

An optimisation geometry framework for the Rayleigh quotient

Jeanne Lefevre
Gipsa Lab, CNRS
Grenoble, France
jeanne.lefevre@gipsa-lab.fr

Jonathan H. Manton
The University of Melbourne,
Parkville, Australia
jmanton@unimelb.edu.au

Nicolas Le Bihan
GIPSA-lab, CNRS
Grenoble, France
nicolas.le-bihan@gipsa-lab.fr

Abstract—This paper briefly introduces optimisation geometry, a method based on family of functions that proposes to solve complex optimisation problems with continuation methods and pre-computed points. As an illustration, the problem of tracking eigenvectors is presented, based on the Rayleigh quotient, and conditions under which the proposed approach is operational are detailed.

Index Terms—global optimisation, optimisation geometry, homotopy method, Rayleigh quotient.

I. INTRODUCTION

The central idea of optimisation geometry is to consider optimisation problems through families of functions rather than individual cost functions [1]. In many applications such as inference or tracking, optimisation problems depend on an observable (some data) which may vary while the core structure of the problem remains unchanged. Classical optimisation would treat every occurrence of a new set of data as a once-off optimisation problem. We argue that the common structure of a problem of optimisation involving different observables should be used to reduce the complexity of the overall problem: for observations $\theta_1, \theta_2, \dots, \theta_n$ compute local minima $x_1^*, x_2^*, \dots, x_n^*$. For instance, a classical inference problem is to fit a given distribution to some data, denoted hereafter θ_0 . This usually involves maximizing the likelihood $\mathcal{L}(x; \theta_0)$ where x is a vector parametrizing the distribution. Suppose the problem is to find the best parameters for θ_0 and θ_1 two independent set of data. We assume independence between the observation as we do not treat the overlap with tracking methods, where it is assumed that θ_0 and θ_1 are close. In the previous question, we are left with two distinct optimisation problems: maximizing $\mathcal{L}(x; \theta_0)$ and $\mathcal{L}(x; \theta_1)$ with respect to x . In optimisation geometry, we propose to see the problem as depending on both x and θ , so that a solution found for θ_0 can be used to determine a solution for θ_1 . Suppose for instance that the structure of \mathcal{L} is such that $\mathcal{L}(x; \theta_1) = \mathcal{L}(h(x); \theta_0)$ where h is an application between vectors of parameters. Then it is obvious that solving the problem in θ_1 can be achieved by finding the antecedent of a maxima x_0^* of $\mathcal{L}(x; \theta_0)$ by the function h . The goal is to exploit the properties of the function h , even though it has no explicit algebraic formulation. This approach allows to relocate the complexity of the problem from the graph of a function in x to the graph of a function in θ . We propose to show in

this paper how optimisation geometry method is compatible with an homotopy method for the problem of estimating and tracking eigenvectors thanks to the Rayleigh quotient [2]. The proposed approach is particularly well suited for functions defined over manifolds and could be generalized to a wide variety of optimisation problems in signal processing, *i.e.* source separation [3] or array processing [4] for example.

II. OPTIMISATION GEOMETRY

This section introduces some general results on optimisation geometry applied to a generic function f , before some illustrations depicting how this can lead to using the same numerical methods as these used by the homotopy community [5]. Let f be a function parametrized by two quantities: x and θ . We refer to x as the parameter variable, while θ is the observation or data. The set of functions $\{f_\theta = f(\cdot; \theta) | \theta \in \Theta\}$ where Θ is any set of possible observations, defines a family of functions of x parametrized by θ .

A. Optimisation as a tracking problem

The idea of optimisation geometry is to use a critical point x_0^* of the function f_{θ_0} to reach a critical point x_1^* of f_{θ_1} . A naive approach would be to use x_0^* as the initialization for an iterative method on f_{θ_1} . This idea is generally supported by perturbation theory results, such as in [6], which shows that under regularity assumptions on f , if θ_0 and θ_1 are in a small enough neighborhood of each other, then x_0^* and x_1^* are also within a small radius of each other. Note however that in the proposed framework, no assumptions are made on the distance between θ_0 and θ_1 . They can be very far away from each other, and so can the critical points of f_{θ_0} and f_{θ_1} .

Our main assumption will be the local existence of a function h so that $f(x, \theta_1) = f(h(x), \theta_0)$. Then, even if h is not explicitly available, it is possible to iteratively follow its graph, initialized at a critical point (x_0^*, θ_0) of f_{θ_0} up to (x_1^*, θ_1) a critical point of f_{θ_1} . The next assumptions are used to prove the local existence of h and the tractability of a tracking problem on the graph of h . Let $f: X \times \Theta \rightarrow \mathbb{R}$ be the cost function defined earlier. We suppose that X is a closed compact manifold and that Θ is a connected Riemannian manifold. Let θ_0, θ_1 be two sets of data in Θ , then there is a smooth path $\theta: [0, 1] \rightarrow \Theta$ that satisfies $\theta(0) = \theta_0$ and $\theta(1) = \theta_1$ because Θ is smooth and connected [7].

Furthermore, it is always possible to choose θ such that the derivative along the curve never vanishes. The geodesic between θ_0 and θ_1 for instance satisfies this condition. The following definition brings up the condition on a subset of a family of functions $\{f_\theta\}$ for h to be nice. They are basically asked to be the “nicest” possible functions on a manifold, *i.e.* Morse functions.

Definition II.1. Let θ be a smooth path in Θ between θ_0 and θ_1 . Let

$$f|_\theta: \begin{array}{ccc} X & \times & \theta([0, 1]) \\ x & , & \theta \end{array} \rightarrow \mathbb{R} \quad \mapsto f(x, \theta) = f_\theta(x)$$

be the restriction of f to the image of θ . $f|_\theta$ is said to be fibre-wise-Morse if it is a smooth function in (x, θ) and for every $\theta \in \theta([0, 1])$, f_θ has only non-degenerate critical points. A critical point of f_θ is called a fibre-wise critical point of $f|_\theta$.

We borrowed the term “fibre” to the fibre-bundles terminology because it is the natural extension for our framework. Similarly, we call a fibre-wise critical point of the function f a point (x^*, θ) such that x^* is a critical point of f_θ . With the notion of fibre-wise Morse function and fibre-wise critical points in hand, we can now state the main result of this paper.

Theorem II.1. If θ is a smooth diffeomorphic path with non-vanishing derivative and $f|_\theta$ is fibre-wise Morse, then the set of fibre-wise critical points \tilde{N} of f is a smooth topologically closed submanifold of dimension one parametrizable by t .

Proof. Let

$$f \circ \theta: \begin{array}{ccc} X & \times & [0, 1] \\ x & , & t \end{array} \rightarrow \mathbb{R} \quad \mapsto f_{\theta(t)}(x)$$

The derivative of $f \circ \theta$ with respect to x is: $\nabla_x f \circ \theta(x, t) = \nabla f_{\theta(t)}(x)$. The derivative of a function h on a manifold is the application from \mathcal{M} to the tangent bundle $T\mathcal{M}$ that sends x on the vector $\nabla h(x)$ in the tangent space at $T_x\mathcal{M}$ [8]. Note that fibre-wise critical points denoted $(\theta^*(t), x^*)$ are the points satisfying

$$\nabla f_{\theta^*(t)}(x^*) = 0_{x^*}$$

where 0_{x^*} is the 0 of the tangent plane $T_{x^*}X$. The set of fibre-wise critical points is then exactly the pre-image of the zero section $\mathbf{0}_{TX} = \{0_x \in TX | x \in X\}$ by the map $\nabla_x f \circ \theta$. According to the pre-image theorem for transverse maps [9], the set $(\nabla_x f \circ \theta)^{-1}(\mathbf{0}_{TX})$ is a closed submanifold of $X \times [0, 1]$ with no boundaries. Note that $\nabla_x f \circ \theta$ is transverse to $\mathbf{0}_{TX}$ because f_θ is fibre-wise Morse. Furthermore, given $(dx, dt) \in T_x X \times [0, 1]$, the kernel of the derivative

$$D[\nabla_x f \circ \theta](x, t) \cdot (dx, dt) = \text{Hess}_x f \circ \theta(x, t) dx + \frac{\partial d}{\partial dt} \nabla_x f \circ \theta(x, t) \cdot dt$$

is made of tangent vectors of the shape $(0, dt)$, which shows that \tilde{N} is one-dimensional and parametrizable by t . \square

A direct consequence of theorem II.1 is that the shape of \tilde{N} is constrained both by f_θ being fibre-wise Morse and by

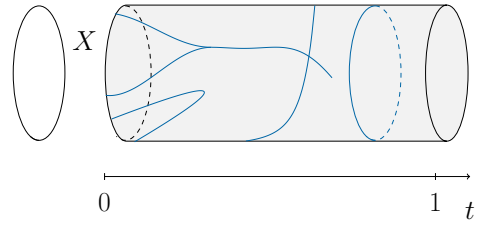


Fig. 1. The domain X is a circle (represented on the left) and the t axis is at the bottom. Every section of the cylinder represents the space X . Every blue point a line above an fixed value of t is a potential critical point of the function $f|_{\theta(t)}$. In this example from left to right: components of \tilde{N} merge, have a turning point with respect to t , cross each other, have a parametrization whose derivative goes to zero and do not admit a parametrization with respect to t (the blue circle). All this situation are prevented if f_θ is fibre-wise Morse.

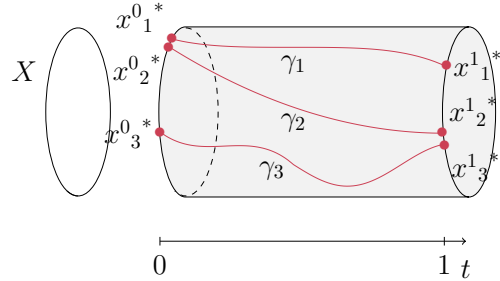


Fig. 2. A possible shape for the set of the fibre-wise critical points \tilde{N} of a fibre-wise Morse function f_θ : the domain space X is a circle, and each t on the horizontal axis can be sent to a unique θ . \tilde{N} is represented by the red lines, it has three connected components in this example. It can be seen that the critical points of f_{θ_0} and f_{θ_1} are related in $X \times [0, 1]$ through \tilde{N}

the topology of X . Figure 1 shows an example of pathological situations that can not happen under our assumptions, while figure 2 shows a possible realization for the set of fibre-wise critical points of a fibre-wise Morse function. Because \tilde{N} is a manifold, it is made of finitely many disjoint components of dimension 1. Because it is one-dimensional, we call a component in \tilde{N} a path. Two paths can not merge or cross. As \tilde{N} is closed and without boundaries, a path can not stop, and because it is parametrizable by t , a tangent to a path can never be orthogonal to the axis $[0, 1]$. In particular, the paths can have no turning points with regard to t .

B. Homotopy methods

Homotopy methods have been known since Poincaré but their interest really came into view in the eighties, with the introduction of the fifth generation of computer whose most important feature was to allow parallel computing [10], [5]. Indeed, homotopy methods first consisted in finding the zeros of polynomials by following curves linking the zeros of a trivial polynomial to the zeros of a more complex polynomial. By definition, the method can only be well-behaved if all curves are distinct, such that different starting points would lead to different endpoints. It is naturally easy to parallelize. On the opposite, with iterative methods such as the Newton descent it is not possible to predict which zero of the polynomial will be reached. As a consequence, in order to find all the roots of

a polynomial, it is necessary to run sequentially the Newton descent, excluding at each step the roots that were already found. It should be noted that the numerical methods adapted to tracking the zeros of polynomials through a homotopy method are not restricted to polynomials and can instead be used as long as the path following problem has some nice properties such as:

- 1) The path γ can be continuously parametrized by a real variable $t \in [0, 1]$.
- 2) The derivative of a parametrization $\gamma(t)$ of the path never vanishes.
- 3) The path γ coincides with the set of attractor of an iterative method in a neighborhood containing γ .

We showed in the previous paragraph that the set of fibre-wise critical points \tilde{N} of a fibre-wise Morse function $f|_{\theta}$ satisfies all three conditions. It is made of several disconnected components, that do not merge or collapse into each other. Hence, one can find separate open neighborhoods around each one of them. Furthermore, the set of fibre-wise critical points are the zeros for the gradient $\nabla_x f_{\theta}$, it is therefore an attractor for any iterative root-finding method. Hence, point 3) is satisfied. Furthermore, if θ is parametrized by $t \in [0, 1]$ then so is \tilde{N} . Smoothness comes from the manifold structure of \tilde{N} , and non-vanishing derivatives of the parametrization is, as seen previously, a consequence of the regular value theorem. This shows how optimisation geometry naturally leads to the same continuation methods as the ones used in homotopy frameworks.

III. OPTIMISATION GEOMETRY APPLIED TO THE RAYLEIGH QUOTIENT

Eigenpairs of matrices are commonly approximated through iterative methods such as power and inverse iteration, QR algorithms or others [11]. For a general matrix with no structure, these methods provide a linear rate of convergence. Algorithms for structured matrices however can reach cubic convergence [11]. For Hermitian matrices, an efficient method is the Rayleigh quotient iteration, which possesses cubic convergence rate to one of the eigenpairs [12]. This section recalls the definition and interest of the Rayleigh quotient and shows how optimisation applied to it leads to manifolds as the natural domain space for computations.

A. Three definitions of the Rayleigh quotient

Let A be a symmetric real matrix with eigenvalues $(\lambda_1, \dots, \lambda_n)$. The Rayleigh quotient with respect to A , denoted $R(\cdot; A)$, is the smooth function:

$$R: \mathbb{R}^n \setminus \{0\} \times M_n(\mathbb{R}) \rightarrow \mathbb{R}$$

$$x \quad ; \quad A \quad \mapsto \quad \frac{x^T A x}{x^T x}$$

The following theorem is a special case of the Courant-Fischer minimax theorem characterizing the eigenvalues of a real symmetric $n \times n$ matrix [13].

Theorem III.1. *The critical points (resp. critical values) for $R(\cdot; A)$ are the eigenvectors (resp. eigenvalues) of A .*

Note that the Rayleigh quotient is scale-invariant, that is:

$$\forall \lambda \in \mathbb{R}, \quad R(x; A) = R(\lambda x; A)$$

In practice, this property raises a problem for numerical methods. For instance, the Newton iteration at x_k on a scale-invariant function yields the Newton iterate $\mathcal{N}(x_k) = 2x_k$ [14]. As a consequence, the Newton method applied to the Rayleigh quotient does not converge to the eigenvectors of the matrix A , unless its initial point is actually an eigenvector. Because of its homogeneity, R is also ill-suited to an optimisation geometry framework. Let A a symmetric matrix be the data, and x the vector-parameter on which we optimize. It is obvious that for any A , $R(\cdot; A)$ is not a Morse function because if x^* is a critical point, then the whole line defined by λx^* , $\lambda \in \mathbb{R}$ is also made of critical points. Hence critical points of the Rayleigh quotient for any matrix A are not isolated, they are necessarily degenerate. In a sense, it is the excess of symmetry in R that generates a problem. To reduce the redundant information, one can change the domain space of the Rayleigh quotient. There are two ways to do so that are briefly introduced hereafter.

B. Rayleigh quotient on the sphere

Rather than defining R over \mathbb{R} , where an infinity of points are sent to the same number, one can define R on the sphere S^{n-1} . By abuse of notation, we consider:

$$R^s: S^{n-1} \times S_n \rightarrow \mathbb{R}$$

$$x, A \quad \mapsto \quad x^T A x.$$

The product $x^T A x$ is not defined if x is a point on the sphere, but this expression should be understood as: $\phi(x)^T A \phi(x)$ where ϕ is an embedding from the sphere to the set of vector with norm 1 in \mathbb{R}^n . The eigenvectors of A can be found by $\phi(x^*)$ where x^* is a critical point of R . Similarly, we define $-x$ the antipodal element to x on the sphere. Then, if $\phi(x^*)$ is an eigenvector of A , x^* and $-x^*$ are critical points of R^s . Furthermore, R^s and R coincide on points where they are both defined: for all x in the image of the embedding ϕ , $R^s(\phi^{-1}(x)) = R(x)$.

C. Rayleigh quotient on the projective space

Similarly, because R sends every line in \mathbb{R}^n to the same number, it is tempting to redefine the domain of R by quotienting \mathbb{R}^n by the relation $x \sim \lambda x$, $\forall \lambda \in \mathbb{R}$. This gives the real projective space RP^{n-1} known as the set of directions in \mathbb{R}^n . It is a compact smooth manifold which, contrarily to the sphere, is not embedded in \mathbb{R}^n . Imagine building the projective space by gluing every point on a sphere with its antipodal. This operation in \mathbb{R}^n leads to a self-intersection, showing it can not be a faithful representation (an embedding) of RP^{n-1} . We define:

$$R^p: RP^{n-1} \times S_n \rightarrow \mathbb{R}$$

$$\bar{x} \quad , \quad A \quad \mapsto \quad \bar{x}^T A \bar{x}$$

where \bar{x} denotes an equivalence class in \mathbb{R}^n : $\bar{x} = \{\lambda x \ ; \ \lambda \in \mathbb{R}\}$. Again, note that R^p and R evaluated on the same element

(modulo the equivalence relation) coincide. Furthermore, if x^* and $-x^*$ are both critical points of R_A^s , \bar{x}^* is the only critical point of R_A^p . These restrictions of the domain of R now enable to give the conditions under which R_A^s or R_A^p are Morse functions.

Theorem III.2. *Let A be a symmetric matrix in $S_n(\mathbb{R})$. The functions R_A^s and R_A^p are Morse if and only if all the eigenvalues of A are distinct.*

Proof. For η a tangent vector to the sphere at x , the second-derivative of R_A^s is [14]:

$$D^2(R_A^s)(x).\eta_x = 2(Id - xx^T)(A\eta_x - \eta_x x^T A x)$$

if v is an eigenvector of A with eigenvalue λ :

$$D^2(R_A^s)(v).\eta_v = 2(Id - vv^T)(A\eta_v - \lambda_v \eta_v)$$

the last expression cancels if and only if η_v is an eigenvector of A for the eigenvalue λ_v . Because all eigenvalues are distinct, it cancels only for vectors proportional to v . As the tangent plane at v on S^{n-1} contains only vectors orthogonal to v in \mathbb{R}^n , the second derivative $D^2(R_A^s)(v)$ is strictly injective. This is true for any eigenvector of A , hence the critical points of R_A^s are not singular.

Similarly, for $\eta_{\bar{x}}$ a tangent vector of PR^{n-1} at \bar{x} , we have

$$D^2 R_A^p(\bar{x}).\eta_{\bar{x}} = 2P_{\bar{x}}^h (A\eta_{\bar{x}} - \lambda_i \eta_{\bar{x}})$$

where $P_{\bar{x}}^h = I - \frac{1}{\|\bar{x}\|^2} \bar{x} \bar{x}^T$ is a non-vanishing matrix. Hence, $D^2 R_A^p(\bar{x}).\eta_{\bar{x}} = 0 \iff \eta_{\bar{x}} = \bar{x}$ which is not possible, as the tangent space $T_{\bar{x}} RP^{n-1} = \{y \in \mathbb{R}^n, x^T y = 0\}$ does not contain \bar{x} . \square

The previous theorem shows that the Rayleigh quotient fits into our optimisation geometry framework as long as there exists a path $\Gamma(t)$ between A and B where all the matrices $\Gamma(t)$ have distinct eigenvalues. In that case only, the function $R_{\Gamma}^{s,p}$ is fibre-wise Morse. We discuss in next section about the possibility to find such a path.

D. A path in the set of symmetric matrices

As mentioned earlier, homotopy methods are essentially path-following method, they rely on the definition of a homotopy map [5] parametrised by a variable t such that $H(0, \cdot)$ is a function whose zeros are known and $H(1, \cdot)$ is the function whose zeros are sought. For instance, a classical map that is built to solve the problem of finding the eigenpairs of a matrix A is [15], [2]:

$$H(t, x, \lambda) = \left(((1-t)D + tA)x - \lambda x, \frac{x^T x - 1}{2} \right)$$

Observe that $H(t, (x, \lambda)) = 0 \iff (\lambda, x)$ is an eigenpair of $(1-t)D + tA$ with $\|x\| = 1$. In particular $H(1, (x, \lambda)) = \left(Ax, \frac{x^T x - 1}{2} \right)$ so that the set $H(1, (\cdot, \cdot))^{-1}(\{0\})$ is exactly made of the eigenpairs of A for eigenvectors of norm 1. Similarly, for $t = 0$, $H(0, (x, \lambda)) = \left(Dx, \frac{x^T x - 1}{2} \right)$ so that the antecedent of 0 are the eigenpairs of the matrix D . The eigenpairs of D are the starting points of the continuation

method used to solve this homotopy problem, they should therefore be known in advance. This can be achieved through precomputations or more simply by choosing D diagonal which is the approach adopted in [15], [2]. Note that the second term of the function is to control the norm of x , so that if H is defined on the sphere or the projective space, it's not needed anymore. It was already noticed in [15] that the homotopy map is well suited to continuation method if and only if all the matrices in the set $\{(1-t)A + tD \mid t \in [0, 1]\}$ have distinct eigenvalues. We find the same conditions as with the optimisation geometry framework, which is logical as the numerical method for homotopy method and optimisation geometry are interchangeable. The following result is shown in [2]:

Proposition III.3. *Let E be the set of n -tuples in \mathbb{C}^n such that $D = \text{diag}(d_1, \dots, d_n)$ satisfies: $(1-t)A + tD$ have distinct eigenvalues for all $t \in [0, 1]$. Then, E is a dense open subset of \mathbb{C}^n of full Lebesgue measure in \mathbb{C}^n .*

From a practical point of view, the previous theorem says that the homotopy method can be initialized with a random diagonal matrix in $\text{diag}(\mathbb{C}^n)$ and it gives a probability one that the homotopy map from D to A built from this matrix is well-behaved. By well-behaved, we mean that the continuation method used to numerically solve the problem is proved to converge in a finite number of steps. Interestingly, looking into the proof of this theorem shows that it is crucial to choose randomly D in \mathbb{C}^n . The set E' of real n -tuples satisfying the condition of the proposition III.3 has Lebesgue measure 0 in \mathbb{C}^n . In optimisation geometry, we wish to use the pre-computed points from previous iteration of an optimisation algorithm, meaning that the starting points are not picked at random. But as we saw, if the starting point is real, it should be chosen carefully to maintain the eigenvalue separation throughout the path. Given the poor reliability of the convex combination in the real case, we propose to examine a path following method based on the Riemannian geodesic between two matrices. Let A and B be two invertible positive symmetric matrices. Then, the riemannian geodesic between A and B is parametrised by: $\Gamma(t) = A^{\frac{1}{2}}(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})^t A^{\frac{1}{2}}$. Note that positiveness was not a criteria until there, but because the set of invertible symmetric matrices is not connected it is convenient to choose A and B positive from here. Figure 3 compares in the 2-dimensional case $A, B \in S_2^+(\mathbb{R})$ the Riemannian geodesic and the convex combination. Figure 4 displays the resulting paths that would be followed by a continuation method in the cylinder $S^1 \times [0, 1]$. It is visible on the figures that the choice of the path in $S_2^+(\mathbb{R})$ greatly influences the path in the set of eigenvectors. This has its importance, as convergence of continuation methods are more or less fast given elements such as the variation of the derivatives on this path, and the absolute value of the derivative. In this precise example, a high derivative results in a curve with almost vertical sections. Hence our prediction is that for the two matrices chosen in figure 3, the straight line gives better numerical results than the Riemannian geodesic, as we can

see the variations on the grey curves are smoother than the variations on the red curve. This result however can not be generalized to every possible couple of matrices.

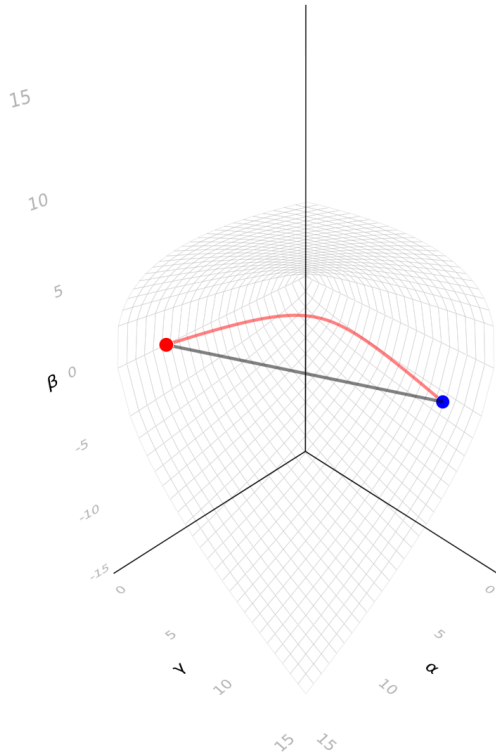


Fig. 3. Three dimensional illustration of paths in the set of positive semi-definite matrices: paths go from A (red dot) at $t = 0$ to B (blue dot) at $t = 1$. Black line is a straight line between the two while red line is the Riemannian geodesic. The wireframe in the background is the limit set of positive semi-definite matrices.

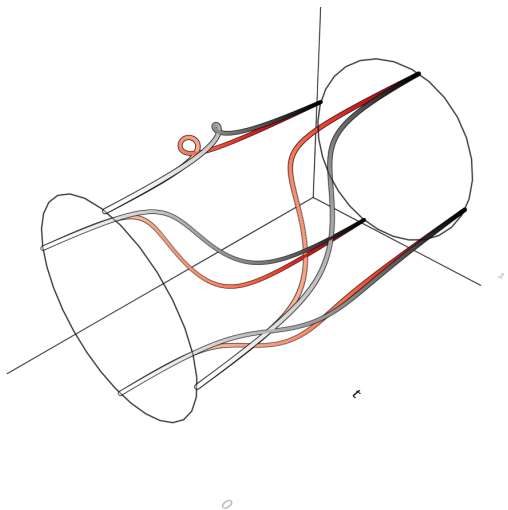


Fig. 4. Paths in the set of one-dimensional eigenvectors: black and red lines are the critical point along respectively the straight line and the Riemannian geodesic between A and B of the Rayleigh quotient. Parametrization of the path is shown both as a color gradient and on the x-axis: left-white for $t = 0$, right-darker for $t = 1$.

IV. CONCLUSION

In this paper we briefly presented the idea of optimisation geometry, and showed how it compares with the work of the homotopy community on the special example on the Rayleigh quotient. We highlighted priority targets for future work, such as determining properties on the separate eigenvalue condition along a Riemannian geodesic in the set of symmetric positive matrices. Future work will consist in bringing more insight and theoretical results on the relation between the choice of the paths in a particular instance of optimisation geometry, and the convergence properties of optimisation methods. Rayleigh quotient is a simple example of how knowledge on the structure of the function, and in particular the data space, is necessary to adapt the path strategy. If probabilistic results are available, deterministic approaches are still an open problem, for the Rayleigh quotient as for many other functions.

REFERENCES

- [1] J. H. Manton, "Optimisation Geometry," in *Mathematical System Theory — Festschrift in Honor of Uwe Helmke on the Occasion of His Sixtieth Birthday*. CreateSpace, 2013, pp. 261–274, <http://arxiv.org/abs/1212.1775>.
- [2] M. T. Chu, "A note on the homotopy method for linear algebraic eigenvalue problems," *Linear Algebra and its Applications*, vol. 105, pp. 225–236, 1988.
- [3] F. Bouchard, J. Malick, and M. Congedo, "Riemannian optimization and approximate joint diagonalization for blind source separation," *IEEE Transactions on Signal Processing*, vol. 66, no. 8, pp. 2041–2054, April 2018.
- [4] S. Choi, T. K. Sarkar, and J. Choi, "Adaptive antenna array for direction-of-arrival estimation utilizing the conjugate gradient method," *Signal Processing*, vol. 45, no. 3, pp. 313 – 327, 1995.
- [5] E. L. Allgower and K. Georg, *Numerical continuation methods: an introduction*. Springer Science & Business Media, 2012, vol. 13.
- [6] L. Blum, F. Cucker, M. Shub, and S. Smale, *Complexity and real computation*. Springer Science & Business Media, 2012.
- [7] J. M. Lee, "Smooth manifolds," in *Introduction to Smooth Manifolds*. Springer, 2013, pp. 1–31.
- [8] J. M. Lee, B. Chow, S.-C. Chu, D. Glickenstein, C. Guenther, J. Isenberg, T. Ivey, D. Knopf, P. Lu, F. Luo *et al.*, "Manifolds and differential geometry," *Topology*, vol. 643, p. 658, 2009.
- [9] V. Guillemin and A. Pollack, *Differential topology*. American Mathematical Soc., 2010, vol. 370.
- [10] R. Wilson and J. Gray, *Mathematical conversations: selections from The mathematical intelligencer*. Springer Science & Business Media, 2001.
- [11] J. W. Demmel, *Applied numerical linear algebra*. Siam, 1997, vol. 56.
- [12] U. Helmke and J. B. Moore, *Optimization and dynamical systems*. Springer Science & Business Media, 2012.
- [13] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU press, 2012, vol. 3.
- [14] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.
- [15] T. Li and N. Rhee, "Homotopy algorithm for symmetric eigenvalue problems," *Numerische Mathematik*, vol. 55, no. 3, pp. 265–280, 1989.