

Riemannian Trust Region Method for Haplotype Assembly

Mohamad Mahdi Mohades and Mohammad Hossein Kahaei

Abstract—In this letter we model the Haplotype assembly problem (HAP) as a maximization problem over an $(n - 1)$ -dimensional sphere. Due to nonconvexity of the feasible set, we propose a manifold optimization approach to solve the mentioned maximization problem. To escape local maxima as well as saddle points we utilize trust region method. Simulation results show that our proposed method is with high accuracy in estimation of Haplotype.

Index Terms—Haplotype assembly, Manifold optimization, Riemannian trust region.

I. INTRODUCTION

HAPLOTYPE is a string of single nucleotide polymorphisms (SNPs) of chromosomes [1]. Haplotypes are useful in studying human evolutionary history and drug discovery and development. Haplotypes can be assembled utilizing sequenced reads, where each read is a fragment of the two chromosomes [2]. Such assembly can be performed by solving mathematical models of sequenced reads. For diploid organisms, each SNP site is considered to be either $+1$ or -1 . This means that the result of haplotype assembly from the sequenced reads ought to be a vector of ± 1 elements. Moreover, the obtained haplotype, say $\mathbf{h}_{n \times 1}$, is corresponding to one of the two chromosomes and $-\mathbf{h}$ is corresponding to the other one.

One mathematical approach to find the haplotype is matrix completion. In this approach a read matrix, say $\mathbf{M}_{m \times n}$, containing sequenced reads is created. The elements of this matrix are either ± 1 corresponding to the reads or \times where there is no read. Then, it is required to replace the \times elements of the matrix \mathbf{M} with $+1$ or -1 so that the rank of the newly generated matrix, say $\overline{\mathbf{M}}$, be one. Then the factorization $\overline{\mathbf{M}} = \overline{\mathbf{c}}_{m \times 1} \overline{\mathbf{h}}_{n \times 1}^T$ gives the haplotype $\mathbf{h}_{n \times 1}$ through applying sign function over $\overline{\mathbf{h}}_{n \times 1}$. When the reads are affected by noise, the sign of some of the entries of the matrix $\mathbf{M}_{m \times n}$ is changed. In this case using the following optimization problem the completed matrix $\overline{\mathbf{M}}$ is obtained. Please note that in HAP, whether to estimate \mathbf{h} or $-\mathbf{h}$, the estimation is accurate.

$$\begin{aligned} \min_{\mathbf{X}} \quad & \|P_{\Omega}(\mathbf{M}) - P_{\Omega}(\mathbf{X})\|_F^2 \\ \text{subject to} \quad & \text{rank}(\mathbf{X}) = 1, \end{aligned} \quad (1)$$

where P_{Ω} is the sampling operator acting as follows

$$P_{\Omega}(\mathbf{Q}) = \begin{cases} P_{\Omega}(q_{ij}) = q_{ij} & (i, j) \in \Omega \\ P_{\Omega}(q_{ij}) = 0 & (i, j) \notin \Omega \end{cases}, \quad (2)$$

and Ω is the set of the observations. There exist some approaches to solve the problem (1). Based on [3], the problem

(1) can be directly solved over the manifold of rank-one matrices. Also, a variation of the formulation of the problem (1) has been solved in [4] over the Cartesian product of Grassmann manifolds. In fact, instead optimizing over the variable \mathbf{X} , the problem is solved over the variables \mathbf{U} , Σ and \mathbf{V} where $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^T$ is the singular value decomposition of \mathbf{X} . Consider the following equivalent optimization problem for the problem (1).

$$\begin{aligned} \min_{\mathbf{X}} \quad & \text{rank}(\mathbf{X}) \\ \text{subject to} \quad & \|P_{\Omega}(\mathbf{M}) - P_{\Omega}(\mathbf{X})\|_F^2 < \delta, \end{aligned} \quad (3)$$

where δ is a given value. In [5] the nonconvex objective function $\text{rank}(\mathbf{X})$ has been replaced by the nuclear norm and the minimization problem turns into a convex one [6]. Moreover, the problem (1) can be modeled as the following optimization problem.

$$\min_{\mathbf{u}, \mathbf{v}} \quad \|P_{\Omega}(\mathbf{M}) - P_{\Omega}(\mathbf{u}_{m \times 1} \mathbf{v}_{n \times 1}^T)\|_F^2, \quad (4)$$

Alternating minimization is utilized to solve the above non-convex problem [7]. In [8], the authors have given an example to show that the proposed objective function of the problem (1) is not a reliable cost function in haplotype assembly. They instead have proposed an objective function to truly model the haplotype assembly problem. Using simulation results they have shown their approach is more accurate than previously proposed algorithms in haplotype assembly.

Apart from the aforementioned optimization approaches, it has been proposed to minimize the following minimum error correction (MEC) function to estimate the haplotype.

$$\text{MEC}(\mathbf{M}, \mathbf{z}) = \sum_{i=1}^m \min(hd(\mathbf{m}_i, \mathbf{z}), hd(\mathbf{m}_i, -\mathbf{z})), \quad (5)$$

where \mathbf{m}_i is the i -th row of \mathbf{M} and

$$hd(\mathbf{m}_i, \mathbf{z}) = \sum_{j|(i,j) \in \Omega} d(m_{ij}, z_j), \quad (6)$$

where $d(\cdot, \cdot)$ equals 0 for the same inputs and 1 otherwise. It is NP-Hard to obtain the optimal solution of the MEC problem (5). Therefore, heuristic methods have been proposed to solve the problem (5). For example, in [9], a heuristic combinatorial approach is proposed to find the haplotype. However, no provable guarantee for a good solution is proposed therein. Another heuristic approaches can be found in [10] and [11]. In [12], the authors have firstly offered the following optimization problem to find the haplotype,

$$\begin{aligned} \max_{\mathbf{x}} \quad & \mathbf{x}^T \mathbf{W} \mathbf{x} \\ \text{subject to} \quad & x_i^2 = 1 \end{aligned} \quad (7)$$

The authors are with the School of Electrical Engineering, Iran University of Science & Technology, Tehran 16846-13114, Iran (e-mail: mohamad_mohaddes@elec.iust.ac.ir; kahaei@iust.ac.ir).

where \mathbf{W} is the adjacency matrix defined to evaluate the similarity of each pair of rows of the read matrix and x_i is the i th entry of \mathbf{x} . Then, they proposed a semidefinite program and solved it rapidly and accurately.

In this paper, we talk about some facts about the haplotype assembly problem (HAP). Then, based on such facts we propose a maximization problem to estimate haplotypes. Note that unlike the maximization problem 7, our formulation directly uses the reads and moreover is not a quadratic form. The proposed problem is defined over the $(n-1)$ -dimensional sphere. We use an algorithm to solve the mentioned problem and prove the convergence of the algorithm. Simulation results illustrate that for a large observation error probability, our method outweighs some of the previously proposed methods in the sense of haplotype estimation accuracy.

The remainder of the paper is organized as follows. In Section II some preliminaries are introduced to state our main result. Section III talks about the main result of the paper. Simulation results are presented in Section IV. Conclusion is presented in Section V.

II. PRELIMINARIES

In Section III we formulate HAP as a manifold optimization problem. To do so, we require some related concepts as discussed in the following.

Definition 1. [13] Let \mathcal{M} be a manifold. The set of all tangent vectors at point $\mathbf{x} \in \mathcal{M}$ is called the tangent space at point \mathbf{x} and denoted by $T_{\mathbf{x}}\mathcal{M}$. Moreover, the disjoint collection of the tangent spaces is called tangent bundle and denoted by $T\mathcal{M}$.

Definition 2. A given subset \mathcal{U} of the manifold \mathcal{M} along with a bijective mapping φ between \mathcal{U} and an open subset of \mathbb{R}^d consist a pair (\mathcal{U}, φ) which is called a d -dimensional chart of \mathcal{M} .

Definition 3. [13] A differentiable bijective mapping between two manifolds is called diffeomorphism provided that its inverse is differentiable too.

Definition 4. [13] A tangent vector field ξ on \mathcal{M} is a smooth function which assigns to each point of the manifold a tangent vector belonging to tangent bundle. The gradient of a real valued smooth function f over a manifold is an example of vector field which is denoted by $\text{grad}f$. For example, tangent vector field ξ over sphere S^{n-1} is specified as:

$$\xi_{\mathbf{x}} = \text{grad}f(\mathbf{x}) = (I - \mathbf{x}\mathbf{x}^T) \text{Grad}f(\mathbf{x}), \forall \mathbf{x} \in S^{n-1}, \quad (8)$$

where $\text{Grad}f(\mathbf{x})$ is the gradient over the Euclidean space and I is the identity matrix [13].

Definition 5. [13] Let the manifold \mathcal{N} be a subset of the manifold \mathcal{M} . When the manifold topology of \mathcal{N} coincides the induced topology of \mathcal{M} , \mathcal{N} is called embedded submanifold.

Definition 6. [13] A manifold endowed with a smoothly varying inner product, is called Riemannian manifold. The inner product over the manifold is denoted by either g or $\langle \cdot, \cdot \rangle$. Moreover, $\langle \cdot, \cdot \rangle_{\mathbf{x}}$ illustrates the restriction of the inner product to the tangent space $T_{\mathbf{x}}\mathcal{M}$. Also, the metric induced

by this norm is called Riemannian distance and denoted by $\text{dist}(\cdot, \cdot)$.

Definition 7. [13] A locally distance minimizing curve over a manifold is called geodesic. Specifically, straight lines over Euclidean spaces are geodesics.

Definition 8. [13] A globally distance minimizing curve over a manifold is called minimizing geodesic. For example, α is the minimizing geodesic between points $\mathbf{x}, \mathbf{y} \in S^{n-1}$ over sphere S^{n-1} as:

$$\alpha(t) = \alpha(0) \cos(\|\dot{\alpha}(0)\|vt) + \dot{\alpha}(0) \frac{1}{v\|\dot{\alpha}(0)\|} \sin(\|\dot{\alpha}(0)\|vt), \quad (9)$$

where $\dot{\alpha}(0) \in T_{\mathbf{x}}S^{n-1}$, $\alpha(0) = \mathbf{x}$, and $\alpha(1) = \mathbf{y}$. For simplicity we consider $v = 1$.

Definition 9. [13] Let \mathcal{M} be a Riemannian manifold whose tangent bundle is $T\mathcal{M}$. Exponential map Exp is a mapping from $T\mathcal{M}$ to \mathcal{M} so that for $v \in T\mathcal{M}$, $\text{Exp}(v)$ is equal to h at time 1; where h is the unique geodesic starting from the base point of v with velocity v at time 0.

Definition 10. [13] Injectivity radius of the manifold \mathcal{M} is defined as follows,

$$i(\mathcal{M}) := \inf_{\mathbf{x} \in \mathcal{M}} \sup \left\{ \varepsilon > 0 : \text{Exp}_{\mathbf{x}}|_{B_{\varepsilon}(0_{\mathbf{x}})} \text{ is diffeomorphism} \right\}, \quad (10)$$

where $\text{Exp}_{\mathbf{x}}|_{B_{\varepsilon}(0_{\mathbf{x}})}$ shows the restriction of the mapping $\text{Exp}_{\mathbf{x}}$ to the ball $B_{\varepsilon}(0_{\mathbf{x}})$. Moreover, $B_{\varepsilon}(0_{\mathbf{x}})$ is a normal neighborhood.

Definition 11. [13] Let (\mathcal{M}, g) own a positive injectivity radius of $i(\mathcal{M})$. Then the real valued function f on \mathcal{M} is Lipschitz continuously differentiable ($L-C^1$) provided that

- 1) f is differentiable,
- 2) $\forall \mathbf{x}, \mathbf{y} \in \mathcal{M}$ with $\text{dist}(\mathbf{x}, \mathbf{y}) < i(\mathcal{M})$, there exists β for which

$$\|P_{\alpha}^{0 \leftarrow 1} \text{grad}f(\mathbf{y}) - \text{grad}f(\mathbf{x})\| \leq \beta \text{dist}(\mathbf{x}, \mathbf{y}), \quad (11)$$

where α is the unique minimizing geodesic satisfying $\alpha(0) = \mathbf{x}$ and $\alpha(1) = \mathbf{y}$. Moreover, $P_{\alpha}^{0 \leftarrow 1}$ is an isometry operator; which translates the tangent vector $\text{grad}f(\mathbf{y}) \in T_{\mathbf{y}}\mathcal{M}$ to the tangent space $T_{\mathbf{x}}\mathcal{M}$ making it possible to differentiate the tangent vectors $\text{grad}f(\mathbf{y})$ and $\text{grad}f(\mathbf{x})$, and is called parallel translation. Also, $\text{dist}(\mathbf{x}, \mathbf{y})$ is obtained by taking infimum over the length of all curves joining \mathbf{x} to \mathbf{y} (see formula (3-30) of [13]). For example, for the compact embedded submanifold S^{n-1} we have

$$\text{dist}(\mathbf{x}, \mathbf{y}) = \int_0^1 \langle \dot{\alpha}(t), \dot{\alpha}(t) \rangle^{\frac{1}{2}} dt, \quad (12)$$

where $\langle a, b \rangle = a^T b$.

Proposition 1. (Lemma 7.4.7 of [13]) Consider $B_{\varepsilon}(\mathbf{x})$ as a normal neighborhood of $\mathbf{x} \in \mathcal{M}$ and ζ as a continuously differentiable tangent vector field over \mathcal{M} . Also, let α be the

unique minimizing geodesic with $\alpha(0) = \mathbf{x}$, $\alpha(1) = \mathbf{y}$, and $\dot{\alpha}(1) = \xi$. Then $\forall \mathbf{y} \in B_\varepsilon(\mathbf{x})$:

$$P_\alpha^{0 \leftarrow 1} \zeta_{\mathbf{y}} = \zeta_{\mathbf{x}} + \nabla_\xi \zeta + \int_0^1 (P_\alpha^{0 \leftarrow \tau} \nabla_{\dot{\alpha}(\tau)} \zeta - \nabla_\xi \zeta) d\tau, \quad (13)$$

where ∇ is Riemannian connection which generalizes the concept of directional derivative of a vector field. Specifically, ∇ for Sphere at point \mathbf{x} is as follows:

$$\nabla_\xi \zeta_{\mathbf{x}} = (I - \mathbf{x}\mathbf{x}^T) D\zeta_{\mathbf{x}}(\mathbf{x})[\xi], \quad (14)$$

where $D\zeta_{\mathbf{x}}(\mathbf{x})[\xi]$ is the conventional directional derivative on a Euclidean space.

Definition 12. [13] The Riemannian Hessian of the real valued function g on the Riemannian manifold \mathcal{M} at $\mathbf{x} \in \mathcal{M}$ is defined as follows:

$$\text{Hess}f(\mathbf{x}) : T_{\mathbf{x}}\mathcal{M} \rightarrow T_{\mathbf{x}}\mathcal{M} : \xi \rightarrow \nabla_\xi \text{grad}f, \forall \xi \in T_{\mathbf{x}}\mathcal{M}. \quad (15)$$

To solve a manifold optimization problem, Riemannian Trust Region (RTR) method can be utilized. RTR methods utilize second order geometry of the cost function which lets them escape saddle points and obtain better results compared to first order line search methods [13]. RTR algorithm to minimize given cost function f over the manifold \mathcal{M} is presented in Table 1. Before presenting RTR algorithm, let us define the trust region subproblem.

Definition 13. [13] Let $f(\mathbf{x})$ be a real valued cost function over the manifold \mathcal{M} , the following quadratic optimization problem is called trust region subproblem,

$$\begin{aligned} \min_{\eta \in T_{\mathbf{x}}\mathcal{M}} \hat{m}_{\mathbf{x}}(\eta) &= f(\mathbf{x}) + \langle \text{grad}f(\mathbf{x}), \eta \rangle_{\mathbf{x}} + 0.5 \langle H[\eta], \eta \rangle_{\mathbf{x}} \\ \text{s.t. } \langle \eta, \eta \rangle_{\mathbf{x}} &\leq \Delta^2 \end{aligned}, \quad (16)$$

where H is some symmetric operator on $T_{\mathbf{x}}\mathcal{M}$.

Algorithm 1: Riemannian trust region algorithm
Requirements: Riemannian manifold (\mathcal{M}, g) ; Cost function f on \mathcal{M} ; initial point $\mathbf{x}_0 \in \mathcal{M}$, Retraction function R , Scalars $\bar{\Delta} > 0$, $\Delta_0 \in (0, \bar{\Delta})$ and $\rho' \in [0, 0.25)$.
for $k = 0, 1, 2, \dots$ do
Step 1: Find η_k by solving Problem 16.
Step 2: Evaluate $\rho_k = \frac{f(\mathbf{x}_k) - f(R_{\mathbf{x}_k}(\eta_k))}{\hat{m}_{\mathbf{x}_k}(0_{\mathbf{x}_k}) - \hat{m}_{\mathbf{x}_k}(\eta_k)}$.
Step 3: If $\rho_k < 0.25$ then $\Delta_{k+1} = 0.25\Delta_k$, elseif $\rho_k > 0.75$ and $\ \eta_k\ = \Delta_k$ then , $\Delta_{k+1} = \min(2\Delta_k, \bar{\Delta})$ else $\Delta_{k+1} = \Delta_k$, end if .
Step 4: If $\rho_k > \rho'$ then $\mathbf{x}_{k+1} = R_{\mathbf{x}}(\eta_k)$ else $\mathbf{x}_{k+1} = \mathbf{x}_k$; end for

Proposition 2. [13] Let $\{\mathbf{x}_k\}$ be a sequence generated by Alg. 1. Also, the following conditions are satisfied,

- 1) Mapping f is C^1 and bounded below at the level set $\{\mathbf{x} \in \mathcal{M} : f(\mathbf{x}) \leq f(\mathbf{x}_0)\}$,
- 2) Mapping f is $L - C^1$,
- 3) There exist $\mu > 0$ and $\delta_\mu > 0$ such that the retraction function $R : T\mathcal{M} \rightarrow \mathcal{M}$ (see Definition 4.1.1 of [13]) satisfies

$$\|\xi\| \geq \mu \text{dist}(\mathbf{x}, R_{\mathbf{x}}\xi) \quad \forall \mathbf{x} \in \mathcal{M}, \forall \xi \in T_{\mathbf{x}}\mathcal{M}, \|\xi\| \leq \delta_\mu, \quad (17)$$

- 4) Mapping $\hat{f} : T\mathcal{M} \rightarrow \mathbb{R} : \xi \rightarrow f(R\xi)$ is radially $L - C^1$, i.e., $\exists \beta_{RL}, \delta_{RL} > 0$ such that

$$\left| \frac{d}{d\tau} \hat{f}_{\mathbf{x}}(\tau\xi) \Big|_{\tau=t} - \frac{d}{d\tau} \hat{f}_{\mathbf{x}}(\tau\xi) \Big|_{\tau=0} \right| \leq \beta_{RL} t, \quad (18)$$

for all $\mathbf{x} \in \mathcal{M}$, $t < \delta_{RL}$ and $\xi \in T_{\mathbf{x}}\mathcal{M}$ with $\|\xi\| = 1$.

- 5) There is a constant β_H such that $\|H_k\| \leq \beta_H$ for all k , where H_k is the symmetric operator defined in Definition 16 at iteration k .

- 6) Any η_k obtained in Step 1 of Alg. 1 satisfies inequality

$$\hat{m}_{\mathbf{x}_k}(0) - \hat{m}_{\mathbf{x}_k}(\eta_k) \geq c_1 \|\text{grad}f(\mathbf{x}_k)\| \min\left(\Delta_k, \frac{\|\text{grad}f(\mathbf{x}_k)\|}{\|H_k\|}\right), \quad (19)$$

for some constant $c_1 > 0$, where $\|H_k\|$ is the operator norm of H_k . Then, the following holds:

$$\lim_{k \rightarrow \infty} \text{grad}f(\mathbf{x}_k) = 0. \quad (20)$$

III. MAIN RESULT

In this section we firstly present some facts in regard of HAP. Then, we propose an optimization problem to estimate haplotypes. Finally, we discuss the convergence of an algorithm for solving the proposed optimization problem.

Let \mathbf{M} be the noiseless read matrix and $\bar{\mathbf{M}}$ be the completion of \mathbf{M} . Then, the following statements hold.

- 1) $\bar{\mathbf{M}}$ is a rank one matrix for which there exists the factorization $\bar{\mathbf{M}} = \bar{\mathbf{c}}_{m \times 1} \bar{\mathbf{h}}_{n \times 1}^T$.
- 2) There is no difference for HAP to estimate \mathbf{h} as $\text{sign}(\bar{\mathbf{h}})$ or $\text{sign}(-\bar{\mathbf{h}})$. In other words, \mathbf{h} is equivalent to $-\mathbf{h}$.
- 3) There exists a unique maximizer for the problem,

$$\max_{\substack{\mathbf{x} \in \mathbb{R}^n \\ \|\mathbf{x}\|_2 \leq 1}} \|\bar{\mathbf{M}}\mathbf{x}\|_1 \quad (21)$$

up to the equivalence of \mathbf{h} and $-\mathbf{h}$.

- 4) It can be verified that (21) is equivalent to the following optimization problem over Sphere S^{n-1} ,

$$\max_{\mathbf{x} \in S^{n-1}} \|\bar{\mathbf{M}}\mathbf{x}\|_1. \quad (22)$$

We therefore propose the next optimization problem to estimate the haplotype for the noisy HAP,

$$\max_{\mathbf{x} \in S^{n-1}} \|(\mathbf{P}_\Omega(\mathbf{M}))\mathbf{x}\|_1, \quad (23)$$

Please note that, even though objective function of Problem 21 is convex, we intend to maximize the objective function. Therefore, the solution of the optimization problem cannot be trivially obtained through convex optimization approaches. Moreover, due to Alg. 1 is a descent algorithm, we rewrite Problem (23) in the following form:

$$\min_{\mathbf{x} \in S^{n-1}} -\|(\mathbf{P}_\Omega(\mathbf{M}))\mathbf{x}\|_1. \quad (24)$$

Let us make the objective function of (24) differentiable to easily use smooth optimization approaches for finding the solution. For this purpose, we propose the subsequent problem,

$$\min_{\mathbf{x} \in S^{n-1}} f(\mathbf{x}) = -\sum_{i=1}^m \left((\mathbf{M}_{\Omega_i} \mathbf{x})^2 + \varepsilon \right)^{\frac{1}{2}}, \quad (25)$$

where \mathbf{M}_{Ω_i} is the i th row of the matrix $\mathbf{P}_{\Omega}(\mathbf{M})$ and ε is a very small positive value. Please note that for $\varepsilon = 0$, the objective function of (25) turns into the objective function of (24).

Theorem 1. *For the optimization problem (25), Conditions of Proposition 2 are satisfied and consequently Alg. 1 converges.*

Proof. We require to prove that the conditions of Proposition 2 are satisfied for $f(\mathbf{x})$ in problem (25).

Condition 1; It is easy to see that $f(\mathbf{x})$ is continuous. Let (\mathcal{U}, φ_i) be a chart for S^{n-1} , then f is differentiable over S^{n-1} if $f \circ \varphi_i^{-1}$ is differentiable [13]. Let $\mathbf{x} = (x_1, \dots, x_n) \in S^n$, then, a straightforward choice for φ_i is to choose φ_i so that $\varphi_i(\mathbf{x}) = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$, when $x_i \neq 0$. Accordingly, it is easy to see that $f \circ \varphi_i^{-1}$ is differentiable and subsequently f is differentiable. Therefore, $f(\mathbf{x})$ is C^1 . Moreover, for $\mathbf{x} \in S^{n-1}$ it is obvious that $f(\mathbf{x})$ is lower bounded.

Condition 2; Let $\alpha(t)$ be the geodesic defined in Definition (8) with $v = 1$, and distance function be as Equation (12). Now, let $\xi(0) = \dot{\alpha}(0) = \text{grad}f(\mathbf{x})$. Being aware of the fact that $\mathbf{x}^T \rho = 0$, $\forall \rho \in T_{\mathbf{x}}S^{n-1}$, specifically for $\rho = \text{grad}f(\mathbf{x})$, it is easy to verify that,

$$\text{dist}(\mathbf{x}, \mathbf{y}) = \|\text{grad}f(\mathbf{x})\|_2. \quad (26)$$

Then, by considering ζ as the gradient vector field in equality (13), we have:

$$\begin{aligned} \|\mathbf{P}_{\alpha}^{0 \leftarrow 1} \zeta_{\mathbf{y}} - \zeta_{\mathbf{x}}\| &= \left\| \int_0^1 \mathbf{P}_{\alpha}^{0 \leftarrow \tau} \nabla_{\dot{\alpha}(\tau)} \zeta d\tau \right\| \\ &\leq \int_0^1 \|\nabla_{\dot{\alpha}(\tau)} \zeta\|_2 d\tau, \end{aligned} \quad (27)$$

where we used the isometry property of $\mathbf{P}_{\alpha}^{0 \leftarrow \tau}$. Using Equation (14), we have:

$$\nabla_{\dot{\alpha}(\tau)} \zeta = \left(I - \alpha(\tau) \alpha(\tau)^T \right) \lim_{t \rightarrow 0} \left(\frac{\zeta_{\alpha(\tau) + t \dot{\alpha}(\tau)} - \zeta_{\alpha(\tau)}}{t} \right). \quad (28)$$

By calculating the Euclidean gradient for the cost function of Problem (25) as

$$\text{Grad}f(\alpha(\tau)) = - \sum_{i=1}^m (\mathbf{M}_{\Omega_i}^T \mathbf{M}_{\Omega_i} \alpha(\tau)) \left((\mathbf{M}_{\Omega_i} \alpha(\tau))^2 + \varepsilon \right)^{-\frac{1}{2}} \quad (29)$$

it is not difficult to verify that

$$\lim_{t \rightarrow 0} \left(\frac{\zeta_{\alpha(\tau) + t \dot{\alpha}(\tau)} - \zeta_{\alpha(\tau)}}{t} \right) = \sum_{i=1}^m \left(\frac{-\mathbf{M}_{\Omega_i}^T \mathbf{M}_{\Omega_i} \dot{\alpha}(\tau)}{\left((\mathbf{M}_{\Omega_i} \alpha(\tau))^2 + \varepsilon \right)^{1/2}} \right). \quad (30)$$

Then, using Equality (8), Cauchy-Shwarz inequality and considering that $\|\dot{\alpha}(\tau)\|_2 = \|\text{grad}f(\mathbf{x})\|_2$ and $\|(I - \alpha(\tau) \alpha(\tau)^T)\|_F \leq n$, we will have:

$$\int_0^1 \|\nabla_{\dot{\alpha}(\tau)} \zeta\|_2 d\tau \leq n \sum_{i=1}^m \left\| \frac{(\mathbf{M}_{\Omega_i}^T \mathbf{M}_{\Omega_i})}{\sqrt{\varepsilon}} \right\|_F \|\text{grad}f(\mathbf{x})\|_2. \quad (31)$$

The aforementioned argument along with Equality (26) show that f is $L - C^1$ with $\beta = n \sum_{i=1}^m \left\| \frac{(\mathbf{M}_{\Omega_i}^T \mathbf{M}_{\Omega_i})}{\sqrt{\varepsilon}} \right\|_F$.

Condition 3; Let $R_{\mathbf{x}}$ be the restriction of retraction function R to the tangent space $T_{\mathbf{x}}S^{n-1}$ which can be defined as $R_{\mathbf{x}}(\xi) = \frac{\mathbf{x} + \xi}{\|\mathbf{x} + \xi\|_2}$, $\forall \xi \in T_{\mathbf{x}}S^{n-1}$ [13]. It is not difficult to verify that for a given $\xi \in T_{\mathbf{x}}S^{n-1}$ the geodesic $\alpha(t)$, defined in Formula (9), satisfies $\alpha(0) = \mathbf{x}$, $\alpha(1) = R_{\mathbf{x}}(\xi)$ when $\dot{\alpha}(0) = \frac{\xi}{\|\xi\|_2} \cos^{-1}(\|\mathbf{x} + \xi\|_2^{-1})$. Then, using Equation (12) we have:

$$\text{dist}(\mathbf{x}, R_{\mathbf{x}}(\xi)) = \|\dot{\alpha}(0)\|_2 = \left| \cos^{-1} \left(\left(1 + \|\xi\|_2^2 \right)^{-\frac{1}{2}} \right) \right|. \quad (32)$$

Now, let us find δ_{μ} and μ for sake of inequality (17). It is easy to numerically verify that for $\mu = 1$ and any arbitrary $\delta_{\mu} \geq 0$, inequality (17) holds.

Condition 4; Let us evaluate radially $L - C^1$ property of \hat{f} . We have:

$$\frac{d}{d\tau} \hat{f}_{\mathbf{x}}(\tau \xi) = \frac{\partial \gamma}{\partial \tau} \frac{\partial}{\partial \gamma} f(\gamma), \quad (33)$$

where $\gamma = \frac{(\mathbf{x} + \tau \xi)}{\|\mathbf{x} + \tau \xi\|_2}$. Now, using the fact that \mathbf{x} is orthogonal to ξ , $\|\mathbf{x}\|_2 = \|\xi\|_2 = 1$ and some simple calculus, for any arbitrary $t > 0$ we obtain that:

$$\begin{aligned} \left| \frac{d}{d\tau} \hat{f}_{\mathbf{x}}(\tau \xi) \Big|_{\tau=t} - \frac{d}{d\tau} \hat{f}_{\mathbf{x}}(\tau \xi) \Big|_{\tau=0} \right| &\leq \\ &t \left(\sum_{i=1}^m \|\mathbf{M}_{\Omega_i}\|_2^3 + \|\mathbf{M}_{\Omega_i}\|_2^2 \right). \end{aligned} \quad (34)$$

Meaning that radially $L - C^1$ condition has been satisfied with $\beta_{RL} = t \left(\sum_{i=1}^m \|\mathbf{M}_{\Omega_i}\|_2^3 + \|\mathbf{M}_{\Omega_i}\|_2^2 \right)$.

Condition 5; Let us consider the symmetric operator H_k be Hessian of function f at point \mathbf{x}_k . Then, based on Definition 12, the fact that $\|H_k\| \leq \|H_k\|_F$ and also an argument discussed in verifying Condition 2, we just require to prove the boundedness of $\|\text{grad}f(\mathbf{x}_k)\|$ (see Inequality (31)). Based on Equation (8) and using the Euclidean gradient of function f , as defined in Equation (29), we will have:

$$\begin{aligned} \|\text{grad}f(\mathbf{x})\|_2 &= \|(I - \mathbf{x}\mathbf{x}^T) \text{Grad}f(\mathbf{x})\|_2 \\ &\leq \frac{n}{\sqrt{\varepsilon}} \sum_{i=1}^m \|\mathbf{M}_{\Omega_i}\|_2^2. \end{aligned} \quad (35)$$

This results in

$$\|H_k\| \leq \|H_k\|_F \leq \beta_H = \frac{n^2}{\varepsilon} \left(\sum_{i=1}^m \|\mathbf{M}_{\Omega_i}\|_2^2 \right)^2. \quad (36)$$

Condition 6; It is shown in [13] that using the truncated conjugate gradient method (see Alg. 11 of [13]), Inequality (19) is satisfied with $c_1 = 1/2$.

The proof is complete. \blacksquare

IV. SIMULATION RESULTS

Hamming distance of the original haplotype and its estimation, denoted by hd , is our criterion to evaluate the performance of different methods. Please note that in HAP, it does not matter that the original haplotype is either \mathbf{h} or $-\mathbf{h}$. Accordingly, to obtain hd we calculate the Hamming distance of the estimated haplotype with both \mathbf{h} and $-\mathbf{h}$ and choose the minimum one. Simulations are performed using synthetic data created by random generation of bipolar vectors $\mathbf{h}_{n \times 1}$ and $\mathbf{c}_{m \times 1}$ to construct $\mathbf{M} = \mathbf{c}_{m \times 1} \mathbf{h}_{n \times 1}^T$. The observation set Ω is randomly produced with the probability of observation pd . Also, some observed samples are erroneous after changing their original sign; we show the set of erroneous samples by Ω_E , where $\Omega_E \in \Omega$. Fig. 1 is depicted for $m = 250$, $n = 300$, and $0.25 \leq pd \leq 0.7$. Moreover, $|\Omega_E|/|\Omega|$ is set to 0.35, where $|\cdot|$ denotes the cardinality of a set. As seen, our method outperforms the other approaches by generating lower hd .

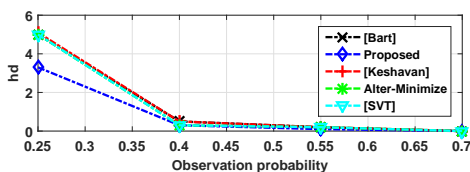


Fig. 1. Hamming distance.

V. CONCLUSION

In this letter we proposed a new method for haplotype estimation. We properly modeled HAP over an $(n-1)$ -dimensional Sphere. We also discussed the convergence of a Riemannian trust region method. Simulation results confirmed our method outperforms some of the other haplotype assembly methods.

REFERENCES

- [1] R. Schwartz *et al.*, “Theory and algorithms for the haplotype assembly problem,” *Communications in Information & Systems*, vol. 10, no. 1, pp. 23–38, 2010.
- [2] V. Bansal, A. L. Halpern, N. Axelrod, and V. Bafna, “An mcmc algorithm for haplotype assembly from whole-genome sequence data,” *Genome research*, vol. 18, no. 8, pp. 1336–1346, 2008.
- [3] B. Vandereycken, “Low-rank matrix completion by riemannian optimization,” *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1214–1236, 2013.
- [4] R. H. Keshavan, A. Montanari, and S. Oh, “Matrix completion from a few entries,” *IEEE transactions on information theory*, vol. 56, no. 6, pp. 2980–2998, 2010.
- [5] J.-F. Cai, E. J. Candès, and Z. Shen, “A singular value thresholding algorithm for matrix completion,” *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [6] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Foundations of Computational mathematics*, vol. 9, no. 6, p. 717, 2009.
- [7] C. Cai, S. Sanghavi, and H. Vikalo, “Structured low-rank matrix factorization for haplotype assembly,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 647–657, 2016.
- [8] M. M. Mohades, S. Majidian, and M. H. Kahaei, “Haplotype assembly using manifold optimization and error correction mechanism,” *IEEE Signal Processing Letters*, vol. 26, pp. 868–872, June 2019.
- [9] V. Bansal and V. Bafna, “Hapcut: an efficient and accurate algorithm for the haplotype assembly problem,” *Bioinformatics*, vol. 24, no. 16, pp. i153–i159, 2008.
- [10] Z. Puljiz and H. Vikalo, “Decoding genetic variations: Communications-inspired haplotype assembly,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, vol. 13, no. 3, pp. 518–530, 2016.
- [11] A. Hashemi, B. Zhu, and H. Vikalo, “Sparse tensor decomposition for haplotype assembly of diploids and polyploids,” *BMC genomics*, vol. 19, no. 4, p. 191, 2018.
- [12] S. Das and H. Vikalo, “Sdhap: haplotype assembly for diploids and polyploids via semi-definite programming,” *BMC genomics*, vol. 16, no. 1, p. 260, 2015.
- [13] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.