Amplitude Modulated Riemannian Optimization for QAP

Timothee Leleu TIMOTHEE.LELEU@NTT-RESEARCH.COM

NTT Research, Sunnyvale, CA, USA and Stanford University, CA, USA

Aron Vizkeleti
Notre Dame University, IN, USA

Samuel Reifenstein
NTT Research, Sunnyvale, CA, USA

Abstract

We introduce a novel approach for solving the Quadratic Assignment Problem (QAP) by combining Riemannian optimization and control of soft-spin amplitude. By reformulating the QAP as an optimization problem on the Stiefel manifold, we leverage its geometric structure to define Riemannian gradients over continuous variables while inherently satisfying orthogonality constraints. Additional permutation matrix constraints are enforced using auxiliary variables within a descent-ascent framework, ensuring that solutions remain within the feasible set. Numerical simulations demonstrate the effectiveness of our method in finding globally optimal solutions.

1. Problem formulation

In this work, we aim to achieve global optimization of constrained non-convex problems using a combination of Riemannian optimization[12] and the modulation of soft-spin amplitudes, reminiscent of descent-ascent or primal-dual method in Lagragian optimization[14, 26]. Specifically, we tackle the following problem:

$$\underset{x \in \mathcal{S}}{\text{minimize}} \quad f(x) \tag{1}$$

where $f : \mathcal{S} \subseteq \mathbb{R} \to \mathbb{R}$ is a non-convex function defined on a set of discrete elements. Such problems are typical of NP-Hard combinatorial optimization. Relaxation to an continuous state is useful in many approximate or heuristic algorithms, such as semi-definite programming[8], analog-state solvers[9, 11, 14, 26], and message-passing algorithms[17]. Algorithms based on relaxation have shown competitive performance against discrete-state solvers when $\mathcal{S}$ is simple enough, such as binary states without constraints, as seen in QUBO solvers[9, 14, 26], motivating their implementation on special-purpose analog hardware[26]. However, the generalization to more complex discrete sets, such as permutation matrices, has been more elusive. Here, we show that the combination of Riemannian optimization and dynamic correction of soft-spin amplitude similar to ascent-descent (or primal-dual) algorithm allows extending the recent high-performance solvers to a broader set of applications.

Although our methodology is general, in this work we apply it specifically to the Quadratic Assignment Problem (QAP) in which $\mathcal{S}$ is a set of permutations and f is a quadratic function. The QAP is itself quite general, as it models a variety of important problems such as the Traveling Salesman Problem, the Facility Layout Problem, and the Graph Partitioning Problem.

2. Related works

2.1. Quadratic Assignment Problem

The seminal work on solving constrained non-convex optimization problems using gradient descent on a quadratic cost function with augmented Lagrangian penalty terms [11] was quickly demonstrated to underperform [29] when compared to alternative Markov Chain Monte Carlo (MCMC)-based approaches [3, 4, 28]. This led to the development of various other heuristic methods, including tabu search [16, 24, 27], genetic algorithms [6], ant colony optimization [7, 25], iterated local search [22], alternative neural network-based approaches [10, 13, 23], and other meta-heuristics [5].

2.2. Riemannian optimization

While continuous-state dynamics offer the distinctive advantage of providing a smooth trajectory from the origin to the solution space, a significant challenge in methods based on continuous-state spaces lies in connecting the subspaces of solutions that satisfy the imposed constraints. Techniques such as the valid subspace method [2] aim to identify linear subspaces connecting feasible solutions. Riemannian optimization [12] generalizes this concept to arbitrary manifolds, enabling optimization over curved spaces that inherently satisfy the constraints. Current state-of-the-art approaches focus on developing efficient algorithms tailored to specific manifolds which are crucial in applications like machine learning, signal processing, and computer vision [1, 18].

2.3. Amplitude control methods

To achieve improved mapping between continuous and discrete states, an alternative to the Lagrangian penalty approach is the introduction of auxiliary variables, which ensure that the continuous state converges to a discrete one. This technique has been utilized in gradient-descent-ascent methods [30] and primal-dual Lagrangian approaches [26]. Notably, amplitude control of soft spin dynamics has shown strong performance in solving QUBO [14, 15, 19, 26] and Boolean satisfiability problems [20]. However, few studies have combined primal-dual inspired methods with Riemannian optimization for non-convex combinatorial optimization [21], presenting a promising direction for future research.

3. Method

3.1. Riemmanian optimization

In the following, we show how to reformulate the QAP problem into a Riemannian optimization problem. The QAP problem with n sites is defined as follows:

$$\min_{P} \quad f(P) := \mathrm{Tr}(DPT^T P^T) + \mathrm{Tr}(P^T C) \tag{2a}$$

$$\text{s.t.} \quad P \text{ is a permutation matrix.} \tag{2b}$$

where: $D \in \mathbb{R}^{n\times n}$ is the distance matrix; $T \in \mathbb{R}^{n\times n}$, the interaction matrix (e.g., representing a cyclic graph in the case of TSP with $C = 0$); C, the linear cost of placing a facility at site j.

We reformulate QAP as a Riemannian optimization problem on the square Stiefel manifold $\mathcal{O}$ to breakdown constraint in relation to the algorithm's design:

$$\min_{Z \in \mathbb{R}^{n\times n}} \quad f_Z(Z) := \text{Tr}(DZT^TZ^T) + \text{Tr}(Z^TC) \tag{3a}$$

s.t.

$$Z^TZ = I \ (Z \in \mathcal{O}), \tag{3b}$$

$$Z \in \{0,1\}^{n\times n}. \tag{3c}$$

3.2. Amplitude modulated Riemannian gradient

We consider the relaxation of Z to $Y \in \mathbb{R}^{n\times n}$ with Y a matrix of "soft-spins" and define the following coupled ordinary differential equations in Y constituting the algorithm for solving QAP as formulated in eqs. (3a-3b):

$$\frac{dY}{dt} = -\frac{2}{a^2}\nabla_{\mathcal{O}} f_Y^c(Y,E) \qquad \text{(Riemannian gradient descent)}, \tag{4}$$

$$\frac{dE}{dt} = -\xi((2(Y+b)-a)^2 - a^2)E \qquad \text{(Amplitude modulation)}\ , \tag{5}$$

where the Riemannian gradient on the square Stiefel manifold is given as follows (see supplementary materials):

$$\nabla_{\mathcal{O}} f_c(Y,E) = Y\{Y, \nabla f_c(Y,E)\}, \tag{6}$$

where $\{A,B\}$ is the Lie bracket $\{A,B\} = \frac{1}{2}(A^TB - B^TA)$ and $\nabla f_Y^c(Y,E)$ the Euclidian gradient of the cost function f_Y^c. The function f_Y^c is given as follows:

$$f_Y^c(Y) = \epsilon E \circ [f_Y(Y) + q(f_Y^1(Y) + f_Y^2(Y))], \tag{7}$$

where $f_Y(Y)$ is equal to f_Z evaluated at $Z = \frac{Y+b}{a}$. The symbol $\circ$ denotes the Hadamard product. To help guide the direction of the Euclidian gradient, the cost function is penalized by the auxiliary costs $f_Y^1(Y)$ and $f_Y^2(Y)$ that correspond to the constraints of a doubly stochastic matrix. That is, $f_Y^1(Y)$ and $f_Y^2(Y)$ are equal to $f_Z^1(Z) = (Z\mathbf{1}-\mathbf{1})^T(Z\mathbf{1}-\mathbf{1})$ and $f_Z^2(Z) = (Z^T\mathbf{1}-\mathbf{1})^T(Z^T\mathbf{1}-\mathbf{1})$, respectively ($\mathbf{1}$ is the vector of all ones). These constraints are satisfied at the steady-state, but adding these penalty terms in the cost function improve numerical stability. The parameters q, ϵ, ξ are positive constants.

Note that when Z is orthogonal ($Z^TZ = I$), Y is such that $Y^TY = a^2I + b(bN - 2a)J$. We choose in the following $b = \frac{2a}{N}$, in which case we have $Y^TY = a^2I$.

The dynamics of auxiliary variables E (see eq. (5)) ensure that constraint of a binary state (see eq. (3c)) is satisfied at the steady-state. The Riemannian gradient of eq. (4) satisfy condition of orthogonality (see eq. (3b)) at all times. When simulating the ODEs, a retraction operator is applied at fixed intervals to correct for numerical errors. This is done using a scaled Newton-Schulz Iteration as follows (I is the identity matrix):

$$Y(t+\mathrm{d}t) = \frac{1}{2} Y(t)\left(3I - Y(t)^{\top} Y(t)\right). \tag{8}$$

4. Experiments

To illustrate the effectiveness of combining the Riemannian gradient with amplitude control, numerical simulations of the convergence to a permutation matrix is shown Fig. 1.

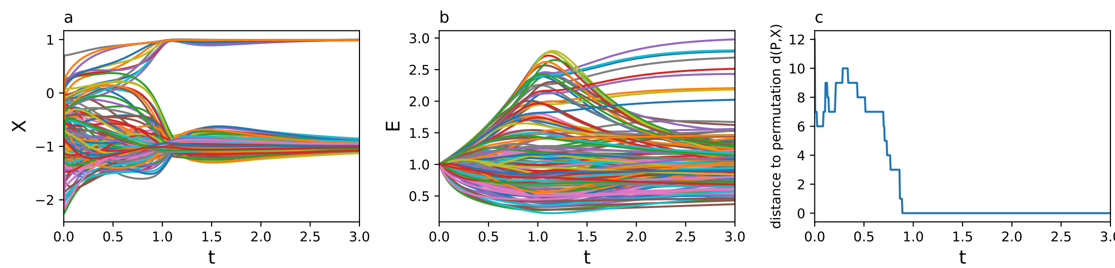

Figure 1: Numerical simulations of ODEs shown in eqs. (4-6) showing $X(t)$ (a), $Y(t)$ (b), and the distance to the closest permutation matrix. The matrix-state quickly converges from a random orthogonal matrix to a permutation matrix staying on the Stiefel manifold. Each colored line correspond to $Y_{ij}(t)$. $n = 12$.

In practice, the parameters are set such that dynamics exhibits chaotic behavior. We evaluate the algorithm on the QAP library[1]. The algorithm is solely deterministic and optimal solutions are found after chaotic transient dynamics, as shown in Fig. 2 and 3.

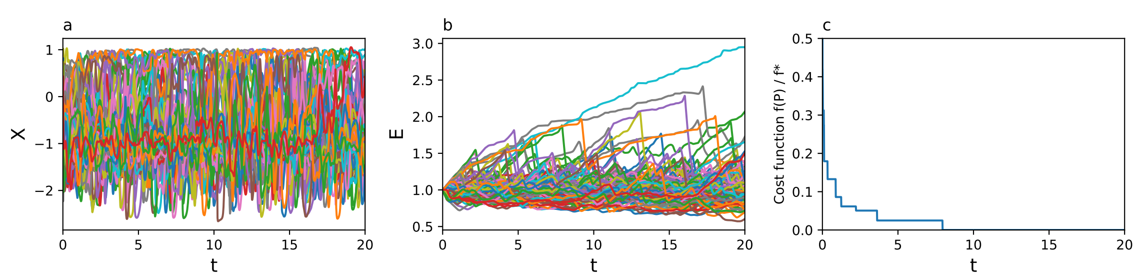

Figure 2: (a) and (b) are the same as in Fig. 1. 144 variables are shown per panel. (c) Ratio $f(P)/f^*$ to the optimal cost f^*. $n = 12$ (Tai12a). The optimal solution is found after approximately 10 normalized time units.

1. https://coral.ise.lehigh.edu/data-sets/qaplib/

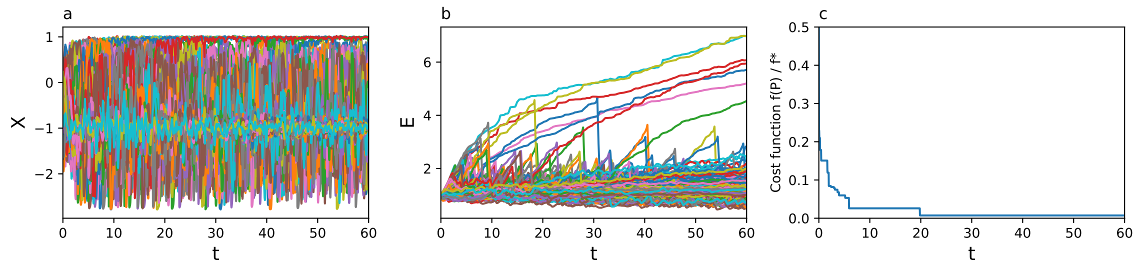

Figure 3: The same as Fig. 3 for $n = 20$ (Tai20a).

We conducted ablation experiments to evaluate the impact of the Riemannian gradient on the quality of the solutions obtained. In Fig. 4, the density of the cost $f(P)$ for solutions found with and without auxiliary variables e (i.e., $E = 1$ and $\xi = 0$) and the Riemannian gradient is compared. The highest solution quality is achieved when auxiliary variables are used in conjunction with the Riemannian gradient.

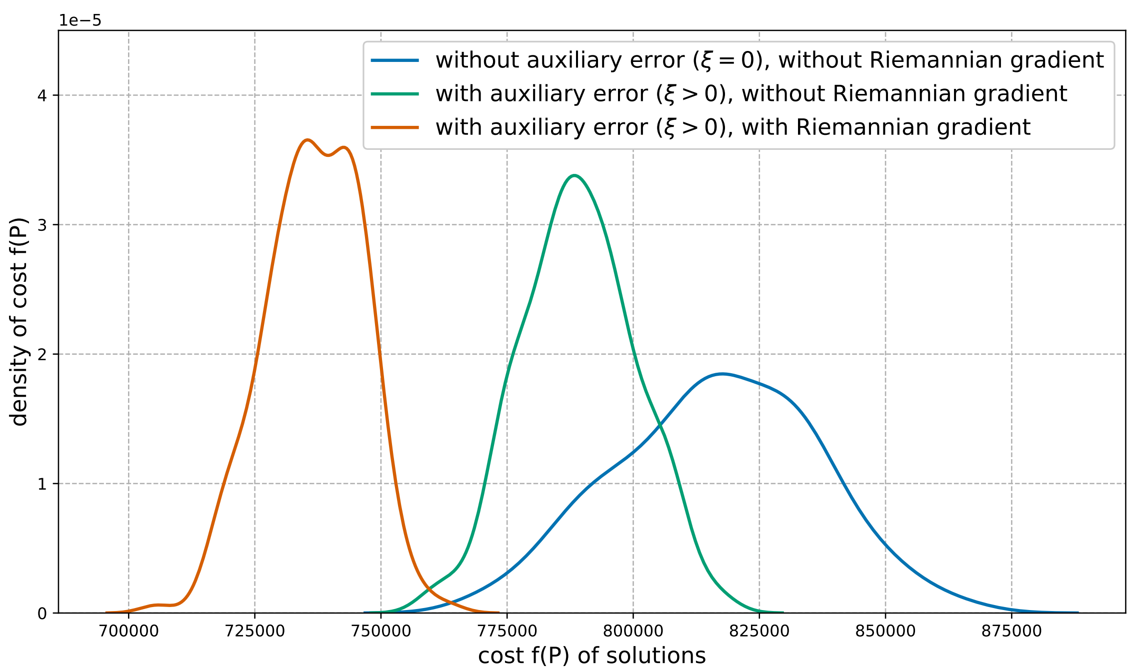

Figure 4: Density of the cost $f(P)$ of solutions found with and without auxiliary variables correcting amplitude heterogeneity (i.e., $E = 1$ and $\xi = 0$) and the Riemannian gradient. In the case without Riemannian, a simple Euclidient gradient is used instead. $n = 20$ (Tai20a).

5. Discussion

The proposed framework demonstrates a promising application of differential geometry and continuous flow for solving NP-Hard combinatorial optimization problems. The algorithm described in this paper is particularly suited for special-purpose analog hardware and is highly parallelizable on

GPUs, given its reliance on matrix-matrix multiplications. Future work will focus on optimizing the GPU implementation to assess the scalability of the approach on large-scale QAP problems.

References

[1] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.

[2] Sreeram VB Aiyer, Mahesan Niranjan, and Frank Fallside. A theoretical investigation into the performance of the hopfield model. *IEEE transactions on neural networks*, 1(2):204–215, 1990.

[3] Rainer E Burkard and Franz Rendl. A thermodynamically motivated simulation procedure for combinatorial optimization problems. *European Journal of Operational Research*, 17(2): 169–174, 1984.

[4] David T Connolly. An improved annealing scheme for the qap. *European Journal of Operational Research*, 46(1):93–100, 1990.

[5] Jonathan Duque, Danny A Múnera, Daniel Díaz, and Salvador Abreu. Solving qap with auto-parameterization in parallel hybrid metaheuristics. In *International Conference on Optimization and Learning*, pages 294–309. Springer, 2021.

[6] Charles Fleurent, Jacques A Ferland, et al. Genetic hybrids for the quadratic assignment problem. *Quadratic assignment and related problems*, 16:173–187, 1993.

[7] Luca Maria Gambardella, É D Taillard, and Marco Dorigo. Ant colonies for the quadratic assignment problem. *Journal of the operational research society*, 50(2):167–176, 1999.

[8] Michel X Goemans and David P Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995.

[9] Hayato Goto, Kosuke Tatsumura, and Alexander R Dixon. Combinatorial optimization by simulating adiabatic bifurcations in nonlinear hamiltonian systems. *Science advances*, 5(4): eaav2372, 2019.

[10] Mikio Hasegawa, Tohru Ikeguchi, Kazuyuki Aihara, and Kohji Itoh. A novel chaotic search for quadratic assignment problems. *European Journal of Operational Research*, 139(3):543–556, 2002.

[11] John J Hopfield and David W Tank. "neural" computation of decisions in optimization problems. *Biological cybernetics*, 52(3):141–152, 1985.

[12] Jiang Hu, Xin Liu, Zai-Wen Wen, and Ya-Xiang Yuan. A brief introduction to manifold optimization. *Journal of the Operations Research Society of China*, 8:199–248, 2020.

[13] Shin Ishii and Hirotaka Niitsuma. λ-opt neural approaches to quadratic assignment problems. *Neural computation*, 12(9):2209–2225, 2000.

[14] Timothée Leleu, Yoshihisa Yamamoto, Peter L McMahon, and Kazuyuki Aihara. Destabilization of local minima in analog spin systems by correction of amplitude heterogeneity. *Physical review letters*, 122(4):040607, 2019.

[15] Timothée Leleu, Farad Khoyratee, Timothée Levi, Ryan Hamerly, Takashi Kohno, and Kazuyuki Aihara. Scaling advantage of chaotic amplitude control for high-performance combinatorial optimization. *Communications Physics*, 4(1):266, 2021.

[16] Alfonsas Misevicius. An implementation of the iterated tabu search algorithm for the quadratic assignment problem. *OR spectrum*, 34(3):665–690, 2012.

[17] Andrea Montanari. Optimization of the sherrington–kirkpatrick hamiltonian. *SIAM Journal on Computing*, (0):FOCS19–1, 2021.

[18] Magnus Rattray, David Saad, and Shun-ichi Amari. Natural gradient descent for on-line learning. *Physical review letters*, 81(24):5461, 1998.

[19] Sam Reifenstein, Satoshi Kako, Farad Khoyratee, Timothée Leleu, and Yoshihisa Yamamoto. Coherent ising machines with optical error correction circuits. *Advanced Quantum Technologies*, 4(11):2100077, 2021. doi: https://doi.org/10.1002/qute.202100077. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/qute.202100077`.

[20] Sam Reifenstein, Timothee Leleu, Timothy McKenna, Marc Jankowski, Myoung-Gyun Suh, Edwin Ng, Farad Khoyratee, Zoltan Toroczkai, and Yoshihisa Yamamoto. Coherent sat solvers: a tutorial. *Advances in Optics and Photonics*, 15(2):385–441, 2023.

[21] Tuhin Sahai, Adrian Ziessler, Stefan Klus, and Michael Dellnitz. Continuous relaxations for the traveling salesman problem. *Nonlinear Dynamics*, 97:2003–2022, 2019.

[22] Thomas Stützle. Iterated local search for the quadratic assignment problem. *European journal of operational research*, 174(3):1519–1539, 2006.

[23] Toshihiro Tachibana and Masaharu Adachi. Gpu acceleration of method for solving optimization problems using chaotic neural networks. *IEICE Proceedings Series*, 47(B3L-D-4), 2015.

[24] Éric Taillard. Robust taboo search for the quadratic assignment problem. *Parallel computing*, 17(4-5):443–455, 1991.

[25] E-G Talbi, Olivier Roux, Cyril Fonlupt, and Denis Robillard. Parallel ant colonies for the quadratic assignment problem. *Future Generation Computer Systems*, 17(4):441–449, 2001.

[26] Sri Krishna Vadlamani, Tianyao Patrick Xiao, and Eli Yablonovitch. Physics successfully implements lagrange multiplier optimization. *Proceedings of the National Academy of Sciences*, 117(43):26639–26650, 2020.

[27] Thé Van Thé Van Luong, Lakhdar Loukil, Nouredine Melab, and El-Ghazali Talbi. A gpu-based iterated tabu search for solving the quadratic 3-dimensional assignment problem. In *ACS/IEEE International Conference on Computer Systems and Applications-AICCSA 2010*, pages 1–8. IEEE, 2010.

[28] Mickey R Wilhelm and Thomas L Ward. Solving quadratic assignment problems by 'simulated annealing'. *IIE transactions*, 19(1):107–119, 1987.

[29] GV Wilson and GS Pawley. On the stability of the travelling salesman problem algorithm of hopfield and tank. *Biological Cybernetics*, 58(1):63–70, 1988.

[30] Guodong Zhang, Yuanhao Wang, Laurent Lessard, and Roger B Grosse. Near-optimal local convergence of alternating gradient descent-ascent for minimax optimization. In *International Conference on Artificial Intelligence and Statistics*, pages 7659–7679. PMLR, 2022.

Supplementary materials

Riemannian gradient descent: some explanations

Let $\mathcal{M}$ be a Riemannian manifold with a Riemannian metric g_x at each point $x \in \mathcal{M}$. For a smooth function $f : \mathcal{M} \to \mathbb{R}$, the Riemannian gradient of f at a point $x \in \mathcal{M}$, denoted by $\nabla_x f \in T_x\mathcal{M}$, is the unique vector in the tangent space $T_x\mathcal{M}$ that satisfies:

$$g_x(\nabla_x f, v) = df_x(v) \quad \forall v \in T_x\mathcal{M}$$

where $df_x : T_x\mathcal{M} \to \mathbb{R}$ is the differential of f at x, representing the directional derivative of f in the direction of $v \in T_x\mathcal{M}$, given as follows:

$$df_x(v) = \lim_{t \to 0} \frac{f(\gamma(t)) - f(x)}{t}$$

where $\gamma(t)$ is a smooth curve on the manifold $\mathcal{M}$ such that $\gamma(0) = x$ and $\dot{\gamma}(0) = v$. Intuitively, $df_x(v)$ represents the rate of change of f at x in the direction of the tangent vector v. $g_x(u, v)$ is the Riemannian metric, providing the inner product between two tangent vectors u and v in $T_x\mathcal{M}$.

In local coordinates $(x^1, x^2, \ldots, x^n)$ around a point $x \in \mathcal{M}$, the Riemannian metric is represented by the positive-definite metric tensor $g_{ij}(x)$. The Riemannian gradient in local coordinates is computed by:

$$\nabla_x f = g_x^{-1} \nabla f(x)$$

where g_x^{-1} is the inverse of the metric tensor $g_{ij}(x)$, $\nabla f(x)$ is the Euclidean gradient of f in these local coordinates. In the case of the Stiefel manifold, the metric tensor is simply the identity matrix.

Retraction Operator

The retraction operator $\mathcal{R}_x(v)$ is a map that projects a tangent vector $v \in T_x\mathcal{M}$ back to the manifold $\mathcal{M}$. It provides an approximation of the exponential map, which moves along geodesics on the manifold, but is often computationally simpler. The retraction ensures that after a gradient descent step in the tangent space, the next point lies on the manifold. Mathematically, it is defined as:

$$\mathcal{R}_x(v) : T_x\mathcal{M} \to \mathcal{M}, \quad \mathcal{R}_x(v) \approx \exp_x(v)$$

where $\exp_x(v)$ is the exponential map, which follows the geodesic curve starting at x in the direction v. The exponential map $\exp_x(v)$ takes a tangent vector $v \in T_x\mathcal{M}$ at a point $x \in \mathcal{M}$ and maps it to a point on the manifold $\mathcal{M}$ by following the geodesic emanating from x in the direction of v.

Inner Product on the Stiefel Manifold

On the Stiefel manifold $\mathcal{O}(n, n)$, the Riemannian metric at a point $X \in \mathcal{O}(n, n)$ is the standard Euclidean inner product between two tangent vectors $A, B \in T_X\mathcal{O}(n, p)$. The inner product is defined as:

$$g_X(A, B) = \mathrm{Tr}(A^\top B),$$

where $\mathrm{Tr}(\cdot)$ denotes the matrix trace, and $A^\top$ is the transpose of A.

To derive the tangent space condition, we recall that the Stiefel manifold consists of matrices $X \in \mathbb{R}^{n\times n}$ with orthonormal columns, i.e., $X^\top X = I_n$. Let $X(t)$ be a curve on the Stiefel manifold such that $X(0) = X$ and $\dot{X}(0) = V \in T_X\mathcal{O}(n,n)$. Differentiating the orthogonality condition $X(t)^\top X(t) = I_n$ with respect to t at $t = 0$, we obtain:

$$\frac{d}{dt}\left(X(t)^\top X(t)\right)\Big|_{t=0} = 0,$$

which leads to:

$$X^\top V + V^\top X = 0.$$

This is the condition that any tangent vector $V \in T_X\mathcal{S}(n,p)$ must satisfy.

Symmetric and Antisymmetric Components

From the tangent space condition $X^\top V + V^\top X = 0$, we see that $X^\top V$ must be antisymmetric. Given the Euclidean gradient $\nabla f(X)$, we need to project it onto the tangent space to ensure that the resulting vector satisfies the tangent space condition.

We can achieve this by removing the symmetric part of $X^\top G$ for any matrix $G \in \mathbb{R}^{n\times n}$. The projection of G onto the tangent space is given by:

$$\mathrm{Proj}_{T_X\mathcal{S}(n,p)}(G) = G - X\,\mathrm{sym}(X^\top G),$$

where $\mathrm{sym}(X^\top G)$ is the symmetric part of $X^\top G$, defined as:

$$\mathrm{sym}(X^\top G) = \frac{1}{2}(X^\top G + G^\top X).$$

Thus, the Riemannian gradient on the Stiefel manifold is:

$$\nabla_x f = \nabla f(X) - X\,\mathrm{sym}(X^\top \nabla f(X)), \tag{9}$$

$$\nabla_x f = X\{X, \nabla f(X)\}. \tag{10}$$

This ensures that the Riemannian gradient lies in the tangent space by removing the symmetric component of $X^\top \nabla f(X)$.