Superlinear and Quadratic Convergence of Riemannian Interior Point Methods

05001384 筑波大学大学院システム情報工学研究科 *頼 志堅 RAI Shiken
01703540 筑波大学システム情報系 吉瀬章子 YOSHISE Akiko

1. Introduction

We consider the following problem:

$$\min_{x\in\mathbb{M}} \quad f(x) \qquad \text{s.t.} \quad h(x)=0, \text{ and } g(x)\geq 0, \tag{RCOP}$$

where $\mathbb{M}$ is a finite dimensional Riemannian manifold, $f:\mathbb{M}\to\mathbb{R}, h:\mathbb{M}\to\mathbb{R}^l$, and $g:\mathbb{M}\to\mathbb{R}^m$ are C^2 functions on $\mathbb{M}$. This problem is called Riemannian Constrained Optimization Problem (RCOP). Such problems feature naturally in applications. For example, the matrix factorization with nonnegative constrains on fixed-rank manifold; k-means via low-rank SDP as a constrained optimization problem on the Stiefel manifold.

Contributions: In this manuscript, we extend the classical primal-dual interior point algorithms from the Euclidean setting to Riemannian setting, named Riemannian Interior Point (RIP) methods for (RCOP). Under some mild assumptions, we establish the locally superlinear/quadratic convergence for RIP, and the superlinear convergence for quasi-Newton type of RIP. Those are the generalizations of classical local convergent theory of interior point methods for nonlinear programming, proposed by El-Bakry et al. [1], and Yamashita and Yabe [2].

2. Interpretation: KKT vector field

The Lagrangian of (RCOP) is $\mathcal{L}(x,y,z)=f(x)-y^Th(x)-z^Tg(x)$, where $y\in\mathbb{R}^l$, $z\in\mathbb{R}^m$. $\mathcal{L}(\cdot,y,z)$ is a scalar function on $\mathbb{M}$, and its Riemannian gradient $\operatorname{grad}_x\mathcal{L}(x,y,z)$ is equal to

$$\operatorname{grad} f(x)-\sum_{i=1}^{l} y_i \operatorname{grad} h_i(x)-\sum_{i=1}^{m} z_i \operatorname{grad} g_i(x),$$

where $\operatorname{grad} f(x)$, $\{\operatorname{grad} h_i(x)\}$, $\{\operatorname{grad} g_i(x)\}$ are Riemannian gradients for components of f,h,g, respectively. The active set $\mathcal{A}(x)=\{i: g_i(x)=0, i=1,\ldots,m\}$ consists of indices of the active constraints at $x\in\mathbb{M}$. The Riemannian version of KKT conditions [3] for (RCOP) are given by

$$\begin{cases} \operatorname{grad}_x\mathcal{L}(x,y,z)=0,\\ h(x)=0,\\ g(x)\geq 0,\\ Zg(x)=0,\\ z\geq 0. \end{cases} \tag{1}$$

With a slack variable $s:=g(x)$, the above KKT conditions can be written as

$$F(w):=\begin{pmatrix} \operatorname{grad}_x\mathcal{L}(x,y,z)\\ h(x)\\ g(x)-s\\ ZSe \end{pmatrix}=0, \tag{2}$$

and $(s,z)\geq 0$, where $w:=(x,y,s,z)\in\mathscr{M}:=\mathbb{M}\times\mathbb{R}^l\times\mathbb{R}^m\times\mathbb{R}^m$. Remark that we get a vector field on Riemannian product manifold $\mathscr{M}$, i.e.,

$$F:\mathscr{M}\to T\mathscr{M}\equiv T\mathbb{M}\times T\mathbb{R}^l\times T\mathbb{R}^m\times T\mathbb{R}^m,$$

where $T\mathscr{M}:=\bigsqcup_{w\in\mathscr{M}}T_w\mathscr{M}$ denotes the tangent bundle of $\mathscr{M}$ with the tangent space $T_w\mathscr{M}\equiv T_x\mathbb{M}\times\mathbb{R}^l\times\mathbb{R}^m\times\mathbb{R}^m$ under the identifications $T_v\mathcal{E}\equiv\mathcal{E}$ for any vector space $\mathcal{E}$ and any $v\in\mathcal{E}$.

Definition 2.1 *The vector field F on $\mathscr{M}$ defined in (2) is called KKT vector field of (RCOP).*

3. Motivations: generalized Newton

Let $\mathcal{M}$ be a general Riemannian manifold. The generalized Newton method has been studied under the Riemannian setting and aims to find a point $p\in\mathcal{M}$ such that $F(p)=0$. Let ∇ be a Riemannian connection on $\mathcal{M}$. The covariant derivative of vector filed F assigns each point $p\in\mathcal{M}$ a linear operator $\nabla F(p):T_p\mathcal{M}\to T_p\mathcal{M}$. Then the Riemannian Newton iterate is stated as follows.

(Step 1) Compute the direction $v \in T_p\mathcal{M}$ as a solution of linear system $\nabla F(p)v = -F(p)$.
(Step 2) Compute $p_+ := R_p(v)$, where R denotes a retraction on $\mathcal{M}$ (a tool for moving on $\mathcal{M}$).

We have known that if p^* is a solution and $\nabla F(p^*)$ is nonsingular then under some mild conditions of covariant derivative ∇F, the local superlinear/quadratic convergence holds. Thus, the *nonsingularity* of covariant derivative at solution is essential if Newton method is to be applied for (2). The standard Riemannian assumptions for (RCOP) are as follows:

(A1) Existence. There exists w^* satisfying (1).
(A2) Smoothness. f, g, h are C^2; Hessian of their components are Lipschitz continuous at x^*.
(A3) Regularity. $\{\operatorname{grad} h_i(x^*)\} \cup \{\operatorname{grad} g_i(x^*) :$ for $i \in \mathcal{A}(x^*)\}$ is linearly independent in $T_{x^*}\mathbb{M}$.
(A4) Strict Complementarity. $(z^*)_i > 0$ if $g_i(x^*) = 0$ for all $i = 1, \cdots, m$.
(A5) Second Order Sufficiency. For all nonzero $\xi \in T_{x^*}\mathbb{M}$ such that $\langle \xi, \operatorname{grad} h_i(x^*)\rangle = 0$ for all i and $\langle \xi, \operatorname{grad} g_i(x^*)\rangle = 0$ for $i \in \mathcal{A}(x^*)$, one has $\langle \operatorname{Hess}_x \mathcal{L}(w^*)\xi, \xi\rangle > 0$, where $\operatorname{Hess}_x \mathcal{L}(w)$ denotes the Hessian of the scalar function $\mathcal{L}(\cdot, y, z)$.

The following result motivates the usage of Riemannian Newton method for solving (2).

Proposition 3.1 *If the assumptions (A1)-(A5) hold then* $\nabla F(w^*)$ *is nonsingular.*

3.1. Riemannian Interior Point (RIP)

As observed in usual Euclidean setting, to keep the iterates sufficiently far from the boundary, we introduce the perturbed complementary equation for some positive number $\mu > 0$, i.e.,

$$F_\mu(w) := F(w) - \mu\hat{e}(w), \text{ and } \hat{e}(w) := (0_x, 0, 0, e),$$

with zero element 0_x in $T_x\mathbb{M}$ and all ones $e \in \mathbb{R}^m$.

Now, we propose the prototype algorithms for RIP, or called Newton RIP, and its quasi-Newton version as follows.

(Step 0) Let R be a retraction on $\mathbb{M}$. Let $w_0 \in \mathscr{M}$ with $(s_0, z_0) > 0$, for $k = 0, 1, 2, \ldots$, do:
(Step 1) Choose the barrier parameter $\mu_k > 0$.
(Step 2) Solve the following linear system,

$$\nabla F(w_k)\Delta w_k = -F_{\mu_k}(w_k).$$

(Step 3) Choose γ_k with $0 < \hat{\gamma} \leq \gamma_k \leq 1$ for a constant $\hat{\gamma}$ and compute the step size

$$\alpha_k = -\gamma_k / \min\left((S_k)^{-1}\Delta s, (Z_k)^{-1}\Delta z, -\gamma_k\right).$$

(Step 4) Update: $w_{k+1} = \bar{R}_{w_k}(\alpha_k \Delta w_k)$, where $\bar{R}$ is the retraction on $\mathscr{M}$ that constructed by R.

Moreover, if the linear operator B_k on $T_{w_k}\mathscr{M}$ is constructed as the approximation of $\nabla F(w_k)$ then the quasi-Newton RIP is given if in (Step 2) we solve

$$B_k \Delta w_k = -F_{\mu_k}(w_k).$$

About convergent results, we prove the locally superlinear/quadratic convergence for RIP, and the local and superlinear convergence for quasi-Newton RIP. Due to limited space, their statements are omitted here.

4. Conclusion

In this article, we proposed a Riemannian version of classical interior point methods and establish some local convergent theories. To our knowledge, this article is the first study to apply the interior point method to the constrained optimization on manifolds.

参考文献

[1] El-Bakry, A. S., Richard A. Tapia, T. Tsuchiya, and Yin Zhang. "On the formulation and theory of the Newton interior-point method for nonlinear programming." *Journal of Optimization theory and Applications* 89, no. 3 (1996): 507-541.

[2] Yamashita, Hiroshi, and Hiroshi Yabe. "Superlinear and quadratic convergence of some primal-dual interior point methods for constrained optimization." *Mathematical Programming* 75, no. 3 (1996): 377-397.

[3] Yang, Wei Hong, Lei-Hong Zhang, and Ruyi Song. "Optimality conditions for the nonlinear programming problems on Riemannian manifolds." *Pacific Journal of Optimization* 10, no. 2 (2014): 415-434.