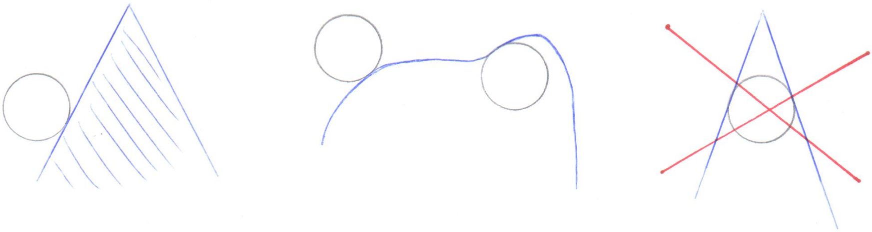
- *Supporting principle*: for all unit normals $p \in \mathcal{N}(S, x)$, $\|p\| = 1$, $S \cap B_R(x + Rp) = \emptyset$.

Weakly convex sets (proximally smooth sets)

A closed set Q is proximally smooth with constant R, if (equivalent definitions)

- Distance function $\rho(x, S) = \inf_{a \in S} \|x - a\|$ is continuously differentiable in tube $U_S(R) = \{x \in \mathbb{R}^n \mid 0 < \rho(x, S) < R\}$
- Metric projection $P_S : U_S(R) \to S$ is a single-valued continuous function,
- *Supporting principle*: for all unit normals $p \in \mathcal{N}(S, x)$, $\|p\| = 1$, $S \cap B_R(x + Rp) = \emptyset$.

Examples



Normal cone of proximal normals:

$$\mathcal{N}(S,x) = \{p \in \mathbb{R}^n : \exists \delta > 0, P_S(x+\delta p) = x\}$$

Known results

Convex functions on weakly convex sets:
linear convergence rate of projected gradient.
(Balashov, 2017)

Stationary point condition

$$\min_{x \in S} f(x)$$

$$\mathcal{N}(S, x) = \{p \in \mathbb{R}^n : \exists \delta > 0, P_S(x + \delta p) = x\}$$

Theorem (Necessary optimality condition)

If $x^* \in \arg\min_{x \in S} f(x)$, *then*

$$-f'(x^*) \in \mathcal{N}(S, x^*)$$

Stationary points: $\Omega = \{x : -f'(x) \in \mathcal{N}(S, x)\}$

Problem:

$$\min_{x \in \mathbb{R}^n} f(x)$$

Method:

$$x_{k+1} = x_k - \gamma f'(x_k)$$

Step-size:

$$0 < \gamma < \frac{2}{L_1}$$

Method: Projected gradient descent

Problem:

$$\min_{x \in S} f(x), \quad x_0 \in S, \text{ convex}$$

Method:

$$x_{k+1} = P_S(x_k - \gamma f'(x_k))$$

Step-size:

$$0 < \gamma < \frac{2}{L_1}$$

Method: Projected gradient descent

Weakly convex set

Problem:

$$\min_{x \in S} f(x), \quad x_0 \in S$$

Method:

$$x_{k+1} = P_S(x_k - \gamma f'(x_k))$$

Step-size:

$$0 < \gamma < \min\left\{\frac{1}{L_1}, \frac{R}{L_0}\right\}$$

Weak convergence of projected gradient method

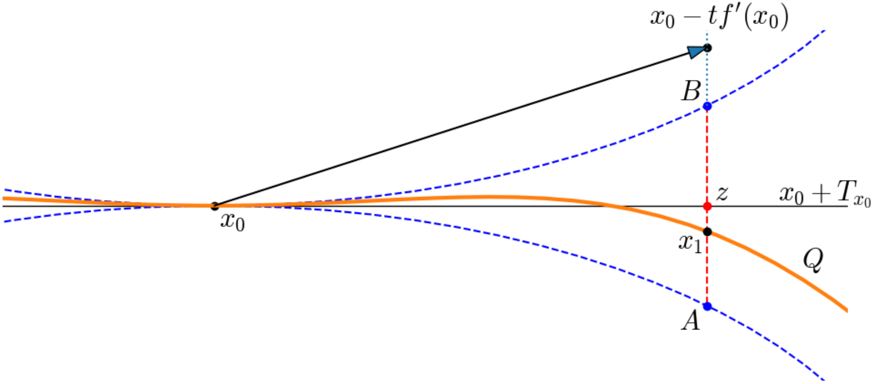
Within $N = O(\frac{1}{\varepsilon^2})$ there is a $k \leq N$:

$$\rho(-f'(x_k), \mathcal{N}(S, x_k)) < \varepsilon.$$

Algorithm without projections

Quasi-projection

$$S = Q = \{x : g(x) = 0\}, \; g : \mathbb{R}^n \to \mathbb{R}$$



Towards stronger convergence conditions

Smooth weakly convex manifolds without edge.

Tangent hyperplane[2] $T_x = \{v : M_x v = 0\}$ and projection operator to the hyperplane is

$$P_{T_x} = I - M_x^T (M_x M_x^T)^{-1} M_x$$

Stationarity condition $-f'(x) \in \mathcal{N}(S, x)$ may be rewritten as

$$P_{T_x} f'(x) = 0$$

[2] For $S = \{x : g_i(x) = 0,\ i = 1, ..., m\}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $M_x = g'(x)$

Ležanski-Polyak-Łojasievicz

Unconstrained minimization:
Ležanski-Polyak-Łojasievicz condition (LPL, PL)

$$f(x_k) - f^* \le \frac{1}{\mu}\|f'(x_k)\|^2$$

Ležanski-Polyak-Łojasievicz

Unconstrained minimization:

Ležanski-Polyak-Łojasievicz condition (LPL, PL)

$$f(x_k) - f^* \leq \frac{1}{\mu}\|f'(x_k)\|^2$$

Constrained optimization:

Ležanski-Polyak-Łojasievicz condition (LPL)

$$f(x_k) - f^* \leq \frac{1}{\mu}\|P_{T_x} f'(x_k)\|^2$$

Key properties

- Never achieved on whole S!
  A patch: consider a sublevel set $\{x : f(x) \leq f(x_0)\}$.
- **Linear convergence of projected gradient method!**
- Generalization to different degree (cf. Kurdyka-Łojasievicz and Ležanski)

$$f(x_k) - f^* \leq \frac{1}{\mu}\|P_{T_x} f'(x_k)\|^{\alpha}$$

Examples: strongly convex function and quadratic form on unit sphere

1. Strongly convex function with constant $\varkappa$ on proximally smooth set (on sublevel set) has LPL property if $\frac{L_0}{\varkappa} < R$.
2. Quadratic function on unit sphere (assume $\lambda_1 < \lambda_2 \leq \dots \leq \lambda_n$) has LPL property (on set $\{(x, e_1) \geq \tau, \|x\| = 1\}$)
with
$$\mu = 4\tau^2(\lambda_2 - \lambda_1)$$

Other ways to beat non-convexity

Full-step conditional gradient method

Frank-Wolfe

$$x_{k+1} = \arg\min_{x\in Q} \left\{ f(x_k) + (f'(x_k), x - x_k) + \frac{L_1}{2}||x - x_k||^2 \right\} =$$
$$= P_Q\left(x_k - \frac{1}{L_1} f'(x_k)\right),$$

"Almost linear functions": $L \approx 0$

$$x_{k+1} = \arg\min_{x\in Q}(f'(x_k), x)$$

Known convergence for strongly convex sets with radius r.

Existence condition

The set is the *boundary* of a strongly convex set.

Theorem (Gradient domination condition 1)

Let B be a strongly convex set of radius r, If $-f'(x^*) \in \mathcal{N}(B, x^*)$ and $\frac{\|f'(x^*)\|}{L_1} > r$ then x^* is the strict global minimum of the function f on the set B (and its boundary $Q = \partial B$).

A variant: Sphere of radius r: $\|f'(0)\| > 2L_1$.

Full-step Frank-Wolfe convergence condition

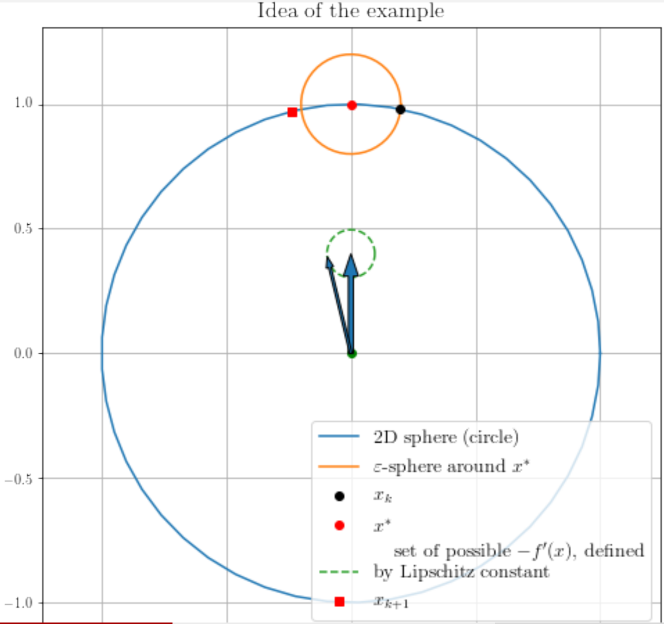
B - strongly convex set of radius r, x^* is a stationary point and $m = \frac{\|f'(x^*)\|}{r\,L_1} > 1$

Theorem (Convergence near optimum)

If $\|x_0 - x_*\| < \theta_m \frac{\|f'(x^*)\|}{L_1}$, then full-step Frank-Wolfe method converges to the minimum with linear rate.

$$\theta_m = \begin{cases} \frac{2}{m^2}\sqrt{m^2 - 1}, \; m \in (1, \sqrt{2}], \\ 1, \; m > \sqrt{2}. \end{cases}$$

Idea of proof: contraction (from the contrary)



Wrap-up

- Convergence and convergence rate of projected gradient methods on weakly convex sets
- Full-step Frank-Wolfe method convergence on the boundary of a convex set (under conditions)

Plans

- Stiefel manifolds
- Applications
- Generalizations

Thank you!

References

- F. H. Clarke, R. J Stern and P. R. Wolenski, Proximal smoothness and the lower-C^2 property, J. Convex Analysis, 2 (1995), 117-144. MR 96j:49014

- M. Balashov, About the gradient projection algorithm for a strongly convex function and a proximally smooth set, J. Convex Analysis, 24 (2017), 493–500.

- M. Balashov, B. Polyak, A. Tremba, Gradient projection and conditional gradient methods for constrained nonconvex minimization, arXiv:1906.11580 (submitted)