Implement Note of Global RIPM
Matrix submanifold

LAI ZHIJIAN

October 4, 2022

Chapter 1

Algorithm for submanifold

We consider the following problem,

$$\begin{array}{ll} \min\limits_{X \in \mathbb{M}} & f(X) \\ \text{s.t.} & h(X) = O_{p\times q}, \text{ and } g(X) \leqslant O_{n\times k}, \end{array} \tag{RCOP}$$

where $\mathbb{M}$ is a d-dimensional Riemannian submanifold of $\mathbb{R}^{r\times s}$ and $f : \mathbb{M} \to \mathbb{R}, h : \mathbb{M} \to \mathbb{R}^{p\times q}$, and $g : \mathbb{M} \to \mathbb{R}^{n\times k}$ are smooth on the manifold.

- All the inner products are Frobenius inner product, $\langle \cdot, \cdot \rangle$.
- For any X, $T_X\mathbb{M} \subseteq \mathbb{R}^{r\times s}$, $0_X = O_{r\times s}$, and $\langle \cdot, \cdot \rangle_X = \langle \cdot, \cdot \rangle$.
- The codomain of equality constraint h is $\mathbb{R}^{p\times q}$; inequality constraint g is $\mathbb{R}^{n\times k}$.
- The number of inequality constraints is m. (here, $m = nk$. Is it always that $m = \dim \mathbb{R}^{n\times k}$?)

1.1 Formulation of Interior Point Method on Submanifold

1.1.1 New notations

- For each $X \in \mathbb{M}$, we define the linear maps $H_X : \mathbb{R}^{p\times q} \to T_X\mathbb{M}$,

$$H_X(Y) := \sum_{i,j}^{p,q} Y_{ij} \operatorname{grad} h_{ij}(X) = \operatorname{grad}_X \langle Y, h(X) \rangle, \tag{1.1}$$

  and $\bar{H}_X : \mathbb{R}^{p\times q} \to \mathbb{R}^{r\times s}$,

$$\bar{H}_X(Y) := \sum_{i,j}^{p,q} Y_{ij} \operatorname{egrad} h_{ij}(X) = \operatorname{egrad}_X \langle Y, h(X) \rangle, \tag{1.2}$$

  Here, $X \mapsto \langle Y, h(X) \rangle : \mathbb{M} \subseteq \mathbb{R}^{r\times s} \to \mathbb{R}$. Then,

$$H_X(Y) = \operatorname{Proj}_X\left(\bar{H}_X(Y)\right), \text{ i.e., } H_X = \operatorname{Proj}_X \circ \bar{H}_X. \tag{1.3}$$

- We have that $\bar{H}_X^* : \mathbb{R}^{r\times s} \to \mathbb{R}^{p\times q}$,

$$\bar{H}_X^*(V) = \begin{pmatrix} \langle \operatorname{egrad} h_{11}(X), V \rangle & \cdots & \langle \operatorname{egrad} h_{1q}(X), V \rangle \\ \vdots & \ddots & \vdots \\ \langle \operatorname{egrad} h_{p1}(X), V \rangle & \cdots & \langle \operatorname{egrad} h_{pq}(X), V \rangle \end{pmatrix}, \tag{1.4}$$

  and

$$H_X^* = \bar{H}_X^*\big|_{T_X\mathbb{M}} : T_X\mathbb{M} \to \mathbb{R}^{p\times q}.$$

Proof. In fact, for $\Delta X \in T_X\mathbb{M} \subseteq \mathbb{R}^{r\times s}$,

$$H_X^*(\Delta X) = \begin{pmatrix} \langle \operatorname{grad} h_{11}(X), \Delta X\rangle & \cdots & \langle \operatorname{grad} h_{1q}(X), \Delta X\rangle \\ \vdots & \ddots & \vdots \\ \langle \operatorname{grad} h_{p1}(X), \Delta X\rangle & \cdots & \langle \operatorname{grad} h_{pq}(X), \Delta X\rangle \end{pmatrix}. \tag{1.5}$$

Since for any $Y \in \mathbb{R}^{p\times q}$, $\Delta X \in T_X\mathbb{M}$,

$$\langle H_X(Y), \Delta X\rangle = \left\langle \operatorname{Proj}_X\left(\bar{H}_X(Y)\right), \Delta X\right\rangle = \left\langle \bar{H}_X(Y), \operatorname{Proj}_X(\Delta X)\right\rangle = \left\langle \bar{H}_X(Y), \Delta X\right\rangle,$$

we obtain

$$\langle H_X(Y), \Delta X\rangle = \left\langle Y, \bar{H}_X^*(\Delta X)\right\rangle,$$

which means that $H_X^* = \bar{H}_X^*$ on subspace $T_X\mathbb{M}$. □

- We often conflate notation for $Y \in \mathbb{R}^{p\times q}$ and $\Delta Y \in \mathbb{R}^{p\times q}$.
- For each $X \in \mathbb{M}$, we define the linear maps $G_X : \mathbb{R}^{n\times k} \to T_X\mathbb{M}$,

$$G_X(Z) := \sum_{i,j}^{p,q} Z_{ij} \operatorname{grad} g_{ij}(X) = \operatorname{grad}_X\langle Z, g(X)\rangle, \tag{1.6}$$

and $\bar{G}_X : \mathbb{R}^{n\times k} \to \mathbb{R}^{r\times s}$,

$$\bar{G}_X(Z) := \sum_{i,j}^{p,q} Z_{ij} \operatorname{egrad} g_{ij}(X) = \operatorname{egrad}_X\langle Z, g(X)\rangle, \tag{1.7}$$

Here, $X \mapsto \langle Z, g(X)\rangle : \mathbb{M} \subseteq \mathbb{R}^{r\times s} \to \mathbb{R}$. Then,

$$G_X(Z) = \operatorname{Proj}_X\left(\bar{G}_X(Z)\right), \text{ i.e., } G_X = \operatorname{Proj}_X \circ \bar{G}_X. \tag{1.8}$$

- We have that $\bar{G}_X^* : \mathbb{R}^{r\times s} \to \mathbb{R}^{n\times k}$,

$$\bar{G}_X^*(V) = \begin{pmatrix} \langle \operatorname{egrad} g_{11}(X), V\rangle & \cdots & \langle \operatorname{egrad} g_{1k}(X), V\rangle \\ \vdots & \ddots & \vdots \\ \langle \operatorname{egrad} g_{n1}(X), V\rangle & \cdots & \langle \operatorname{egrad} g_{nk}(X), V\rangle \end{pmatrix}, \tag{1.9}$$

and

$$G_X^* = \bar{G}_X^*|_{T_X\mathbb{M}} : T_X\mathbb{M} \to \mathbb{R}^{n\times k}.$$

In fact, for $\Delta X \in T_X\mathbb{M} \subseteq \mathbb{R}^{r\times s}$,

$$G_X^*(\Delta X) = \begin{pmatrix} \langle \operatorname{grad} g_{11}(X), \Delta X\rangle & \cdots & \langle \operatorname{grad} g_{1k}(X), \Delta X\rangle \\ \vdots & \ddots & \vdots \\ \langle \operatorname{grad} g_{n1}(X), \Delta X\rangle & \cdots & \langle \operatorname{grad} g_{nk}(X), \Delta X\rangle \end{pmatrix}. \tag{1.10}$$

- We often conflate notation for $Z \in \mathbb{R}^{n\times k}$ and $\Delta Z \in \mathbb{R}^{n\times k}$.

How to find the $\bar{H}_X, H_X$ and $H_X^*, \bar{H}_X^*$ in practice? (Similar for G.)

1. We can either give H_X directly, or use $H_X = \operatorname{Proj}_X \circ \bar{H}_X$.
2. If we will find $\bar{H}_X$, it is convenient to use formulation

$$\bar{H}_X(Y) = \operatorname{egrad}_X\langle Y, h(X)\rangle, \tag{1.11}$$

   that is Euclidean gradient of function $X \mapsto \langle Y, h(X)\rangle$. (We seem Y as parameter.) To this end, we can either use definition of Euclidean gradient or tools "Matrix Calculus."
3. Since $H_X^* = \bar{H}_X^*\big|_{T_X\mathbb{M}}$, it is sufficient to find $\bar{H}_X^*$ (In the code, we just let H_X^* equal $\bar{H}_X^*$ if we can ensure that the input is a tangent vector at X.). We strongly recommend to use the definition of "adjoint operator" directly through the formulation of $\bar{H}_X$, instead of (1.4).

1.1.2 Lagrangian

The Lagrangian function of (RCOP) is

$$\mathcal{L}(X,Y,Z) = f(X) + \langle Y, h(X)\rangle + \langle Z, g(X)\rangle, \tag{1.12}$$

where $Y \in \mathbb{R}^{p\times q}$ and $Z \in \mathbb{R}^{n\times k}$ are Lagrange multipliers.

- Here, $X \mapsto \mathcal{L}(X,Y,Z) : \mathbb{M} \subseteq \mathbb{R}^{r\times s} \to \mathbb{R}$. Then,

$$\begin{aligned}
\operatorname{egrad}_X \mathcal{L}(X,Y,Z) &= \operatorname{egrad} f(X) + \sum_{i,j}^{p,q} Y_{ij} \operatorname{egrad} h_{ij}(X) + \sum_{i,j}^{n,k} Z_{ij} \operatorname{egrad} g_{ij}(X) && (1.13)\\
&= \operatorname{egrad} f(X) + \operatorname{egrad}_X\langle Y, h(X)\rangle + \operatorname{egrad}_X\langle Z, g(X)\rangle && (1.14)\\
&= \operatorname{egrad} f(X) + \bar{H}_X(Y) + \bar{G}_X(Z) && (1.15)\\
&\in \mathbb{R}^{r\times s}. && (1.16)
\end{aligned}$$

and

$$\begin{aligned}
\operatorname{grad}_X \mathcal{L}(X,Y,Z) &= \operatorname{grad} f(X) + \sum_{i,j}^{p,q} Y_{ij} \operatorname{grad} h_{ij}(X) + \sum_{i,j}^{n,k} Z_{ij} \operatorname{grad} g_{ij}(X) && (1.17)\\
&= \operatorname{grad} f(X) + \operatorname{grad}_X\langle Y, h(X)\rangle + \operatorname{grad}_X\langle Z, g(X)\rangle && (1.18)\\
&= \operatorname{grad} f(X) + H_X(Y) + G_X(Z) && (1.19)\\
&= \operatorname{Proj}_X\left(\operatorname{egrad}_X \mathcal{L}(X,Y,Z)\right) && (1.20)\\
&\in T_X\mathbb{M}. && (1.21)
\end{aligned}$$

- Again, $X \mapsto \mathcal{L}(X,Y,Z) : \mathbb{M} \subseteq \mathbb{R}^{r\times s} \to \mathbb{R}$. Then,

$$\begin{aligned}
\operatorname{ehess}_X \mathcal{L}(W) &\equiv \operatorname{ehess}_X \mathcal{L}(X,Y,Z) && (1.22)\\
&= \operatorname{ehess} f(X) + \sum_{i,j}^{p,q} Y_{ij} \operatorname{ehess} h_{ij}(X) + \sum_{i,j}^{n,k} Z_{ij} \operatorname{ehess} g_{ij}(X) && (1.23)\\
&= \operatorname{ehess} f(X) + \operatorname{ehess}_X\langle Y, h(X)\rangle + \operatorname{ehess}_X\langle Z, g(X)\rangle. && (1.24)
\end{aligned}$$

Note that $\operatorname{ehess}_X \mathcal{L}(W) : \mathbb{R}^{r\times s} \to \mathbb{R}^{r\times s}$,

$$\operatorname{ehess}_X \mathcal{L}(W)[U] \overset{\text{def}}{=} \mathcal{D}(X \mapsto \operatorname{egrad}_X \mathcal{L}(X,Y,Z))(X)[U]. \tag{1.25}$$

And

$$\begin{aligned}
\operatorname{hess}_X \mathcal{L}(W) &= \operatorname{hess} f(X) + \sum_{i,j}^{p,q} Y_{ij} \operatorname{hess} h_{ij}(X) + \sum_{i,j}^{n,k} Z_{ij} \operatorname{hess} g_{ij}(X) && (1.26)\\
&= \operatorname{hess} f(X) + \operatorname{hess}_X\langle Y, h(X)\rangle + \operatorname{hess}_X\langle Z, g(X)\rangle. && (1.27)
\end{aligned}$$

How to find $\operatorname{hess}_X \mathcal{L}(W)$ **?**

1. If not given directly, then we use `M.ehess2rhess(x, egrad, ehess, u)`: converts the Euclidean gradient and Hessian of f at x along a tangent vector u to the Riemannian Hessian of f at x along u on the manifold. It requests $\operatorname{egrad}_X \mathcal{L}(X,Y,Z)$ and $\operatorname{ehess}_X \mathcal{L}(X,Y,Z)[U]$.

1.1.3 KKT Vector Field and $\nabla F(W)$

- The Riemannian versions of the KKT conditions for (RCOP) are given by

$$\begin{cases} \operatorname{grad}_X \mathcal{L}(X,Y,Z) = 0_X, \\ h(X) = O_{p\times q}, \\ g(X) \leqslant O_{n\times k}, \\ Z \odot g(X) = O_{n\times k}, \\ Z \geqslant O_{n\times k}. \end{cases} \tag{1.28}$$

  With slack variables $S := -g(X) \in \mathbb{R}^{n\times k}$, the above KKT conditions can be written as

$$F(W) := \begin{pmatrix} \operatorname{grad}_X \mathcal{L}(X,Y,Z) \\ h(X) \\ g(X) + S \\ Z \odot S \end{pmatrix} = \begin{pmatrix} 0_X \\ O_{p\times q} \\ O_{n\times k} \\ O_{n\times k} \end{pmatrix}, \tag{1.29}$$

  and $(Z,S) \geqslant 0$, where $W := (X,Y,Z,S) \in \mathscr{M} := \mathbb{M} \times \mathbb{R}^{p\times q} \times \mathbb{R}^{n\times k} \times \mathbb{R}^{n\times k}$. We generate a vector field F on the Riemannian product manifold $\mathscr{M}$, i.e.,

$$F : \mathscr{M} \to T\mathscr{M} \equiv T\mathbb{M} \times T\mathbb{R}^{p\times q} \times T\mathbb{R}^{n\times k} \times T\mathbb{R}^{n\times k}, \tag{1.30}$$

  where $T\mathscr{M}$ denotes the tangent bundle of $\mathscr{M}$, and

$$T_W\mathscr{M} \equiv T_X\mathbb{M} \times \mathbb{R}^{p\times q} \times \mathbb{R}^{n\times k} \times \mathbb{R}^{n\times k}. \tag{1.31}$$

- Some concepts on $\mathscr{M}$.
  - For any $\xi, \zeta \in T_W\mathscr{M}$, the Riemannian product metric is defined as

$$\langle \xi, \zeta \rangle_W := \langle \xi_X, \zeta_X \rangle + \langle \xi_Y, \zeta_Y \rangle + \langle \xi_S, \zeta_S \rangle + \langle \xi_Z, \zeta_Z \rangle,$$

    where $\xi = (\xi_X, \xi_Y, \xi_S, \xi_Z)$ and $\zeta = (\zeta_X, \zeta_Y, \zeta_S, \zeta_Z)$.
  - Accordingly, the induced norm $\|\xi\|_W := \sqrt{\langle \xi, \xi \rangle_W}$ satisfies

$$\|\xi\|_W^2 = \|\xi_X\|_F^2 + \|\xi_Y\|_F^2 + \|\xi_S\|_F^2 + \|\xi_Z\|_F^2. \tag{1.32}$$

    Here, $\|\cdot\|_F$ denotes the usual Frobenius norm.
  - The Riemannian distance on $\mathscr{M}$ is defined as

$$d(W_1, W_2) := \sqrt{d^2(X_1, X_2) + \|Y_1 - Y_2\|_F^2 + \|S_1 - S_2\|_F^2 + \|Z_1 - Z_2\|_F^2}. \tag{1.33}$$

  - For any $W \in \mathscr{M}$ and $\xi \in T_W\mathscr{M}$,

$$\bar{R}_W(\xi) := (R_X(\xi_X), Y + \xi_Y, S + \xi_S, Z + \xi_Z) \tag{1.34}$$

    defines a retraction on $\mathscr{M}$. $\bar{R}$ is the exponential map on $\mathscr{M}$ if R is the exponential map on $\mathbb{M}$.

Remark 1.1.1. In Manopt, some manifolds are not equipped with a distance function. For example, `fixedrankembeddedfactory` and `stiefelfactory`. Instead, we use the distance of its ambient space. We can have the above structure by going through `productmanifold` to construct $\mathscr{M}$.

- Given any $W \in \mathscr{M}$, for the KKT vector field F defined in (4.20), the linear operator

$$\nabla F(W) : T_W\mathscr{M} \to T_W\mathscr{M}$$

is given by

$$\nabla F(W)[\Delta W] = \begin{pmatrix} \operatorname{hess}_X \mathcal{L}(W)[\Delta X] + \sum_{i,j}^{p,q} \Delta Y_{ij} \operatorname{grad} h_{ij}(X) + \sum_{i,j}^{n,k} \Delta Z_{ij} \operatorname{grad} g_{ij}(X) \\ \langle \operatorname{grad} h_{ij}(X), \Delta X \rangle, \text{ for } 1 \leqslant i \leqslant p, 1 \leqslant j \leqslant q \\ \langle \operatorname{grad} g_{ij}(X), \Delta X \rangle + \Delta S_{ij}, \text{ for } 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant k \\ Z \odot \Delta S + S \odot \Delta Z \end{pmatrix} \tag{1.35}$$

where $\Delta W = (\Delta X, \Delta Y, \Delta Z, \Delta S) \in T_X\mathbb{M} \times \mathbb{R}^{p\times q} \times \mathbb{R}^{n\times k} \times \mathbb{R}^{n\times k}$.
Next, we obtain a compact form of (4.26) as below,

$$\nabla F(W)[\Delta W] = \begin{pmatrix} \operatorname{hess}_X \mathcal{L}(W)[\Delta X] + H_X(\Delta Y) + G_X(\Delta Z) \\ H_X^*(\Delta X) \\ G_X^*(\Delta X) + \Delta S \\ Z \odot \Delta S + S \odot \Delta Z \end{pmatrix}. \tag{1.36}$$

Moreover, its adjoint is

$$\nabla F(W)^*[\Delta W] = \begin{pmatrix} \operatorname{hess}_X \mathcal{L}(W)[\Delta X] + H_X(\Delta Y) + G_X(\Delta Z) \\ H_X^*(\Delta X) \\ G_X^*(\Delta X) + S \odot \Delta S \\ Z \odot \Delta S + \Delta Z \end{pmatrix}. \tag{1.37}$$

- If we define a merit function $\varphi : \mathscr{M} \to \mathbb{R}$ by

$$\varphi(W) := \|F(W)\|_W^2, \tag{1.38}$$

then

$$\operatorname{grad} \varphi(W) = 2\nabla F(W)^*[F(W)]. \tag{1.39}$$

In order to calculate $\operatorname{grad} \varphi(W)$, one needs to know $\nabla F(W)^*$.

1.1.4 Condensed form of perturbed Newton Equation

The perturbed Newton Equation is

$$\nabla F(W)[\Delta W] = -F(W) + \rho \hat{J}, \tag{1.40}$$

where $\rho > 0, \hat{J} := \hat{J}(W) = \begin{pmatrix} 0_X \\ O_{p\times q} \\ O_{n\times k} \\ J_{n\times k} \end{pmatrix}$ with all-ones matrix $J_{n\times k}$. In later, $\rho = \sigma\mu$.

Define some notations

$$F(W) \equiv \begin{pmatrix} F_X \\ F_Y \\ F_Z \\ F_S \end{pmatrix} \equiv \begin{pmatrix} \operatorname{grad}_X \mathcal{L}(X, Y, Z) \\ h(X) \\ g(X) + S \\ Z \odot S \end{pmatrix}. \tag{1.41}$$

By compact form (4.27) and notations above, we need to solve the following linear system on $T_X\mathbb{M} \times \mathbb{R}^{p\times q} \times \mathbb{R}^{n\times k} \times \mathbb{R}^{n\times k}$:

$$\begin{pmatrix} \operatorname{hess}_X \mathcal{L}(W)[\Delta X] + H_X(\Delta Y) + G_X(\Delta Z) \\ H_X^*(\Delta X) \\ G_X^*(\Delta X) + \Delta S \\ Z \odot \Delta S + S \odot \Delta Z \end{pmatrix} = \begin{pmatrix} -F_X \\ -F_Y \\ -F_Z \\ -F_S + J_{n\times k} \end{pmatrix}. \tag{1.42}$$

We suppose that $(Z, S) > 0$. Then system can be written in condensed form on $T_X\mathbb{M} \times \mathbb{R}^{p\times q}$:

$$\mathcal{T}(\Delta X, \Delta Y) := \begin{pmatrix} \mathcal{A}_W(\Delta X) + H_X(\Delta Y) \\ H_X^*(\Delta X) \end{pmatrix} = \begin{pmatrix} \boldsymbol{c} \\ \boldsymbol{q} \end{pmatrix}. \tag{1.43}$$

where

$$\begin{aligned} \mathcal{A}_W &:= \operatorname{hess}_X \mathcal{L}(W) + G_X \circ S^{\circ(-1)} \odot Z \odot G_X^*, \\ \boldsymbol{c} &:= -F_X - G_X[S^{\circ(-1)} \odot (Z \odot F_Z + \rho J_{n\times k} - F_S)], \\ \boldsymbol{q} &:= -F_Y. \end{aligned} \tag{1.44}$$

Moreover,

$$\Delta Z = S^{\circ(-1)} \odot [Z \odot (G_X^*[\Delta X] + F_Z) + \rho J_{n\times k} - F_S] \tag{1.45}$$

$$\Delta S = Z^{\circ(-1)} \odot (\rho J_{n\times k} - F_S - S \odot \Delta Z). \tag{1.46}$$

Here, $U \mapsto Z \odot U : \mathbb{R}^{n\times k} \to \mathbb{R}^{n\times k}$. $Z \odot (\cdot)$ is self-adjoint on $\mathbb{R}^{n\times k}$ because for all $U, W \in \mathbb{R}^{n\times k}$,

$$\langle Z \odot U, W\rangle = \langle U \odot Z, W\rangle = \langle U, W \odot Z\rangle = \langle U, Z \odot W\rangle. \tag{1.47}$$

- How to solve linear system (4.34) becomes critical.
  1. $G_X S^{\circ(-1)} \odot Z \odot G_X^*$ is self-adjoint on $T_X\mathbb{M}$.
  2. $\mathcal{A}_W$ is self-adjoint on $T_X\mathbb{M}$.
  3. $\mathcal{T}$ is self-adjoint on $T_X\mathbb{M} \times \mathbb{R}^{p\times q}$ with the inner product $\langle \Delta X, \Delta X'\rangle + \langle \Delta Y, \Delta Y'\rangle$.
- From the discussion above, if $(Z, S) > 0$ holds, then the operator $\nabla F(W)$ in (4.27) is nonsingular if and only if $\mathcal{T}$ in (4.34) is nonsingular.
  A necessary condition for $\mathcal{T}$ to be nonsingular is injectivity of H_X. Note that H_X is injective if and only if $\operatorname{rank} H_X := \dim \operatorname{span} \{\operatorname{grad} h_{ij}(X)\}_{ij} = \dim \mathbb{R}^{p\times q}$.

1.2 Global Line Search RIP Algorithm

1.2.1 Fundamental Description

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.,

$$W(\alpha) := \bar{R}_W(\alpha \Delta W), \tag{1.48}$$

for some step length $\alpha > 0$. By introducing

$$W(\alpha) = (X(\alpha), Y(\alpha), Z(\alpha), S(\alpha)),$$

we have

$$X(\alpha) = R_X(\alpha\Delta X), Y(\alpha) = Y + \alpha\Delta Y, Z(\alpha) = Z + \alpha\Delta Z, S(\alpha) = S + \alpha\Delta S.$$

For a given starting point $W_0 = (X_0, Y_0, Z_0, S_0)$ with $X_0 \in \mathbb{M}, (Z_0, S_0) > 0$, let

$$\tau_1 := \frac{\min(Z_0 \odot S_0)}{\langle Z_0, S_0\rangle/m}, \quad \tau_2 := \frac{\langle Z_0, S_0\rangle}{\|F(w_0)\|}. \tag{1.49}$$

Then $0 < \tau_1 \leqslant 1$ and $0 < \tau_2 \leqslant \sqrt{m}$. Define

$$f^I(\alpha) := \min(Z(\alpha) \odot S(\alpha)) - \gamma\tau_1 \langle Z(\alpha), S(\alpha)\rangle / m, \tag{1.50}$$

$$f^{II}(\alpha) := \langle Z(\alpha), S(\alpha)\rangle - \gamma\tau_2 \|F(W(\alpha))\|, \tag{1.51}$$

where $\gamma \in (0,1)$ is a constant. We note that the functions $f^i(\alpha), i = I, II$, depend on the iteration count k, though for simplicity we choose not to write explicitly this dependency. For $i = I, II$, define

$$\alpha^i := \max_{\alpha \in (0,1]} \left\{\alpha : f^i(t) \geqslant 0, \text{ for all } t \in (0, \alpha]\right\}, \tag{1.52}$$

i.e., α^i are either one or the smallest positive root for the functions $f^i(\alpha)$ in $(0,1]$.

We have some observations as below:

$$\|F(W)\|_W^2 = \|\operatorname{grad}_X L(W)\|_X^2 + \|h(X)\|_2^2 + \|g(X) + S\|_2^2 + \|ZSe\|_2^2. \tag{1.53}$$

For any nonnegative $S, Z \in \mathbb{R}^{n \times k}$, one has

$$\|Z \odot S\|_F \leqslant \langle Z, S\rangle = \|Z \odot S\|_1 \leqslant \sqrt{m}\, \|Z \odot S\|_F\,. \tag{1.54}$$

Hence,

$$\|Z \odot S\|_F / \sqrt{m} \leqslant \frac{\langle Z, S\rangle}{\sqrt{m}} \leqslant \|Z \odot S\|_F \leqslant \|F(W)\|_W. \tag{1.55}$$

Now, we describe the globally convergent Riemannian interior point method.

Algorithm 1 (Global RIP Algorithm).

(Step 0) Choose $W_0 = (X_0, Y_0, Z_0, S_0)$ such that $(Z_0, S_0) > 0, \theta \in (0,1)$, and $\beta \in (0, 1/2]$. Set $k = 0, \gamma_{k-1} \in (1/2, 1)$, and $\varphi_0 = \varphi(W_0)$. For $k = 0, 1, 2, \ldots$, do the following steps.

(Step 1) Test for convergence: if $\varphi_k \leqslant \epsilon_{\text{exit}}$, stop.

(Step 2) Choose $\sigma_k \in (0,1)$; for w_k, compute the perturbed Newton direction Δw_k with

$$\mu_k \in [\langle Z_k, S_k\rangle / m, \|F(w_k)\| / \sqrt{m}] \tag{1.56}$$

and by

$$\nabla F(W)[\Delta W] = -F(W) + \sigma_k \mu_k \hat{J}. \tag{1.57}$$

(Step 3) Step length selection.

(3a) *Centrality conditions*: Choose $1/2 < \gamma_k < \gamma_{k-1}$; compute $\alpha^i, i = I, II$, from (4.43); and let

$$\bar{\alpha}_k = \min(\alpha^I, \alpha^{II}). \tag{1.58}$$

(3b) *Sufficient decreasing*: Let $\alpha_k = \theta^t \bar{\alpha}_k$, where t is the smallest nonnegative integer such that α_k satisfies

$$\varphi(\bar{R}_{W_k}(\alpha_k \Delta W_k)) - \varphi(W_k) \leqslant \alpha_k \beta \langle \operatorname{grad} \varphi_k, \Delta W_k\rangle\,. \tag{1.59}$$

(Step 4) Let $W_{k+1} = \bar{R}_{W_k}(\alpha_k \Delta W_k)$ and $k \leftarrow k + 1$. Go to Step 1.