Computational and Applied Mathematics (2023) 42:164
https://doi.org/10.1007/s40314-023-02302-0

A collection of efficient retractions for the symplectic Stiefel manifold

H. Oviedo[1] · R. Herrera[2]

Received: 6 February 2023 / Revised: 31 March 2023 / Accepted: 10 April 2023 /
Published online: 27 April 2023
© The Author(s) under exclusive licence to Sociedade Brasileira de Matemática Aplicada e Computacional 2023

Abstract
This article introduces a new map on the symplectic Stiefel manifold. The operation that requires the highest computational cost to compute the novel retraction is a inversion of size $2p$-by-$2p$, which is much less expensive than those required for the available retractions in the literature. Later, with the new retraction, we design a constraint preserving gradient method to minimize smooth functions defined on the symplectic Stiefel manifold. To improve the numerical performance of our approach, we use the non-monotone line-search of Zhang and Hager with an adaptive Barzilai–Borwein type step-size. Our numerical studies show that the proposed procedure is computationally promising and is a very good alternative to solve large-scale optimization problems over the symplectic Stiefel manifold.

Keywords Symplectic Stiefel manifold · Riemannian gradient method · Riemannian optimization · Symplectic matrix

Mathematics Subject Classification 65K05 · 70G45 · 90C48

1 Introduction

The aim of this work is to design a fast Riemannian gradient method to find local minimizers of the following optimization problem with skew–symmetric constraints

$$\min_{X\in\mathbb{R}^{2n\times 2p}} \mathcal{F}(X) \quad \text{s.t.} \quad X^\top J_{2n} X = J_{2p}, \tag{1}$$

Communicated by yimin wei.

✉ H. Oviedo
harry.oviedo@uai.cl

R. Herrera
rherrera@cimat.mx

[1] Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Av. Diag. Las Torres 2640, 7941169 Santiago de Chile, Metropolitan Region, Chile

[2] Departamento de Matemáticas, Centro de Investigación en Matemáticas A. C., Jalisco s/n, Valenciana, 36023 Guanajuato, Guanajuato, Mexico

where $p \leq n$, $\mathcal{F} : \mathbb{R}^{2n\times 2p} \to \mathbb{R}$ is a bounded below and continuously differentiable function, and $J_{2m} := [0, I_m; -I_m, 0]$, where I_m denotes the m-by-m identity matrix for any positive integer m. The constraints set $Sp(2p, 2n) = \{X \in \mathbb{R}^{2n\times 2p} : X^\top J_{2n} X = J_{2p}\}$ of problem (1) is so-called the *symplectic Stiefel manifold*, Gao et al. (2021). In fact, if we equip $Sp(2p, 2n)$ with a positive-definite inner product $\langle \cdot, \cdot \rangle_X$ on the tangent space $T_X Sp(2p, 2n)$ at each point X, then the pair $(Sp(2p, 2n), \langle \cdot, \cdot \rangle_X)$ becomes a closed embedded Riemannian sub-manifold of $\mathbb{R}^{2n\times 2p}$, whose dimension is equal to $4np - p(2p - 1)$, see Proposition 3.1 in Gao et al. (2021). In the special case $p = n$, the symplectic Stiefel manifold reduces to the *symplectic group* and denoted by $Sp(2n)$, which is a Lie group.

Besides differential and symplectic geometries, symplectic matrices arise in many fields, such as optical systems (Fiori 2016), optimal control of quantum symplectic gates (Wu et al. 2008), optimization problems on the set of symplectic matrices (Fiori 2016; Gao et al. 2021; Peng and Mohseni 2016; Son et al. 2021), Procrustes problem (Zhao 2022), matrix decompositions (Benner et al. 1999; Salam and Al-Aidarous 2014), computation of eigenvalues of (skew-)Hamiltonian matrices (Benner and Fabender 1997, 1998; Del Buono et al. 2008; Van Loan 1984), trace minimization problems (Son et al. 2021) and symplectic principal component analysis (Lei et al. 2016; Lei and Meng 2016; Parra 1995), which motivates the study of this class of matrices. In particular, the geometry and topology of the symplectic Stiefel manifolds and their compact counterparts, the quaternionic Stiefel manifolds, have been of interest in differential geometry for a long time.

Most of the optimization methods on an Euclidean space, perform a line search after computing a descent search direction using a straight line as parameterization (Nocedal and Wright 2006). By contrast, in the Riemannian optimization context, the concept of a straight line is substituted with a curve (not necessarily a geodesic) over a given Riemannian manifold (Absil et al. 2009; Hu et al. 2020). In particular, the *retractions* are the most pragmatic approach to constructing these curves. Roughly, a retraction is a smooth mapping that sends tangent vectors back to the manifold, and defines an appropriate curve for searching a next iterate point on the corresponding manifold. The rigorous definition of retraction appears in Absil et al. (2009).

The Riemannian gradient method equipped with a non-monotone globalization technique and the Barzilai–Borwein step-size (Iannazzo and Porcelli 2018), has been an efficient alternative to solve several manifold constrained optimization problems with real applications, for example see Iannazzo and Porcelli (2018), Jiang and Dai (2015), Oviedo (2022), Oviedo et al. (2021), Oviedo et al. (2019), Son et al. (2021), Wen and Yin (2013). Recently, some constraint preserving gradient methods were proposed based on the geometrical study of the symplectic Stiefel manifold (Gao et al. 2021; Son et al. 2021). In particular, in Gao et al. (2021) the authors developed two Riemannian gradient procedures to solve the optimization problem (1). The first approach uses a globally defined quasi-geodesic on the constraint set. Nevertheless, it is necessary to compute two matrix exponentials to evaluate that quasi-geodesic, which is computationally restrictive. With the goal of designing a less computationally costly iterative scheme, Gao et al. (2021) built another gradient method based on the Cayley transform by generalizing the works of Wen and Yin (2013). Specifically, in Gao et al. (2021) they introduced the following retraction map

$$R_X(\xi_X) = \left(I_{2n} - \frac{1}{2}A_X J_{2n}\right)^{-1}\left(I_{2n} + \frac{1}{2}A_X J_{2n}\right)X, \tag{2}$$

for all $X \in Sp(2n, 2p)$ and $\xi_X = A_X J_{2n} X$ in the tangent space of $Sp(2p, 2n)$ at X. With this retraction, the authors (Gao et al. 2021; Son et al. 2021), designed a Riemannian gradient method combined with chosen of Barzilai–Borwein step-size. Additionally, the sophisticated retraction (2) is equivalent to

$$R_X(\xi_X) = X + U\left(I_{4p} + \frac{1}{2}V^\top J_{2n}^\top U\right)^{-1} V^\top J_{2n} X, \tag{3}$$

where $A_X = UV^\top$, with $U, V \in \mathbb{R}^{2n\times 4p}$ are two matrices associated with $\xi_X = A_X J_{2n} X$, (for more details see Proposition 5.4 in Gao et al. (2021)). Observe that this last formula is more advantageous than (2), when $p < \frac{n}{2}$. However, both formulae (2)–(3) require inverting a large-size matrix when $p \geq n/2$, and n large.

In this paper, we introduce a collection of efficient retraction mapping on the symplectic Stiefel manifold, where each member of the collection requires inverting a matrix of smaller size than the required by the Cayley based retraction. More precisely, each member needs to invert a matrix of size $2p \times 2p$. The new general retraction is constructed following the descriptions presented in Jiang and Dai (2015). Thus, the new retraction can be seen as a generalization of the curve developed in Jiang and Dai (2015) for the Stiefel manifold. With the purpose of evaluating the numerical performance of the proposed family of retractions, we design a novel Riemannian gradient method, which uses the Zhang and Hager non-monotone globalization strategy (Zhang and Hager 2004) combined with a step-size developed in Oviedo et al. (2021), to speed up the numerical behavior of the proposal. Our preliminary numerical experiments suggest that our proposal is numerically superior to some existing Riemannian gradient methods in the state-of-the-art.

The following sections are organized as follows: Sect. 2 reviews the geometry of the symplectic Stiefel manifold by summarizing the concepts contained in Gao et al. (2021), Son et al. (2021). Section 3 describes the new approach in detail. In Sect. 4 we present some numerical results to illustrate the numerical performance of our proposal. Finally, in Sect. 5 we provide the conclusions of this work.

2 Geometric tools associated with the symplectic Stiefel manifold

This section contains some background concepts and notation required later on. The notions summarized in this section can also be found in Absil et al. (2009), Gao et al. (2021), Son et al. (2021).

Let m be a positive integer, the matrix J_{2m}, defined at the beginning of Sect. 1, has the following properties

$$J_{2m}^\top = -J_{2m}, \quad J_{2m}^\top J_{2m} = I_{2m}, \quad J_{2m}^2 = -I_{2m}, \quad J_{2m}^{-1} = J_{2m}^\top.$$

Now, given a square matrix $A \in \mathbb{R}^{m\times m}$, $\mathtt{sym}(A)$ denotes the symmetric part of A that is, $\mathtt{sym}(A) = 0.5(A + A^\top)$. The set of symmetric matrices of size $m \times m$ will be denoted by $\mathcal{S}_{\text{sym}}(2p)$. $\mathcal{S}_+(m)$ will denotes the set of m-by-m positive definite matrices with real entries. The trace of A is defined as the sum of the diagonal elements which we denote by $tr(A)$. The standard inner product of two matrices $A, B \in \mathbb{R}^{m\times n}$ is given by $\langle A, B\rangle := \sum_{i,j} a_{ij}b_{ij} = tr(A^\top B)$. The Frobenius norm is defined by $\|A\|_F = \sqrt{\langle A, A\rangle}$. Let $X \in Sp(2p, 2n)$, the tangent space of the symplectic Stiefel manifold at X is given by Gao et al. (2021)

$$T_X Sp(2p, 2n) = \{Z \in \mathbb{R}^{2n\times 2p} : Z^\top J_{2n} X + X^\top J_{2n} Z = 0\}.$$

A characterization of the tangent space (Gao et al. 2021), crucial for this manuscript, is

$$T_X Sp(2p, 2n) = \{S J_{2n} X : S^\top = S,\ S \in R^{2n\times 2n}\}.$$

The tangent bundle of $Sp(2p, 2n)$ is defined as the disjoint union of all the tangent spaces, i.e. $TSp(2p, 2n) = \cup_{X \in Sp(2p,2n)} T_X Sp(2p, 2n)$. Let $\mathcal{F} : \mathbb{R}^{2n\times 2p} \to \mathbb{R}$ be a differentiable function, we denote by $\nabla\mathcal{F}(X) := (\frac{\partial \mathcal{F}(X)}{\partial X_{ij}})$ the matrix of partial derivatives of $\mathcal{F}$ (the Euclidean gradient of $\mathcal{F}$). Let $\Phi : Sp(2p, 2n) \to \mathbb{R}$ be a smooth function defined on the symplectic Stiefel manifold, then the Riemannian gradient of Φ at $X \in Sp(2p, 2n)$, denoted by $\text{grad}\Phi(X)$, is the unique vector in $T_X Sp(2p, 2n)$ satisfying

$$\mathcal{D}\Phi(X)[\xi_X] := \lim_{\tau\to 0} \frac{\Phi(\gamma(\tau)) - \Phi(\gamma(0))}{\tau} = \langle \text{grad}\Phi(X), \xi_X\rangle, \quad \forall \xi_X \in T_X Sp(2p, 2n),$$

where $\gamma : [0, \tau_{\max}] \to Sp(2p, 2n)$ is any curve on manifold that verifies $\gamma(0) = X$ and $\dot{\gamma}(0) = \xi_X$.

In addition, at $X \in Sp(2p, 2n)$, the canonical-like metric (Son et al. 2021) associated with the symplectic Stiefel manifold is defined as

$$\langle \xi_X, \eta_X \rangle_c := tr\left(\xi_X^\top \left(\frac{1}{\rho} J_{2n} X (J_{2n} X)^\top - (J_{2n} X J_{2p} X^\top J_{2n}^\top - J_{2n})^2\right) \eta_X\right),$$

where $\xi_X, \eta_X \in T_X Sp(2p, 2n)$ and $\rho > 0$. Under the canonical-like metric, the Riemannian gradient of $\mathcal{F}$ has the following closed expression, see Gao et al. (2021),

$$\text{grad}_\rho \mathcal{F}(X) = A_X J_{2n} X, \tag{4}$$

where $A_X = 2\mathtt{sym}(H_X \nabla\mathcal{F}(X)(X J_{2p})^\top)$, $H_X = I_{2p} + \frac{\rho}{2} X X^\top + J_{2n} X (X^\top X)^{-1} X^\top J_{2n}$, and $\rho > 0$. Similar to the unconstrained optimization theory, in the Riemannian setting, the points where the Riemannian gradient is zero are candidates to be local minimizers of an optimization problem defined over a corresponding manifold. Therefore, $\text{grad}_\rho \mathcal{F}(X) = 0$ is the first-order necessary optimality condition for the problem of interest (1).

3 A computationally efficient family of retractions on *Sp*(2*p*, 2*n*)

In this section, we introduce a new constraint preserving approach to tackle the optimization problem (1). The new approach can be regarded as a generalization of the feasible curve constructed by Jiang and Dai in Jiang and Dai (2015). Afterwards, we will introduce a very efficient retraction for the symplectic Stiefel manifold, based on the new feasible curve.

Before introducing the new curve on $Sp(2p, 2n)$, let us define

$$P_X := I_{2n} + X J_{2p} X^\top J_{2n}, \tag{5}$$

for all $X \in Sp(2p, 2n)$. For $X \in Sp(2p, 2n)$ and $Z \in \mathbb{R}^{2n\times 2p}$ be an arbitrary matrix, it is easy to prove that $P_X Z \in T_X Sp(2p, 2n)$. Hence, the matrix P_X is a projection operator on the tangent space of $Sp(2p, 2n)$ at X. In addition, observe that this matrix satisfies $X^\top J_{2n} P_X = 0$, for all $X \in Sp(2p, 2n)$.

Now, given a tangent vector $Z \in T_X Sp(2p, 2n)$, we will construct a feasible curve by using the following formulation

$$Y(\tau) = (\, X R(\tau) + \tau P_X Z \,) S(\tau), \tag{6}$$

where $S : [0, \tau_{\max}] \to \mathbb{R}^{2p\times 2p}$ is an invertible curve (at least locally), that is, there exists $\tau_{\max} > 0$ such that the matrix $S(\tau)$ is invertible for all $\tau \in [0, \tau_{\max}]$; and $R : [0, \tau_{\max}] \to \mathbb{R}^{2p\times 2p}$.

Since our aim is to build a retraction map on $Sp(2n, 2p)$, we require that $Y(\tau)$ satisfies the following properties:

(A) The curve must pass through the point X, i.e. $Y(0) = X$.
(B) Local rigidity: $\dot{Y}(0) = Z$.
(C) Feasibility: $Y(\tau)^\top J_{2n} Y(\tau) = J_{2p}$, for all $\tau \in [0, \tau_{\max}]$.

Note that if we impose the initial conditions $S(0) = I_{2p}$ and $R(0) = I_{2p}$ then property (A) is guaranteed. By differentiating $Y(\tau)$ and using these initial conditions, we have the following relation

$$\dot{Y}(0) = X(\,\dot{R}(0) + \dot{S}(0)\,) + P_X Z. \tag{7}$$

Then, we can assure property (B) by imposing

$$\dot{R}(0) + \dot{S}(0) = J_{2p}^\top X^\top J_{2n} Z. \tag{8}$$

In the rest of this construction, we select the following correlative model between the curves $S(\tau)$ and $R(\tau)$,

$$R(\tau) = 2I_{2p} - S(\tau)^{-1}. \tag{9}$$

Note that this correspondence is exactly the one used in Jiang and Dai (2015).

Differentiating both sides of (9) we obtain $\dot{R}(\tau)S(\tau) + R(\tau)\dot{S}(\tau) = 2\dot{S}(\tau)$, which together with the initial conditions imply that $\dot{R}(0) = \dot{S}(0)$. Combining this last relation with (8) we arrive at

$$\dot{S}(0) = \frac{1}{2} J_{2p}^\top X^\top J_{2n} Z. \tag{10}$$

On the other hand, it follows form property (C) that

$$S(\tau)^\top (\,R(\tau)^\top X^\top + \tau Z^\top P_X^\top\,) J_{2n} (\,X R(\tau) + \tau P_X Z\,) S(\tau) = J_{2p}, \tag{11}$$

or equivalently

$$R(\tau)^\top J_{2p} R(\tau) + \tau^2 Z^\top J_{2n} P_X Z = S(\tau)^{-\top} J_{2p} S(\tau)^{-1}. \tag{12}$$

In addition, from (9) we have

$$R(\tau)^\top J_{2p} R(\tau) = 4J_{2p} - 2J_{2p} S(\tau)^{-1} - 2S(\tau)^{-\top} J_{2p} + S(\tau)^{-\top} J_{2p} S(\tau)^{-1}. \tag{13}$$

In view of the equations (12) and (13), we obtain $J_{2p} S(\tau)^{-1} + S(\tau)^{-\top} J_{2p} = 2J_{2p} + \frac{\tau^2}{2} Z^\top J_{2n} P_X Z$, which implies that $J_{2p} S(\tau)^{-1} - (J_{2p} S(\tau)^{-1})^\top = 2J_{2p} + \frac{\tau^2}{2} Z^\top J_{2p} P_X Z$. This last equation suggests that

$$J_{2p} S(\tau)^{-1} = J_{2p} + \frac{\tau^2}{4} Z^\top J_{2n} P_X Z + L(\tau), \tag{14}$$

where $L : [0, \tau_{\max}] \to \mathbb{R}^{2p \times 2p}$ must satisfy $L(0) = 0$ and $L(\tau)$ be symmetric for all $\tau \in [0, \tau_{\max}]$. Notice that the equation (14) is equivalent to

$$S(\tau) = \left(J_{2p} + \frac{\tau^2}{4} Z^\top J_{2n} P_X Z + L(\tau) \right)^{-1} J_{2p}. \tag{15}$$

Observe that if we find a formula for the curve $L(\tau)$, then the curves $S(\tau)$ and $R(\tau)$ will be completely determined. Therefore, this description reduces to finding a suitable curve $L(\tau)$.

Again, by differentiating (15) an evaluating at $\tau = 0$, we get $\dot{S}(0) = -J_{2p}^{\top}\dot{L}(0)$. Substituting this last result in (10) we arrive at

$$\dot{L}(0) = -\frac{1}{2}X^{\top}J_{2n}Z. \tag{16}$$

Therefore, we simply have to select a curve $L(\tau)$ such that $L(0) = 0$, $\dot{L}(0) = -\frac{1}{2}X^{\top}J_{2n}Z$ and $L(\tau)^{\top} = L(\tau)$.

In summary, we can write the curve $Y(\tau)$ in a simple equation as follows

$$Y(\tau) = (2X + \tau P_X Z)\left(I_{2p} - J_{2p}L(\tau) + \frac{\tau^2}{4}J_{2p}^{\top}Z^{\top}J_{2n}P_X Z\right)^{-1} - X, \tag{17}$$

where $L : [0, \infty) \to \mathcal{S}_{\mathrm{sym}}(2p)$ is any smooth curve satisfying the conditions $L(0) = 0$ and $\dot{L}(0) = -\frac{1}{2}X^{\top}J_{2n}Z$.

Finally, this formulation of $Y(\tau)$ suggests the following collection of retractions on $Sp(2p, 2n)$. Let $W : T_X Sp(2p, 2n) \to \mathcal{S}_{\mathrm{sym}}(2p)$ be a smooth function verifying that $W(0_X) = 0$ and $DW(0_X)[\xi_X] = -\frac{1}{2}X^{\top}J_{2n}\xi_X$, for all $\xi_X \in T_X Sp(2p, 2n)$. Then the curve $Y(\tau)$ is given by

$$Y(\tau) = (2X + \tau P_X Z)\left(I_{2p} - J_{2p}W(\tau Z) + \frac{\tau^2}{4}J_{2p}^{\top}Z^{\top}J_{2n}P_X Z\right)^{-1} - X, \tag{18}$$

and the corresponding retraction is

$$\mathcal{R}_X(\xi_X) = (2X + P_X\xi_X)\left(I_{2p} - J_{2p}W(\xi_X) + \frac{1}{4}J_{2p}^{\top}\xi_X^{\top}J_{2n}P_X\xi_X\right)^{-1} - X, \tag{19}$$

which is obtained from (18) by substituting τZ by $\xi_X \in T_X Sp(2p, 2n)$. Here the mapping $W(\cdot)$ plays the role of $L(\cdot)$ in equation (17). In fact, we are using the equality $L(\tau) = W(\tau Z)$. Clearly different selections of the mapping $W(\cdot)$ in (19) lead to different retractions on $Sp(2p, 2n)$. Thus, our proposal constitutes an infinite uncountable collection of retractions for the symplectic Stiefel manifold. We confirm right away that (19) is indeed a retraction.

Lemma 1 *Let $W : T_X Sp(2p, 2n) \to \mathcal{S}_{sym}(2p)$ be a smooth function verifying that $W(0_X) = 0$ and $DW(0_X)[\xi_X] = -\frac{1}{2}X^{\top}J_{2n}\xi_X$, for all $\xi_X \in T_X Sp(2p, 2n)$. Then the function $\mathcal{R} : TSp(2p, 2n) \to Sp(2p, 2n)$ defined in (19) is a retraction on $Sp(2p.2n)$.*

Proof To prove this lemma, we need to demonstrate that the map (19) satisfies Definition 4.1.1 contained in Absil et al. (2009). Firstly, notice that $\mathcal{R}_X(\cdot)$ is a smooth mapping because it is made up of products and subtractions of differentiable functions. In addition, for the construction of the curve $Y(\tau)$ in (17), we have that the mapping $\mathcal{R}(\cdot)$ preserves the manifold structure of $Sp(2p, 2n)$, i.e. $\mathcal{R}_X(\xi_X) \in Sp(2p, 2n)$, for all $X \in Sp(2p, 2n)$ and $\xi_X \in T_X Sp(2p, 2n)$. Secondly, let X in $Sp(2n, 2p)$, note that

$$\mathcal{R}_X(0_X) = 2X\left(I_{2p}\right)^{-1} - X = X. \tag{20}$$

On the other hand, the directional derivative of $\mathcal{R}_X(\cdot)$ at 0_X in direction $Z \in T_X Sp(2p, 2n)$ verifies that

$$\begin{aligned}\mathcal{D}\mathcal{R}_X(0_X)[Z] &= \lim_{\tau\to 0}\frac{\mathcal{R}_X(0_X + \tau Z) - \mathcal{R}_X(0_X)}{\tau}\\ &= \lim_{\tau\to 0}\frac{\mathcal{R}_X(\tau Z) - X}{\tau}\end{aligned}$$

$$
\begin{aligned}
&= \lim_{\tau \to 0} \frac{Y(\tau) - Y(0)}{\tau} \\
&= \dot{Y}(0) = Z.
\end{aligned}
$$

where $Y(\tau) := \mathcal{R}_X(\tau Z)$ and the last identity is obtained by the construction of the curve $Y(\tau)$. Therefore, we conclude that the mapping defined in (19) is a retraction. □

Similar to the retraction based on the Cayley transform (2), the new family of retractions (19) is not globally defined, since the matrix $M(\tau) := I_{2p} - J_{2p} W(\tau Z) + \frac{\tau^2}{4} J_{2p}^\top Z^\top J_{2n} P_X Z$ may be singular for some values of $\tau \in \mathbb{R}$. However, since I_{2p} is a non-singular matrix, and $M(\tau)$ is a continuous function on τ (it is a polynomial function), there always exists a neighborhood of $\tau = 0$ such that $M(\tau)$ is non-singular. Therefore, the new general retraction (19) is well-defined at least locally. In addition, compared to the Cayley based retraction, this new functional provides a computationally more efficient scheme, because the retraction (19) requires inverting a smaller-size matrix $2p \times 2p$, while the Cayley transform needs to invert a matrix of size $2n \times 2n$ (or $4p \times 4p$ in the best case, see equation (3)). This interesting property makes retraction (19) more attractive for solving large-scale optimization problems over $Sp(2p, 2n)$.

3.1 Connections with the Cayley transform based retraction and with scaled gradient methods

In this subsection, we establish a connection between the retraction based on the Cayley transform (2) and the general retraction presented in (19). In particular, we show that the family of retractions (19) includes the Cayley retraction as one of its members when $n = p$ and we provide some comments to justify that this result is also valid for $p < n$.

Firstly, note that the Cayley retraction (2) can be rewritten as

$$
\begin{aligned}
R_X^{\text{Cayley}}(\tau \xi_X) &= \left(I_{2n} - \frac{\tau}{2} S_X J_{2n}\right)^{-1} \left[\left(I_{2n} - \frac{\tau}{2} S_X J_{2n}\right) + \tau S_X J_{2n}\right] X \\
&= X + \tau \left(I_{2n} - \frac{\tau}{2} S_X J_{2n}\right)^{-1} \xi_X,
\end{aligned} \tag{21}
$$

where $\xi_X = S_X J_{2n} X$ and $S_X \in \mathcal{S}_{\text{sym}}(2n)$. This last equality gives us an interesting interpretation of the Riemannian linear-search methods based on the Cayley retraction for optimization over $Sp(2p, 2n)$. In particular, if we take $\xi_X = -\text{grad}_\rho \mathcal{F}(X) = -A_X J_{2n} X$, with A_X as in (4), then equation (21) becomes

$$
R_X^{\text{Cayley}}(\tau \xi_X) = X - \tau \left(I_{2n} - \frac{\tau}{2} A_X J_{2n}\right)^{-1} \text{grad}_\rho \mathcal{F}(X),
$$

which corresponds to a Riemannian gradient method. This last formula suggests that the gradient method obtained using the Cayley transform-based retraction can be interpreted as a scaled gradient method, employing $(I_{2n} - \frac{\tau}{2} A_X J_{2n})^{-1}$ as the scale matrix. This observation provides an interesting connection with the scaled gradient methods, which are very well-known in the context of unconstrained optimization.

On the flip side, notice that with the special choice $W : T_X Sp(2p, 2n) \to \mathcal{S}_{\text{sym}}(2p)$ given by $W(\xi_X) = -\frac{1}{2} X^\top J_{2n} \xi_X$, we can achieve the two properties required in the definition of our proposal (19). In fact, if $\xi_X \in T_X Sp(2p, 2n)$ then we have $\xi_X^\top J_{2n} X = -X^\top J_{2n} \xi_X$, which implies that $W(\xi_X)^\top = W(\xi_X)$, for all $\xi_X \in T_X Sp(2p, 2n)$. In addition, from the definition of $W(\cdot)$ we have directly that $W(0_X)=0$ and

$$\mathcal{D}W(0_X)[\xi_X]=\lim_{t\to 0}\frac{W(0_X+t\xi_X)-W(0_X)}{t}=\lim_{t\to 0}\frac{W(t\xi_X)}{t}=-\frac{1}{2}X^\top J_{2n}\xi_X.$$

Thus the particular selection $W(\xi_X)=-\frac{1}{2}X^\top J_{2n}\xi_X$ is a candidate for deriving a retraction, by substituting it into (19). Replacing this function in (19) and rearranging the terms, we arrive at

$$\mathcal{R}_X(\xi_X)=(\,2X+P_X\xi_X\,)\left(I_{2p}+\frac{1}{4}(\xi_X J_{2p})^\top J_{2n}(2X+P_X\xi_X)\right)^{-1}-X. \qquad (22)$$

Now let us assume that $n=p$. In this case the set $Sp(2p,2n)=Sp(2n)\equiv\{X\in\mathbb{R}^{2n\times 2n}:X^\top JX=XJX^\top=J\}$ is a Lie group (Gao et al. 2021). In this case we will omit the subscripts in matrix J since it is no longer necessary to refer to p and n separately. Since $n=p$, it is straightforward to verify that $P_X=0$, for all $X\in Sp(2n)$, with this result the functional (22) reduces to

$$\mathcal{R}_X(\xi_X)=2X\left(I_{2p}+\frac{1}{2}J^\top\xi_X^\top JX\right)^{-1}-X. \qquad (23)$$

Since $\xi_X\in T_X Sp(2n)$ then we have the equality $\xi_X^\top JX=-X^\top J\xi_X$ and we also have that $\xi_X=S_X JX$, for some symmetric matrix $S_X\in\mathbb{R}^{2n\times 2n}$.

On the other hand, let us denote by $H=JX$. Since $X\in Sp(2n)$, we have $X^\top JX=J$, which implies that $J^\top X^\top JX=J^\top J$, or equivalently, $(XJ)^\top JX=I$, which means that $H^{-1}=(XJ)^\top$. Considering the equation (23) and that $\xi_X\in T_X Sp(2n)$ we obtain

$$\begin{aligned}\mathcal{R}_X(\xi_X)&=2X\left(I-\frac{1}{2}J^\top X^\top JS_X(JX)\right)^{-1}-X\\&=2X\left(I+\frac{1}{2}J^\top X^\top J^\top S_X(JX)\right)^{-1}-X\\&=2X\left(I+\frac{1}{2}J^\top(JX)^\top S_X(JX)\right)^{-1}-X\\&=2X\left(I+\frac{1}{2}(HJ)^\top S_X H\right)^{-1}-X\\&=\left[2X-X\left(I+\frac{1}{2}(HJ)^\top S_X H\right)\right]\left(I+\frac{1}{2}(HJ)^\top S_X H\right)^{-1}\\&=X\left(I-\frac{1}{2}(HJ)^\top S_X H\right)\left(I+\frac{1}{2}(HJ)^\top S_X H\right)^{-1}. \qquad (24)\end{aligned}$$

Additionally, observe that $(HJ)^\top(JH)=J^\top H^\top JH=J^\top X^\top J^\top JJX=J^\top(X^\top JX)=I$. Finally, merging this identity with (24) we obtain that

$$\begin{aligned}\mathcal{R}_X(\xi_X)&=X\left((HJ)^\top(JH)-\frac{1}{2}(HJ)^\top S_X H\right)\left((HJ)^\top(JH)+\frac{1}{2}(HJ)^\top S_X H\right)^{-1}\\&=X(HJ)^\top\left(J-\frac{1}{2}S_X\right)HH^{-1}\left(J+\frac{1}{2}S_X\right)^{-1}(HJ)^{-\top}\\&=X(HJ)^\top\left(J-\frac{1}{2}S_X\right)\left(J+\frac{1}{2}S_X\right)^{-1}(J^\top H^{-1})^\top\end{aligned}$$

$$
\begin{aligned}
&= X(HJ)^\top \left(J - \frac{1}{2}S_X\right)\left(J + \frac{1}{2}S_X\right)^{-1}(J^\top (XJ)^\top)^\top \\
&= X(HJ)^\top \left(J - \frac{1}{2}S_X\right)\left(J + \frac{1}{2}S_X\right)^{-1} XJ^2 \\
&= -XJ^\top H^\top \left(J - \frac{1}{2}S_X\right)\left(J + \frac{1}{2}S_X\right)^{-1} X \\
&= (XJX^\top)J^\top \left(J - \frac{1}{2}S_X\right)\left(J + \frac{1}{2}S_X\right)^{-1} X \\
&= JJ^\top \left(J - \frac{1}{2}S_X\right)\left(J + \frac{1}{2}S_X\right)^{-1} X \\
&= \left(J - \frac{1}{2}S_X\right)\left(J + \frac{1}{2}S_X\right)^{-1} X \\
&= \left(J - \frac{1}{2}S_X J^\top J\right)\left(J + \frac{1}{2}S_X J^\top J\right)^{-1} X \\
&= \left(J + \frac{1}{2}S_X JJ\right)\left(J - \frac{1}{2}S_X JJ\right)^{-1} X \\
&= \left(I + \frac{1}{2}S_X J\right)\left(I - \frac{1}{2}S_X J\right)^{-1} X \\
&= \left(I - \frac{1}{2}S_X J\right)^{-1}\left(I + \frac{1}{2}S_X J\right) X, \qquad (25)
\end{aligned}
$$

where the last equality is obtained using the fact that $(I - Q)(I + Q) = (I + Q)(I - Q)$ for any square matrix Q. Notice that this formula is just the retraction based on the Cayley transform (2). It is important to mention that for the case $p < n$, the retraction (22) obtained from our general scheme (19) with the particular selection $W(\xi_X) = -\frac{1}{2}X^\top J_{2n}\xi_X$ is still equivalent to the Cayley retraction (2). To prove that, one can first apply the Sherman-Morrison-Woodbury Formula to the function (2), which leads to the form (3). If one then inverts the center block-matrix of (3) via the Schur complement, the reduced form of eq. (19) is obtained. This calculation is executed in a parallel independent work presented in Bendokat and Zimmermann (2021), see Proposition 5.2. Therefore, our proposal (19) can be seen as a generalized version of the Cayley retraction even in the case when $n \neq p$.

3.2 A novel Riemannian gradient algorithm

Evidently, the formula (19) defines a family of retractions on $Sp(2p, 2n)$, where each member of the family is obtained by a particular choice of the mapping $W(\cdot)$. As we saw in the previous subsection, one possibility is to take $W(\xi_X) = -\frac{1}{2}X^\top J_{2n}\xi_X$, for all $\xi_X \in T_X Sp(2p, 2n)$. However, this is not the unique function that satisfies the properties required for $W(\cdot)$. In fact, let $B \in \mathbb{R}^{2n\times 2n}$ be any symmetric matrix, the following selection

$$
W(\xi_X) = -\frac{1}{2}X^\top J_{2n}\xi_X + \frac{1}{4}\xi_X^\top B\xi_X, \qquad (26)
$$

also satisfies the required properties. Due to the existence of infinitely many symmetric matrices, we clearly see that (19) effectively constitutes an uncountable collection of retractions.

Another valid possibility is to use

$$W(\xi_X) = -\frac{1}{2}\sum_{i=1}^{N}(X^\top J_{2n}\xi_X)^i, \tag{27}$$

with $N \in \mathbb{N}$.

Now, focusing on (26) with $B = \sigma I_{2n}$ and $\sigma > 0$, we propose a non-monotone Riemannian gradient method to address the minimization problem (1). Specifically, we propose to construct a feasible sequence $\{X_k\}$ using the following curvilinear search iterative scheme, starting at $X_0 \in Sp(2p, 2n)$,

$$X_{k+1} = U_k\left(I_{2p} - \frac{\tau_k}{4}(Z_k J_{2p})^\top(\, J_{2n}U_k - \tau_k\sigma Z_k\,)\right)^{-1} - X_k, \tag{28}$$

where $Z_k = \mathrm{grad}_\rho\mathcal{F}(X_k)$, $U_k = 2X_k - \tau_k P_{X_k} Z_k$ and $\tau_k > 0$ is the step-size. Here, observe that the reason for taking $B = \sigma I_{2n}$ is simply to save the computation of one matrix product. To determine the k-th step-size, in this work, we use the adaptive ODH step-size originally introduced in Oviedo et al. (2021), which is given by

$$\tilde{\tau}_k^{ODH} = \begin{cases} \tau_k^{ODH2} & \text{if } \tau_k^{ODH2} \le \kappa\tau_k^{ODH1}; \\ \tau_k^{ODH1} & \text{otherwise,} \end{cases} \tag{29}$$

where $\kappa \in (0, 1)$, τ_k^{ODH1} and τ_k^{ODH2} are the following spectral step-sizes (Oviedo et al. 2021),

$$\tau_k^{ODH1} = \frac{\|S_{k-1}\|_F^2 + \theta}{\mid\langle S_{k-1}, Y_{k-1}\rangle\mid + \theta\frac{\|Y_{k-1}\|_F^2}{|\langle S_{k-1}, Y_{k-1}\rangle|}}, \tag{30}$$

and

$$\tau_k^{ODH2} = \frac{\theta\frac{\|S_{k-1}\|_F^2}{|\langle S_{k-1}, Y_{k-1}\rangle|} + \mid\langle S_{k-1}, Y_{k-1}\rangle\mid}{\theta + \|Y_{k-1}\|_F^2}, \tag{31}$$

where $\theta = 4np$, $S_{k-1} = X_k - X_{k-1}$ and $Y_{k-1} = \mathrm{grad}_\rho\mathcal{F}(X_k) - \mathrm{grad}_\rho\mathcal{F}(X_{k-1})$. As the term $\mid\langle S_{k-1}, Y_{k-1}\rangle\mid$ can be equal to zero or very close to zero, in our procedure, we include a safeguard that guarantees that the k-th step-size is neither too small nor too large. In particular, we use $\tau_k^{AODH} = \max(\min(\tilde{\tau}_k^{ODH}, \tau_M), \tau_m)$, where $0 < \tau_m < \tau_M < \infty$. In addition, since the step-size τ_k^{AODH} alone does not guarantee a sufficient decrease in the objective function value at every iteration, it may invalidate the convergence of the proposed method. However, this issue can be solved by incorporating a globalization strategy that regulates the step-size τ_k^{AODH} only when necessary (Di Serafino et al. 2018; Raydan 1997). In this work, we use the non-monotone line-search globalization technique developed by Zhang and Hager in Zhang and Hager (2004). Now we are ready to present the proposed iterative algorithm in detail, see Algorithm 1.

Algorithm 1 Riemannian gradient method.

```
Require: $X_0 \in Sp(2n, 2p), 0 < \tau_m < \tau_M < \infty, \eta \in [0, 1), c_1, \rho, \epsilon, \delta, \tau, \kappa \in (0, 1), Q_0 = 1, C_0 = \mathcal{F}(X_0)$,
  $k = 0$.
1: while $\|\mathrm{grad}_\rho \mathcal{F}(X_k)\|_F > \epsilon$ do
2:   while $\mathcal{F}(\mathcal{R}_{X_k}(-\tau \mathrm{grad}_\rho \mathcal{F}(X_k))) > C_k - c_1 \tau \|\mathrm{grad}_\rho \mathcal{F}(X_k)\|_F^2$ do
3:     $\tau = \delta\tau$,
4:   end while
5:   $X_{k+1} = \mathcal{R}_{X_k}(-\tau \mathrm{grad}_\rho \mathcal{F}(X_k))$, according to (28),
6:   Compute $\tau_k^{ODH}$ according to (29),
7:   $\tau = \max(\min(\tilde{\tau}_k^{ODH}, \tau_M), \tau_m)$,
8:   $Q_{k+1} = \eta Q_k + 1$ and $C_{k+1} = (\eta Q_k C_k + \mathcal{F}(X_{k+1}))/Q_{k+1}$.
9:   $k \leftarrow k + 1$.
10: end while
```

The convergence of the retraction-based Riemannian gradient method combined with the non-monotone Zhang–Hager globalization strategy has been demonstrated in Hu et al. (2020), Oviedo (2022). However, these convergence results are valid only for globally defined retractions. Therefore, the theorems contained in Hu et al. (2020); Oviedo (2022) do not apply directly to our Algorithm 1. Fortunately, the fact that the new retraction is not globally defined does not prevent us from adapting the convergence and complexity results of Absil et al. (2009), Hu et al. (2020), Oviedo (2022). In fact, using the same proof as in Gao et al. (2021) (see Theorem 5.6 and Corollary 5.7) proposed for the Riemannian gradient method using the retraction based on the Cayley transformation (which is also not globally defined), we have the following theoretical result.

Theorem 1 *Let $\mathcal{F} : Sp(2p, 2n) \to \mathbb{R}$ be a smooth and bounded below function. Let $\{X_k\}$ be an infinite sequence of matrices generated by Algorithm 1. Then any accumulation point X_* of $\{X_k\}$ is a stationary point of $\mathcal{F}$, i.e., $\|grad_\rho \mathcal{F}(X_*)\|_F = 0$.*

4 Numerical experiments

In this section, we show the efficiency of Algorithm 1 applying it to two different groups of experiments, considering the solution of the nearest symplectic matrix problem and also the trace minimization problem over the symplectic Stiefel manifold. We implement all the simulations in Matlab (version 2017b) with double precision on a machine intel(R) CORE(TM) i7–8750 H, CPU 2.20 GHz with 1TB HD and 16GB RAM. For comparative purposes, we test our Algorithm 1 equipped with the retraction (28) against the Riemannian gradient method based on the Cayley transformation (Cayley), the Riemannian gradient method based on the quasi-geodesic approach (Qgeodesic) (Gao et al. 2021),[1] and with our Algorithm 1 using the retraction given by (22). We will refer to this last iterative scheme as “Cayley2” because this is equivalent to the retraction based on the Cayley transform, see Bendokat and Zimmermann (2021). However, this algorithm is not equivalent to the presented in Bendokat and Zimmermann (2021) since Algorithm 1 uses a different strategy to select the step-size. For the *Cayley* and *Qgeodesic* gradient methods we use the default parameters. In Algorithm 1 and *Cayley2* we use the following default values: $\rho = 1$, $\tau_m =$ 1e-15, $\tau_M =$ 1e+15, $\eta = 0.85$, $c_1 =$ 1e-4, $\delta = 0.2$, $\sigma = 0.2$, $\tau =$ 1e-3 and $\kappa = 0.65$. In all

[1] The Riemannian gradient methods Cayley and Qgeodesic can be downloaded from https://github.com/opt-gaobin/spopt.

the experiments, we adopt the following expressions as the stopping criterion: the iterations stops if the algorithms find a matrix $\hat{X} \in Sp(2n, 2p)$ such that $\|\mathrm{grad}_\rho \mathcal{F}(\hat{X})\|_F < \epsilon$, or if the corresponding algorithm exceeds N iterations, where the values of N and ϵ will be specified for each experiment in the following subsections. The implementation of our algorithm is available in https://www.mathworks.com/matlabcentral/fileexchange/127084-retractions-for-the-symplectic-stiefel-manifold.

Throughout this section, we use the following notation: *Time*, *Iter*, *Grad*, *Feasi* denote the average total computing time in seconds, the average number of iterations, the average residual $\|\mathrm{grad}_\rho \mathcal{F}(X)\|_F$ and the average feasibility error $\|\hat{X}^\top J_{2n}\hat{X} - J_{2p}\|_F$, respectively. In all the experiment, we solve thirty independent instances for each pair (n, p) and then we report all these mean values. When we do not specify how the initial point $X_0 \in Sp(2p, 2n)$ was designed, it will be understood that we randomly generated X_0 following the strategy suggested in the subsection 6.1 in Gao et al. (2021).

4.1 The nearest symplectic matrix problem

Given an arbitrary matrix $A \in \mathbb{R}^{2n\times 2p}$, the nearest symplectic matrix problem refers to computing the symplectic matrix $X^* \in Sp(2p, 2n)$ closest to A in Frobenius norm. This problem is formulated mathematically as follows

$$\min_{X\in\mathbb{R}^{2n\times 2p}} \|X - A\|_F^2 \quad s.t. \qquad X^\top J_{2n} X = J_{2p}. \tag{32}$$

As a first experiment, we apply the methods on problem (32) considering random data. In particular, given (n, p), the matrix $A \in \mathbb{R}^{2n\times 2p}$ is assembled as $A = \bar{A}/\|\bar{A}\|_2$, where $\bar{A} \in \mathbb{R}^{2n\times 2p}$ is a matrix whose entries are sampled from the standard Gaussian distribution. Additionally, in these experiments we use $N = 1000$ and $\epsilon = 1\text{e-}5$ for all the algorithms. Table 1 reports the numerical results associated to this first test varying $p \in \{100, 200, 300, 400, 500, 600\}$ and a fixed $n = 1000$. As shown in Table 1, all the methods obtained estimates of a solution of problem (32) with the required accuracy. Furthermore, we notice that as p approaches n our proposals Algorithm 1 and Cayley2 converge more quickly (in terms of computational time) than the other two methods.

Secondly, we test the three methods on the solution of problem (32) but now using real data. We consider 30 large sparse and square matrices taken from the SuiteSparse Matrix Collection (Davis and Hu 2011).[2] Since all these data matrices are square we truncate their columns to obtain the matrices A's of appropriate size. In this experiment, we fix $p = 50$ and the matrix A will be determined as $A = \bar{A}/\|\bar{A}\|_\infty$, where $\bar{A} = M(:, 1 : 2p)$ using Matlab notation and M is the original matrix taken form the SuiteSparse matrix collection. For this second set of experiments, we use $N = 1000$ and $\epsilon = 1\text{e-}4$ in the stop criteria of the algorithms. To make this computational comparison reproducible,, we construct the starting point with the following Matlab commands (omitting the multiplication symbol):

$$\texttt{randn}('\mathrm{seed}', 1);\ W = \texttt{randn}(2p, 2p);\ W = W'W + 0.1\mathrm{eye}(2p);$$
$$E = \texttt{expm}([W(p+1 : \mathrm{end}, :); -W(1 : p, :)]);$$

and

$$X_0 = [E(1 : p, :); \texttt{zeros}(n-p, 2p); E(p+1 : \mathrm{end}, :); \texttt{zeros}(n-p, 2p)].$$

[2] The SuiteSparse Matrix Collection tool-box is available in https://sparse.tamu.edu/.

Table 1 Numerical results related to randomly generated nearest symplectic matrix problems

p	100	200	300	400	500	600
Cayley						
Iter	34.2	39.6	42.9	50.3	58	68.5
Time	1.99	6.48	15.81	33.91	61.06	90.53
NrmG	6.69e−6	5.90e−6	7.91e−6	7.70e−6	7.09e−6	7.15e−6
Feasi	2.31e−13	2.65e−13	4.15e−13	1.37e−12	6.88e−11	7.50e−11
Qgeodesic						
Iter	33.6	39.4	45.4	50.8	61.3	69.8
Time	2.46	10.26	31.68	69.88	138.73	242.47
NrmG	7.08e−6	6.38e−6	6.44e−6	6.93e−6	6.39e−6	6.99e−6
Feasi	3.09e−12	1.30e−11	2.25e−11	3.01e−11	7.01e−11	8.54e−11
Cayley2						
Iter	34.1	39.3	45.5	48.3	58.9	69
Time	2.13	7.09	16.52	28.96	51.74	85.45
NrmG	6.86e−6	5.45e−6	6.87e−6	6.23e−6	8.79e−6	7.01e−6
Feasi	1.70e−12	7.23e−12	1.32e−11	1.65e−11	3.45e−11	4.24e−11
Algorithm 1						
Iter	34.4	39.4	43.5	48.6	56.6	65.9
Time	2.21	7.68	15.52	29.14	48.75	79.20
NrmG	6.79e−6	5.11e−6	6.05e−6	5.65e−6	6.75e−6	7.21e−6
Feasi	1.67e−12	7.01e−12	1.29e−11	1.62e−11	3.39e−11	4.15e−11

The numerical results associated with this second test set are contained in Table 2. On the one hand, we note that the Cayley method implemented in Gao et al. (2021) does not achieve convergence for the following instances: *bodyy4*, *crystm03*, *Trefethen*_20000, *dw4096*, *coater2*. This is because if we use Cayley retraction based on the scheme (3), then the method fails to converge because the error in the feasibility of the iterates deteriorates, which possibly occurs due to the numerical instability of the Sherman–Morrison–Woodbury formula used in the matrix inversion of this procedure. On the other hand, we cannot directly use the Cayley retraction given by (2) since the size of the matrix A, for these 5 instances, is very large. However, Cayley2 does not suffer from numerical instability for any of the instances. Thus, our implementation of the Riemannian gradient method based on (22) is more robust than the implementation provided in Gao et al. (2021). In addition, the geodesic-based gradient method, Cayley2 and Algorithm 1 achieve the desired precision in the gradient norm for all the instances. Comparing all the methods, we clearly see that our proposals Algorithm 1 and Cayley2 are more efficient than the Qgeodesic and the original Cayley procedure, since our proposals obtain local minimum estimates in less iterations and less computational time, for most instances. In fact, the Qgeodesic method only wins for the instance *bcsstk21*. Additionally, we note that Algorithm 1 and Cayley2 show very similar numerical performance.

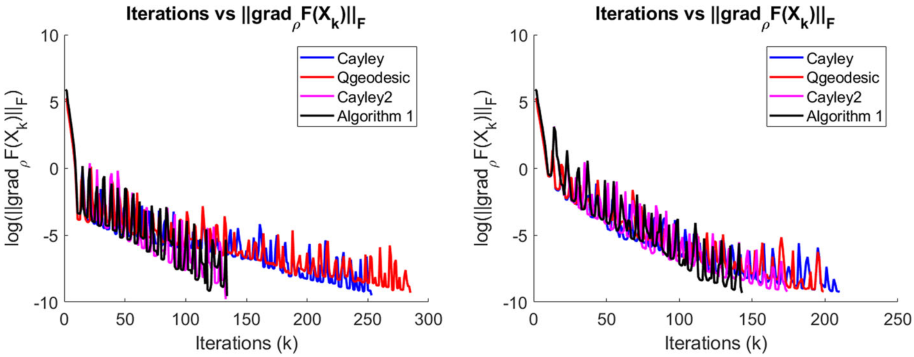
To visualize the good performance of our methods, we present in Fig. 1 the convergence history of the four methods for the instances *psmigr_1*, *nasa4704*. For these data, we observe that all the methods progressively decrease the norm of the Riemannian gradient. And among these four methods, our proposals are superior for these particular instances.

Table 2 Solving the nearest symplectic matrix problem for 30 instances in the SuiteSparse matrix collection

		Cayley				Qgeodesic				Algorithm 1				Cayley2			
Name	2n	Iter	Time	Grad	Feasi	*Iter*	Time	Grad	Feasi	*Iter*	Time	Grad	Feasi	*Iter*	Time	Grad	Feasi
1138_bus	1138	497	30.08	9.96e−5	7.60e−13	658	11.33	8.83e−5	1.30e−11	599	5.84	9.57e−5	1.13e−11	458	4.30	8.57e−5	2.72e−11
bcsstk08	1074	373	22.64	9.87e−5	7.61e−13	394	5.52	9.77e−5	6.18e−12	313	2.88	9.94e−5	1.82e−11	409	3.72	8.98e−5	6.15e−11
bcsstk10	1086	306	17.14	8.68e−5	7.59e−13	334	4.54	9.83e−5	3.31e−12	329	3.02	8.46e−05	6.88e−12	343	3.06	9.70e−5	5.31e−12
bcsstk16	4884	97	122.33	6.19e−5	7.57e−13	82	4.96	9.42e−5	7.38e−13	54	3.20	9.48e−5	6.45e−13	52	4.36	3.72e−5	7.78e−13
bcsstk21	3600	473	307.24	8.82e−5	7.57e−13	371	15.96	9.82e−5	7.22e−12	373	20.32	6.98e−5	1.60e−11	450	22.01	7.93e−5	1.25e−11
bcsstk27	1224	420	28.83	8.52e−5	7.58e−13	307	4.19	8.79e−5	6.45e−12	291	2.98	9.42e−5	9.24e−12	319	3.25	9.35e−5	9.61e−12
bodyy4	17546	failed	failed	failed	failed	217	35.89	9.74e−5	3.73e−12	78	21.10	6.57e−5	4.25e−12	80	18.79	6.93e−5	5.49e−12
crystm03	24696	failed	failed	failed	failed	91	18.31	9.03e−5	9.00e−13	56	14.63	6.52e−5	9.27e−13	56	16.40	8.36e−5	7.73e−13
Trefethen_20000	20000	failed	failed	failed	failed	132	22.59	9.11e−5	5.54e−12	44	9.55	6.83e−5	8.93e−12	42	9.28	7.89e−5	9.09e−12
ash608	608	57	0.38	8.14e−5	2.38e−13	51	0.59	5.31e−5	6.19e−13	67	0.34	2.53e−5	4.94e−13	65	0.32	8.57e−5	5.00e−13
bcspwr08	1624	49	0.89	5.17e−5	1.19e−5	57	1.49	6.22e−5	3.27e−13	40	0.82	5.29e−5	2.45e−13	40	0.82	5.42e−5	2.29e−13
can_1072	1072	68	0.62	4.66e−5	4.77e−4	78	1.10	8.89e−5	5.57e−13	50	0.45	8.88e−05	3.06e−13	46	0.42	8.65e−5	2.99e−13
dwt_2680	2680	85	2.51	3.10e−5	3.97e−11	69	2.54	6.93e−5	4.67e−13	62	2.15	7.18e−5	3.19e−13	52	1.78	9.59e−05	3.18e−13
illc1850	1850	33	0.69	8.22e−5	5.89e−14	32	0.93	8.44e−5	7.27e−13	26	0.69	2.60e−5	4.11e−13	26	0.60	2.58e−5	4.11e−13
psmigr_1	3140	252	139.14	7.51e−5	7.59e−13	284	11.21	9.12e−5	7.56e−12	133	5.56	7.12e−5	1.49e−11	132	5.50	5.67e−5	1.68e−11
psmigr_2	3140	54	1.85	8.86e−5	2.33e−9	57	2.43	6.15e−5	5.93e−13	52	2.07	6.72e−5	4.10e−13	48	1.90	2.35e−5	4.11e−13
psmigr_3	3140	31	1.11	9.89e−5	8.87e−14	32	1.38	3.53e−5	6.58e−13	28	1.25	5.76e−5	4.06e−13	28	1.11	6.24e−5	4.18e−13

Table 2 continued

Name	2n	Cayley				Qgeodesic				Algorithm 1				Cayley2			
		Iter	Time	Grad	Feasi	*Iter*	Time	Grad	Feasi	*Iter*	Time	Grad	Feasi	*Iter*	Time	Grad	Feasi
saylr4	3564	496	313.52	1.00e−4	7.59e−13	524	22.48	9.94e−5	4.15e−12	431	18.96	9.49e−5	1.15e−11	480	20.77	9.94e−5	7.97e−12
epb1	14734	46	7.46	2.84e−5	1.42e−12	45	6.45	8.29e−5	7.09e−13	39	6.66	7.35e−5	5.08e−13	42	7.12	4.25e−5	5.19e−13
dw4096	8192	failed	failed	failed	failed	625	59.27	9.86e−5	6.54e−12	426	41.61	9.85e−5	1.77e−11	394	39.04	8.06e−5	1.71e−11
olm5000	5000	334	442.46	9.96e−5	7.57e−13	424	24.59	9.62e−5	4.35e−12	403	26.07	9.99e−5	8.40e−12	387	24.92	1.00e−4	9.02e−12
rdb5000	5000	48	2.52	8.02e−5	3.40e−11	58	3.45	1.73e−5	7.16e−13	50	3.13	5.00e−5	5.41e−13	46	2.82	9.93e−5	5.49e−13
nasa4704	4707	209	252.15	9.95e−5	7.57e−13	198	11.24	9.98e−5	2.03e−12	142	8.48	9.01e−5	4.41e−12	173	10.23	9.94e−5	3.31e−12
nasa2910	2910	336	154.45	9.62e−5	7.57e−13	296	11.18	9.89e−5	6.72e−12	202	8.11	5.47e−5	1.22e−11	172	7.06	9.69e−5	9.09e−12
fs_760_3	760	274	6.95	9.39e−5	7.61e−13	319	3.32	9.18e−5	6.40e−12	339	2.37	8.06e−5	4.25e−12	266	1.90	9.61e−5	6.18e−12
jagmesh4	1440	58	0.96	9.09e−5	1.79e−12	57	1.31	5.17e−5	7.05e−13	52	1.05	7.29e−5	5.11e−13	46	0.89	8.41e−5	5.13e−13
lshp2614	2614	52	1.48	9.50e−5	2.19e−12	50	1.73	9.50e−5	7.20e−13	56	1.96	3.75e−5	5.14e−13	47	1.67	8.28e−5	5.15e−13
lshp3466	3466	56	1.95	5.29e−5	2.06e−10	52	2.33	9.67e−5	7.44e−13	54	2.42	8.43e−5	5.01e−13	55	2.46	4.25e−5	5.07e−13
coater2	9540	failed	failed	failed	failed	174	18.86	9.19e−5	7.93e−12	66	7.83	9.20e−5	9.47e−12	72	8.50	8.39e−5	1.63e−11
ex29	2870	85	35.76	6.50e−5	7.56e−13	95	5.78	2.60e−5	1.58e−12	78	2.95	9.90e−5	7.08e−13	78	2.85	7.66e−5	7.21e−13



(a) psmigr_1 (b) nasa4704

Fig. 1 Convergence history of Cayley, Quasi-geodesic, Cayley2 and Algorithm 1, from the same initial point, for the nearest symplectic matrix problem with $p = 50$

4.2 Symplectic eigenvalues computation via trace minimization

Let $A \in \mathbb{R}^{2n\times 2n}$ be a positive definite matrix. It follows from the Williamson's theorem (Williamson 1936) that there exists a symplectic matrix $V \in Sp(2n)$ such that

$$V^\top A V = \begin{bmatrix} D & 0 \\ 0 & D \end{bmatrix} \tag{33}$$

where $D = \mathtt{diag}(d_1, d_2, \ldots, d_n) \in \mathbb{R}^n$ is a diagonal matrix with positive entries. The diagonal entries d_i's are uniquely determined by A and characterise the orbits of $\mathcal{S}_+(2n)$ under the action of the Lie group $Sp(2n)$. These numbers are known as the symplectic eigenvalues of A. This kind of eigenvalues are of great importance in quantum mechanics (Dutta et al. 1995), Hamiltonian dynamics (Arnol'd 2013; Benner and Fabender 1997; Van Loan 1984), symplectic principal component analysis (Lei et al. 2016; Parra 1995), in symplectic topology (Hofer and Zehnder 2012), and in the more recent subject of quantum information; see e.g., (Eisert et al. 2008; Hiroshima 2006). It is known that the symplectic eigenvalues of A are related to the following constrained optimization problem

$$\min_{X\in\mathbb{R}^{2n\times 2p}} \mathcal{F}(X) := tr(X^\top A X) \quad \text{s.t.} \quad X^\top J_{2n} X = J_{2p}. \tag{34}$$

In particular, if X_* is a solution of the trace minimization problem (34) then $\mathcal{F}(X_*) = 2\sum_{i=1}^{p} d_i$, this result appears in Son et al. (2021), see Theorem 4.1.

In this subsection, we consider problem (34) to evaluate the effectiveness of the following methods: Cayley (Gao et al. 2021), Algorithm 1 and the symplectic Lanczos method developed in Amodio (2003), denote by *SympLanczos*. In particular, we follow the same design of the experiment reported in Son et al. (2021), Section 6.1. Specifically, the matrix A is given by $A = Q\mathtt{diag}(D, D)Q^\top$, where $D = \mathtt{diag}(d_1, d_2, \ldots, d_n)$ and $Q = KL(n/2, 1.2, -\sqrt{n/5})$, where $L(n/2, 1.2, -\sqrt{n/5}) \in Sp(2n)$ is the symplectic Gauss transformation defined in Fabender (2001), and $K \in \mathbb{R}^{2n\times 2n}$ is constructed in the same way as in Son et al. (2021). Observe that by construction of matrix A, its p smallest symplectic eigenvalues are $1, 2, \ldots, p-1, p$. To compare the precision of the estimated p smallest

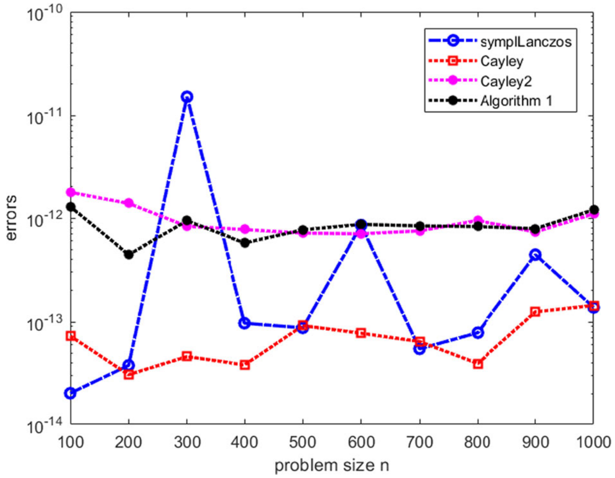


Fig. 2 The behavior of normalized residuals obtained by solving the trace minimization problem (34) with a matrix A with known symplectic eigenvalues

Table 3 The 5 smallest symplectic eigenvalues of a 2000-by-2000 positive definite matrix A, computed by the three algorithms

SympLanczos	Cayley	Algorithm 1	Cayley2
1.000000000000028	1.000000000000043	1.000000000000090	1.000000000000108
1.999999999999998	1.999999999999984	2.000000000000125	2.000000000000101
3.000000000000045	3.000000000000041	3.000000000000259	3.000000000000250
3.999999999999971	3.999999999999964	4.000000000000288	4.000000000000304
5.000000000000033	5.000000000000009	5.000000000000436	5.000000000000334

symplectic eigenvalues $\tilde{d}_1, \ldots, \tilde{d}_p$ by the algorithms, we compute the normalized residual

$$\frac{\left\| A\tilde{X}_{1:p} - J_{2n}\tilde{X}_{1:p}\begin{bmatrix} 0 & -\tilde{D}_{1:p} \\ \tilde{D}_{1:p} & 0 \end{bmatrix}\right\|_F}{\|A\tilde{X}_{1:p}\|_F},$$

where $\tilde{X}_{1:p}$ is the symplectic eigenvector matrix related to the symplectic eigenvalues $\tilde{D}_{1:p} = \mathtt{diag}(\tilde{d}_1, \ldots \tilde{d}_p)$ obtained for each method. In this experiments, we set $p = 5$ and solve (34) for different values of n in the range between 100 and 1000. For this experiment, we use $N = 5000$ and $\epsilon = 1\text{e-}9$ in the stop criteria of all the algorithms. In Fig. 2 we plot the errors obtained by all the methods for each of the values of n. Additionally, in Table 3, we report the eigenvalues computed by the four methods for $n = 1000$. As shown in Table 3, the four methods obtain estimates of the 5 smallest symplectic eigenvalues very close to the real ones.

5 Final remarks

We have presented a first-order iterative approach for minimizing smooth nonlinear functions defined on the domain of symplectic matrices $Sp(2n, 2p)$. The proposed procedure is a Riemannian gradient method, which uses a new retraction with low computational cost to preserve the feasibility of each point. The designed retraction can be regarded as a generalization of the feasible curve introduced by Jiang and Dai in Jiang and Dai (2015). The operation that requires the highest computational cost, to evaluate the new retraction, is a matrix inversion of size $2p$-by-$2p$, which is significantly less expensive than the operations required by the existing methods in the literature. To improve the numerical performance of the proposal, we consider an adaptive Barzilai–Borwein-type step-size (Oviedo et al. 2021). To guarantee the convergence of the method to critical points of the objective function (in the purely Riemannian sense), we adopt the globalization strategy of Zhang and Hager (2004).

The numerical experiments carried out indicate that the new algorithm is suitable for solving large-scale and sparse, as well as small and dense, symplectic Stiefel manifold constrained optimization problems. Moreover, we notice that the our proposal is more efficient than the two Riemannian gradient methods recently developed in Gao et al. (2021), solving trace minimization problems and computing the projection of an arbitrary matrix onto the symplectic Stiefel manifold.

Finally, since the proposed approach can be regarded as an extension of the Jiang and Dai's feasible curve for the Stiefel manifold in Jiang and Dai (2015), which unifies many constraint preserving schemes including the geodesic, projection based retraction, polar decomposition, and the QR retraction; it will remain as future work examining whether the new general approach can also unify other existing feasible curves for the symplectic Stiefel manifold, as for example the recently introduced SR retraction on $Sp(2p, 2n)$ (Gao et al. 2022).

Acknowledgements The first author was financially supported by the Faculty of Engineering and Sciences, Universidad Adolfo Ibáñez, through the FES startup package for scientific research. The second author was financially supported in part by CONACYT (Mexico), Grants 256126.

Data Availability The data used in this research is freely accessible. This can be downloaded at https://sparse.tamu.edu/.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Absil PA, Mahony R, Sepulchre R (2009) Optimization algorithms on matrix manifolds. Princeton University Press, Princeton. https://doi.org/10.1515/9781400830244

Amodio P (2003) A symplectic Lanczos-type algorithm to compute the eigenvalues of positive definite Hamiltonian matrices. In: Peter MAS, David A, Alexander VB, Yuriy EG, Jack JD, Albert YZ (eds) Lecture notes in computer science, vol 2658, pp 139–148. https://doi.org/10.1007/3-540-44862-4_16

Arnol'd VI (2013) Mathematical methods of classical mechanics. Springer, New York. https://doi.org/10.1007/978-1-4757-1693-1

Bendokat T, Zimmermann R (2021) The real symplectic Stiefel and Grassmann manifolds: metrics, geodesics and applications. arXiv:2108.12447. Accessed 30 Jan 2023

Benner P, Fabbender H (1997) An implicitly restarted symplectic Lanczos method for the Hamiltonian eigenvalue problem. Linear Algebra Appl 263:75–111. https://doi.org/10.1016/S0024-3795(96)00524-1

Benner P, Faβbender H, Watkins DS (1999) SR and SZ algorithms for the symplectic (butterfly) eigenproblem. Linear Algebra Appl 287(1-3):41–76. https://doi.org/10.1016/S0024-3795(98)10090-3

Benner P, Faβbender H (1998) The symplectic eigenvalue problem, the butterfly form, the SR algorithm, and the Lanczos method. Linear Algebra Appl 275:19-47. https://doi.org/10.1016/S0024-3795(97)10049-0

Davis TA, Hu Y (2011) The university of Florida sparse matrix collection. ACM Trans Math Softw 38(1):1–25. https://doi.org/10.1145/2049662.2049663

Del Buono N, Lopez L, Politi T (2008) Computation of functions of Hamiltonian and skew-symmetric matrices. Math Comput Simul 79(4):1284–1297. https://doi.org/10.1016/j.matcom.2008.03.011

Di Serafino D, Ruggiero V, Toraldo G, Zanni L (2018) On the steplength selection in gradient methods for unconstrained optimization. Appl Math Comput 318:176–195. https://doi.org/10.1016/j.amc.2017.07.037

Dutta B, Mukunda N, Simon R (1995) The real symplectic groups in quantum mechanics and optics. Pramana 45(6):471–497. https://doi.org/10.1007/BF02848172

Eisert J, Tyc T, Rudolph T, Sanders BC (2008) Gaussian quantum marginal problem. Commun Math Phys 280(1):263–280. https://doi.org/10.1007/s00220-008-0442-4

Faβbender H (2001) The parameterized SR algorithm for symplectic (butterfly) matrices. Math Comput 70(236):1515–1541. https://doi.org/10.1090/S0025-5718-00-01265-5

Fiori S (2016) A Riemannian steepest descent approach over the inhomogeneous symplectic group: application to the averaging of linear optical systems. Appl Math Comput 283:251–264. https://doi.org/10.1016/j.amc.2016.02.018

Gao B, Son NT, Absil PA, Stykel T (2021) Riemannian optimization on the symplectic Stiefel manifold. SIAM J Optim 31(2):1546–1575. https://doi.org/10.1137/20M1348522

Gao B, Son N, Stykel T (2022) Optimization on the symplectic Stiefel manifold: SR decomposition-based retraction and applications. arXiv:2211.09481

Hiroshima T (2006) Additivity and multiplicativity properties of some Gaussian channels for Gaussian inputs. Phys Rev A 73(1):012330. https://doi.org/10.1103/PhysRevA.73.012330

Hofer H, Zehnder E (2012) Symplectic invariants and Hamiltonian dynamics. Birkhäuser. https://doi.org/10.1007/978-3-0348-8540-9

Hu J, Liu X, Wen ZW, Yuan YX (2020) A brief introduction to manifold optimization. J Oper Res Soc China 8(2):199–248. https://doi.org/10.1007/s40305-020-00295-9

Iannazzo B, Porcelli M (2018) The Riemannian Barzilai-Borwein method with nonmonotone line search and the matrix geometric mean computation. IMA J Numer Anal 38(1):495–517. https://doi.org/10.1093/imanum/drx015

Jiang B, Dai YH (2015) A framework of constraint preserving update schemes for optimization on Stiefel manifold. Math Program 153(2):535–575. https://doi.org/10.1007/s10107-014-0816-7

Lei M, Meng G (2016) Symplectic principal component analysis: a noise reduction method for continuous chaotic systems. In: Noor A (ed) Advances in noise analysis, mitigation and control. IntechOpen, p 23. https://doi.org/10.5772/64410

Lei M, Meng G, Zhang W, Wade J, Sarkar N (2016) Symplectic entropy as a novel measure for complex systems. Entropy 18(11):412. https://doi.org/10.3390/e18110412

Nocedal J, Wright SJ (2006) Numerical optimization. Springer, New York. https://doi.org/10.1007/978-0-387-40065-5

Oviedo H (2022) Global convergence of Riemannian line search methods with a Zhang-Hager-type condition. Numer Algor. https://doi.org/10.1007/s11075-022-01298-8

Oviedo H (2022) Implicit steepest descent algorithm for optimization with orthogonality constraints. Optim Lett 16(6):1773–1797. https://doi.org/10.1007/s11590-021-01801-5

Oviedo H, Dalmau O, Herrera R (2021) Two novel gradient methods with optimal step sizes. J Comput Math 39(3):375–391. https://doi.org/10.4208/jcm.2001-m2018-0205

Oviedo H, Dalmau O, Lara H (2021) Two adaptive scaled gradient projection methods for Stiefel manifold constrained optimization. Numer Algor 87(3):1107–1127. https://doi.org/10.1007/s11075-020-01001-9

Oviedo H, Lara H, Dalmau O (2019) A non-monotone linear search algorithm with mixed direction on Stiefel manifold. Optim Methods Softw 34(2):437–457. https://doi.org/10.1080/10556788.2017.1415337

Parra L (1995) Symplectic nonlinear component analysis. Adv Neural Inf Process Syst 1995:437–443

Peng L, Mohseni K (2016) Symplectic model reduction of Hamiltonian systems. SIAM J Sci Comput 38(1):A1–A27. https://doi.org/10.1137/140978922

Raydan M (1997) The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. SIAM J Optim 7(1):26–33. https://doi.org/10.1137/S1052623494266365

Salam A, Al-Aidarous E (2014) Equivalence between modified symplectic Gram-Schmidt and Householder SR algorithms. BIT Numer Math 54(1):283–302. https://doi.org/10.1007/s10543-013-0441-5

Son NT, Absil PA, Gao B, Stykel T (2021) Computing symplectic eigenpairs of symmetric positive-definite matrices via trace minimization and Riemannian optimization. SIAM J Matrix Anal Appl 42(4):1732–1757. https://doi.org/10.1137/21M1390621

Van Loan C (1984) A symplectic method for approximating all the eigenvalues of a Hamiltonian matrix. Linear Algebra Appl 61:233–251. https://doi.org/10.1016/0024-3795(84)90034-X

Wen Z, Yin W (2013) A feasible method for optimization with orthogonality constraints. Math Program 142(1):397–434. https://doi.org/10.1007/s10107-012-0584-1

Williamson J (1936) On the algebraic problem concerning the normal forms of linear dynamical systems. Am J Math 58(1):141–163. https://doi.org/10.2307/2371062

Wu R, Chakrabarti R, Rabitz H (2008) Optimal control theory for continuous-variable quantum gates. Phys Rev A 77(5):052303. https://doi.org/10.1103/PhysRevA.77.052303

Zhang H, Hager WW (2004) A nonmonotone line search technique and its application to unconstrained optimization. SIAM J Optim 14(4):1043–1056. https://doi.org/10.1137/S1052623403428208

Zhao L (2022) Linear constraint problem of Hermitian unitary symplectic matrices. Linear Multilinear Algebra 70(8):1423–1441. https://doi.org/10.1080/03081087.2020.1762533

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.