日本オペレーションズ・リサーチ学会
2023 年春季研究発表会

Riemannian Interior Point Methods for Constrained Optimization on Manifolds

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 primal-dual interior point method from the Euclidean setting to the Riemannian one. Our method, named the Riemannian interior point method (RIPM), solves the Riemannian constrained optimization problems:

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

where $\mathbb{M}$ is a d-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 optimization with nonnegative orthogonality constraints in machine learning and data sciences.

2. Interpretation of RIPM

The Lagrangian of (RCOP) is $\mathcal{L}(x,y,z) := f(x)+y^\top h(x)+z^\top g(x)$. 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 gradients of components. The KKT conditions [2] for (RCOP) are given by

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

With $s:=-g(x)$, the above conditions become $F(w)=0$ and $(z,s)\ge 0$, where

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

is called *KKT vector field* and $w:=(x,y,z,s)\in\mathscr{M}:=\mathbb{M}\times\mathbb{R}^l\times\mathbb{R}^m\times\mathbb{R}^m$. Here we get a vector field on a product manifold $\mathscr{M}$. The tangent space appears as $T_w\mathscr{M}\cong T_x\mathbb{M}\times\mathbb{R}^l\times\mathbb{R}^m\times\mathbb{R}^m$.

Let $\mathcal{M}$ be a general 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 $p\in\mathcal{M}$ and $F(p)=0$. 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. Solve $\nabla F(p_k)\xi_k=-F(p_k)$ to get Newton direction $\xi_k\in T_{p_k}\mathcal{M}$; then update $p_{k+1}:=R_{p_k}(\xi_k)$, where R is a retraction on $\mathcal{M}$.

If the Newton method is applied to (2), we must formulate the covariant derivative of KKT vector field. For each $x\in\mathbb{M}$, we define a map $H_x:\mathbb{R}^l\to T_x\mathbb{M}$ by $H_x v := \sum_{i=1}^{l} v_i \operatorname{grad} h_i(x)$. Its adjoint operator $H_x^*: T_x\mathbb{M}\to\mathbb{R}^l$ is $H_x^*\xi = [\langle \operatorname{grad} h_1(x),\xi\rangle_x,\ldots,\langle \operatorname{grad} h_l(x),\xi\rangle_x]^\top$. One can define $\mathcal{G}_x$ and $\mathcal{G}_x^*$ verbatim.[1] Now, the covariant derivative $\nabla F(w): T_w\mathscr{M}\to T_w\mathscr{M}$ is given by

$$\nabla F(w)\Delta w = \begin{pmatrix} \operatorname{Hess}_x \mathcal{L}(w)\Delta x+\mathcal{H}_x\Delta y+\mathcal{G}_x\Delta z \\ \mathcal{H}_x^*\Delta x \\ \mathcal{G}_x^*\Delta x+\Delta s \\ Z\Delta s+S\Delta z \end{pmatrix},$$

where $\Delta w=(\Delta x,\Delta y,\Delta z,\Delta s)$ and $\operatorname{Hess}_x \mathcal{L}(w)$ is the Riemannian Hessian of $\mathcal{L}(\cdot,y,z)$. As in Euclidean setting, to keep the iterates sufficiently far from the boundary, we instead solve the *perturbed* Newton equation, i.e., $\nabla F(w)\Delta w=-F(w)+\mu\hat{e}$ where $\hat{e}:=(0,0,0,e)$ and $e\in\mathbb{R}^m$ is all-ones.

3. Algorithm and Convergence

Now, let us describe the prototype algorithm of the Riemannian interior point method.

(Step 0) Let R be a retraction on $\mathbb{M}$. Set $w_0\in\mathscr{M}$ with $(z_0,s_0)>0$, for $k=0,1,2,\ldots$, do:

(Step 1) Choose the barrier parameter $\mu_k>0$.

[1] If $\mathbb{M}\equiv\mathbb{R}^d$, then $\mathcal{H}_x,\mathcal{G}_x$ are expressed as the Jacobian matrices of h,g, and $\mathcal{H}_x^*,\mathcal{G}_x^*$ are their transposes.

(Step 2) Solve the following linear system,

$$\nabla F(w_k)\Delta w_k = -F(w_k) + \mu_k \hat{e}. \qquad (3)$$

(Step 3) Let γ_k with $0 < \hat{\gamma} \le \gamma_k \le 1$ 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.

Under some standard assumptions, we prove the locally superlinear and quadratic convergence of the prototype algorithm. Moreover, based on the above prototype algorithm, in this article we describe a globally convergent Riemannian interior point algorithm with the classical line search [1] and merit function $\varphi(w) = \|F(w)\|^2$. Due to limited space, their statements are omitted here.

4. Matrix-free Implementation

We will focus our attention to the solution of the linear system (3). After some substitutions,[2] we obtain a condensed equation on $T_x\mathbb{M} \times \mathbb{R}^l$:

$$\mathcal{T}(\Delta x, \Delta y) := \begin{pmatrix} \mathcal{A}_w \Delta x + \mathcal{H}_x \Delta y \\ \mathcal{H}_x^* \Delta x \end{pmatrix} = \begin{pmatrix} c \\ q \end{pmatrix}, \quad (4)$$

where

$$\mathcal{A}_w := \mathrm{Hess}_x\, \mathcal{L}(w) + \mathcal{G}_x S^{-1} Z \mathcal{G}_x^*,$$
$$c := -F_x - \mathcal{G}_x S^{-1} (Z F_z + \mu e - F_s),\ q := -F_y.$$

As in Euclidean setting, (4) is essentially a *symmetric indefinite* linear operator equation. Indeed, (4) is often called the saddle point problem in many literature. Now, the challenge of RIPM is how to solve (4) in an efficient manner.

For simplicity, we consider the case of only inequality constraints in (RCOP). Therefore, in (4) Δy vanishes and only a linear operator equation $\mathcal{A}_w \Delta x = c$ needs to be solved. The Riemannian situation leaves us with no explicit matrix form available. A general approach is to first find the representing matrix of $\hat{\mathcal{A}}$ (subscript w omitted). Clearly, this approach is expensive.

[2] In fact, we use $\Delta s = Z^{-1}(\mu e - F_s - S\Delta z)$ and $\Delta z = S^{-1}[Z(\mathcal{G}_x^* \Delta x + F_z) + \mu e - F_s]$.

An ideal approach is to use an iterative method, such as a *Krylov subspace method* (e.g., the conjugate gradients method), on tangent space $T_x\mathbb{M}$ directly. Such a method does not explicitly require a coefficient matrix, and instead needs only a matrix-vector product. More precisely, it only needs to call an abstract linear operator $v \mapsto \mathcal{A}v$. Since the operator $\mathcal{A}$ in (4) is a self-adjoint (or say, symmetric) but indefinite, we use the conjugate residual method to solve it.

5. Numerical experiments

Here, we model two problems:

- nonnegative low-rank matrix approximation. ($\mathbb{M}$ is fixed-rank manifold)
- projection onto nonnegative Stiefel manifold. ($\mathbb{M}$ is Stiefel manifold and/or oblique manifold)

within the framework of (RCOP) and use them to evaluate the performance of our RIPM and various other Riemannian algorithms. Numerical experiments show the stability and efficiency of our method, and the details will be explained in the presentation.

6. Conclusion

We proposed a Riemannian version of the classical interior point method and established its local and global convergence. To our knowledge, this is the first study to apply the primal-dual interior point method to the constrained optimization problem on a Riemannian manifold. Finally, the numerical experiments show the robustness and efficiency of our method.

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.