Some remarks about metric spaces, 1

Stephen William Semmes*

Of course various kinds of metric spaces arise in various contexts and are viewed in various ways. In this brief survey we hope to give some modest indications of this. In particular, we shall try to describe some basic examples which can be of interest.

For the record, by a *metric space* we mean a nonempty set M together with a distance function $d(x, y)$, which is a real-valued function on $M \times M$ such that $d(x, y) \geq 0$ for all $x, y \in M$, $d(x, y) = 0$ if and only if $x = y$, $d(x, y) = d(y, x)$ for all $x, y \in M$, and

$$d(x, z) \leq d(x, y) + d(y, z) \tag{1}$$

for all $x, y, z \in M$. This last property is called the *triangle inequality*, and sometimes it is convenient to allow the weaker version

$$d(x, z) \leq C\,(d(x, y) + d(y, z)) \tag{2}$$

for a nonnegative real number C and all $x, y, z \in M$, which which case $(M, d(x, y))$ is called a quasi-metric space. Another variant is that we may wish to allow $d(x, y) = 0$ to hold sometimes without having $x = y$, in which case we have a semi-metric space, or a semi-quasi-metric space, as appropriate.

A sequence of points $\{x_j\}_{j=1}^{\infty}$ in a metric space M with metric $d(x, y)$ is said to converge to a point x in M if for every $\epsilon > 0$ there is a positive integer L such that

$$d(x_j, x) < \epsilon \quad \text{for all } j \geq L, \tag{3}$$

in which case we write

$$\lim_{j \to \infty} x_j = x. \tag{4}$$

---

*This survey has been prepared in connection with the workshop on discrete metric spaces and their applications at Princeton, August, 2003.

A sequence $\{x_j\}_{j=1}^\infty$ of points in M is said to be a *Cauchy sequence* if for every $\epsilon > 0$ there is a positive integer L such that

$$d(x_j, x_k) < \epsilon \quad \text{for all } j, k \ge L. \tag{5}$$

It is easy to see that every convergent sequence is a Cauchy sequence, and conversely a metric space in which every Cauchy sequence converges is said to be *complete*.

A very basic example of a metric space is the real line $\mathbf{R}$ with its standard metric. Recall that if x is a real number, then the *absolute value* of x is denoted $|x|$ and defined to be equal to x when $x \ge 0$ and to $-x$ when $x < 0$. One can check that

$$|x + y| \le |x| + |y| \tag{6}$$

and

$$|x\,y| = |x|\,|y| \tag{7}$$

when x, y are real numbers, and that the standard distance function $|x - y|$ on $\mathbf{R}$ is indeed a metric.

Let n be a positive integer, and let $\mathbf{R}^n$ denote the real vector space of n-tuples of real numbers. Thus elements x of $\mathbf{R}^n$ are of the form $(x_1, \ldots, x_n)$, where the n coordinates x_j, $1 \le j \le n$, are real numbers. If $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$ are two elements of $\mathbf{R}^n$ and r is a real number, then the sum $x + y$ and scalar product $r\,x$ are defined coordinatewise in the usual manner, by

$$x + y = (x_1 + y_1, \ldots, x_n + y_n) \tag{8}$$

and

$$r\,x = (r\,x_1, \ldots, r\,x_n). \tag{9}$$

If x is an element of $\mathbf{R}^n$, then the standard Euclidean norm of x is denoted $|x|$ and defined by

$$|x| = \Big(\sum_{j=1}^n x_j^2\Big)^{1/2}. \tag{10}$$

One can show that

$$|x + y| \le |x| + |y| \tag{11}$$

holds for all $x, y \in \mathbf{R}^n$, and we shall come back to this in a moment, and clearly we also have that

$$|r\,x| = |r|\,|x| \tag{12}$$

for all $r \in \mathbf{R}$ and $x \in \mathbf{R}^n$, which is to say that the norm of a scalar product of a real number and an element of $\mathbf{R}^n$ is equal to the product of the absolute value of the real number and the norm of the element of $\mathbf{R}^n$. Using these properties, one can check that the standard Euclidean distance $|x-y|$ on $\mathbf{R}^n$ is indeed a metric.

More generally, a *norm* on $\mathbf{R}^n$ is a real-valued function $N(x)$ such that $N(x) \geq 0$ for all $x \in \mathbf{R}^n$, $N(x) = 0$ if and only if $x = 0$,

$$N(r\,x) = |r|\,N(x) \tag{13}$$

for all $r \in \mathbf{R}$ and $x \in \mathbf{R}^n$, and

$$N(x+y) \leq N(x) + N(y) \tag{14}$$

for all $x, y \in \mathbf{R}^n$. If $N(x)$ is a norm on $\mathbf{R}^n$, then

$$d(x,y) = N(x-y) \tag{15}$$

defines a metric on $\mathbf{R}^n$. As for metrics, one can weaken the triangle inequality or relax the condition that $N(x) = 0$ implies $x = 0$ to get quasi-norms, semi-norms, and semi-quasi-norms.

Recall that a subset E of $\mathbf{R}^n$ is said to be *convex* if

$$t\,x + (1-t)\,y \in E \tag{16}$$

whenever x, y are elements of E and t is a real number such that $0 < t < 1$. A real-valued function $f(x)$ on $\mathbf{R}^n$ is said to be convex if and only if

$$f(t\,x + (1-t)\,y) \leq t\,f(x) + (1-t)\,f(y) \tag{17}$$

for all $x, y \in \mathbf{R}^n$ and $t \in \mathbf{R}^n$ with $0 < t < 1$. If $N(x)$ is a real-valued function on $\mathbf{R}^n$ which is assumed to satisfy the conditions of a norm except for the triangle inequality, then one can check that the triangle inequality, the convexity of the closed unit ball

$$\{x \in \mathbf{R}^n : N(x) \leq 1\}, \tag{18}$$

and the convexity of $N(x)$ as a function on $\mathbf{R}^n$, are all equivalent.

For example, if p is a real number such that $1 \leq p < \infty$, then define $|x|_p$ for $x \in \mathbf{R}^n$ by

$$|x|_p = \Big(\sum_{j=1}^{n} |x|^p\Big)^{1/p}, \tag{19}$$

which is the same as the standard norm $|x|$ when $p = 2$. For $p = \infty$ let us set

(20) $$|x|_\infty = \max\{|x_j| : 1 \le j \le n\}.$$

One can check that these define norms on $\mathbf{R}^n$, using the convexity of the function $|r|^p$ on $\mathbf{R}$ when $1 < p < \infty$ to check that the closed unit ball of $|x|_p$ is convex and hence that the triangle inequality holds when $1 < p < \infty$.

Let us now consider a class of metric spaces along the lines of Cantor sets. For this we assume that we are given a sequence $\{F_j\}_{j=1}^\infty$ of nonempty finite sets. We also assume that $\{\rho_j\}_{j=1}^\infty$ is a monotone decreasing sequence of positive real numbers which converges to 0.

For our space M we take the Cartesian product of the F_j's, so that en element x of M is a sequence $\{x_j\}_{j=1}^\infty$ such that $x_j \in F_j$ for all j. We define a distance function $d(x, y)$ on M by setting $d(x, y) = 0$ when $x = y$, and

(21) $$d(x, y) = \rho_j$$

when $x_j \ne y_j$ and $x_i = y_i$ for all $i < j$. One can check that this does indeed define a metric space, and in fact $d(x, y)$ is an *ultrametric*, which is to say that

(22) $$d(x, z) \le \max(d(x, y), d(y, z))$$

for all $x, y, z \in M$.

The classical Cantor set is the subset of the unit interval $[0, 1]$ in the real line obtained by first removing the open subinterval $(1/3, 2/3)$, then removing the the open middle thirds of the two closed intervals which remain, and so on. Alternatively, the classical Cantor set can be described as the set of real numbers t such that $0 \le t \le 1$ and t has ab expansion base 3 whose coefficients are are either 0 or 2. This set, equipped with the standard Euclidean metric, is very similar to the general situation just described with each F_j having two elements and with $\rho_j = 2^{-j}$ for all j, although the metrics are not quite the same.

In general, if $(M, d(x, y))$ is a metric space and E is a nonempty subset of M, then E can be considered as a metric space in its own right, using the restriction of the metric $d(x, y)$ from M to E. Sometimes there may be another metric on E which is similar to the one inherited from the larger space M, and which has other nice properties, as in the case of Cantor sets just described. Another basic instance of this occurs with arcs in Euclidean spaces which are "snowflakes", and which are similar to taking the unit interval $[0, 1]$

in the real line with the metric $|x-y|^a$ for some real number a, $0 < a < 1$, or other functions of the standard distance on $[0,1]$.

A nonempty subset E of a metric space $(M, d(x,y))$ is said to be *bounded* if the real numbers $d(x,y)$, $x, y \in E$, are bounded, in which case the *diameter* of E is denoted $\operatorname{diam} E$ and defined by

$$\operatorname{diam} E = \sup\{d(x,y) : x, y \in E\}. \tag{23}$$

A stronger condition is that E be *totally bounded*, which means that for each $\epsilon > 0$ there is a finite collection $A_1, \ldots, A_k$ of subsets of E such that

$$E \subseteq \bigcup_{j=1}^{k} A_j \tag{24}$$

and

$$d(x,y) < \epsilon \quad x, y \in A_j, \tag{25}$$

$j = 1, \ldots, k$. A basic feature of Euclidean spaces is that bounded subsets are totally bounded, and the generalized Cantor sets described before are totally bounded.

A metric space $(M, d(x,y))$ is *compact* if it is complete, so that every Cauchy sequence converges, and totally bounded. This is equivalent to the standard definitions in terms of open coverings or the existence of limit points. A closed and bounded subset of $\mathbf{R}^n$ is compact, and the generalized Cantor sets described earlier are compact.

Another way that metric spaces arise is to start with a connected smooth n-dimensional manifold M, which is basically a space which looks locally like n-dimensional Euclidean space. At each point p in M one has an n-dimensional tangent space $T_p(M)$, which looks like $\mathbf{R}^n$ as a vector space, and on which one can put a norm. If at each point p in M one can identify $T_p(M)$ with $\mathbf{R}^n$ with its standard norm, then the space is Riemannian, and with general norms the space is of Riemann–Finsler type.

In this type of situation, the length of a nice path in M can be defined by integrating the infinitesimal lengths determined by the norms on the tangent spaces. The distance between two points is defined to be the infimum of the lengths of the paths connecting the two points. It is easy to see that this does indeed define a metric, with the triangle inequality being a consequence of the way that the distance is defined.

A basic example of this is the n-dimensional sphere $\mathbf{S}^n$ in $\mathbf{R}^{n+1}$, defined by

$$\mathbf{S}^n = \{x \in \mathbf{R}^{n+1} : |x| = 1\}. \tag{26}$$

The tangent space of $\mathbf{S}^n$ can be identified with the n-dimensional linear subspace of $\mathbf{R}^{n+1}$ of vectors x which are orthogonal to p as a vector itself. This leads to a Euclidean norm on the tangent spaces, inherited from the one on $\mathbf{R}^{n+1}$.

One can also define distances through paths in other situations. In a nonempty connected graph, every pair of points is connected by a path, the length of a path can be defined as the number of edges traversed, and the distance between two points can be defined as the length of the shortest path between the two points. In many standard fractals, like the Sierpinski gasket and carpet, there are a lot of paths of finite length between arbitrary elements of the fractal, and the infimum of the lengths of these paths defines a metric on the fractal.

Let $(M, d(x, y))$ be a metric space. If A, B are nonempty subsets of M and t is a positive real number, then let us say that A, B are "t-close" if for every $x \in A$ there is a $y \in B$ such that $d(x, y) < t$, and if for every $y \in B$ there is an $x \in A$ such that $d(x, y) < t$. By definition, this relation is symmetric in A and B.

A subset E of M is said to be *closed* if for every sequence $\{z_j\}_{j=1}^{\infty}$ of points in E which converges to a point z in M, we have that $z \in E$. Let us write $\mathcal{S}(M)$ for the set of nonempty closed and bounded subsets of M. If A, B are two elements of $\mathcal{S}(M)$, then the *Hausdorff distance* between A and B is defined to be the infimum of the set of positive real numbers t such that A, B are t-close.

Because A and B are bounded subsets of M, there are positive real numbers t such that A, B are t-close. The restriction to closed sets ensures that if A, B are t-close for all $t > 0$, then $A = B$. If A_1, A_2, A_3 are nonempty subsets of M and t_1, t_2 are positive real numbers such that A_1, A_2 are t_1-close and A_2, A_3 are t_2-close, then one can check that A_1, A_3 are $(t_1 + t_2)$-close, and this implies the triangle inequality for the Hausdorff distance.

From this it follows easily that $\mathcal{S}(M)$, equipped with the Hausdorff distance $D(A, B)$, is indeed a metric space. A basic result states that if M is compact, then $\mathcal{S}(M)$ is compact too. This is not too difficult to show.

Now let us consider a situation with a lot of deep mathematical structure which has been much-studied, involving interplay between algebra, analysis, and geometry.

Fix a positive integer n, which we may as well take to be at least 2. Let $\mathcal{M}_n^+$ denote the set of $n \times n$ real symmetric matrices which are positive-definite and have determinant equal to 1. One can think of this as a smooth hypersurface in the vector space of $n \times n$ real symmetric matrices, and it is a smooth manifold in particular.

We can start with a Riemannian view of this space. For each $H \in \mathcal{M}_n^+$, we can identify the tangent space of $\mathcal{M}_n^+$ at H with the vector space of $n \times n$ real symmetric matrices A such that the trace of $H^{-1} A$ is equal to 0. Thus for each such A, we get a one-parameter family of perturbations of H by taking $H + t\,A$, where t is a real number with small absolute value, so that $H + t\,A$ is still positive definite, and to first order in t these perturbations also have determinant equal to 1.

Conversely, to first order in t, each smooth perturbation of H in $\mathcal{M}_n^+$ is of this form. Now, for each A of this type, we define its norm as an element of the tangent space to $\mathcal{M}_n^+$ to be

$$\Big(\operatorname{tr} H^{-1}\,A\,H^{-1}\,A\Big)^{1/2}. \tag{27}$$

Here $\operatorname{tr} B$ denotes the trace of a square matrix A, and we are using ordinary matrix multiplication in this expression.

When H is the identity matrix, this reduces to the square root of $\operatorname{tr} A^2$, which is a kind of Euclidean norm of A. For general H's, we adapt the norm to H. It is still basically a Euclidean norm, so that we are in the Riemannian case.

Let us consider the transformation on $\mathcal{M}_n^+$ defined by $H \mapsto H^{-1}$. If $H + t\,A$ is a basic first-order deformation of H, as before, then that is transformed to $(H + t\,A)^{-1}$, which is the same as

$$H^{-1} - t\,H^{-1}\,A\,H^{-1} \tag{28}$$

to first order in t. Thus A as a tangent vector at H corresponds to $-H^{-1}\,A\,H^{-1}$ as a tangent vector at H^{-1} under the mapping $H \mapsto H^{-1}$, and it is easy to see that the norm of A as a tangent vector at H is equal to the norm of $-H^{-1}\,A\,H^{-1}$ as a tangent vector at H^{-1}.

Now let T be an $n \times n$ matrix with determinant equal to 1, and let T^* denote its transpose, which is also an invertible $n \times n$ matrix. Associated to T is the mapping

$$H \mapsto T\,H\,T^*, \tag{29}$$

and if A corresponds to a tangent vector at H as before, then $T\,A\,T^*$ is the tangent vector at $T\,H\,T^*$ induced by our mapping. Again one can check that the norm of A as a tangent vector at H is the same as the norm of $T\,A\,T^*$ as a tangent vector at $T\,H\,T^*$.

Assume further that the entries of T are integers. This implies that the inverse of T also has integer entries, by Cramer's rule. The product of two such matrices has the same property, and indeed this defines a nice discrete group of matrices.

This discrete group acts on $\mathcal{M}_n^+$, and we can pass to the corresponding quotient space. That is, we now identify two elements H_1, H_2 in $\mathcal{M}_n^+$ if there is an integer matrix T as above so that

$$H_2 = T\,H_1\,T^*. \tag{30}$$

This relation between $H_1, H_2 \in \mathcal{M}_n^+$ is indeed an equivalence relation, so that we get a nice quotient.

Because of the discreteness of the group of integer matrices with determinant 1, the quotient space is still a smooth manifold, since it looks locally like $\mathcal{M}_n^+$. The norm on the tangent spaces still makes sense as well, because the transformations defining the equivalence relation preserves these norms, as we have seen. Thus this quotient of $\mathcal{M}_n^+$ is still a nice smooth connected Riemannian manifold.

In some situations like this the quotient space turns out to be compact. In this case the quotient space is not compact, but it does have finite volume. More precisely, the notion of volume at the level of the tangent spaces is determined by the norm, and preserved in the present circumstances by the transformations used in the equivalence relation, and the volume of the quotient can be obtained by integrating the infinitesimal volumes.

The group of $n \times n$ matrices with integer entries and determinant 1 is a very interesting special case of discrete groups more generally. Suppose now that Γ is a group and that A is a finite symmetric subset of Γ which generates Γ, so that $\alpha^{-1} \in A$ when $\alpha \in A$ and every element of Γ can be expressed as a finite product of elements of A, with the identity element automatically corresponding to an empty product. This leads to the associated *Cayley graph*, in which two elements γ_1, γ_2 of Γ are considered to be adjacent if γ_2 can be expressed as $\gamma_1\,\alpha$ for some $\alpha \in A$, and to a distance function on Γ which is invariant under left-multiplication in the group.

In various contexts in analysis and geometry, it is natural to consider metric spaces $(M, d(x, y))$ which satisfy a *doubling condition*, which means that there is a constant C so that for each $x \in M$ and positive real number r there is a finite subset F of M with at most C elements such that for each $y \in M$ with $d(y, x) \le r$ there is a $z \in F$ with $d(y, z) \le r/2$. This can automatically be iterated to say that for each $x \in M$, $r > 0$, and positive integer k there is a finite subset F_k of M with at most C^k elements such that for each $y \in M$ with $d(y, x) \le r$ there is a $z \in F_k$ with $d(y, z) \le 2^{-k} r$. In one sense this is a kind of polynomial bound on local complexity, as a function of the radius, although as usual this corresponds to exponential growth in other terms.

In harmonic analysis one often looks at a space or a function on it by looking at behavior at various locations and scales. For instance, if $(M, d(x, y))$ is a metric space, $f(x)$ is a real-valued function on M, x is a point in M, and r is a positive real number, one can consider the quantity

$$\mathrm{osc}_{x,r}(f) = \sup\{|f(y) - f(x)| : y \in M, d(y, x) \le\}. \tag{31}$$

One can look at when this is small, comparisons with powers of r, perhaps estimates away from some kind of singular set, etc.

A real-valued function $f(x)$ on a metric space $(M, d(x, y))$ is said to be 1-*Lipschitz* if

$$|f(x) - f(y)| \le d(x, y) \tag{32}$$

for all $x, y \in M$. For instance, if p is a point in M, then $f(x) = d(x, p)$ is 1-Lipschitz, because

$$d(y, p) \le d(x, y) + d(x, p) \tag{33}$$

and

$$d(x, p) \le d(x, y) + d(y, p) \tag{34}$$

by the triangle inequality. More generally, if A is a nonempty subset of M, and if x is a point in M, then the distance from x to A is denoted $\mathrm{dist}(x, A)$ and defined by

$$\mathrm{dist}(x, A) = \inf\{d(x, y) : y \in A\}, \tag{35}$$

and one can check that $\mathrm{dist}(x, A)$ is a 1-Lipschitz function on M.

In short, there is a fairly rich supply of fairly regular functions on any metric space, which can be used as basic building blocks. At the same time one can look at various measurements of local oscillation and so forth at varying locations and scales, as well as sizes of singular sets where some kind of tricky behavior is concentrated. In this regard note that metric spaces often come equipped as well with some natural way of measuring volumes, which is then another ingredient which mixes with the others in very interesting ways.