

Bargmann invariants, null phase curves, and a theory of the geometric phase

N. Mukunda*

Centre for Theoretical Studies, Indian Institute of Science, Bangalore 560 012, India

Arvind†

Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

E. Ercolessi‡

Dipartimento di Fisica, Università di Bologna, INFN and INFN, Via Irnerio 46, 40126 Bologna, Italy

G. Marmo§

Dipartimento di Scienze Fisiche, Università di Napoli Federico II and INFN, Via Cinzia, 80126 Napoli, Italy

G. Morandi||

Dipartimento di Fisica, Università di Bologna, INFN and INFN, Viale Berti-Pichat 6/2, 40127 Bologna, Italy

R. Simon¶

The Institute of Mathematical Sciences, CIT Campus, Tharamani, Chennai 600 113, India

We present a theory of the geometric phase based logically on the Bargmann invariant of quantum mechanics, and null phase curves in ray space, as the fundamental ingredients. Null phase curves are themselves defined entirely in terms of the (third order) Bargmann invariant, and it is shown that these are the curves natural to geometric phase theory, rather than geodesics used in earlier treatments. The natural symplectic structure in ray space is seen to play a crucial role in the definition of the geometric phase. Logical consistency of the formulation is explicitly shown, and the principal properties of geometric phases are deduced as systematic consequences.

I. INTRODUCTION

The evolution of our understanding of the geometric phase (GP) [1] has brought together many aspects of the basic structure of quantum mechanics, both in Hilbert space and ray space levels. They include both linear vector space features and differential geometric features [2]. During this development, on one hand, the original assumptions of adiabaticity, cyclicity, and unitary evolution were relaxed in stages in significant generalizations [3]. On the other hand, starting from its discovery in an essentially dynamical context, it has gradually become clear that the GP is largely kinematical in content [4]. In the process, important connections to properties of the Bargmann invariants (BI) [5], and even to earlier ideas of Pancharatnam in classical optics [6], have also been established [7].

It is well known that the ray space (complex projective space) associated with the Hilbert space of any quantum sys-

tem carries a natural Riemannian metric—the Fubini-Study metric [8]—as well as a natural symplectic structure [9]. The former determines a corresponding family of geodesics which have played an important role in the GP theory in at least two ways. Initially, they were exploited to show how to define the GP for noncyclic evolution governed by the Schrödinger equation, essentially by converting such an evolution to a cyclic one by adding on a geodesic to connect the end points [10]. Later they were found to be useful in showing that phases of BI's are particular instances of the GP [11]. The ray space symplectic structure has been known to be intimately involved with GP's, specially for cyclic evolution.

Subsequent work has shown that rather than geodesics, the really basic geometric objects needed to connect BI's and GP's are a family of ray space and Hilbert space curves which have been named null phase curves (NPC) [12]. It has been shown that while geodesics are NPC's, the latter form a vastly larger class of curves having little to do with the Fubini-Study metric or the notion of geodesics. They also lead to the most general possible connection between BI's and GP's.

The purpose of this paper is to provide a logical basis for defining the GP and understanding its properties. We will show that the primitive building block is the BI (in particular, the three-vertex BI) in terms of which NPC's can be defined. Once this is in hand, the GP (for the noncyclic case in general) can be defined as a derived object directly as a suitable two-dimensional surface integral of the symplectic two-form

*Email address: nmukunda@cts.iisc.ernet.in

†Permanent address: Department of Physics, Guru Nanak Dev University, Amritsar 143 005, India.

Email address: xarvind@andrew.cmu.edu

‡Email address: ercolessi@bo.infn.it

§Email address: gimarmo@na.infn.it

||Email address: morandi@bo.infn.it

¶Email address: simon@imsc.res.in

on ray space. Thus in this approach, primary importance is given to BI's, NPC's, and ray space symplectic structure. Earlier definitions of the GP, in particular, the kinematic definition, are seen to be immediate consequences of the present one.

We shall make frequent reference to *the* geodesic arc connecting two given points in state space. By this we shall always mean the *shorter* geodesic, which is *unique* assuming that the given pair of points correspond to *nonorthogonal* states.

The contents of this paper are arranged as follows. In Sec. II, we assemble some details of notation relating to the Hilbert and ray spaces of a general quantum-mechanical system. We define in a precise manner various classes of smooth curves needed for further work; recall the natural one-form on Hilbert space and the symplectic two-form on ray space; and then the kinematic definition of the GP. Section III begins with the definition and basic properties of the BI, and in terms of them defines the family of NPC's in Hilbert and ray spaces. Some important formulas connecting the two are also developed. With this preparation, the definition of the GP for a general open (sufficiently smooth) ray space curve is given, and its consistency is demonstrated. The recovery of the kinematic definition of the GP is also shown, and the connections between BI's and GP's, mediated by the uses of NPC's, are brought out. The emphasis in this section is to display the logical structure of ideas. Section IV explores the properties of NPC's from various points of view, emphasizing always that it is these curves that are natural and basic to the structure of the GP. The fact that they are far more numerous than geodesics means that their description is very different from that of the latter; in particular, they cannot be viewed as solutions to any local finite-order ordinary differential equations at all. Examples of (infinitely many) NPC's connecting any two given ray space points; a description of the most general NPC; of submanifolds every curve in which is a NPC; and examples of such submanifolds; are all developed. Section V contains concluding remarks.

There are two appendixes, devoted to basic differential geometry of Hilbert and ray spaces and to a complete description of geodesics, respectively.

II. NOTATIONAL PRELIMINARIES AND KINEMATIC DEFINITION OF GEOMETRIC PHASE

We denote by \mathcal{H} the complex (separable) Hilbert space describing the pure states of some quantum system. The complex dimension of \mathcal{H} may be finite, say $n=1,2,\dots$, or infinite. The inner product and the norm for vectors ψ, ϕ, \dots in \mathcal{H} are written as (ψ, ϕ) and $\|\psi\|$, respectively. The subset of unit vectors in \mathcal{H} is defined by

$$\mathcal{B} = \{\psi \in \mathcal{H} \mid \|\psi\| = 1\} \subset \mathcal{H}. \quad (2.1)$$

For dimension $\mathcal{H}=n$ finite, \mathcal{B} is the real Euclidean sphere S^{2n-1} of real dimension $(2n-1)$. It is both connected and simply connected. The space of unit rays associated with \mathcal{H}

and \mathcal{B} is denoted by \mathcal{R} . It is the quotient of \mathcal{B} with respect to the equivalence relation $\psi \sim e^{i\alpha}\psi$ among unit vectors for all real phases α :

$$\mathcal{R} = \mathcal{B}/U(1). \quad (2.2)$$

Elements of \mathcal{R} are represented by pure state density matrices, or one-dimensional projections, ρ ; and there is a projection π from \mathcal{B} to \mathcal{R} :

$$\pi: \mathcal{B} \rightarrow \mathcal{R}: \psi \in \mathcal{B} \rightarrow \pi(\psi) = \rho_\psi = \psi\psi^\dagger \in \mathcal{R}. \quad (2.3)$$

As is well known, \mathcal{B} is a principal $U(1)$ bundle over \mathcal{R} , which in turn for finite dimensions is the complex projective space CP^{n-1} . The real dimension of \mathcal{R} in that case is $2(n-1)$, and it is also connected and simply connected.

For considerations of geometric phases, null phase curves, and geodesics, we need to deal with continuous parametrized curves $\mathcal{C} \subset \mathcal{B}$, and their projections $C = \pi(\mathcal{C}) \subset \mathcal{R}$, obeying suitable smoothness conditions. They are always directed curves. We describe them as follows:

$$\mathcal{C} = \{\psi(s) \in \mathcal{B} \mid s_1 \leq s \leq s_2\} \subset \mathcal{B};$$

$$C = \pi(\mathcal{C}) = \{\rho(s) = \psi(s)\psi(s)^\dagger \in \mathcal{R} \mid s_1 \leq s \leq s_2\} \subset \mathcal{R}. \quad (2.4)$$

In general, we assume that the parameter s varies over a closed finite interval $[s_1, s_2] \subset \mathbb{R}$, exceptions will be indicated. We permit strictly monotonic reparametrizations of these curves,

$$\begin{aligned} s \rightarrow s' = f(s), \\ \frac{df(s)}{ds} > 0, \end{aligned} \quad (2.5)$$

and impose other smoothness conditions as appropriate and described below. We also permit smooth local phase changes along a curve \mathcal{C} to lead to a new curve \mathcal{C}' ,

$$\mathcal{C}' = \{\psi'(s) = e^{i\alpha(s)}\psi(s) \mid s_1 \leq s \leq s_2\} \subset \mathcal{B} \quad (2.6)$$

having the same image in \mathcal{R} :

$$\pi(\mathcal{C}') = \pi(\mathcal{C}) = C \subset \mathcal{R}. \quad (2.7)$$

Thus, both \mathcal{C} and \mathcal{C}' are lifts, from \mathcal{R} to \mathcal{B} , of C .

We now define three classes of curves, with different smoothness conditions, as follows.

Class I:

$$(\psi(s_1), \psi(s_2)) \neq 0$$

$$\psi(s), \rho(s) \text{ continuous and piecewise once differentiable.} \quad (2.8)$$

Class II:

$$(\psi(s), \psi(s')) \neq 0, \text{ any } s, s' \in [s_1, s_2]$$

$$\psi(s), \rho(s) \text{ continuous once differentiable.} \quad (2.9)$$

Class III:

$$(\psi(s_1), \psi(s_2)) \neq 0$$

$$\psi(s), \quad \rho(s) \text{ continuous twice differentiable.} \quad (2.10)$$

It will turn out that class-I curves are those for which geometric phases can be defined; class-II curves subject to further conditions are null phase curves; and class-III curves obeying suitable differential equations are geodesics. In each case, both $C \subset \mathcal{B}$ and $C = \pi(C) \subset \mathcal{R}$ will be assumed to obey the same smoothness conditions and both will belong to the same class. Similarly, reparametrizations and local phase changes will be assumed to preserve the smoothness properties of each class.

The basic differential geometric objects needed for our purposes, with suitable notations, are given in Appendix A [13]. These are a one-form A on \mathcal{B} ; its exterior derivative two-form dA on \mathcal{B} ; and a closed nondegenerate symplectic two-form ω on \mathcal{R} related to dA via pullback

$$dA = \pi^* \omega. \quad (2.11)$$

The one-form A is essentially defined by giving its integral along any curve $C \subset \mathcal{B}$ of class I:

$$\int_C A = \text{Im} \int_{s_1}^{s_2} ds \left(\psi(s), \frac{d\psi(s)}{ds} \right) = -i \int_{s_1}^{s_2} ds \left(\psi(s), \frac{d\psi(s)}{ds} \right). \quad (2.12)$$

If C happens to be a closed loop with $\psi(s_2) = \psi(s_1)$ so that its projection $\pi(C) = C \subset \mathcal{R}$ is also a closed loop, we have

$$\oint_C A = \int_S dA = \int_S \omega, \quad (2.13)$$

where S is any smooth two-dimensional surface in \mathcal{B} having C as boundary, and S is the image of S in \mathcal{R} with C as boundary:

$$\partial S = C,$$

$$S = \pi(S), \partial S = C. \quad (2.14)$$

The orientations of S and S are determined by the directions of C and C , respectively. It is to be emphasized that ω is not exact, so A is not the pullback via π^* of any one-form on \mathcal{R} .

Since the two-form ω on \mathcal{R} will play a primary role in our definition of the GP, we add the following comments to help better understand its nature. In the finite-dimensional case, dimension $\mathcal{H} = n$, we have mentioned that $\mathcal{R} = \mathbb{C}P^{n-1}$, the complex projective space of (complex) dimension $(n-1)$; and it is well known that these spaces are important canonical examples of symplectic manifolds [9]. This property can however be also grasped from another point of view. The unitary group $U(n)$ acting on \mathcal{H} via its defining representation has for its Lie algebra $\mathfrak{u}(n)$, the set of all Hermitian operators on \mathcal{H} . Upon conjugation by elements of $U(n)$ (which is the adjoint action), $\mathfrak{u}(n)$ is mapped onto itself and broken up into disjoint orbits. Each orbit is essentially a

coset space $U(n)/H$, where H is the stability group of any chosen representative point on the orbit. Most orbits are generic and of real dimension $n(n-1)$, being the coset space $U(n)/U(1) \times U(1) \times \dots \times U(1)$ (n factors). Apart from these, there are several exceptional or singular orbits of various lower dimensions. From the general Kostant-Kirillov-Souriau theory of coadjoint orbits of Lie groups G , it is known that each orbit (generic or exceptional) is a symplectic manifold; the symplectic two-form is obtained by descent from the Maurer-Cartan two-form on the group G itself, i.e., by quotienting with respect to its kernel [14]. In the case at hand, the pure state density operators for an n -dimensional quantum system are elements of $\mathfrak{u}(n)$. They are clearly acted upon transitively by conjugation with elements of $U(n)$; and they form a single nongeneric orbit in the Lie algebra, the stability group H in this case being easily seen to be $U(1) \times U(n-1)$. We can then identify \mathcal{R} with the coset space $U(n)/U(1) \times U(n-1)$, which is indeed of real dimension $2(n-1)$; and according to the general theory of coadjoint orbits of Lie groups, it is a symplectic manifold.

Even without appeal to the Kostant-Kirillov-Souriau (KKS) theory and then limiting ourselves to the nongeneric orbit of pure state density operators, one can directly display the connection between the forms A and dA on \mathcal{B} and the Maurer-Cartan one- and two-forms on $U(n)$.

Denote by g, g', \dots the matrices of the defining representation of $U(n)$. The actions of $U(n)$ on \mathcal{B} and on \mathcal{R} given by

$$g \in U(n): \psi \in \mathcal{B} \rightarrow g\psi \in \mathcal{B} \quad (2.15)$$

and

$$\rho \in \mathcal{R} \rightarrow g\rho g^{-1} \in \mathcal{R} \quad (2.16)$$

are both transitive. Therefore, for any choice of a fiducial vector $\psi_0 \in \mathcal{B}$, we have an onto map $\mu_0: U(n) \rightarrow \mathcal{B}$ given by

$$\mu_0: g \in U(n) \rightarrow g\psi_0 \in \mathcal{B}. \quad (2.17)$$

This map can be used to take any (smooth) curve $\gamma = \{g(s)\} \subset G$ to an image $C = \{\mu_0(g(s)) = \{\psi(s) = g(s)\psi_0\} \subset \mathcal{B}$, and in turn to its projection $C = \{\rho(s) = \psi(s)\psi(s)^\dagger = g(s)\psi_0\psi_0^\dagger g(s)^{-1}\} \subset \mathcal{R}$. In the reverse direction, evidently, the pullback μ_0^* allows us to take A and dA on \mathcal{B} to appropriate forms on $U(n)$. From Eq. (2.12) A on \mathcal{B} has the expression

$$A = -i\psi^\dagger d\psi. \quad (2.18)$$

Combining with Eq. (2.17) we get

$$\mu_0^* A = -i\mu_0^*(\psi^\dagger d\psi) = -i\psi_0^\dagger g^\dagger dg \psi_0 = -i \text{Tr}(\rho_0 g^\dagger dg). \quad (2.19)$$

The expression $g^\dagger dg$ is the matrix of left-invariant Maurer-Cartan one-forms on $U(n)$, so in Eq. (2.19) we have the expected connection between the one-form A on \mathcal{B} and the Maurer-Cartan one-forms on $U(n)$. A particular linear

combination of the latter, determined by the choice of ψ_0 , is picked out. It also follows by taking the exterior derivative of Eq. (2.19) that

$$\mu_0^* dA = -i \operatorname{Tr}(\rho_0 dg^\dagger \wedge dg) = i \operatorname{Tr}(\rho_0 g^\dagger dg \wedge g^\dagger dg), \quad (2.20)$$

so here the matrix of left-invariant Maurer-Cartan two-forms on $U(n)$ appears. Hence, if X_1 and X_2 are two left-invariant vector fields on $u(n)$ associated with Hermitian matrices τ_1 and τ_2 in $u(n)$,

$$i_{X_j} g^\dagger dg = i \tau_j, \quad j=1,2, \quad (2.21)$$

then a short calculation shows

$$\begin{aligned} \mu_0^* dA(X_1, X_2) &= (dA)(\mu_{0*}(X_1), \mu_{0*}(X_2)) \\ &= i_{X_2} i_{X_1} \mu_0^* dA \\ &= i \operatorname{Tr}(\rho_0[\tau_1, \tau_2]). \end{aligned} \quad (2.22)$$

In particular, we can choose τ_1 and τ_2 to be two elements $\rho_1, \rho_2 \in \mathcal{R}$ and then we get in Eq. (2.22) the result $i \operatorname{Tr}(\rho_0[\rho_1, \rho_2])$ which agrees with Eq. (A.7). Thus, the connection between A, dA on \mathcal{B} and the Maurer-Cartan forms on $U(n)$, mentioned in the previous paragraph, are made explicit [15].

We now recall the definition of the GP according to the kinematic approach [4]. If C is a class-I curve with image C , then the GP for C is

$$\varphi_g[C] = \arg(\psi(s_1), \psi(s_2)) - \int_C A. \quad (2.23)$$

Clearly, because of the first term on the right, this phase is defined modulo 2π ; and as implied by the notation it is a functional of C independent of the lift \tilde{C} used to compute the individual terms on the right. Moreover it is unchanged by any permitted reparametrizations.

In a previous work, Eq. (2.23) was adopted as the definition of the GP, and thereafter null phase curves were defined and used in various ways [12]. Our approach here will be to regard null phase curves as primitive objects and to define GP's in terms of them. This will be done in the following section.

If the end points $\psi(s_1), \psi(s_2)$ of C are mutually orthogonal, clearly the GP $\varphi_g[C]$ becomes undefined. This is the reason behind the condition of nonorthogonality of end points in the definition (2.8) of class-I curves. It ensures that for any class-I curve the, GP is well-defined modulo 2π . However, the definition (2.8) does not forbid the possibility that for some $s_0 \in (s_1, s_2)$, $\psi(s_0)$ may be orthogonal to either $\psi(s_1)$ or $\psi(s_2)$ or both. Assume $\psi(s_0)$ is indeed orthogonal to $\psi(s_1)$, and let s_0 be a point on C and C at which these curves are differentiable. If for sufficiently small ϵ both $\psi(s_0 - \epsilon)$ and $\psi(s_0 + \epsilon)$ are not orthogonal to $\psi(s_1)$, then the GP's are defined for the portions of C running from s_1 to $s_0 \pm \epsilon$, and they obey

$$\varphi_g[C \text{ for } s_1 \leq s \leq s_0 + \epsilon] - \varphi_g[C \text{ for } s_1 \leq s \leq s_0 - \epsilon] = \pm \pi, \quad (2.24)$$

which are equivalent modulo 2π . Thus, in such a situation we see that the GP is defined upto just before a point of orthogonality to the initial point, as well as to a point just after; and there is a discontinuity of $\pm \pi$ as we cross that point.

In passing we may mention that even in the context of a real Hilbert space, the GP survives though in a rudimentary form [16]. The dynamical phase is of course absent, however the total phase could be an odd multiple of π . In fact, each time the inner product $(\psi(s_1), \psi(s_2))$ passes through zero, we pick up a contribution $\pm \pi$, just as in Eq. (2.24). The BI $\Delta_3(\psi_1, \psi_2, \psi_3)$ [see below] can also have a nontrivial phase, namely, $\pm \pi$ when it is negative. All these remarks remain valid also in the case of a complex Hilbert space, if we restrict ourselves to the real linear span of a set of vectors taken from an orthonormal basis. The role of such ‘‘real’’ subspaces will become evident in the sequel.

III. NULL PHASE CURVES AND A DEFINITION OF THE GEOMETRIC PHASE

Our aim now is to define NPC in \mathcal{B} and \mathcal{R} as the basic or primitive objects, then define GP's in terms of them, and derive their properties in a logically consistent manner. To begin with, we recall the definition and properties of the third-order Bargmann invariant, as the NPC definition will depend on it. For convenience, we divide this section into further sections.

A. Bargmann invariants (BI)

Given any three mutually nonorthogonal vectors $\psi_1, \psi_2, \psi_3 \in \mathcal{B}$, projecting onto $\rho_1, \rho_2, \rho_3 \in \mathcal{R}$, the third-order BI is defined as

$$\Delta_3(\psi_1, \psi_2, \psi_3) = (\psi_1, \psi_2)(\psi_2, \psi_3)(\psi_3, \psi_1) = \operatorname{Tr}(\rho_1 \rho_2 \rho_3). \quad (3.1)$$

Its key properties are well known: (i) for dimension $\mathcal{H} \geq 2$ it is in general complex; (ii) it is cyclically symmetric; (iii) as the second form shows, it is invariant under independent phase changes in each of the vectors ψ_1, ψ_2, ψ_3 .

Higher-order BI's can be defined in a similar manner. For any m vectors $\psi_1, \psi_2, \dots, \psi_m \in \mathcal{B}$, such that no two successive ones are mutually orthogonal, we have the generally complex m th order BI

$$\begin{aligned} \Delta_m(\psi_1, \psi_2, \dots, \psi_m) &= (\psi_1, \psi_2)(\psi_2, \psi_3) \cdots (\psi_m, \psi_1) \\ &= \operatorname{Tr}(\rho_1 \rho_2 \cdots \rho_m). \end{aligned} \quad (3.2)$$

For $m=2$ of course, $\Delta_2(\psi_1, \psi_2)$ is real non-negative.

B. Null phase curves

A curve $C \subset \mathcal{B}$ of class II, with image $C \subset \mathcal{R}$, will be said to be an NPC if for any three vectors on it, the BI is real positive:

$$\begin{aligned}
\mathcal{C}, \text{CNPC} &\Leftrightarrow \Delta_3(\psi(s), \psi(s'), \psi(s'')) \\
&= \text{Tr}(\rho(s)\rho(s')\rho(s'')) = \text{real positive} \\
&\Leftrightarrow \text{Tr}(\rho(s)[\rho(s'), \rho(s'')]) = 0, \\
&\text{any } s, s', s'' \in [s_1, s_2]. \tag{3.3}
\end{aligned}$$

For convenience, when this condition is obeyed, we refer to both \mathcal{C} and C as NPC's.

It is immediately evident that any connected subset or portion of an NPC, say running from s_3 to s_4 , where $[s_3, s_4] \subset [s_1, s_2]$, is also an NPC.

From Eq. (B13) of Appendix B, it follows that every geodesic in \mathcal{R} (and any lift of it in \mathcal{B}) is an NPC. Since any two points $\rho_1, \rho_2 \in \mathcal{R}$ can definitely be connected by a geodesic [which is moreover unique if $\text{Tr}(\rho_1\rho_2) > 0$], we can say that they can definitely be connected by an NPC. However, as we will see in the following section, provided dimension $\mathcal{H} \geq 3$, NPC's are far more numerous than geodesics: there are infinitely many of them connecting any $\rho_1, \rho_2 \in \mathcal{R}$.

The definition (2.9) of a class-II curve includes the condition that no two vectors along it should be mutually orthogonal. The motivation for this is now understandable: we need the BI appearing in Eq. (3.3) to be nonzero for any triplet of vectors on the curve. It will soon emerge that a somewhat more economical definition of an NPC, which is however fully equivalent to Eq. (3.3), is

$$\begin{aligned}
\mathcal{C}, \text{CNPC} &\Leftrightarrow \Delta_3(\psi(s_0), \psi(s), \psi(s')) = \text{real positive}, \\
&\text{any fixed } s_0 \in [s_1, s_2], \quad \text{any } s, s' \in [s_1, s_2]. \tag{3.4}
\end{aligned}$$

Let now C be an NPC. We derive a fundamental formula for the integral of A along any lift \mathcal{C} of C . For any chosen reference point $s_0 \in [s_1, s_2]$, choose some $\psi_0(s_0) \in \pi^{-1}(\rho(s_0))$. By definition of class II, for any $s \in [s_1, s_2]$ and any choice of $\psi(s) \in \pi^{-1}(\rho(s))$, the scalar product $(\psi_0(s_0), \psi(s))$ is nonzero. Adjust the phase of $\psi(s)$ to get $\psi_0(s) \in \pi^{-1}(\rho(s))$ such that

$$(\psi_0(s_0), \psi_0(s)) = \text{real positive}, \quad \text{any } s \in [s_1, s_2]. \tag{3.5}$$

This gives us a particular lift $\mathcal{C}_0 = \{\psi_0(s) | s_1 \leq s \leq s_2\}$ of C with the property

$$(\psi_0(s), \psi_0(s')) = \text{real positive}, \quad \text{any } s, s' \in [s_1, s_2]. \tag{3.6}$$

We see this by setting $s'' = s_0$ in the definition (3.3) and then using Eq. (3.5). This means that any two points on \mathcal{C}_0 are in phase in the Pancharatnam sense [6], a nonlocal property; and furthermore \mathcal{C}_0 is horizontal, a local property,

$$\mathcal{C}_0: \left(\psi_0(s), \frac{d}{ds} \psi_0(s) \right) = 0, \quad \text{any } s \in [s_1, s_2]. \tag{3.7}$$

Therefore, for the end points of \mathcal{C}_0 and for the integral of A along \mathcal{C}_0 , we have

$$\begin{aligned}
\arg(\psi_0(s_1), \psi_0(s_2)) &= 0, \\
\int_{\mathcal{C}_0} A &= 0. \tag{3.8}
\end{aligned}$$

Now let $\mathcal{C} = \{\psi(s) = e^{i\alpha(s)}\psi_0(s)\}$ be a general lift of C obtained from \mathcal{C}_0 by a (sufficiently smooth) local phase transformation. For \mathcal{C} we find in place of Eq. (3.8),

$$\arg(\psi(s_1), \psi(s_2)) = \alpha(s_2) - \alpha(s_1),$$

$$\int_{\mathcal{C}} A = -i \int_{s_1}^{s_2} ds \left(\psi(s), \frac{d\psi(s)}{ds} \right) = \int_{s_1}^{s_2} ds \frac{d\alpha(s)}{ds}, \tag{3.9}$$

that is,

$$\int_{\mathcal{C}} A = \arg(\psi(s_1), \psi(s_2)). \tag{3.10}$$

This is the basic property of NPC's that we will use repeatedly.

Concerning the construction of the particular ‘‘Pancharatnam lift’’ \mathcal{C}_0 of the NPC $C = \{\rho(s) = \psi(s)\psi(s)^\dagger\} \subset \mathcal{R}$, we may add the following remark. The lift \mathcal{C}_0 is completely determined once a choice of $\psi_0(s_0)$ at the reference point s_0 is made. Any alteration of $\psi_0(s_0)$ by a phase leads to a rigid or constant phase change of all points along \mathcal{C}_0 . One can now see that the rule (3.5) to determine $\psi_0(s)$ for general s has the following quite explicit solution:

$$\psi_0(s) = N(s)\rho(s)\psi_0(s_0) = \psi(s)e^{-i\arg(\psi_0(s_0), \psi(s))}, \tag{3.11}$$

where $N(s)$ is a real positive normalization factor,

$$N(s) = |(\psi_0(s_0), \psi(s))|^{-1}. \tag{3.12}$$

The vector $\psi_0(s)$ in Eq. (3.11) is clearly invariant under changes in phase of $\psi(s)$. Then both Eqs. (3.5) and (3.6) are obeyed by the expression (3.11):

$$\begin{aligned}
(\psi_0(s_0), \psi_0(s)) &= N(s)(\psi_0(s_0), \rho(s)\psi_0(s_0)) \\
&= |(\psi_0(s_0), \psi(s))| > 0, \\
(\psi_0(s), \psi_0(s')) &= N(s)N(s')(\rho(s)\psi_0(s_0), \rho(s')\psi_0(s_0)) \\
&= N(s)N(s')\text{Tr}(\rho(s_0)\rho(s)\rho(s')) > 0, \tag{3.13}
\end{aligned}$$

since C is given to be an NPC.

Now let C be an NPC from ρ_1 to ρ_2 , and C' an NPC from ρ_2 to ρ_1 . Choose vectors $\psi_1 \in \pi^{-1}(\rho_1)$, $\psi_2 \in \pi^{-1}(\rho_2)$, and lifts $\mathcal{C}, \mathcal{C}'$ of C, C' from ψ_1 to ψ_2 and ψ_2 to ψ_1 , respectively. The unions $\mathcal{C} \cup \mathcal{C}' \subset \mathcal{R}, \mathcal{C} \cup \mathcal{C}' \subset \mathcal{B}$ are closed loops. Let $S \subset \mathcal{R}$ be any smooth two-dimensional surface with boundary $\partial S = \mathcal{C} \cup \mathcal{C}'$. Combining Eq. (2.11) with the basic property (3.10) for both \mathcal{C} and \mathcal{C}' , we find

$$\mathcal{C}, \text{NPC } \rho_1 \text{ to } \rho_2, \quad \mathcal{C}', \text{NPC } \rho_2 \text{ to } \rho_1, \quad \partial S = \mathcal{C} \cup \mathcal{C}',$$

$$\int_S \omega = \oint_{C \cup C'} A = \arg \Delta_2(\psi_1, \psi_2) = 0. \quad (3.14)$$

It is perhaps worth emphasizing that while the curves C, C' and their lifts $\mathcal{C}, \mathcal{C}'$ are each of class II, since they are NPC's, the closed loops $C \cup C'$ and $\mathcal{C} \cup \mathcal{C}'$ may not be of class II, and in any case they are not expected to be NPC's. They are of course class I which requires only piecewise once differentiability.

The generalization of Eq. (3.14) to a string of three or more successive NPC's, altogether forming a closed loop, involves the phase of a nontrivial BI. Thus for any $m \geq 3$, if $\rho_1, \rho_2, \dots, \rho_m \in \mathcal{R}$ and $C_{j,j+1}$ are NPC's from ρ_j to ρ_{j+1} for $j=1, 2, \dots, m$ (with $\rho_{m+1} = \rho_1$), with lifts $\mathcal{C}_{j,j+1}$ running from ψ_j to ψ_{j+1} , we find using Eq. (3.10) repeatedly

$$\partial S = C_{12} \cup C_{23} \cup \dots \cup C_{m1}, \quad C_{j,j+1} \text{ NPC's } \rho_j \text{ to } \rho_{j+1};$$

$$\int_S \omega = \sum_{j=1}^m \int_{\mathcal{C}_{j,j+1}} A = \arg \Delta_m(\psi_1, \psi_2, \dots, \psi_m). \quad (3.15)$$

It is because only $\Delta_2(\psi_1, \psi_2)$ is known to be always real positive that we obtain a vanishing right-hand side in the result (3.14).

C. Definition of the GP

With this preparation we are able to define the GP $\varphi_g[C]$ for any class-I curve $C \subset \mathcal{R}$ from ρ_1 to ρ_2 . We choose any NPC $C' \subset \mathcal{R}$ from ρ_2 to ρ_1 , so that $C \cup C'$ is a class-I closed loop in \mathcal{R} , and then choose any two-dimensional surface $S \subset \mathcal{R}$ with $\partial S = C \cup C'$. Then $\varphi_g[C]$ is defined as the integral of ω over S [17]:

$$C \text{ class I } \rho_1 \text{ to } \rho_2, \quad C' \text{ NPC } \rho_2 \text{ to } \rho_1, \quad \partial S = C \cup C';$$

$$\varphi_g[C] = - \int_S \omega. \quad (3.16)$$

For consistency, we must show that the integral involved here is independent of the choice of the NPC C' . Pending that, we see immediately upon comparing Eq. (3.14) with the definition (3.16) that for any NPC, the GP vanishes,

$$C \text{ is NPC} \Rightarrow \varphi_g[C] = 0. \quad (3.17)$$

The proof of the consistency of the definition (3.16) also rests on the result (3.14). First, we introduce an item of notation. For any curves $\mathcal{C}, \mathcal{C}'$, we denote by $\tilde{\mathcal{C}}, \tilde{\mathcal{C}'}$ the reversed curves obtained by traversing them backwards. We note then that the NPC property is preserved while the GP changes sign:

$$C, \mathcal{C} \text{ NPC} \Leftrightarrow \tilde{C}, \tilde{\mathcal{C}} \text{ NPC},$$

$$\varphi_g[\tilde{C}] = - \varphi_g[C]. \quad (3.18)$$

Now turning to the consistency of Eq. (3.16), let C'' be any other NPC from ρ_2 to ρ_1 , which could have been used in

place of C' to compute $\varphi_g[C]$. Choose any surface $S' \subset \mathcal{R}$ with boundary $\partial S' = \tilde{C}' \cup C''$. Then $S \cup S'$ has boundary $\partial(S \cup S') = C \cup C''$. Using Eq. (3.14) for the pair \tilde{C}', C'' we have

$$\partial S = C \cup C', \quad \partial S' = \tilde{C}' \cup C'', \quad \partial(S \cup S') = C \cup C'';$$

$$\varphi_g[C] = - \int_S \omega = - \int_S \omega - \int_{S'} \omega = - \int_{S \cup S'} \omega. \quad (3.19)$$

Here we used the additivity property for integrals of ω over (nonoverlapping) surfaces S and S' . Thus, the consistency of the definition (3.16) is established.

All of the above is applicable for a general class-I curve $C \subset \mathcal{R}$ which could be open, i.e., $\rho_2 \neq \rho_1$. In case $\rho_2 = \rho_1$ and C is a closed loop, there is no need to append any NPC to it before computing its GP. We choose any lift \mathcal{C} also in the form of a closed loop and any two-dimensional surface $S \subset \mathcal{R}$ with $\partial S = C$, and directly have

$$\partial S = C, \quad \partial \mathcal{C} = \partial \tilde{\mathcal{C}} = 0;$$

$$\varphi_g[C] = - \int_S \omega = - \oint_{\mathcal{C}} A. \quad (3.20)$$

Going back to the general definition (3.16) of $\varphi_g[C]$ for an open C , we now compare with Eq. (3.20) for closed C and draw the conclusion

$$C \text{ class I } \rho_1 \text{ to } \rho_2, \quad C' \text{ NPC } \rho_2 \text{ to } \rho_1, \quad \partial(C \cup C') = 0;$$

$$\varphi_g[C] = \varphi_g[C \cup C']. \quad (3.21)$$

The kinematic definition (2.23) for $\varphi_g[C]$ is immediately recovered from the present definition (3.16). With reference to the latter, let \mathcal{C} be any lift of C from any $\psi_1 \in \pi^{-1}(\rho_1)$ to any $\psi_2 \in \pi^{-1}(\rho_2)$, and let \mathcal{C}' , an NPC, be any lift of the NPC C' from ψ_2 to ψ_1 . Then using Eqs. (2.11) and (3.10), we obtain

$$\varphi_g[C] = - \int_S \omega = - \oint_{\mathcal{C} \cup \mathcal{C}'} A$$

$$= - \int_{\mathcal{C}} A - \int_{\mathcal{C}'} A = \arg(\psi_1, \psi_2) - \int_{\mathcal{C}} A, \quad (3.22)$$

which is Eq. (2.23).

D. The BI-GP connections

There are two important formulas connecting BI's and GP's. Both of them can be derived from the definition (3.16) with the property (3.10) for NPC's.

Let $\rho_1, \rho_2, \rho_3 \in \mathcal{R}$ be images of $\psi_1, \psi_2, \psi_3 \in \mathcal{B}$, no two in either triplet being mutually orthogonal. Join them pairwise by NPC's: C_{12} from ρ_1 to ρ_2 , and C_{23} from ρ_2 to ρ_3 , and C_{31} from ρ_3 to ρ_1 , their lifts \mathcal{C}_{12} from ψ_1 to ψ_2 , \mathcal{C}_{23} from ψ_2 to ψ_3 , and \mathcal{C}_{31} from ψ_3 to ψ_1 . Now both $C_{12} \cup C_{23} \cup C_{31}$ and

$C_{12} \cup C_{23} \cup C_{31}$ are closed loops, so we can use Eq. (3.20) to get

C 's, C 's NPC's;

$$\begin{aligned} \varphi_g[C_{12} \cup C_{23}] &= \varphi_g[C_{12} \cup C_{23} \cup C_{31}] \\ &= - \oint_{C_{12} \cup C_{23} \cup C_{31}} A = - \int_{C_{12}} A - \int_{C_{23}} A - \int_{C_{31}} A \\ &= - \arg \Delta_3(\psi_1, \psi_2, \psi_3). \end{aligned} \quad (3.23)$$

This is in fact the most general connection between GP's and (phases of) BI's, as discussed elsewhere [12]. It takes the known connection (3.15) between NPC's and BI's one step further and brings in the GP. Equation (3.23) goes with Eq. (3.15) for $m=3$. For $m \geq 4$, we have, using the notations of Eq. (3.15),

C 's, C 's NPC's;

$$\begin{aligned} \varphi_g[C_{12} \cup C_{23} \cup \dots \cup C_{m1}] &= - \int_{C_{12}} A - \int_{C_{23}} A - \dots - \int_{C_{m1}} A \\ &= - \arg \Delta_m(\psi_1, \psi_2, \dots, \psi_m). \end{aligned} \quad (3.24)$$

The second formula brings out the role of BI's in showing the nonadditivity of GP's and its derivation exploits Eq. (3.23). For triplets of points $\rho_1, \rho_2, \rho_3 \in \mathcal{R}$, $\psi_1, \psi_2, \psi_3 \in \mathcal{B}$ as before, let C_{12} and C_{23} be any class-I curves, not necessarily NPC's from ρ_1 to ρ_2 and ρ_2 to ρ_3 , respectively. Next let $C'_{21}, C'_{32}, C'_{31}$ be NPC's from ρ_2 to ρ_1 , ρ_3 to ρ_2 , and ρ_3 to ρ_1 , respectively. These are needed to define the GP's which appear below. Finally, we choose two-dimensional surfaces S_1, S_2, S_3 with boundaries

$$\begin{aligned} \partial S_1 &= C_{12} \cup C'_{21}, \quad \partial S_2 = C_{23} \cup C'_{32}, \\ \partial S_3 &= \tilde{C}'_{21} \cup \tilde{C}'_{32} \cup C'_{31}, \\ \partial(S_1 \cup S_2 \cup S_3) &= C_{12} \cup C_{23} \cup C'_{31}. \end{aligned} \quad (3.25)$$

Then repeatedly using the definition (3.16) and at the last step appealing to the result (3.23), we find

$$\begin{aligned} &\varphi_g[C_{12} \cup C_{23}] - \varphi_g[C_{12}] - \varphi_g[C_{23}] \\ &= - \int_{S_1 \cup S_2 \cup S_3} \omega + \int_{S_1} \omega + \int_{S_2} \omega = - \int_{S_3} \omega \\ &= \varphi_g[\tilde{C}'_{21} \cup \tilde{C}'_{32} \cup C'_{31}] \\ &= - \arg \Delta_3(\psi_1, \psi_2, \psi_3), \\ \text{i.e., } \varphi_g[C_{12} \cup C_{23}] &= \varphi_g[C_{12}] + \varphi_g[C_{23}] \\ &\quad - \arg \Delta_3(\psi_1, \psi_2, \psi_3). \end{aligned} \quad (3.26)$$

This is a known result, the purpose here was to derive it as a logical consequence of Eq. (3.16).

An exception to the general lack of additivity expressed by Eq. (3.26) occurs when we take $\rho_3 = \rho_2$ because BI $\Delta_3(\psi_1, \psi_2, \psi_1) = \Delta_2(\psi_1, \psi_2)$ is real positive. In this case, we have

$$\partial S = C_{12} \cup C_{21};$$

$$\begin{aligned} \varphi_g[C_{12} \cup C_{21}] &= \varphi_g[C_{12}] + \varphi_g[C_{21}] \\ &= \varphi_g[C_{12}] - \varphi_g[\tilde{C}_{21}] \\ &= - \int_S \omega. \end{aligned} \quad (3.27)$$

This will be used in the sequel.

IV. EXAMPLES AND PROPERTIES OF NULL PHASE CURVES

We have mentioned that for any two points $\rho_1, \rho_2 \in \mathcal{R}$, there is a geodesic connecting them, which is unique when $\text{Tr}(\rho_1 \rho_2) > 0$, and that geodesics are NPC's. In this section, we explore NPC's from several points of view, so as to visualize them better. We will show by explicit construction that for dimension $\mathcal{H} \geq 3$, given $\rho_1, \rho_2 \in \mathcal{R}$ with $\text{Tr}(\rho_1 \rho_2) > 0$, there are infinitely many (in a quite nontrivial sense) NPC's connecting ρ_1 to ρ_2 . We follow this up by developing an explicit analytical description, as far as is possible, of the most general NPC from ρ_1 to ρ_2 . Finally, we explore the differential geometric properties and characterization of smooth submanifolds $M \subset \mathcal{R}$ with the property that every continuous once-differentiable curve $C \subset M$ is an NPC, and give examples of such submanifolds.

It is instructive to see how the condition $\dim \mathcal{H} \geq 3$ for the existence of nontrivial NPC's arises. For $\dim \mathcal{H} = 2$, the ray space \mathcal{R} is the Poincaré sphere S^2 . If now three points $\rho_1, \rho_2, \rho_3 \in \mathcal{R}$ correspond to respective unit vectors $\hat{n}_1, \hat{n}_2, \hat{n}_3 \in S^2$, then, as is known, $\arg \text{Tr}(\rho_1 \rho_2 \rho_3)$ is one-half of the solid angle subtended at the center of S^2 by the spherical triangle with vertices $\hat{n}_1, \hat{n}_2, \hat{n}_3$ [18]. Thus, for condition (3.3) to be obeyed for any three points on an NPC, this solid angle must always vanish, so the NPC must be contained within some great circle. More explicitly, ρ is expressible in terms of its representative point $\hat{n} \in S^2$ as

$$\rho = \frac{1}{2} [1 + \hat{n} \cdot \vec{\sigma}], \quad (4.1)$$

and then

$$\text{Tr}\{\rho(s)[\rho(s'), \rho(s'')]\} = \frac{i}{2} \hat{n}(s) \cdot \hat{n}(s') \times \hat{n}(s''). \quad (4.2)$$

Parametrizing the \hat{n} 's with spherical polar angles θ, ϕ on S^2 in the usual way, and assuming with no loss of generality $\hat{n}(s) = (0, 0, 1)$, the vanishing of $\hat{n}(s) \cdot \hat{n}(s') \times \hat{n}(s'')$ for all independent s', s'' amounts to

$$\frac{x_2(s')}{x_1(s')} = \tan \phi(s') = \text{const.} \quad (4.3)$$

Hence, the NPC condition corresponds, after rotating to the configuration $\hat{n}(s) = (0,0,1)$, to $\phi = \text{const}$; and we find that NPC's are great circle arcs or geodesics on S^2 . Given any two nonantipodal points on S^2 , then an NPC connecting them is either the corresponding geodesic, or it may explore some more extended portion of the corresponding great circle. The vast generalization involved in going from geodesics to NPC's really shows up only for $\dim \mathcal{H} \geq 3$.

For convenience, as in the preceding section, the present one is also divided into further sections.

A. Examples of null phase curves

Let two distinct points $\rho_1, \rho_2 \in \mathcal{R}$ with $\text{Tr}(\rho_1 \rho_2) > 0$ be given. We will construct examples of class-II curves $C \subset \mathcal{R}$ from ρ_1 to ρ_2 which are NPC's. Since by Eq. (2.9) every point $\rho(s) \in C$ must obey $\text{Tr}[\rho_1 \rho(s)] > 0$, it follows that C must lie entirely in the neighborhood $\mathcal{R}(\rho_1) \subset \mathcal{R}$ of ρ_1 defined in the manner of Eq. (A.8). We can therefore use a local description of $\mathcal{R}(\rho_1)$ as set up in Appendix A.

Let $\psi_1 \in \pi^{-1}(\rho_1)$ and choose $\psi_2 \in \pi^{-1}(\rho_2)$ such that (ψ_1, ψ_2) is real positive. Introduce an angle θ_0 by

$$(\psi_1, \psi_2) = \cos \theta_0, \quad \theta_0 \in (0, \pi/2). \quad (4.4)$$

Let us when convenient write $\psi_1 = e_1$. According to Eq. (A.10) we can express ψ_2 in the form

$$\begin{aligned} \psi_2 &= e_1 \cos \theta_0 + e_2 \sin \theta_0, \\ (e_1, e_2) &= 0, \quad \|e_2\| = 1, \end{aligned} \quad (4.5)$$

so $\psi_1 = e_1, e_2$ form an orthonormal pair. As in Appendix A we supplement e_1, e_2 by further vectors $e_3, e_4, \dots \in \mathcal{B}$, terminating with e_n if dimension $\mathcal{H} = n$ is finite, such that $\{\psi_1 = e_1, e_2, e_3, \dots\}$ is an orthonormal basis for \mathcal{H} . Let $\mathcal{C} = \{\psi(s) | s_1 \leq s \leq s_2\}$ be a lift of C from ψ_1 to ψ_2 . For $\psi(s)$, we write

$$\begin{aligned} \psi(s) &= x_1(s)e_1 + x_2(s)e_2 + x_3(s)e_3 + \dots, \\ x_1(s) &\neq 0, \quad x_1(s)\cos \theta_0 + x_2(s)\sin \theta_0 \neq 0, \\ |x_1(s)|^2 + |x_2(s)|^2 + |x_3(s)|^2 + \dots &= 1, \quad s_1 \leq s \leq s_2. \end{aligned} \quad (4.6)$$

The coefficients $x_1(s), x_2(s), \dots$ must be continuous once differentiable. At the end points, they have real values

$$\begin{aligned} x(s_1) &= (1, 0, 0, \dots), \\ x(s_2) &= (\cos \theta_0, \sin \theta_0, 0, 0, \dots). \end{aligned} \quad (4.7)$$

Now choose any integer $m \in (3, 4, \dots, n)$ and consider the real unit sphere $S^{m-1} \subset \mathbb{R}^m$. Assume $x(s)$ to have real components, with $x_{m+1}(s) = x_{m+2}(s) = \dots = 0$ for $s \in [s_1, s_2]$. Thus the first m components of $x(s)$ describe a moving point on S^{m-1} , with $x_1(s)$ and $x_1(s)\cos \theta_0$

$+x_2(s)\sin \theta_0$ nonzero throughout. Let us further limit ourselves when $s_1 < s < s_2$ to vectors on S^{m-1} with all components strictly positive. That is, generalizing the positive quadrant and octant in two and three dimensions, we define

$$S_+^{m-1} = \{x \in S^{m-1} | x_1, x_2, \dots, x_m > 0\} \subset S^{m-1}, \quad (4.8)$$

and choose

$$x(s) \in S_+^{m-1}, \quad s_1 < s < s_2. \quad (4.9)$$

Then the vectors

$$\psi(s) = x_1(s)e_1 + x_2(s)e_2 + x_3(s)e_3 + \dots + x_m(s)e_m \quad (4.10)$$

obey

$$(\psi(s), \psi(s')) = x(s) \cdot x(s') = \text{real positive}, \quad s, s' \in [s_1, s_2]. \quad (4.11)$$

Condition (3.3) for \mathcal{C}, C to be an NPC is clearly satisfied, so we have succeeded in constructing infinitely many NPC's from ρ_1 to ρ_2 . In this construction, the integer $m \in (3, 4, \dots, n)$, and the vectors e_3, e_4, \dots, e_m forming along with $\psi_1 = e_1$ and e_2 an orthonormal set in \mathcal{B} may each be freely chosen; and then $x(s)$ for $s_1 < s < s_2$ is any once-differentiable curve on S_+^{m-1} obeying the boundary conditions (4.7) at the end points.

This great profusion of NPC's as compared to geodesics, available only when dimension $\mathcal{H} \geq 3$, is an indication that the former are not solutions to any system of local ordinary differential equations with some boundary conditions, in the way familiar with geodesics. It is also clear from the deductive development in Sec. III that it is NPC's that are basic to the theory of the GP, and in a sense it is incidental that geodesics are NPC's. These remarks lead to the following interesting questions: given any two distinct nonorthogonal points $\rho_1, \rho_2 \in \mathcal{R}$, how can we describe in a constructive sense the most general NPC from ρ_1 to ρ_2 ; and how can we characterize a smooth submanifold $M \subset \mathcal{R}$ if it has the property that every continuous once-differentiable curve $C \subset M$ is an NPC? The latter kind of question is clearly not meaningful in the case of geodesics. We will find that here again the third-order BI plays a key role.

B. Description of a general null phase curve

Now we develop a description of the most general NPC $C = \{\rho(s) | s_1 \leq s \leq s_2\}$ connecting two given points $\rho_1, \rho_2 \in \mathcal{R}$ with $\text{Tr}(\rho_1 \rho_2) > 0$. We assume vectors $\psi_{1,2} \in \pi^{-1}(\rho_{1,2})$ obeying Eqs. (4.4) and (4.5) have been chosen. Let $\mathcal{C}_0 = \{\psi_0(s) | s \in [s_1, s_2]\}$ be the particular lift of C , from ψ_1 to ψ_2 , obeying the condition (3.6) so

$$\psi_0(s_1) = \psi_1, \quad \psi_0(s_2) = \psi_2,$$

$$(\psi_0(s'), \psi_0(s)) = \text{real positive}, \quad s', s \in [s_1, s_2]. \quad (4.12)$$

We expand $\psi_0(s)$ as

$$\begin{aligned}\psi_0(s) &= x_1(s)e_1 + x_2(s)e_2 + \chi(s), \\ (e_1, \chi(s)) &= (e_2, \chi(s)) = 0, \quad s \in [s_1, s_2].\end{aligned}\quad (4.13)$$

At this stage, $x_1(s)$ and $x_2(s)$ are complex continuous once-differentiable functions of s , while $\chi(s)$ is a continuous once-differentiable vector in the subspace $\mathcal{H}_\perp(\psi_1, \psi_2) \subset \mathcal{H}$ orthogonal to the pair ψ_1, ψ_2 i.e., to e_1, e_2 . At the end points, we have

$$\begin{aligned}x_1(s_1) &= 1, \quad x_2(s_1) = \chi(s_1) = 0; \\ x_1(s_2) &= \cos \theta_0, \quad x_2(s_2) = \sin \theta_0, \quad \chi(s_2) = 0.\end{aligned}\quad (4.14)$$

Now we draw out step by step the implications of the real positivity condition (4.12), and of $\psi_0(s) \in \mathcal{B}$ for all s . From the real positivity of $(\psi_1, \psi_0(s))$ and $(\psi_2, \psi_0(s))$, we get

$$\begin{aligned}x_1(s) &= \text{real positive}, \\ x_1(s)\cos \theta_0 + x_2(s)\sin \theta_0 &= \text{real positive}, \quad s \in [s_1, s_2].\end{aligned}\quad (4.15)$$

These imply

$$\begin{aligned}x_2(s) &= \text{real}, \\ x_1(s)\cos \frac{\theta_0}{2} + x_2(s)\sin \frac{\theta_0}{2} &= \text{real positive}, \quad s \in [s_1, s_2].\end{aligned}\quad (4.16)$$

Thus, $x_1(s)$ can never vanish, while $x_2(s)$ could vanish, as it does at $s = s_1$, or even sometimes be negative. Next from the normalization of $\psi_0(s)$, we have

$$\|\psi_0(s)\| = 1 \Leftrightarrow x_1(s)^2 + x_2(s)^2 + \|\chi(s)\|^2 = 1, \quad s \in [s_1, s_2].\quad (4.17)$$

We therefore parametrize $x_1(s)$ and $x_2(s)$ by

$$\begin{aligned}x_1(s) &= \sigma(s)\cos \theta(s), \quad x_2(s) = \sigma(s)\sin \theta(s), \\ 0 < \sigma(s) &\leq 1, \quad s \in [s_1, s_2]; \\ \sigma(s_1) &= \sigma(s_2) = 1, \quad \theta(s_1) = 0, \quad \theta(s_2) = \theta_0.\end{aligned}\quad (4.18)$$

Both $\sigma(s)$ and $\theta(s)$ are continuous once differentiable, and for the norm of $\chi(s)$ we have

$$\|\chi(s)\| = (1 - \sigma(s)^2)^{1/2} \in [0, 1].\quad (4.19)$$

The positivity conditions (4.15) lead to the allowed range for $\theta(s)$,

$$-\pi/2 + \theta_0 < \theta(s) < \pi/2,\quad (4.20)$$

which exceeds $\pi/2$ in extent. The permitted region in the x_1 - x_2 plane is thus a segment OAB of the unit disk subtending an angle $(\pi - \theta_0)$ at the center, as shown in Fig. 1. The open arc A to B is included, while the end points A, B , and the radii OA, OB are excluded. At this point, $\sigma(s)$ and $\theta(s)$,

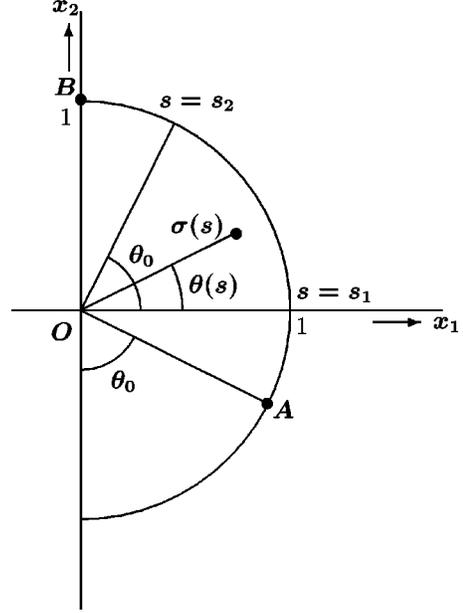


FIG. 1. The allowed region in the x_1 - x_2 plane is the segment OAB of the unit disc, excluding the radii OA and OB .

hence $x_1(s)$ and $x_2(s)$, can be chosen freely subject to the bounds and conditions given in Eqs. (4.18) and (4.20).

Now we turn to the more comprehensive positivity condition (4.12),

$$\begin{aligned}(\psi_0(s'), \psi_0(s)) &= \text{real positive} \\ &\Leftrightarrow \sigma(s')\sigma(s)\cos(\theta(s') - \theta(s)) \\ &\quad + (\chi(s'), \\ \chi(s)) &= \text{real positive}, \\ s', s &\in [s_1, s_2].\end{aligned}\quad (4.21)$$

An immediate conclusion is that for all s' and $s, (\chi(s'), \chi(s))$ is real. One can show quite easily that this means the following: there is some orthonormal basis $\{e_3, e_4, \dots\}$ for the subspace $\mathcal{H}_\perp(\psi_1, \psi_2) \subset \mathcal{H}$ such that

$$\begin{aligned}\chi(s) &= \sum_{r=3,4,\dots} x_r(s)e_r, \\ x_r(s) &= \text{real},\end{aligned}$$

$$(\chi(s'), \chi(s)) = \sum_{r=3,4,\dots} x_r(s')x_r(s) = \text{real},$$

$$\|\chi(s)\|^2 = \sum_{r=3,4,\dots} x_r(s)^2 = 1 - \sigma(s)^2 \in [0, 1].\quad (4.22)$$

The orthonormal set $\{e_3, e_4, \dots\}$, defined of course upto a real orthogonal transformation, depends in general on the particular NPC C from ρ_1 to ρ_2 that we are describing; it need not be the same for all NPC's from ρ_1 to ρ_2 . The information on $\|\chi(s)\|$ leads via the Cauchy-Schwartz inequality to a bound on the magnitude of the last term on the right-hand side of the condition (4.21):

$$|(\chi(s'), \chi(s))| \leq \{(1 - \sigma(s')^2)(1 - \sigma(s)^2)\}^{1/2}. \quad (4.23)$$

Since, as already mentioned, the range for $\theta(s)$ in Eq. (4.20) exceeds $\pi/2$ in extent, the difference $\theta(s') - \theta(s)$ can sometimes exceed $\pi/2$ in magnitude; this would make the first term in the inequality (4.21) negative. In that case, $(\chi(s'), \chi(s))$ must be positive and large enough to compensate for this, while still subject to Eq. (4.23). We are thus led to an interesting nonlocal condition on $\sigma(s)$ and $\theta(s)$:

$$\frac{\sigma(s')\sigma(s)}{\{(1 - \sigma(s')^2)(1 - \sigma(s)^2)\}^{1/2}} |\cos(\theta(s') - \theta(s))| \leq 1$$

if $|\theta(s') - \theta(s)| > \pi/2$. (4.24)

This condition, being nonlocal, cannot be translated in any simple way to a further restriction on the so-far allowed ranges for $\sigma(s)$ and $\theta(s)$, but must be carried along as a nontrivial condition to be obeyed by them. For such allowed choices of $\sigma(s)$ and $\theta(s)$, $\chi(s)$ must then be chosen as in Eq. (4.22) ensuring that the inequality (4.21) is obeyed.

This is the extent to which an explicit constructive description of a general NPC from ρ_1 to ρ_2 can be given. Admittedly, it is much less “complete” than the description we can give for a geodesic, again an indication that NPC’s form a much larger family of curves than do geodesics. We can now see that the examples of NPC’s given in the preceding section correspond to the special simplifying assumption that the functions $x_1(s), x_2(s), x_3(s), \dots$ are all non-negative. Then $|\theta(s') - \theta(s)|$ never exceeds $\pi/2$, and the nonlocal conditions (4.21) and (4.24) are automatically obeyed as even $(\chi(s'), \chi(s))$ is throughout real non-negative.

C. Submanifolds of null phase curves

Now we enlarge the scope of our analysis of NPC’s and ask the question: how can we characterize a smooth submanifold $M \subset \mathcal{R}$ if every once-differentiable curve $C \subset M$ is to be an NPC? For brevity let us call such a submanifold an NPM. To answer this question we first assemble the basic formula which generalizes Eq. (3.26) and connects GP’s, BI’s, and two-dimensional surface integrals of the two-form ω . Let $C \subset \mathcal{R}$ be any closed class-I curve, and $S \subset \mathcal{R}$ any two-dimensional surface with $\partial S = C$. Let ρ_1, ρ_2, ρ_3 be any three pairwise nonorthogonal points chosen in sequence along C . Denote the successive portions of C from ρ_1 to ρ_2 , ρ_2 to ρ_3 and ρ_3 to ρ_1 by C_{12}, C_{23} , and C_{31} , respectively. Then we have the relation

$$\begin{aligned} \varphi_g[C = C_{12} \cup C_{23} \cup C_{31}] &= \varphi_g[C_{12}] + \varphi_g[C_{23}] + \varphi_g[C_{31}] \\ &\quad - \arg \Delta_3(\psi_1, \psi_2, \psi_3) \\ &= - \int_S \omega \end{aligned} \quad (4.25)$$

for any $\psi_j \in \pi^{-1}(\rho_j)$, $j=1,2,3$. If as a special case we set $\rho_3 = \rho_2$ here, recall the definition (3.16) and the fact that $\Delta_2(\psi_1, \psi_2)$ is real positive, we recover Eq. (3.27). Thus, the

latter is a particular special case of Eq. (4.25). However, in the reverse direction, of course, it is not possible to derive Eq. (4.25) from Eq. (3.27). Now Eq. (4.25) can be generalized to the case where we choose any $m \geq 4$ points $\rho_1, \rho_2, \dots, \rho_m \in C$, located one after the other in a sequence, with no two consecutive ones being orthogonal. Then in an obvious notation we have

$$\begin{aligned} \varphi_g[C = C_{12} \cup C_{23} \cup \dots \cup C_{m1}] &= \varphi_g[C_{12}] + \varphi_g[C_{23}] + \dots + \varphi_g[C_{m1}] \\ &\quad - \arg \Delta_m(\psi_1, \psi_2, \dots, \psi_m) \\ &= - \int_S \omega, \quad m \geq 4. \end{aligned} \quad (4.26)$$

This is also derivable from Eq. (4.25) which is the primitive relation of this kind. The proofs of Eqs. (4.25) and (4.26) follow the pattern of the arguments in Sec. III. In these relations, if all the segments $C_{12}, C_{23}, \dots, C_{m1}$ are NPC’s, we recover Eqs. (3.23) and (3.24).

Now let $M \subset \mathcal{R}$ be a connected, simply connected smooth submanifold with dimension $M \geq 2$ in the real sense [19]. The corresponding identification map is $i_M: M \hookrightarrow \mathcal{R}$. The pullback to M of the symplectic two-form ω on \mathcal{R} is

$$\omega_M = i_M^* \omega. \quad (4.27)$$

This is of course closed but may well be degenerate. In the extreme case of an isotropic submanifold it vanishes,

$$M \text{ isotropic} \Leftrightarrow \omega_M = 0. \quad (4.28)$$

If in addition $\dim M = (n-1)$, it is a Lagrangian submanifold.

Assume now that M is an NPM. Every once-differentiable curve $C \subset M$ must then be of class II and we get the following two consequences:

$$M \text{ is an NPM} \Rightarrow M \text{ is isotropic, } \omega_M = 0; \quad (4.29a)$$

$$M \text{ is an NPM} \Rightarrow \text{for any } \rho_j = \pi(\psi_j) \in M, \quad j=1,2,3, \\ \Delta_3(\psi_1, \psi_2, \psi_3) \text{ is real positive.} \quad (4.29b)$$

The first follows from Eq. (3.14) after specializing to $C, C', S \subset M$. The second then follows from the first upon use of Eq. (4.25), again specializing to $C_{12}, C_{23}, C_{31}, S \subset M$. From Eq. (4.29b) we also see by taking $\rho_3 = \rho_2$ that any two points in M are nonorthogonal. This means that every once-differentiable curve $C \subset M$ is of class II.

Conversely, if we start by assuming Eq. (4.29b) for M , we deduce

$$\begin{aligned} &\Delta_3(\psi_1, \psi_2, \psi_3) \text{ real positive,} \\ &\text{any } \rho_j = \pi(\psi_j) \in M, \quad j=1,2,3 \\ &\Rightarrow \text{Tr}(\rho_1 \rho_2) > 0, \quad \text{any } \rho_1, \rho_2 \in M; \end{aligned} \quad (4.30a)$$

$$\Rightarrow M \text{ is an NPM}; \quad (4.30b)$$

$$\Rightarrow M \text{ is isotropic, } \omega_M = 0. \quad (4.30c)$$

These three statements are not independent since, by Eqs. (4.29), (4.30b) implies Eq. (4.30c). In any case, we see from Eqs. (4.29b) and (4.30b) together that the necessary and sufficient condition for M to be an NPM is that $\Delta_3(\psi_1, \psi_2, \psi_3)$ be real positive if $\rho_j = \pi(\psi_j) \in M$, $j=1,2,3$. It then follows that the GP for an open or closed class-I curve lying entirely in an NPM is zero. Isotropy of M is thus only a necessary but not a sufficient condition to ensure this property for M . It is therefore instructive to see how far one can go on the basis of isotropy alone. We now examine this point.

Let M be a connected, simply connected submanifold in \mathcal{R} , such that $\text{Tr}(\rho_1\rho_2) > 0$ for any $\rho_1, \rho_2 \in M$. As in the preceding section for an NPC, we construct a lift of M to a submanifold $\mathcal{M}_0 \subset \mathcal{B}$, in the spirit of the Pancharatnam lift. Choose any point $\rho_0 \in M$ and then any $\psi_0 \in \pi^{-1}(\rho_0) \subset \mathcal{B}$. The lift \mathcal{M}_0 is then completely and uniquely defined by the rule [generalization of Eq. (3.11)]

$$\begin{aligned} \rho \in M \rightarrow \psi &= \frac{\rho\psi_0}{\sqrt{\text{Tr}(\rho_0\rho)}} \in \mathcal{M}_0, \\ \pi(\psi) &= \rho, \end{aligned} \quad (4.31)$$

so \mathcal{M}_0 as a subset of \mathcal{B} is displayed as

$$\mathcal{M}_0 = \{\rho\psi_0 / \sqrt{\text{Tr}(\rho_0\rho)}, \rho \in M\} \subset \mathcal{B}. \quad (4.32)$$

Of course, $\rho_0 \in M$ is lifted to $\psi_0 \in \mathcal{M}_0$. This lift $M \rightarrow \mathcal{M}_0$ is characterized by the fact that each vector $\psi \in \mathcal{M}_0$ is in phase with ψ_0 in the Pancharatnam sense, from Eq. (4.31),

$$(\psi_0, \psi) = \frac{(\psi_0, \rho\psi_0)}{\sqrt{\text{Tr}(\rho_0\rho)}} = \sqrt{\text{Tr}(\rho_0\rho)} = \text{real} > 0. \quad (4.33)$$

However, if we take two general vectors $\psi, \psi' \in \mathcal{M}_0$, we find that their inner product is, in general, complex:

$$\begin{aligned} (\psi', \psi) &= (\rho'\psi_0, \rho\psi_0) / \sqrt{\text{Tr}(\rho_0\rho')\text{Tr}(\rho_0\rho)} \\ &= \text{Tr}(\rho_0\rho'\rho) / \sqrt{\text{Tr}(\rho_0\rho')\text{Tr}(\rho_0\rho)}, \\ \arg(\psi', \psi) &= \arg \Delta_3(\psi_0, \psi', \psi). \end{aligned} \quad (4.34)$$

Thus, whether or not general pairs of points in \mathcal{M}_0 are in phase depends entirely on whether the BI's $\Delta_3(\psi_0, \psi', \psi)$ for triplets of points in M are real positive or complex. This is the sense in which the lift $M \rightarrow \mathcal{M}_0$ is as nearly a Pancharatnam lift as possible in a general case.

Now let us impose the condition that M be isotropic,

$$\omega_M =: i_M^* \omega = 0. \quad (4.35)$$

Then Eq. (3.27) shows that if C_{12}, C'_{12} are any two class-I curves in M from any $\rho_1 \in M$ to any $\rho_2 \in M$, the GP's are the same;

$$\varphi_g[C'_{12}] = \varphi_g[C_{12}]. \quad (4.36)$$

Furthermore, it is clear that this statement exhausts the content of isotropy. Since the lift $M \rightarrow \mathcal{M}_0$ is unique (given ρ_0 and ψ_0), we can faithfully transcribe this statement to \mathcal{M}_0 , which is more convenient since we then deal with vectors. Denote by ψ_1 and ψ_2 the lifts of ρ_1 and ρ_2 to \mathcal{M}_0 , and by C_{12}, C'_{12} the (unique) lifts of C_{12}, C'_{12} to \mathcal{M}_0 . Then the full content of Eq. (4.36) is expressed as follows: for any two points $\psi_1, \psi_2 \in \mathcal{M}_0$, and for any class-I curves C_{12}, C'_{12} in \mathcal{M}_0 connecting them, we have

$$\begin{aligned} \arg(\psi_1, \psi_2) - \int_{C_{12}} A &= \arg(\psi_1, \psi_2) - \int_{C'_{12}} A, \\ \text{i.e., } \int_{C_{12}} A &= \int_{C'_{12}} A. \end{aligned} \quad (4.37)$$

This means that (as we have assumed simple connectedness) the pullback of A from \mathcal{B} to \mathcal{M}_0 is exact. Denoting the relevant identification map as $i_{\mathcal{M}_0}: \mathcal{M}_0 \hookrightarrow \mathcal{B}$, we have

$$\omega_M = 0 \Leftrightarrow i_{\mathcal{M}_0}^* A = df \quad (4.38)$$

for some $f \in \mathcal{F}(\mathcal{M}_0)$. For emphasis we repeat that isotropy of M allows us to conclude that $i_{\mathcal{M}_0}^* A$ is exact, and allows (ψ, ψ') for $\psi, \psi' \in \mathcal{M}_0$ to have a nontrivial phase.

If at this point we assume in addition that M is an NPM, we immediately see that we have much stronger conclusions;

$$\begin{aligned} (\psi, \psi') &= \text{real positive, any } \psi, \psi' \in \mathcal{M}_0, \\ i_{\mathcal{M}_0}^* A &= 0. \end{aligned} \quad (4.39)$$

These results imply that now \mathcal{M}_0 is truly a Pancharatnam lift of M , and they show very effectively the extent to which the NPM property goes beyond isotropy.

To round out this discussion we give some examples of submanifolds $M \subset \mathcal{R}$, connected and simply connected, possessing the NPM property. To begin with, we use a construction similar to that used in Sec. IV A to construct families of NPC's connecting any two given nonorthogonal points $\rho_1, \rho_2 \in \mathcal{R}$. For any $m \geq 3$ (upto n in case $\dim \mathcal{H} = n$ is finite), let $e_r, r=1, 2, \dots, m$, be an orthonormal set of vectors in \mathcal{H} . We first define a submanifold $\mathcal{M} \subset \mathcal{B}$ as consisting of all real normalized "positive" linear combinations of the e_r :

$$\begin{aligned} \mathcal{M} = \left\{ \psi(x) = \sum_{r=1}^m x_r e_r \in \mathcal{B} \mid x_r \text{ real positive,} \right. \\ \left. \sum_{r=1}^m x_r^2 = 1 \right\} \subset \mathcal{B}, \end{aligned} \quad (4.40)$$

and then take the projection to get M ,

$$M = \pi(\mathcal{M}) \subset \mathcal{R}. \quad (4.41)$$

Both \mathcal{M} and M are of real dimension $(m-1)$, in fact \mathcal{M} is essentially S_+^{m-1} of Eq. (4.8). By construction we see that

$$(\psi(x'), \psi(x)) = x' \cdot x = \sum_{r=1}^m x'_r x_r = \text{real positive}, \quad (4.42)$$

and therefore for any three points in M ,

$$\Delta_3(\psi(x), \psi(x'), \psi(x'')) = x \cdot x' x' \cdot x'' x'' \cdot x = \text{real positive}. \quad (4.43)$$

This ensures that every once-differentiable curve $C \subset M$, obtained by setting $x = x(s)$ for suitable functions $x_r(s)$, is of class II and also an NPC.

After this abstract example of an NPM, we give two others involving explicit families of Schrödinger wave functions, which are simple but quite relevant. The context is the family of coherent states of a system of N identical simple harmonic oscillators. Starting with the normalized vacuum state $|0\rangle$ with the wave function

$$|0\rangle \rightarrow \psi_0(X) = \pi^{-N/4} \exp(-\frac{1}{2} X \cdot X),$$

$$X \cdot X = \sum_{j=1}^N x_j^2, \quad (4.44)$$

the spatial translates of $|0\rangle, |Y\rangle$ say, have wave functions

$$|Y\rangle \rightarrow \psi_Y(X) = \pi^{-N/4} \exp[-\frac{1}{2} (X - Y) \cdot (X - Y)]. \quad (4.45)$$

Taken for all $Y \in \mathcal{R}^n$, these states constitute a submanifold \mathcal{R}^n in the manifold \mathcal{R}^{2n} of all coherent states; and the former clearly form an NPM. Clearly, the image of this manifold of states under any unitary transformation will also be an NPM. In particular, the submanifold of all momentum translates of the vacuum state form an NPM.

The second example is in the context of the manifold $Sp(2N, R)/U(N)$ of all squeezed vacuum states obtained as the orbit of the state (4.44) under the unitary action of the group $Sp(2N, R)$ of all real linear canonical transformations [20]. These states can be characterized by a pair of real symmetric $N \times N$ matrices u, v with $u > 0$, and so they form an $N(N+1)$ -dimensional submanifold in Hilbert space. The corresponding normalized wave functions are

$$|u, v\rangle \rightarrow \psi_{(u,v)}(X) = \pi^{-N/4} (\det u)^{1/4} \exp\{-\frac{1}{2} X^T (u + iv) X\}, \quad (4.46)$$

and for general u, v these are complex. This set of wave functions constitute a generalization of the Poincaré upper-half plane, with $Sp(2N, R)$ acting on u, v through (matrix) fractional linear transformations. In this set if we now limit ourselves to those with $v=0$, all the wave functions $\psi_{(u,0)}(X)$ are real and all scalar products among them are real positive. Thus we have another example of an NPM. Again the image of this submanifold under any unitary transformation retains this property.

V. CONCLUDING REMARKS

We have developed an approach to the theory of the GP in quantum mechanics, in which the basic ingredient is the three-vertex BI of quantum mechanics. This invariant leads to the definition of NPC's in the Hilbert and ray spaces of quantum systems. In turn, this lets us define the GP associated with any suitable (open or closed) ray space curve as an area integral of the ray space symplectic form. The emphasis has been on the logical basis and consistency of the entire development, and the crucial role of NPC's. We have shown that it is the BI which is the truly fundamental concept underlying all the others. In the course of the development, we have taken care to define with precision the classes of ray and Hilbert space curves that one must work with for each purpose.

NPC's are a vast and important generalization of the more familiar family of geodesics in Hilbert and ray spaces, and it turns out that they truly belong to the theory of the GP. It happens to be true that geodesics are instances of NPC's; however the latter are far more numerous by any measure, and are intrinsically of a quite different nature. Thus, length of a curve and its minimization are not at all the relevant concepts in arriving at NPC's. This makes their description considerably more difficult than of geodesics, for which a differential equation treatment is available. They have deep properties of a nonlocal nature. The examples of NPC's and NPM's, and their general properties brought out in our discussion, should help in aiding our understanding this important class of quantum mechanical objects.

The Pancharatnam lift \mathcal{M}_0 of an NPM M is characterized by the two properties (4.39). The first actually implies the second. Its structure suggests the following nonlocal operation or construction: pass from the collection of vectors $\psi, \psi', \dots \in \mathcal{M}_0$ to its real linear hull, i.e., form all real linear combinations of any numbers of vectors in \mathcal{M}_0 (and then normalize them to get results in \mathcal{B}). This much enlarged collection of vectors in \mathcal{B} is clearly associated with a real subspace of \mathcal{H} all inner products among whose vectors are real (but of course not anymore always positive). The consideration of NPM's leads in a natural way to associate real linear subspaces in \mathcal{H} within which the Hermitian scalar product of \mathcal{H} reduces to a real symmetric scalar product. Such a subspace is clearly $\pi^* \omega$ isotropic, and we are led to consider trying to characterize NPM's via such associated subspaces.

It is useful to view all this also from another perspective. The Fubini-Study metric and the symplectic form on the quantum-mechanical ray space both originate from the Kahler form, as its real and imaginary parts, respectively [21]. While the geodesics stem from the metric, the BI, NPC's, and GP are all more naturally related to the symplectic structure. Hence, it is that NPC's and not geodesics, form the principal notion in the GP context.

We have seen in Sec. IV that it is only when $\dim \mathcal{H} \geq 3$ that the true differences between NPC's and geodesics emerge. The situations in which the GP has been traditionally studied in detail have involved two-dimensional symplectic manifolds: CP^1 or S^2 corresponding to a two-level system

such as polarized light or spin-1/2 particles; the two-dimensional plane \mathcal{R}^2 as in the coherent states of a harmonic oscillator; and the timelike two-dimensional upper half unit hyperboloid in $(2+1)$ -dimensional space (equivalently the unit disc or the upper half complex plane), as in the squeezed vacuum states of an oscillator. In all these cases, there is simply no extra room in these minimal situations for the differences between NPC's and geodesics to show up. Thus, it is understandable that the geodesics were thought to be intrinsically relevant to GP discussions, but this status rightfully belongs to the NPC's.

We have examined elsewhere the conditions under which constrained geodesics, i.e., geodesics among curves restricted to lie within a given ray space submanifold could turn out to be NPC's [12]. Important examples when this happens, and the corresponding BI-GP connections which are of physical interest, have been given. The more comprehensive account of NPC's and NPM's presented in this paper should enable us to study the connections to constrained geodesics in a more definite manner. We intend to return to this and related problems elsewhere.

APPENDIX A: BASIC DIFFERENTIAL GEOMETRY OF RAY SPACE, LOCAL DESCRIPTIONS

For the convenience of the reader we collect here some basic definitions relating to the unit sphere $\mathcal{B} \subset \mathcal{H}$, the ray space $\mathcal{R} = \pi(\mathcal{B})$, defined in Eqs. (2.1) and (2.3), and geometric objects associated with them. Convenient local coordinate descriptions of some of them are also given. For definiteness, we may assume that dimension $\mathcal{H} = n$ is finite.

The space $\mathcal{B} = S^{2n-1}$ is a differentiable manifold. At any point $\psi_0 \in \mathcal{B}$, the tangent space is

$$T_{\psi_0}\mathcal{B} = \{\phi \in \mathcal{H} | \text{Re}(\psi_0, \phi) = 0\}. \quad (\text{A1})$$

This is clearly a real linear vector space of dimension $(2n-1)$. The connection one-form A defined on \mathcal{B} is specified at each $\psi_0 \in \mathcal{B}$ as a linear functional on $T_{\psi_0}\mathcal{B}$ is

$$\phi \in T_{\psi_0}\mathcal{B}: A_{\psi_0}(\phi) = \text{Im}(\psi_0, \phi) = -i(\psi_0, \phi). \quad (\text{A2})$$

Therefore, the horizontal subspace of $T_{\psi_0}\mathcal{B}$ is defined as

$$\begin{aligned} H_{\psi_0}\mathcal{B} &= \{\phi \in T_{\psi_0}\mathcal{B} | A_{\psi_0}(\phi) = 0\} \\ &= \{\phi \in \mathcal{H} | (\psi_0, \phi) = 0\} \subset T_{\psi_0}\mathcal{B}. \end{aligned} \quad (\text{A3})$$

This is of course a real linear vector space of dimension $2(n-1)$, but it is in a natural sense a complex linear vector space of dimension $(n-1)$, namely, the subspace of \mathcal{H} orthogonal to ψ_0 .

The two-form dA is, at each $\psi_0 \in \mathcal{B}$, an antisymmetric bilinear functional on $T_{\psi_0}\mathcal{B}$,

$$\phi, \phi' \in T_{\psi_0}\mathcal{B}: (dA)_{\psi_0}(\phi, \phi') = 2\text{Im}(\phi, \phi'). \quad (\text{A4})$$

Next we turn to the ray space \mathcal{R} . At a point $\rho_0 = \psi_0\psi_0^\dagger \in \mathcal{R}$, the tangent space can be defined and then described explicitly in terms of $H_{\psi_0}\mathcal{B}$:

$$T_{\rho_0}\mathcal{R} = \{B = \text{linear operator on } \mathcal{H} | B^\dagger = B, \text{Tr}B = 0,$$

$$\{B, \rho_0\} = B\} = \{B = \phi\psi_0^\dagger + \psi_0\phi^\dagger | \phi \in H_{\psi_0}\mathcal{B}\}. \quad (\text{A5})$$

This is a real linear vector space of dimension $2(n-1)$. In the latter form, if we change the representative vector $\psi_0 \in \pi^{-1}(\rho_0)$ by a phase, we must keep track of the change in B if ϕ is unchanged, or alternatively change ϕ by a compensating phase to keep B the same.

The two-form dA on \mathcal{B} is the pullback of a two-form ω on \mathcal{R} ,

$$dA = \pi^*\omega. \quad (\text{A6})$$

As an antisymmetric bilinear functional on $T_{\rho_0}\mathcal{R}$, ω is specified by

$$\omega_{\rho_0}(B, B') = -i\text{Tr}(\rho_0[B, B']) = 2\text{Im}(\phi, \phi'), \quad (\text{A7})$$

where $\phi, \phi' \in H_{\psi_0}\mathcal{B}$ correspond to B, B' , respectively, as in the second line of Eq. (A5).

All the above definitions and expressions are coordinate independent and intrinsic. Now we give local coordinate descriptions for some of them, which are sometimes useful. For given $\rho_0 \in \mathcal{R}$, and some chosen $\psi_0 \in \pi^{-1}(\rho_0)$, we define an open neighborhood $\mathcal{R}(\rho_0)$ by

$$\mathcal{R}(\rho_0) = \{\rho \in \mathcal{R} | \text{Tr}(\rho_0\rho) > 0\} \subset \mathcal{R}. \quad (\text{A8})$$

Thus, what are excluded from $\mathcal{R}(\rho_0)$ are projections ρ onto vectors in \mathcal{B} which are orthogonal to ψ_0 , that is, onto vectors in H_{ψ_0} . The corresponding open subset of \mathcal{B} is denoted by $\mathcal{B}(\psi_0)$, though in fact it is determined by ρ_0 :

$$\mathcal{B}(\psi_0) = \pi^{-1}[\mathcal{R}(\rho_0)] = \{\psi \in \mathcal{B} | (\psi_0, \psi) \neq 0\} \subset \mathcal{B}. \quad (\text{A9})$$

We can give an explicit formula for any $\rho \in \mathcal{R}(\rho_0)$ as follows:

$$\begin{aligned} \rho \in \mathcal{R}(\rho_0) &\Leftrightarrow \rho = \rho(\phi) \\ &= [\phi + (1 - \|\phi\|^2)^{1/2}\psi_0] \\ &\quad \times [\phi + (1 - \|\phi\|^2)^{1/2}\psi_0]^\dagger, \\ \phi &\in H_{\psi_0}\mathcal{B}, \quad \|\phi\| < 1; \end{aligned} \quad (\text{A10})$$

and then

$$\text{Tr}(\rho_0\rho(\phi)) = (\psi_0, \rho(\phi)\psi_0) = 1 - \|\phi\|^2 > 0. \quad (\text{A11})$$

Thus, points in $\mathcal{R}(\rho_0)$ are in one-to-one correspondence with, and are coordinatized by points inside the unit sphere in the subspace $H_{\psi_0}\mathcal{B}$ of \mathcal{H} . By adding a phase factor $e^{i\alpha}$, we get a local description for $\mathcal{B}(\psi_0)$ as

$$\psi \in \mathcal{B}(\psi_0) \Leftrightarrow \psi = \psi(\phi, \alpha) = e^{i\alpha}(\phi + (1 - \|\phi\|^2)^{1/2}\psi_0),$$

$$\phi \in H_{\psi_0}\mathcal{B}, \quad \|\phi\| < 1, \quad 0 \leq \alpha < 2\pi. \quad (\text{A12})$$

Let $\{\psi_0 = e_1, e_2, \dots\}$ be any orthonormal basis for \mathcal{H} . We expand $\phi \in H_{\psi_0}\mathcal{B}$ as

$$\phi = \frac{1}{\sqrt{2}} \sum_{r=2,3,\dots} (\beta_r - i\gamma_r) e_r, \quad (\text{A13})$$

so that the condition $\|\phi\| < 1$ becomes

$$\sum_{r=2,3,\dots} (\beta_r^2 + \gamma_r^2) < 2. \quad (\text{A14})$$

Then $\{\beta_2, \beta_3, \dots, \gamma_2, \gamma_3, \dots\}$ subject to (A14) are real local coordinates, $2(n-1)$ in number over $\mathcal{R}(\rho_0) \subset \mathcal{R}$; while $\{\alpha, \beta_2, \beta_3, \dots, \gamma_2, \gamma_3, \dots\}$ are real local coordinates over $\mathcal{B}(\psi_0) \subset \mathcal{B}$. In these charts we have the explicit expressions

$$A = d\alpha + \frac{1}{2} \sum_{r=2,3,\dots} (\gamma_r d\beta_r - \beta_r d\gamma_r),$$

$$dA = \pi^* \omega = \sum_{r=2,3,\dots} d\gamma_r \wedge d\beta_r. \quad (\text{A15})$$

APPENDIX B: PROPERTIES OF GEODESICS IN RAY SPACE

We provide here a brief account of the definition, differential equations, and main properties of geodesics in ray space \mathcal{R} . Since they will be found to obey second-order ordinary differential equations, it is appropriate to work with curves of class III. Let then

$$C = \{\rho(s) \in \mathcal{R} | s_1 \leq s \leq s_2\} \subset \mathcal{R} \quad (\text{B1})$$

running from $\rho(s_1) = \rho_1$ to $\rho(s_2) = \rho_2$ be a curve of class III, and assume $\text{Tr}(\rho_1 \rho_2) > 0$. Let

$$\mathcal{C} = \{\psi(s) \in \mathcal{B} | \rho(s) = \psi(s)\psi(s)^\dagger, s_1 \leq s \leq s_2\} \subset \mathcal{B} \quad (\text{B2})$$

be any (class III) lift of C . The length of C is the functional

$$\mathcal{L}[C] = \int_{s_1}^{s_2} ds \|u_\perp(s)\|,$$

$$u_\perp(s) = u(s) - (\psi(s), u(s))\psi(s) \in H_{\psi(s)}\mathcal{B};$$

$$u(s) = \frac{d}{ds} \psi(s) \in T_{\psi(s)}\mathcal{B}. \quad (\text{B3})$$

This is both reparametrization and local phase transformation invariant. We make infinitesimal changes $\delta\psi(s)$ in $\psi(s)$ along \mathcal{C} , inducing changes $\delta\rho(s)$ along C , vanishing at the end points s_1 and s_2 . Requiring $\delta\mathcal{L}[C] = 0$ we arrive at a differential equation which must be obeyed if $\mathcal{L}[C]$ is to be stationary and C is to be a geodesic:

$$\delta\mathcal{L}[C] = 0 \Leftrightarrow \left(\frac{d}{ds} - (\psi(s), u(s)) \right) \frac{u_\perp(s)}{\|u_\perp(s)\|} = f(s)\psi(s),$$

$$f(s) \text{ arbitrary real}. \quad (\text{B4})$$

This equation is naturally covariant with respect to both reparametrizations and local phase transformations. These properties can now be exploited to successively specialise the choice of the lift \mathcal{C} and its parametrization, and thereby simplify the differential equation (B4) [22]. By a suitable local phase transformation we can assume that \mathcal{C} is a horizontal lift of C , so we can replace Eq. (B4) by the simpler system

$$\frac{d}{ds} \frac{u(s)}{\|u(s)\|} = f(s)\psi(s),$$

$$(\psi(s), u(s)) = 0,$$

$$f(s) \text{ real}. \quad (\text{B5})$$

Next we can use the reparametrization freedom to switch from the originally given parameter to an affine parameter. This makes $\|u(s)\|$ constant along C , and for the affinely parametrized horizontal lift \mathcal{C} we have in place of Eq. (B5):

$$\frac{d^2 \psi(s)}{ds^2} = f(s)\psi(s), f(s) \text{ real},$$

$$\|\psi(s)\| = 1, \left(\psi(s), \frac{d\psi(s)}{ds} \right) = 0,$$

$$\left\| \frac{d\psi(s)}{ds} \right\| = \text{const}. \quad (\text{B6})$$

A brief analysis shows that this problem is fully equivalent to a second-order ordinary differential equation with suitable initial conditions at $s = s_1$,

$$\frac{d^2 \psi(s)}{ds^2} = - \left\| \frac{d\psi(s)}{ds} \right\|^2 \psi(s),$$

$$\|\psi(s_1)\| = 1, \left(\psi(s_1), \left(\frac{d\psi(s)}{ds} \right)_{s_1} \right) = 0. \quad (\text{B7})$$

Finally, we make a scale change and shift of origin in the parameter to make $s_1 = 0$ and $\|d\psi(s)/ds\| = 1$. This completely exhausts the freedom of reparametrizations, and at this point we may denote the parameter by a special symbol ℓ . For the so-defined and parametrized lift \mathcal{C} of a geodesic C in \mathcal{R} , we have the solution to Eq. (B7) in the form

$$\psi(\ell) = \psi(0)\cos \ell + \dot{\psi}(0)\sin \ell,$$

$$\|\psi(0)\| = \|\dot{\psi}(0)\| = 1, \quad (\psi(0), \dot{\psi}(0)) = 0. \quad (\text{B8})$$

This solution is determined by one orthonormal pair of vectors, and every geodesic in \mathcal{R} has such a lift in \mathcal{B} . The value

of the length functional $\mathcal{L}[C]$ in Eq. (B3), from the starting point $\ell=0$ to a general point ℓ , is the parameter ℓ itself:

$$\mathcal{L}[\text{geodesic } C \text{ from } \ell=0 \text{ to } \ell] = \int_0^\ell d\ell = \ell. \quad (\text{B9})$$

Going back to the original curve C in Eq. (B1), we identify $\psi(0) = \psi_1 \in \pi^{-1}(\rho_1)$, while $\psi_2 \in \pi^{-1}(\rho_2)$ has to be suitably chosen. A careful analysis of the explicit solution (B8) shows us that the following results hold.

(i) For given $\rho_1, \rho_2 \in \mathcal{R}$ obeying $\text{Tr}(\rho_1 \rho_2) > 0$, there is a unique geodesic connecting them.

(ii) If we make any choice of $\psi_1 \in \pi^{-1}(\rho_1)$, we can determine uniquely $\psi_2 \in \pi^{-1}(\rho_2)$ by the condition

$$(\psi_1, \psi_2) = \text{real positive}. \quad (\text{B10})$$

(iii) A horizontal lift \mathcal{C} of this geodesic is given by Eq. (B8) by choosing

$$\begin{aligned} \psi(0) &= \psi_1, \\ \dot{\psi}(0) &= \frac{\psi_2 - (\psi_1, \psi_2)\psi_1}{(1 - (\psi_1, \psi_2)^2)^{1/2}}. \end{aligned} \quad (\text{B11})$$

(iv) The end points of \mathcal{C} correspond to the parameter values $\ell=0$ for ψ_1 , $\ell = \cos^{-1}(\psi_1, \psi_2) \in [0, \pi/2)$ for ψ_2 .

(v) The length of this geodesic is $\cos^{-1}(\psi_1, \psi_2)$ and is strictly less than $\pi/2$.

(vi) For any two points $\ell, \ell' \in [0, \pi/2)$ we have

$$(\psi(\ell), \psi(\ell')) = \cos(\ell - \ell') = \text{real positive}, \quad (\text{B12})$$

and so for any three points $\ell, \ell', \ell'' \in [0, \pi/2)$, we have

$$\begin{aligned} \Delta_3(\psi(\ell), \psi(\ell'), \psi(\ell'')) &= \text{Tr}(\rho(\ell)\rho(\ell')\rho(\ell'')) \\ &= \text{real positive}. \end{aligned} \quad (\text{B13})$$

(vii) If we take the limiting case $\text{Tr}(\rho_1 \rho_2) = 0$, corresponding to $(\psi_1, \psi_2) = 0$, the geodesic distance from ρ_1 to ρ_2 becomes exactly $\pi/2$. However there is now no unique geodesic connecting ρ_1 to ρ_2 since, for any $\alpha \in [0, 2\pi)$, the curve

$$\psi(\ell) = \psi_1 \cos \ell + e^{i\alpha} \psi_2 \sin \ell \quad (\text{B14})$$

is a solution of the variational problem $\delta\mathcal{L}[C] = 0$, i.e., of Eq. (B7), and its projection in \mathcal{R} runs from ρ_1 to ρ_2 as ℓ varies over $[0, \pi/2]$.

These results are very similar to what are known about geodesics or great circle arcs on S^2 , which is the ray space \mathcal{R} when dimension $\mathcal{H} = 2$. Their lengths, in usual units, never exceed π . For nonantipodal points, there is a unique geodesic with length strictly less than π . For antipodal points we have a “ 2π worth” of geodesics, each of length π .

In a qualitative sense it is easy to see that in a ray space \mathcal{R} of any dimension, two points ρ_1 and ρ_2 can never get very far from one another, since

$$\text{Tr}(\rho_1 - \rho_2)^2 = 2[1 - \text{Tr}(\rho_1 \rho_2)] \leq 2. \quad (\text{B15})$$

The properties of geodesics described above are consistent with this fact.

-
- [1] M.V. Berry, Proc. R. Soc. London, Ser. A **392**, 45 (1984); A very useful reprint collection covering the period upto about 1989 is *Geometric Phases in Physics*, edited by J. Shapere and F. Wilczek, (World Scientific, Singapore, 1989).
- [2] B. Simon, Phys. Rev. Lett. **51**, 2167 (1983).
- [3] Y. Aharonov and J. Anandan, Phys. Rev. Lett. **58**, 1593 (1987); J. Samuel and R. Bhandari, *ibid.* **60**, 2339 (1988).
- [4] N. Mukunda and R. Simon, Ann. Phys. (N.Y.) **228**, 205 (1993); **228**, 269 (1993).
- [5] V. Bargmann, J. Math. Phys. **5**, 862 (1964).
- [6] S. Pancharatnam, Proc.-Indian Acad. Sci., Sect. A **44A**, 247 (1956).
- [7] S. Ramaseshan and R. Nityananda, Curr. Sci. **55**, 1225 (1986); N. Mukunda and R. Simon, Ref. [4] above.
- [8] D. Page, Phys. Rev. A **36**, 3479 (1987); S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1969), Vol. II, Chap. 9.
- [9] A concise account may be found, for example, V.I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer-Verlag, Berlin, 1978), Appendix III.
- [10] J. Samuel and R. Bhandari, Ref. [3] above.
- [11] N. Mukunda and R. Simon in Ref. [4] above; R. Simon and N. Mukunda, Phys. Rev. Lett. **70**, 880 (1993). For the use of BI's in the theory of off-diagonal geometric phases introduced by F. Pistolesi and N. Manini, *ibid.* **85**, 1585 (2000); **85**, 3067 (2000), see N. Mukunda, Arvind, S. Chaturvedi, and R. Simon, Phys. Rev. A **65**, 012102 (2001).
- [12] E.M. Rabei, Arvind, N. Mukunda, and R. Simon, Phys. Rev. A **60**, 3397 (1999).
- [13] These are described also in, for example, Refs. [2–4] above.
- [14] For discussions from the viewpoint of physical applications see, for instance, A.P. Balachandran, G. Marmo, B.S. Skagerstam, and A. Stern, *Applications to Particle Dynamics* (Springer-Verlag, Berlin, 1983); *Classical Topology and Quantum States* (World Scientific, Singapore, 1991).
- [15] E. Ercolessi, G. Marmo, G. Morandi, and N. Mukunda, Int. J. Mod. Phys. A **16**, 5007 (2001).
- [16] For a recent work on GP's in real Hilbert spaces see, for instance, J. Samuel and A. Dhar, Phys. Rev. Lett. **87**, 260401 (2001).
- [17] This is to be compared with J. Samuel and R. Bhandari in Ref. [3], where ρ_2 is connected back to ρ_1 using the unique geodesic running between them.
- [18] This may be regarded as the essential content of the discovery of S. Pancharatnam in Ref. [6] above.
- [19] For discussions of symplectic manifolds and their submanifolds see, for instance, R. Abraham and J.E. Marsden, *Foundations of Mechanics* (Benjamin-Cummings, Reading, MA, 1978); P. Liberman and C.M. Marle, *Symplectic Geometry and Analytical Mechanics* (Reidel, Dordrecht, 1987); G. Marmo,

E.J. Saletan, A. Simoni, and B. Vitale, *Dynamical Systems—A Differential Geometric Approach to Symmetry and Reduction* (Wiley, New York, 1985).

[20] R. Simon, E.C.G. Sudarshan, and N. Mukunda, *Phys. Rev. A* **37**, 3028 (1988); Arvind, B. Dutta, N. Mukunda, and R. Simon, *Pramana, J. Phys.* **45**, 471 (1995).

[21] For an exposition of Kaehler structures see, for instance, A. Weil, *Introduction a l'Etude des Varietes Kaehleriennes* (Hermann, Paris, 1971); S.S. Chern, *Complex Manifolds Without Potential Theory* (Springer-Verlag, New York, 1979).

[22] For a detailed discussion see, for instance, N. Mukunda and R. Simon, Ref. [4] above.