Miscelánea Matemática **33** (2001) 21–29 SMM

El espacio físico es geométrico

Gerardo Hernández
Sección de Metodología y Teoría de la Ciencia
Cinvestav, IPN
ghernan@mail.cinvestav.mx

Introducción.

Hace algunos meses, en un seminario de biología teórica se discutía sobre los conceptos de espacio y tiempo en biología. Estas nociones aparecen de forma distinta en los diferentes contextos de la investigación biológica; sin embargo, al escuchar a los colegas físicos hablar sobre estos temas aparece una constante, ausente por completo en todas las áreas biológicas (y en las económicas, según vimos): toda noción de espacio físico (es decir, del espacio al que aluden los físicos) es de naturaleza geométrica. El propósito de este artículo es justificar esta afirmación. Cabe señalar, sin embargo, que este hecho resulta ser tan natural para los físicos, que no es fácil encontrar entre ellos razón alguna para justificarlo. Pues bien, si no hay razones inherentes a la teoría física para geometrizar su espacio, quizá valdría la pena buscar en las raíces históricas de esta disciplina su empeño por geometrizar el espacio para entender lo que ocurre en él, y esta búsqueda lleva ineludiblemente a las ideas de Demócrito y de Aristóteles, de Newton y Leibniz, de Clifford y Einstein. Claramente la discusión propiamente geométrica en física sólo surge en el siglo XIX, cuando aparecen las geometrías no euclidianas, pues antes de entonces la única posibilidad era la geometría euclidiana y, si la posibilidad es única, la duda no existe. Pero tampoco la historia parece ofrecer una explicación convincente. Es cierto que las geometrías no euclidianas aparecen como descripciones alternativas de nuestro espacio; también es cierto que en algún momento se buscó alguna prueba empírica para establecer a qué tipo de geometría respondía nuestro familiar espacio de tres dimensiones, pero la discusión se ha llevado a planos de teorización física donde los espacios involucrados poco tienen de familiar o están al alcance siquiera de nuestras nociones cotidianas, y esos espacios son también geométricos.

En este trabajo ofrecemos una respuesta al porqué de la ubicuidad de nociones geométricas en los espacios que son del interés de los físicos. La respuesta es muy simple: las ecuaciones más importantes de la física son de segundo orden, y la segunda derivada es de naturaleza geométrica. La primera observación es aceptada por todos los físicos que conocemos, la segunda es el contenido de este trabajo. Los conceptos matemáticos involucrados son los de transporte paralelo, conexión y derivada covariante, y están ampliamente cubiertos en todo texto de geometría diferencial. Algunos textos, sin embargo, adoptan un enfoque más bien axiomático, lo que ofrece una pulcritud realmente ejemplar, pero también oculta o suprime la motivación geométrica. En esta nota se proporcionan argumentos heurísticos a favor de nuestra afirmación sobre la naturaleza geométrica de la segunda derivada. Los detalles técnicos, muy importantes, pueden consultarse en cualquier texto de geometría diferencial; creemos que los textos [He, KoNo, SiTh] y [Wo] son muy buenos expositores con distintos enfoques. Desde el punto de vista intuitivo preferimos [Cha].

En un principio fue Newton.

Vale a pena aclarar que el atribuir a la segunda derivada la naturaleza geométrica de la física envía nuestra historia a la obra de Newton (a fines del silo XVII) y no a las geometrías no euclidianas (del siglo XIX.) Pero no recurrimos a Newton para encontrar la fuente de la discusión geométrica; todo lo contrario, pues con él la naturaleza geométrica de la segunda derivada quedó oculta. Aunque la primera formulación matemática de la teoría física se encuentra en Newton (no nos referimos al uso de herramientas matemáticas, porque eso ya se había hecho antes, sino de una formulación organizada y sistemática), el problema es que Newton no es fácil de leer, y para simplificar las cosas es mejor seguir la lectura inteligente y clara que hace Feynman de Newton. Feynman [GoGo] explica la forma en que Newton deriva el movimiento elíptico de los planetas alrededor del sol usando como bases la primera y segunda ley, y la presencia de una fuerza central.

Luego de considerar un planeta moviéndose en dirección arbitraria y el efecto de la fuerza central, Feynman llega a un diagrama poligonal para representar la órbita del planeta y continúa: “Aquí tenemos el núcleo central el que se deducirá todo: tenemos cambios iguales de velocidad cuando la órbita se mueve mediante ángulos iguales. Trazo ahora en este diagrama una pequeña línea que representa las velocidades. A diferencia del otro diagrama, estas líneas no son la línea completa de

A a B porque en aquel diagrama eran proporcionales a las velocidades ya que los tiempos eran iguales y la longitud partida por tiempos iguales representaba las velocidades [...] Así pues, estas líneas representan la sucesión de velocidades. Es dífícil apreciar qué ha cambiado en este diagrama. Dibujaré por lo tanto, otro diagrama (que llamaré diagrama de velocidades) a escala ampliada, sólo para entendernos [Figura 1.a]. Las líneas representadas son, se supone, las mismas. Esta representa el movimiento por segundo (o por un intervalo de tiempo dado) de una partícula en A. Esta representa el movimiento que tendría una partícula desde el comienzo en un intervalo de tiempo dado. A todas les doy un origen común [Figura 1.b]. A todas les estoy dando un origen común, para poder comparar las velocidades. Así pues, tengo una serie de velocidades para la sucesión de estos puntos". [GoGo]

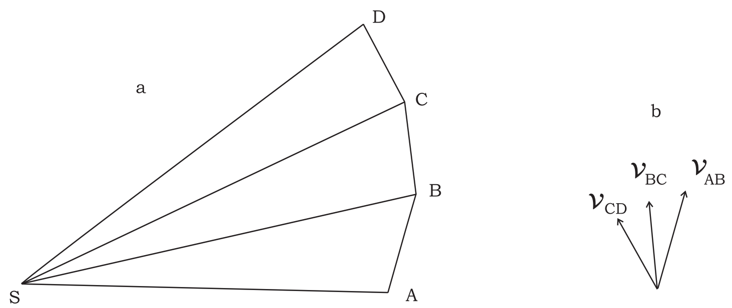

Figura 1.

Justamente el truco está en el paso de la figura 1.a a la figura 1.b.

Geometría del traslado.

Veamos qué sucede en esa transición. Consideremos un subconjunto abierto U de $\mathbb{R}^2$, con coordenadas (x, y). Podemos tomar el abierto más simple que podamos imaginar, conexo y simplemente conexo, pues la topología no va a jugar ningún papel relevante; después de todo, vamos a hablar de la derivada, que es una noción local (infinitesimal.) Ahora tomemos una curva $\sigma : I \to U$, donde I es un intervalo que contiene a 0. La derivada al tiempo 0 es entonces el vector $\dot{\sigma}(0)$ con origen en $\sigma(0) = p$. Claro que existen muchas curvas con $0 \mapsto p$ y cuya derivada en p coincide con la derivada de σ, así que podemos definir un vector tangente a U en p como este conjunto de curvas. Es decir, definimos un vector tangente en p como la clase de equivalencia de curvas cuya derivada es igual en p, y lo denotamos por $[\sigma]_p$ o simplemente por

$\dot{\sigma}(0)$. El conjunto de todas estas clases de equivalencia se denota por T_pU y tiene una estructura de espacio vectorial, donde cada vector está representado por la derivada de una curva que pasa por p, de suerte que $T_pU = \{p\} \times \mathbb{R}^2$. Como esto lo podemos hacer para cada punto de U, definimos el haz tangente a U, esto es, el conjunto de los espacios tangentes a cada punto de U como $TU = \bigsqcup T_pU$, que es simplemente $U \times \mathbb{R}^2$.

Obsérvese también que podemos definir este vector tangente de otra manera: supongamos que tenemos definida una función real $f : U \to \mathbb{R}$, entonces el vector tangente $[c]_p$ actúa sobre esta función como

$$[c]_p(f) = \frac{d}{dt}(f \circ c)\Big|_{t=0},$$

que corresponde a la derivada direccional de f en la dirección $[c]_p$.

(Esto puede extenderse a variedades más generales que $\mathbb{R}^2$, tomando V como un abierto de una variedad diferenciable, y una curva $c : I \to V$. Luego, la estructura diferencial nos dice que existe un homeomorfismo $\phi : V \to U \subset \mathbb{R}^2$. Claramente c define una curva en U con la composición $\sigma = \phi \circ c : I \to U$. De hecho, ϕ asocia a cada $c(t)$ en V un punto $(x(t), y(t))$ de U, de manera que aún cuando no tiene sentido definir la derivada de c como una diferencia (¿qué sentido tiene la diferencia $c(h) - c(0)$ en un espacio topológico?), podemos calcular la diferencia $\sigma(h) - \sigma(0)$ para determinar la derivada de σ como de costumbre y definir el vector tangente a la curva c en el punto $c(0)$ como $\dot{\sigma}(0)$. Claramente la expresión depende de la carta coordenada, pero como los cambios de coordenadas son suaves, podemos ir de una expresión a otra mediante una transformación lineal, así que el vector tangente está bien definido y sólo se necesita de la estructura diferenciable. Para evitar la complicación dada por las cartas coordenadas hemos optado por ir directamente al abierto U.)

Veamos ahora qué sucede para la segunda derivada. En este caso debemos definir la diferencia $\dot{\sigma}(h) - \dot{\sigma}(0)$. Ambos son vectores, así que la diferencia podría tener sentido; pero $\dot{\sigma}(h)$ es un vector tangente a la curva en $q = \sigma(h)$, mientras que $\dot{\sigma}(0)$ es tangente en el punto $p = \sigma(0)$. Para que la diferencia tenga sentido, debemos buscar la manera en que el vector $\dot{\sigma}(h)$ sea "transportado" (¡para cada h!) al punto $\sigma(0)$, y claramente existen muchas formas de transportar este vector. Por ejemplo, en cada punto de U asignamos una base de $\mathbb{R}^2$, $p \mapsto (b_1(p), b_2(p))$. Podríamos decir que un vector en p es el equivalente a otro en q si sus coordenadas con respecto a las bases $(b_1(p), b_2(p))$ y $(b_1(q), b_2(q))$ son las mismas. Otra manera: El vector tangente Y es constante en U si

$Y(p) = a_1b_1(p) + a_2b_2(p)$, para todo punto p en U, donde las a's son constantes.

Claramente, en el caso euclidiano, para cada p en U asignamos los vectores de la base estándar $(e_1(p), e_2(p)) = (e_1, e_2)$, que resultan de considerar las curvas $c_1 : t \mapsto (t+x, y)$ y $c_2 : t \mapsto (x, t+y)$, y sus vectores velocidad $e_1 = \dot{c}_1(0)$ y $e_2 = \dot{c}_1(0)$. (Observe que de acuerdo a nuestra segunda definición, estos vectores tangentes se ven como $e_1 = \partial/\partial x$, y $e_2 = \partial/\partial y$.) De manera que llevar todos los vectores a un mismo origen en el argumento de Feynman consiste precisamente en transportar a los vectores velocidad con sus mismas coordenadas en la base estándar. Con este procedimiento, la aceleración es simplemente la convencional "derivada de la derivada". Un ejercicio interesante sería el ver todas las posibles configuraciones que toma una curva de velocidades dada al tomar distintas asignaciones $p \mapsto (b_1(p), b_2(p))$. Como podemos reemplazar la elección de una base $(b_1(p), b_2(p))$ por la aplicación $B(p)$ que satisface $B(p)(e_i) = b_i(p)$, $i = 1, 2$, la asignación de una base a cada punto equivale a una aplicación $B : U \to GL(2)$. Podemos observar que si dados dos aplicaciones de este tipo B y $\tilde{B}$ existe un $g \in GL(2)$ tal que $B = g \cdot \tilde{B}$, entonces las curvas de velocidad sólo varían con cada cambio de coordenadas.

Pero no es la idea el otorgar un peso distinto a cada punto del espacio. Más bien se trata de buscar una idea de transporte que dependa de la curva, de la trayectoria de la partícula, más que de los puntos por los que pase la particula. No se trata de un problema de bases, sino de identificación de distintos espacios tangentes; de definir un isomorfismo entre ellos.

Primero notemos que la derivada de la curva $\sigma(t) = (x(t), y(t))$, se ve como una curva en el haz tangente $U \times \mathbb{R}^2$, $\dot{\sigma}(t) = (x(t), y(t), x'(t), y'(t))$, donde x' y y' son las derivadas de las funciones reales x y y, y son las coordenadas del vector tangente con respecto a la base estándar. Como la derivada de una curva en U es una curva en $U \times \mathbb{R}^2$, la segunda derivada es una curva en $U \times \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2$. Esta forma de estudiar la segunda derivada tiene que ver con unos objetos llamados "sprays" (ver por ejemplo [La].) Nosotros abordaremos este problema con un enfoque un poco más directo.

Tomemos el intervalo $[0, t_0] \subset I$. Supongamos que en cada vector tangente α_0 en el punto $\sigma(0)$, $(\sigma(0), \alpha_0) \in U \times \mathbb{R}^2$, podemos definir un único campo vectorial $\alpha(t)$ tal que $\alpha(0) = \alpha_0$, de manera que "levantamos" la curva original σ en U, a una curva $\tilde{\sigma}$ en $U \times \mathbb{R}^2$, de forma tal que cuando proyectamos vía $\pi : U \times \mathbb{R}^2 \to U$ obtenemos nuevamente

σ, $\pi(\tilde{\sigma}(t)) = \sigma(t)$, lo que significa que tiene la forma $\tilde{\sigma}(t) = (\sigma(t), \alpha(t))$ y que pasa por $(\sigma(0), \alpha_0)$. Con este levantamiento de la curva podemos definir el trasladado paralelo del vector tangente α_0 en el punto $\sigma(t_0)$ como el vector $\alpha(t_0)$. Claramente, como hemos supuesto que la curva $\alpha(t)$ es única, define una aplicación lineal biyectiva $T_{\sigma(0)}U \mapsto T_{\sigma(t)}U$, para cada $t \in [0, t_0]$. Denotemos a este mapeo con la misma letra que representa a la curva pues depende de ella, $\sigma_0^{t_0} : T_{\sigma(0)}U \to T_{\sigma(t_0)}U$. Con esta transformación ya podemos definir la derivada de cualquier campo vectorial $Y(t)$ a lo largo de σ como

$$\nabla_{\dot{\sigma}_t} Y = \lim_{h\to 0} \frac{1}{h}\left[\sigma_t^{t+h}(Y(t+h)) - Y(t)\right] .$$

(Aquí se usa $\nabla_{\dot{\sigma}}$ para hacer una referencia explícita a la curva considerada, pero no parece que en la definición de trasladado paralelo se usara la derivada de σ, sólo se usó la curva misma y su "levantamiento" $\tilde{\sigma}$. La razón es simple pero laboriosa. En realidad no usamos cualquier levantamiento, sino uno especial llamado levantamiento horizontal, que depende de la conexión usada. La idea de usar una conexión es hacer uso del teorema de existencia y unicidad de la solución de ecuaciones diferenciales ordinarias. En realidad no levantamos directamente la curva, sino un campo vectorial, pero este nuevo campo levantado vive en el horrible espacio $TTU = U \times \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2$. Ahora bien, este espacio es demasiado grande para los levantamientos de campos vectoriales en U, que viven en $U \times \mathbb{R}^2$. Por eso se escogen subespacios de TTU que se copian de manera isomorfa en el tangente TU, pues para cada campo vectorial $X : U \to TU$ generamos uno y sólo uno "levantado" $\tilde{X} : TU \to TTU$. La elección de estos subespacios es precisamente la elección de una conexión, quien a su vez determina, por el mencionado teorema de ecuaciones diferenciales, un único levantamiento de la curva. El término "conexión" viene del hecho de que podemos "conectar" los distintos espacios tangentes, como lo hicimos con nuestros trasladados paralelos. Si levantamos (horizontalmente) el campo vectorial que restringido a la curva coincide con el vector velocidad de la curva e integramos, obtenemos el levantamiento de la curva, de ahí que la derivada ∇ dependa de $\dot{\sigma}$.)

Veamos rápidamente un ejemplo para poder abordar el problema original del espacio. Supongamos que para una curva σ el propio vector velocidad es paralelo:

$$\nabla_{\dot{\sigma}}\dot{\sigma} = 0 .$$

Estas curvas reciben el nombre de geodésicas y juegan el mismo papel que juegan las líneas rectas en el plano ordinario. Note que $\nabla_{\dot{\sigma}}\dot{\sigma}$

es precisamente la derivada segunda. En este ejemplo veremos que no siempre el hecho de que esta segunda derivada sea cero implica que la curva es una recta. Con la conexión euclidiana es cierto que las geodésicas son líneas rectas, pero es un caso muy especial.

Claramente, si el vector $Y(t)$ es paralelo, $g{\cdot}Y(t)$ también es paralelo, donde $g \in GL(2)$. Ahora bien, todos sabemos que el plano hiperbólico de Poincaré tiene como geodésicas a los semicírculos con centro en el eje $y = 0$, y a las líneas verticales. La siguiente figura ilustra la forma en que el traslado paralelo depende de la curva.

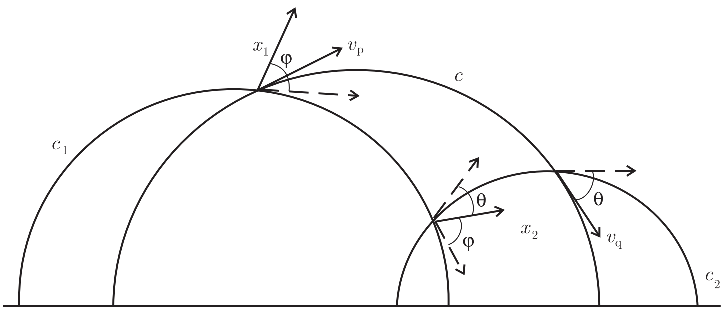

Figura 2.

El vector velocidad v_q de la curva c se traslada a lo largo de la misma hasta v_p, pues c es geodésica. Si lo trasladamos por la curva c_2 se llega hasta el vector x_2 que conserva con respecto al vector velocidad de c_2 la misma relación, ángulo θ, que tenía el vector v_q. El procedimiento se repite con respecto a la curva c_1, ángulo ϕ, y se llega al vector x_1. Claramente notamos que v_p no coincide con x_1. Hemos omitido consideraciones métricas, que en esta conexión sí se aplicarían, porque hemos evitado escrupulosamente toda referencia a ellas para que no se confunda la geometría (que es de lo que hemos estado hablando) con ideas estrictamente métricas (que son solamente una parte de la geometría); sin embargo, es posible darse cuenta de que podemos usar distintas geodésicas para llevar el vector v_q hasta el punto p con cualquier dirección posible. Éste es un buen ejemplo de que un abierto conexo y simplemente conexo (el semiplano $y > 0$!) puede tener una geometría realmente interesante. Una exposición de lo interesante que puede ser el semiplano de Poincaré aparece en [Riv], donde también se ofrece una motivación física a este espacio.

Leyes físicas y geometría.

Los operadores de segundo orden como el laplaciano y el hessiano son también nociones geométricas. Requieren del concepto de conexión y no son definibles con la estuctura diferenciable solamente, aunque se requiera más espacio (papel) para abordarlos. De manera que la formulación de los principios (aquí podemos mencionar la convexidad de la función de entropía, por ejemplo, que involucra al hessiano) y leyes de la física más importantes, que se expresan con ecuaciones de segundo orden llevan implícita una noción geométrica. De aquí podemos sugerir una interpretación de la primera y segunda leyes de Newton.

Como sabemos, la primera ley establece que en ausencia de fuerzas un objeto se mueve en línea recta y con rapidez constante. La segunda ley dice que las fuerzas aceleran los objetos, es decir, hacen variar su velocidad. Algunos han sugerido que la segunda ley contiene a la primera pues si la fuerza no existe, la aceleración es cero y por lo tanto la velocidad es constante. Existen diversas opiniones sobre este punto entre los físicos; algunos argumentan que de ningún modo la primera ley es un caso especial de la segunda; los argumentos son de naturaleza física y sin duda tiene mucho sentido lo que afirman. Creo que desde el punto de vista matemático se trata de algo más simple. La segunda ley es un postulado físico y dice que la aceleración (segunda derivada del desplazamiento) es proporcional a la fuerza; la primera ley dice en qué geometría debe entenderse esa aceleración. Newton dijo que en el sentido euclidiano y nada más se oculta en su primera ley.

El autor agradece a Juan José Rivaud el haber ejercido una presión considerable para que se escribiese este trabajo. Las críticas, comentarios y sugerencias de los árbitros ayudaron mucho a mejorar el manuscrito oiginal.

Referencias

[Cha] L. Charlap, *Bieberbach groups and flat maifolds*, New York: Springer, 1986.

[GoGo] D. Goodstein and J. Goodstein, *La conferencia perdida de Feynman. El movimiento de los planetas alrededor del Sol*, Barcelona: Tusquets, 1999.

[He] S. Helgason, *Differential geometry, Lie groups and symmetric spaces*, Boston: Academic Press, 1978.

[La] S. Lang, *Differential and Riemannian Manifolds*, New York: Springer, 1995.

[KoNo] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, New York: John Wiley and Sons, Vol. I, 1991.

[Riv] J. J. Rivaud, *El mundo de hiperbólico*, Notas.

[SiTh] I. M. Singer and J. A. Thorpe, *Lecture notes on elementary topology and geometry*, Illinois: Scott, Foresman and Company, 1967.

[Wo] J. A. Wolf, *Spaces of constant curvature*, Delaware: Publish or Perish, 1984.