日本オペレーションズ・リサーチ学会
2022年 秋季研究発表会

On the Global Convergence of Riemannian Interior Point Method

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

1. Introduction

We extend the classical globally convergent primal-dual interior point algorithms [1] from the Euclidean setting to the Riemannian one. Our method, named the Riemannian interior point method (RIPM), is for solving Riemannian constrained optimization problems:

$$\begin{aligned} \min_{x\in\mathbb{M}} \quad & f(x) \\ \text{s.t.} \quad & h(x)=0, \text{ and } g(x)\leq 0, \end{aligned} \tag{RCOP}$$

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 smooth functions. Such problem has wide applications, e.g., the matrix factorization with nonnegative constrains on fixed-rank manifold.

2. Formulation of RIPM

KKT Vector Field

The Lagrangian of (RCOP) is $\mathcal{L}(x,y,z) := f(x)+y^\top h(x)+z^\top g(x)$ with multipliers y, z. Let $\operatorname{grad}_x \mathcal{L}(x,y,z)$ be the Riemannian gradient of $\mathcal{L}(\cdot,y,z):\mathbb{M}\to\mathbb{R}$, which 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 of components of f, h, g, respectively. The Riemannian version of KKT conditions [2] for (RCOP) are given by

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

By $s:=-g(x)$, the above is equivalent to

$$F(w)=0 \quad \text{and} \quad (z,s)\geq 0, \tag{2}$$

where

$$F(w) := \begin{pmatrix} \operatorname{grad}_x \mathcal{L}(x,y,z) \\ h(x) \\ g(x)+s \\ ZSe \end{pmatrix}, \tag{3}$$

is called KKT vector field of (RCOP) and $w := (x,y,s,z)\in\mathscr{M} := \mathbb{M}\times\mathbb{R}^l\times\mathbb{R}^m\times\mathbb{R}^m$. 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 $T\mathscr{M}$ denotes tangent bundle of $\mathscr{M}$ and tangent space $T_w\mathscr{M}\equiv T_x\mathbb{M}\times\mathbb{R}^l\times\mathbb{R}^m\times\mathbb{R}^m$.

Covariant Derivative $\nabla F(w)$

Let $\mathcal{M}$ be a general Riemannian manifold with Levi-Civita connection ∇. The generalized Newton method aims to find the singularity of a vector field $F:\mathcal{M}\to T\mathcal{M}$, i.e., a point $p\in\mathcal{M}$ such that $F(p)=0$. The covariant derivative of 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 Newton iterate is stated as follows.

(Step 1) Solve $\nabla F(p_k)[\xi_k]=-F(p_k)$ to get Newton direction $\xi_k\in T_{p_k}\mathcal{M}$.

(Step 2) Compute $p_{k+1}:=R_{p_k}(\xi_k)$, where R denotes a retraction on $\mathcal{M}$.

If the generalized Newton method is applied to (3), we must formulate the covariant derivative of KKT vector field F. For each $x\in\mathbb{M}$, we define a linear map $H_x:\mathbb{R}^l\to T_x\mathbb{M}$ by

$$H_x[v] := \textstyle\sum_{i=1}^{l} v_i \operatorname{grad} h_i(x).$$

Then, its adjoint $H_x^*: T_x\mathbb{M}\to\mathbb{R}^l$ is

$$H_x^*[\xi]=\left[\langle \operatorname{grad} h_1(x),\xi\rangle_x,\ldots,\langle \operatorname{grad} h_l(x),\xi\rangle_x\right]^\top.$$

One can define G_x and G_x^* verbatim. For $w \in \mathscr{M}$, the covariant derivative of KKT vector field F of (3) is $\nabla F(w) : T_w\mathscr{M} \to T_w\mathscr{M}$, and $\nabla F(w)[\Delta w]$ is given by

$$\begin{pmatrix} \operatorname{Hess}_x \mathcal{L}(w)[\Delta x] + H_x[\Delta y] + G_x[\Delta z] \\ H_x^*[\Delta x] \\ G_x^*[\Delta x] + \Delta s \\ Z\Delta s + S\Delta z \end{pmatrix},$$

where $\Delta w = (\Delta x, \Delta y, \Delta z, \Delta s) \in T_w\mathscr{M}$, and $\operatorname{Hess}_x \mathcal{L}(w)$ is equal to

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

with Hess being Riemannian Hessian on $\mathbb{M}$.

Perturbed Newton Equation

As in Euclidean setting, we introduce the perturbed KKT Vector Field: $F_\rho(w) := F(w) - \rho\hat{e}, \quad \hat{e} := (0,0,0,e)$, with barrier parameter $\rho > 0$. Here, $e \in \mathbb{R}^m$ is all-ones. Then it yields the perturbed Newton equation:

$$\nabla F(w)\Delta w = -F(w) + \rho\hat{e}.$$

3. Global Line Search RIPM

For a starting point $w_0 = (x_0, y_0, z_0, s_0)$ with $x_0 \in \mathbb{M}, (z_0, s_0) > 0$, let $\tau_1 := \frac{\min(Z_0S_0e)}{z_0^\top s_0/m}, \tau_2 := \frac{z_0^\top s_0}{\|F(w_0)\|}$. At a current point $w = (x, y, z, s)$ and direction $\Delta w = (\Delta x, \Delta y, \Delta z, \Delta s)$, the next iterate is calculated along a curve on $\mathscr{M}$, i.e., $w(\alpha) := \bar{R}_w(\alpha\Delta w)$, for some step length α. By introducing $w(\alpha) = (x(\alpha), y(\alpha), z(\alpha), s(\alpha))$, then $x(\alpha) = R_x(\alpha\Delta x)$, $y(\alpha) = y + \alpha\Delta y$, $z(\alpha) = z + \alpha\Delta z$, $s(\alpha) = s + \alpha\Delta s$. Define two centrality functions,

$$f^I(\alpha) := \min(Z(\alpha)S(\alpha)e) - \gamma\tau_1 z(\alpha)^\top s(\alpha)/m,$$
$$f^{II}(\alpha) := z(\alpha)^\top s(\alpha) - \gamma\tau_2\|F(w(\alpha))\|,$$

with constant $0 < \gamma < 1$. For $i = I, II$, define

$$\alpha^i := \max_{\alpha\in(0,1]} \left\{\alpha : f^i(t) \geq 0, \text{ for all } t \in (0, \alpha]\right\}.$$

Define a merit function $\varphi : \mathscr{M} \to \mathbb{R}$ by $\varphi(w) := \|F(w)\|^2$. Now, we are ready to describe our algorithm.

(**Step 0**) Choose $w_0 = (x_0, y_0, z_0, s_0)$ with $(z_0, s_0) > 0, \theta \in (0,1)$, and $\beta \in (0, 1/2]$. Set $k = 0, \gamma_{k-1} \in (1/2, 1)$, and $\varphi_0 = \varphi(w_0)$.
(**Step 1**) If $\varphi_k \leq \epsilon_{\text{exit}}$, stop.
(**Step 2**) Choose $\sigma_k \in (0,1)$, $\mu_k = z_k^\top s_k/m$; at w_k, compute direction Δw_k by

$$\nabla F(w)[\Delta w] = -F(w) + \sigma_k\mu_k\hat{e}.$$

(**Step 3**) Step length selection.
(3a) Centrality conditions: Choose $1/2 < \gamma_k < \gamma_{k-1}$; compute $\alpha^i, i = I, II$; and let $\bar{\alpha}_k = \min(\alpha^I, \alpha^{II})$.
(3b) Sufficient decreasing: Let $\alpha_k = \theta^t\bar{\alpha}_k$, where t is the smallest nonnegative integer such that α_k satisfies $\varphi(\bar{R}_{w_k}(\alpha_k\Delta w_k)) - \varphi(w_k) \leq \alpha_k\beta\,\langle \operatorname{grad}\varphi_k, \Delta w_k\rangle_{w_k}$.
(**Step 4**) Update: $w_{k+1} = \bar{R}_{w_k}(\alpha_k\Delta w_k)$.

Convergence

Under some standard assumptions, we prove the global convergence of RIPM with classical linear search. Due to limited space, their statements are omitted here.

4. Conclusion

In this article, we proposed a Riemannian version of classical globally convergent primal-dual interior point algorithms. Through the numerical experiments, we confirmed its validity on sphere manifold, Stiefel manifold, fixed-rank manifold and so on.

参考文献

[1] El-Bakry, A. S., Tapia, R. A., Tsuchiya, T., & Zhang, Y. (1996). On the formulation and theory of the Newton interior-point method for nonlinear programming. *Journal of Optimization theory and Applications*, 89(3), 507-541.

[2] Yang, W. H., Zhang, L. H., & Song, R. (2014). Optimality conditions for the nonlinear programming problems on Riemannian manifolds. *Pacific Journal of Optimization*, 10(2), 415-434.