Metric Spaces and Elementary Topology

E. Speer

Introduction to Mathematics at Rutgers

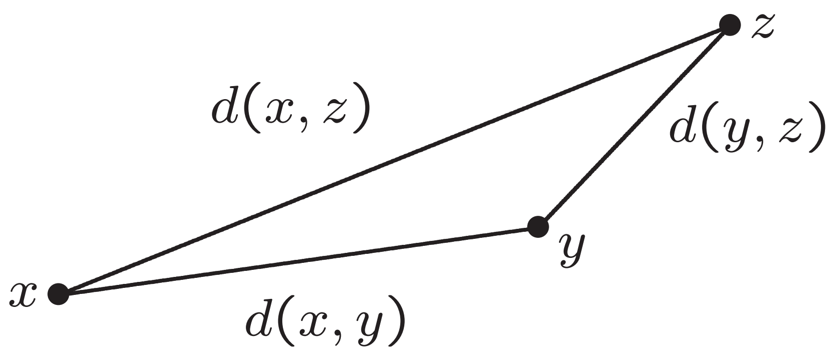
September 2007

I. BASICS

A. Definition

A *metric space* is a pair (X, d), where

- X is a set;
- $d : X \times X \to \mathbb{R}$ is a *metric* on X; $d(x, y)$ is interpreted as the distance between the points $x, y \in X$, and must satisfy
  1. $d(x, y) \geq 0$ for $x, y \in X$, with $d(x, y) = 0$ if and only if $x = y$;
  2. $d(x, y) = d(y, x)$ for $x, y \in X$ (symmetry);
  3. $d(x, z) \leq d(x, y) + d(y, z)$ for $x, y, z \in X$ (the triangle inequality).



We usually write X rather than (X, d) when no confusion can arise.

B. Examples

- The Euclidean space $\mathbb{R}^n$: If $x, y \in \mathbb{R}^n$ with $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, then the *Euclidean distance* between x and y is

$$d_2(x, y) = \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{1/2} .$$

If we define the *length* of a vector $z = (z_1, \ldots, z_n) \in \mathbb{R}^n$ to be $\|z\| = \left(\sum_i z_i^2\right)^{1/2}$, then $d_2(x, y) = \|x - y\|_2$.

- A normed vector space V: If V is a real vector space then a *norm* $\|\cdot\|$ on V is a mapping from V to $\mathbb{R}$ such that

  1. $\|x\| \geq 0$, with $\|x\| = 0$ iff $x = 0$;
  2. $\|c\, x\| = |c|\|x\|$, $x \in V$, $c \in \mathbb{R}$;
  3. $\|x + y\| \leq \|x\| + \|y\|$.

Then $d(x, y) = \|x - y\|$ is a metric on V.

- $C([0,1])$: We let $C([0,1]) \equiv C$ denote the space of all continuous functions $f : [0,1] \to \mathbb{R}$, and define the *uniform norm* $\|\cdot\|_\infty$ on $C([0,1])$ by

$$\|f\|_\infty = \max_{t \in [0,1]} |f(t)|.$$

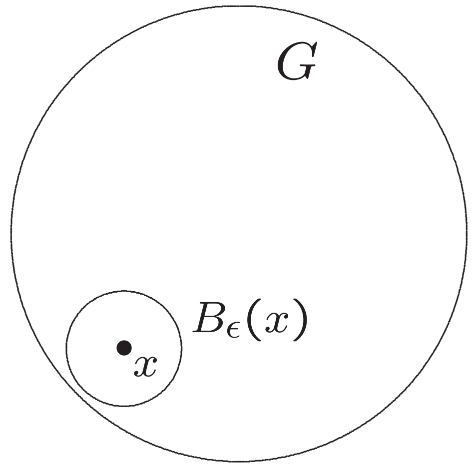
Thus C becomes a metric space:

$$d(f,g) = \max_{t \in [0,1]} |f(t) - g(t)|.$$

Intuitively, two continuous functions in C are close if they are close at all points $x \in [0,1]$, i.e., *uniformly close*.

- Subspaces: If X is a metric space and Y an arbitrary subset of X then Y is also a metric space, with the metric inherited from X.

II. TOPOLOGY



A. Open sets

- For $\epsilon > 0$ and $x \in X$ the set

$$B_\epsilon(x) = \{y \in X \mid d(x, y) < \epsilon\}$$

is called the *ϵ-ball* centered at x.

- A set $G \subset X$ is *open* if for every $x \in X$ there is an $\epsilon > 0$ such that $B_\epsilon(x) \subset G$.

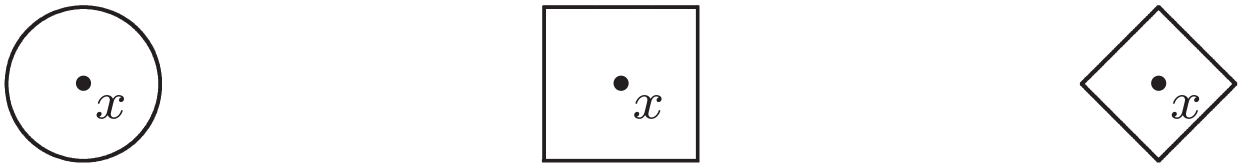
- The family of open sets has the properties:
  1. The empty set $\emptyset$, and X itself, are open;
  2. The union of an arbitrary family of open sets is open;
  3. The intersection of a finite number of open sets is open.

- Such a family of open sets defines a *topology*, as we will see below. Thus a metric gives a topology, but different metrics can give the same topology.

- Example: Equivalent metrics on $\mathbb{R}^n$: We may define other metrics on $\mathbb{R}^n$, e.g., those obtained from the norms

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \qquad \text{and} \qquad \|x\|_\infty = \max_{1 \le i \le n} |x_i|,$$

All these are equivalent in the sense that they define the same open sets—i.e., the same topology, although the balls they define look quite different: in $\mathbb{R}^2$ we have



d_2 $\qquad$ d_∞ $\qquad$ d_1

- We may also use the metric, again equivalent to d_2:

$$d(x, y) = \frac{d_2(x, y)}{1 + d_2(x, y)},$$

which does not come (directly) from a norm. Notice that in the last example the distance between two points is always less than 1. The balls defined with d are the same as those defined with d_2, and hence the topologies are certainly the same.

B. Topology in general—look for the yellow boxes

- If X is a set and $\mathcal{T}$ a collection of subsets of X (called *open sets*) with the above three properties:

  1. The empty set $\emptyset$, and X itself, are open;
  2. The union of an arbitrary family of open sets is open;
  3. The intersection of a finite number of open sets is open;

  then $\mathcal{T}$ is called a *topology* for X and $(X, \mathcal{T})$ (or just X) is a *topological space*.

- For later reference we mention here two special cases:
  - A topological space X is *Hausdorff* if for $x, y \in X$, with $x \neq y$, there are disjoint open sets $G_x, G_y \in \mathcal{T}$ with $x \in G_x$ and $y \in G_y$.
  - A space X is *first countable* if for each $x \in X$ there is a family $(G_{x,n})_{n=1}^{\infty}$ of open sets, all containing x, such that if an open set G contains x then $G_{x,n} \subset G$ for some n.

- A metric space is Hausdorff (if $d(x, y) = a$ take $G_x = B_{a/2}(x)$, $G_y = B_{a/2}(y)$) and first countable (take $G_{x,n} = B_{1/n}(x)$).

C. Closed sets and closure

- **Limit points:** A point x of X is a *limit point* of a set $S \subset X$ if every open set G containing x intersects S in at least one point other than x itself. A limit point of S may or may not itself belong to S.

- **Closed sets:** A set $F \subset X$ is *closed* if its complement $X \setminus F$ is open. Equivalently, F is closed if it contains all its limit points.

  From the basic properties of open sets, we see immediately that the empty set and X itself are closed, that an arbitrary intersection of closed sets is closed, and that a finite union of closed sets is closed.

- **Closure:** If S is a set in a topological space X, the *closure* of S, $\bar{S}$, is the smallest closed set containing S, or equivalently:

  * The intersection of all closed sets containing S;

  * The union of S and all its limit points.

D. Sequences

- A sequence $(x_n)_{n=1}^{\infty}$ of points in a topological space X *converges* to a point $x \in X$ if for every open set G containing x there is an $N \in \mathbb{N}$ such that if $n \geq N$ then $x_n \in G$; in this case we write $\lim_{n\to\infty} x_n = x$ and say that (x_n) is *convergent*.

- If X is a metric space, $\lim_{n\to\infty} x_n = x$ iff $\lim_{n\to\infty} d(x_n, x) = 0$.

- If X is first countable—in particular, a metric space—then many topological concepts can be expressed in terms of sequences:
  - A point $x \in X$ is a limit point of the set $S \subset X$ iff there is a sequence (x_n) of distinct points with $x_n \in S$ and $\lim_{n\to\infty} x_n = x$.
  - A set $F \subset X$ is closed iff whenever (x_n) is a convergent sequence of points from F, $\lim_{n\to\infty} x_n \in F$.

III. Continuity

- Continuity in metric spaces: If X and Y are metric spaces, then $f : X \to Y$ is *continuous* if $f(x)$ and $f(y)$ are close together whenever x and y are. More precisely, f is *continuous* if and only if for every $x \in X$ and every $\epsilon > 0$ there is a $\delta > 0$ such that $f(B_\delta(x)) \subset B_\epsilon(f(x))$.

- Continuity via sequences: If X and Y are topological spaces, with X first countable—in particular, a metric space—then $f : X \to Y$ is continuous iff for every $x \in X$, $\lim_{n\to\infty} f(x_n) = f(x)$ whenever $\lim_{n\to\infty} x_n = x$.

- Continuity in general: Suppose that X and Y are topological spaces. Then a function $f : X \to Y$ is *continuous* if the inverse image $f^{-1}(G)$ of every open set $G \subset Y$ is an open subset of X.

- Examples: We are all familiar with continuous and discontinuous functions from $\mathbb{R}$ to $\mathbb{R}$. Somewhat more interesting are functions defined on C.

- Indefinite integration: Define $T : C \to C$ by $T(f)(x) = \int_0^x f(t)\,dt$. One checks that $T(f)$ is well defined and lies in C. Then T is itself continuous, since if $\|f - g\|_\infty < \epsilon$ then

$$\begin{aligned} |T(f)(x) - T(g)(x)| &= \left| \int_0^x [f(t) - g(t)]\,dt \right| \\ &\le \int_0^x |f(t) - g(t)|\,dt < \int_0^x \epsilon\,dt = \epsilon x \le \epsilon \end{aligned}$$

since $0 \le x \le 1$, and hence $\|T(f) - T(g)\| < \epsilon$.

- Differentiation: Let $C^1 \subset C$ denote the set of functions in C which have continuous (first) derivatives; C^1 is a metric space (with the subspace topology) and we can define the operator $D : C^1 \to C$ by $D(f) = f'$. But D is not continuous, since the sequence (f_n) defined by $f_n = n^{-1} \sin \pi n x \in C^1$ satisfies $\|f_n\|_\infty = n^{-1}$ and hence $\lim_{n\to\infty} f_n = 0$, but $Df_n = \pi \cos n\pi x$ satisfies $\|Df_n\|_\infty = \pi$.

IV. COMPLETENESS

A. Definition

- Cauchy sequences: Suppose that X is a metric space. A sequence (x_n) of points of X is a *Cauchy sequence* if

$$\lim_{n,m\to\infty} d(x_n, x_m) = 0,$$

which means that for every $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that $d(x_n, x_m) < \epsilon$ whenever $n, m \geq N$.

- Completeness: A metric space X is *complete* if every Cauchy sequence (x_n) in X is convergent: $\lim_{n\to\infty} x_n = x$ for some $x \in X$.

- Completion: If (X, d_X) is a metric space then one may construct a complete metric space (Y, d_Y) such that $X \subset Y$, the restriction of the metric d_Y to X is d_X and the closure of X is Y. Y is (essentially) uniquely determined and is called the *completion* of X.

- Example: The set $\mathbb{Q}$ of rational numbers, with the usual metric (inherited from $\mathbb{R}$), is not complete. For example, if

$$x_n = \max\{k/n \mid k \in \mathbb{N}, (k/n)^2 < 2\}$$

then (x_n) is a Cauchy sequence with no limit in $\mathbb{Q}$.

- The set $\mathbb{R}$ of real numbers is complete, essentially because the $\mathbb{R}$ has the *least upper bound property*: every set of real numbers which is bounded above has a least upper bound. $\mathbb{R}^n$ is also complete.

- The completion of $\mathbb{Q}$ is $\mathbb{R}$.

- Example: The space $C = C([0,1])$ is complete. To see this, let (f_n) be a Cauchy sequence of functions in C. Then for any fixed $x \in [0,1]$, $(f_n(x))$ is a Cauchy sequence in $\mathbb{R}$, and hence convergent; this means that there is a function $f : [0,1] \to \mathbb{R}$ with $\lim_{x\to\infty} f_n(x) = f(x)$ for every $x \in [0,1]$. The fact that $\lim_{m,n\to\infty} \|f_n - f_m\|_\infty = 0$ easily implies that also $\lim_{n\to\infty} \|f_n - f\| = 0$, and the result follows from the fact that the uniform limit of continuous functions is continuous.

B. The Contraction Mapping Principle

Let (X, d) be a metric space. A mapping $f : X \to X$ is a *contraction* if there exists a number α, with $0 \leq \alpha < 1$, such that for all $x, y \in X$,

$$d(f(x), f(y)) \leq \alpha\, d(x, y) \tag{$*$}$$

It is easy to see that a contraction must be continuous.

A *fixed point* of f is a point $x \in X$ such that $f(x) = x$.

Theorem: *Let (X, d) be a complete metric space and $f : X \to X$ be a contraction. Then f has a unique fixed point x in X.*

Proof: Choose $\alpha < 1$ so that f satisfies $(*)$. Let x_0 be any point of X and inductively define $x_n = f(x_{n-1})$ for $n = 1, 2, \ldots$. We claim that for any $n \geq 0$, $d(x_n, x_{n+1}) \leq \alpha^n d(x_0, x_1)$; this is trivial for $n = 0$ and follows by induction for general n:

$$d(x_n, x_{n+1}) = d(f(x_{n-1}), f(x_n)) \leq \alpha\, d(x_{n-1}, x_n) \leq \alpha^n d(x_0, x_1)$$

Proof (continued): But then if $m \geq n \geq 0$, then the triangle inequality yields

$$\begin{aligned} d(x_n, x_m) &\leq \sum_{i=0}^{m-n+1} d(x_{n+i}, x_{n+i+1}) \\ &\leq \sum_{i=0}^{m-n+1} \alpha^{n+i} d(x_0, x_1) \;\leq\; \frac{\alpha^n}{1-\alpha} d(x_0, x_1), \end{aligned}$$

so that $\lim_{n,m\to\infty} d(x_n, x_m) = 0$ and (x_n) is Cauchy. Let $x = \lim_{n\to\infty} x_n$; then from the continuity of f, x is a fixed point of f:

$$f(x) = \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} x_{n+1} = x.$$

If y is any fixed point then $d(x, y) = d(f(x), f(y)) \leq \alpha\, d(x, y)$, which is possible only if $d(x, y) = 0$, i.e., $x = y$. ■

The contraction mapping principle is often applied in the following form.

Corollary *Suppose that X is a complete metric space, $B \subset X$ a ball, and $f : X \to X$ a mapping such that $f(B) \subset B$ and f is a contraction when restricted to B. Then f has a fixed point in B.*

V. COMPACTNESS

A. Basics

- **Definition:** Suppose that X is a topological space and K is a subset of X. A family $\{G_\alpha \mid \alpha \in A\}$ of open sets in X is an *open cover* of K if $K \subset \bigcup_{\alpha \in A} G_\alpha$. K is *compact* if every open cover of K contains a finite subcover, i.e., if for some $\alpha_1, \ldots, \alpha_n \in A$, $K \subset \bigcup_{1 \le i \le n} G_{\alpha_i}$.

- A compact subset of a Hausdorff space is closed.

- A subspace K of a metric space X is compact iff it is *sequentially compact*: every sequence (x_n) of points in K contains a convergent subsequence.

- A subset of $\mathbb{R}^n$ is compact iff it is *closed* and is *bounded* in the usual metric d_2. (But this is not true in the "equivalent" metric d defined by $d(x,y) = \dfrac{d_2(x,y)}{1 + d_2(x,y)}$, in which all sets are bounded.)

B. Functions on a compact set

- Suppose that K is compact and $f : K \to \mathbb{R}$ is continuous. Then:

  - f is *bounded:* for some $M > 0$, $|f(x)| \le M$ for all $x \in K$.
  - f has a *maximum* and *minimum* on K: for some $x_1, x_2 \in K$, $f(x_1) \le f(x) \le f(x_2)$ for all $x \in K$.

- If K is a compact metric space and f is as above then:
  - f is *uniformly continuous:* for every $\epsilon > 0$ there is a $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ if $x, y \in X$ with $d(x, y) < \delta$.

- Proof of boundedness: Let $S_n = (-n, n) \subset \mathbb{R}$. Each S_n is open, so that the sets $G_n = f^{-1}(S_n)$ form an open cover of K, which must have a finite subcover. Since $G_n \subset G_m$ if $n < m$, $K \subset G_N$ for some N, i.e., $|f(x)| < N$ for all $x \in K$.

- Ascoli-Arzela Theorem: If K is a compact metric space and (f_n) is a sequence of functions with $f_n : K \to \mathbb{R}$ which is

  - *uniformly bounded*: for some $M > 0$ and all $n \in \mathbb{N}$ and $x \in K$, $|f_n(x)| \leq M$; and
  - *uniformly equicontinuous*: for every $\epsilon > 0$ there is a $\delta > 0$ such that for all $n \in \mathbb{N}$, $|f_n(x) - f_n(y)| < \epsilon$ if $d(x, y) < \delta$;

then (f_n) contains a subsequence converging uniformly to some continuous function $f : K \to \mathbb{R}$.

- A more abstract version, given for $K = [0, 1]$: Let S be a subset of $C = C([0, 1])$ whose elements are

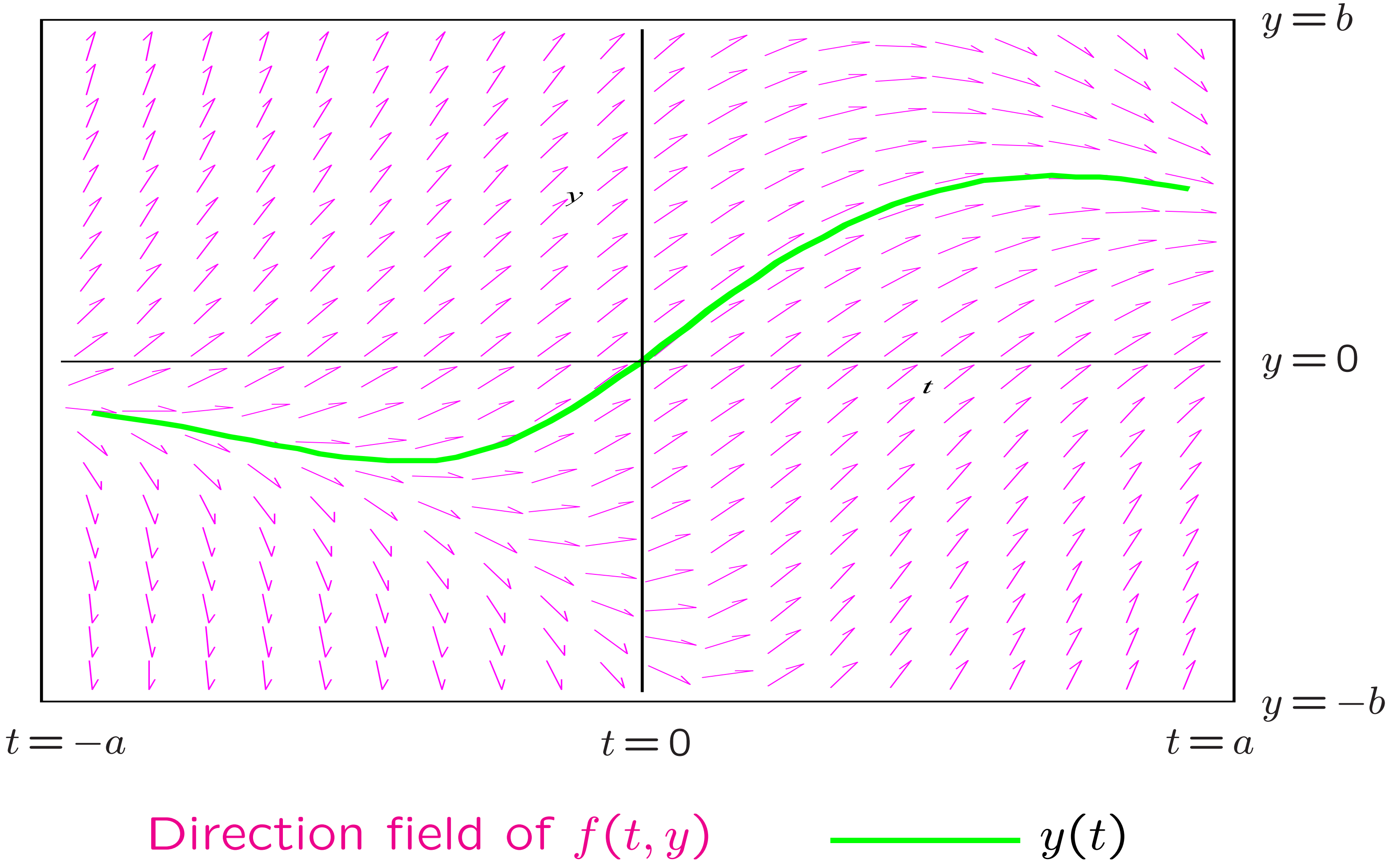
  - *uniformly bounded*: for some $M > 0$ and all $f \in S$ and $x \in [0, 1]$, $|f(x)| \leq M$; and
  - *uniformly equicontinuous*: given $\epsilon > 0$ there is a $\delta > 0$ such that for all $f \in S$, $|f(x) - f(y)| < \epsilon$ if $d(x, y) < \delta$.

Then the closure of S in C is compact.

- Example: We will show that the initial value problem

$$y'(t) = f(t, y), \qquad y(t_0) = y_0, \tag{2}$$

with f defined and continuous in some neighborhood of (t_0, y_0), has a solution $y(t)$ on some interval $(t_0 - a, t_0 + a)$. In a picture:



- Proof sketch: For simplicity take $t_0 = y_0 = 0$. Suppose that f is continuous and satisfies $|f| \le M$ in the rectangle $[-a, a] \times [-b \times b]$, and if necessary decrease a so that

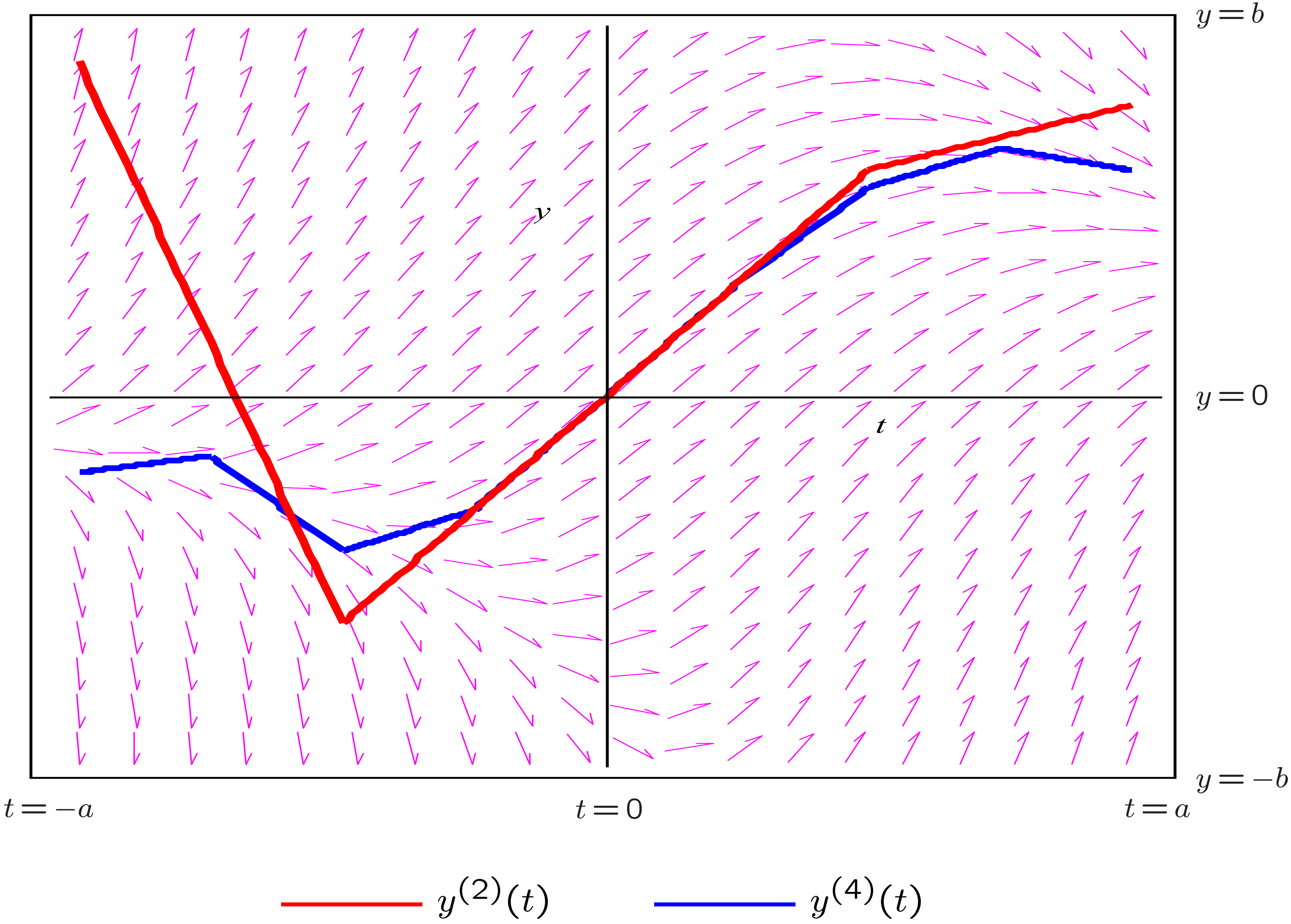
$$Ma \le b. \tag{3}$$

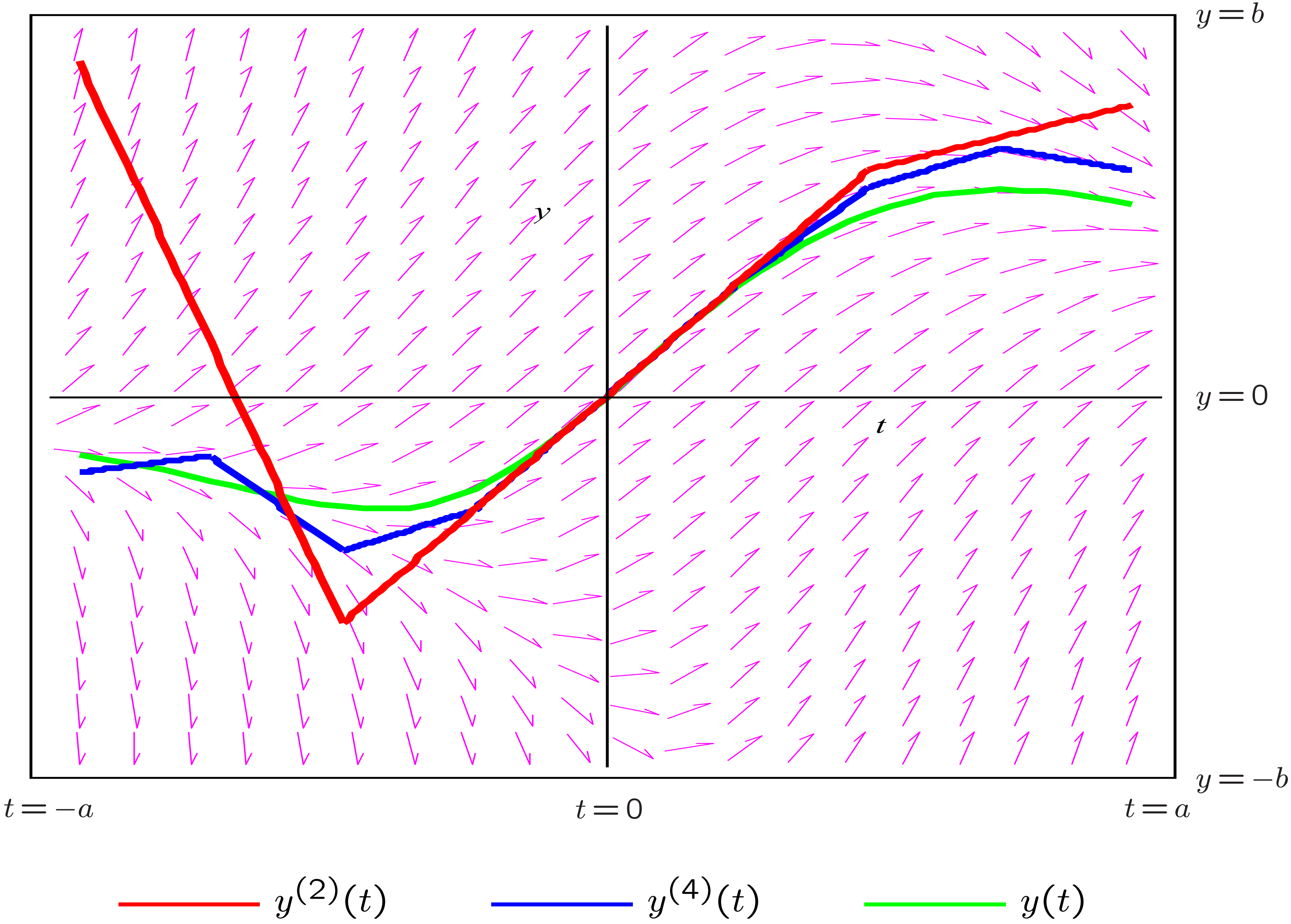
For each integer $n \ge 1$ construct an approximate solution $y^{(n)}(t)$ on $[-a, a]$ using Euler's method.

Let $t_k = ka/n$, $k = 0, \ldots, n$. $y^{(n)}(t)$ is obtained, for $t \ge 0$, by starting at $(0, 0)$ and following a straight line of slope $f(0, 0)$ until $t = t_1$; if the resulting y value is $y_1^{(n)} = y^{(n)}(t_1)$ we then follow a straight line of slope $f(t_1, y_1^{(n)})$ until $t = t_2$, and so on. More precisely, for $0 \le t \le a$ and with $y_k^{(n)} = y^{(n)}(t_k)$, $k = 0, \ldots, n$, we define

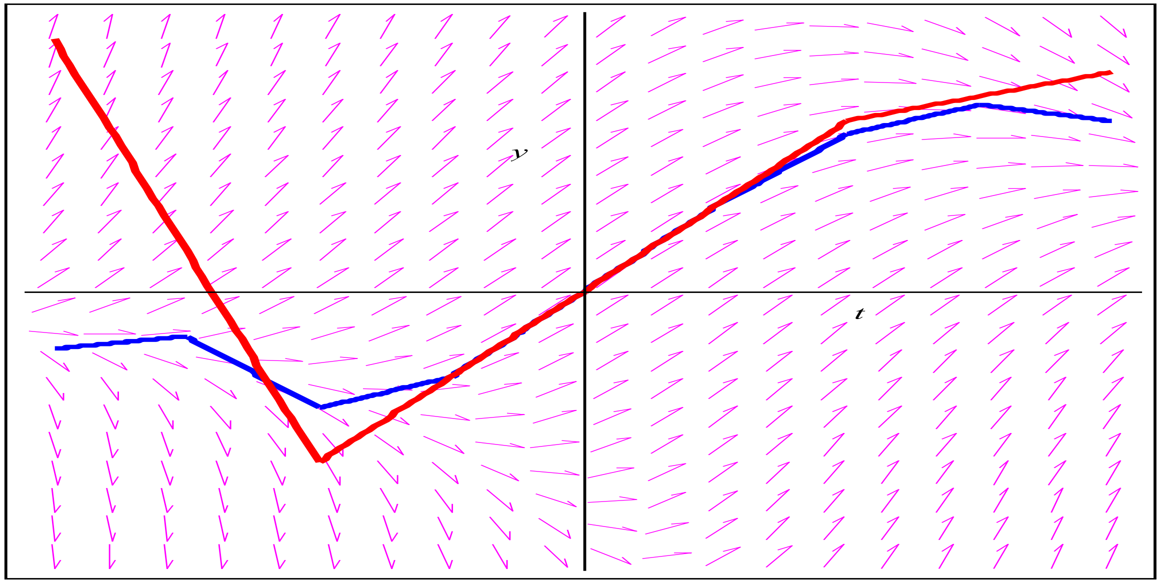
$$\begin{aligned}
y^{(n)}(t) &= f(t_0, y_0)\, t \;=\; f(0, 0)\, t, && 0 \le t \le t_1;\\
y^{(n)}(t) &= y_1^{(n)} + f(t_1, y_1^{(n)}))\, (t - t_1), && t_1 < t \le t_2;\\
&\;\;\vdots\\
y^{(n)}(t) &= y_{n-1}^{(n)} + f(t_{n-1}, y_{n-1}^{(n)})\, (t - t_{n-1}), && t_{n-1} < t \le a.
\end{aligned}$$

Note that $y_k^{(n)}$ is defined in the k^{th} equation here and then used in the $(k+1)^{\text{st}}$ equation. The definition for $-a \le t < 0$ is similar. Here is a plot of $y^{(2)}(t)$ and $y^{(4)}(t)$ and then a plot of $y^{(2)}(t)$, $y^{(4)}(t)$, and $y(t)$:


$y = b$
y
t
$y = 0$
$y = -b$
$t = -a$
$t = 0$
$t = a$
$y^{(2)}(t)$
$y^{(4)}(t)$



$y=b$
$y=0$
$y=-b$
$t=-a$
$t=0$
$t=a$
y
t
$y^{(2)}(t)$
$y^{(4)}(t)$
$y(t)$



Each of the line segments comprising $y^{(n)}$ has slope at most M, so

$$\left|y^{(n)}(t') - y^{(n)}(t)\right| \leq M|t' - t| \tag{4}$$

for $t', t \in [-a, a]$, and hence for $t \in [-a, a]$,

$$\left|y^{(n)}(t)\right| \leq \left|y^{(n)}(0)\right| + M|t - 0| = 0 + M|t| \leq Ma \leq b, \tag{5}$$

which guarantees that the solutions $y^{(n)}(t)$ never leave the rectangle. Now (5) is uniform boundedness and (4) is uniform equicontinuity, so by Ascoli/Arzela there is a convergent subsequence $y^{(n_k)}$, and one easily checks that this satisfies the initial value problem.

- Remark: If $f(t,y)$ satisfies a Lipchitz condition,

$$|f(t,y) - f(t,y')| \leq C|y - y'| \quad \text{for some } C > 0,$$

then an alternative proof of the existence of a solution of the initial value problem can be given using the contraction mapping principle. In this case the solution is unique. Without the Lipschitz condition the solution may not be unique and one needs a proof such as the one given here. The nonuniqueness of the solution corresponds to the fact that we do not prove here that the sequence $y^{(n)}$ converges, but only that a subsequence converges; in theory different subsequence limits are possible.

VI. CONNECTEDNESS

• Definition: A subset S of a topological space X is *connected* if there do not exist two open subsets of X, G_1 and G_2, such that $S \subset G_1 \cup G_2$ and $S \cap G_1$ and $S \cap G_2$ are disjoint and are both not empty.

• Example: Suppose that S is a subset of $\mathbb{R}$ that contains both positive and negative numbers, but not zero. Then S is not connected, as shown by consideration of the open subsets $G_1 = (0, \infty)$ and $G_2 = (-\infty, 0)$ of $\mathbb{R}$.

By the same reasoning we see that if a connected subset $S \subset \mathbb{R}$ contains distinct points x and y then it must also contain all intermediate points z and conclude that the only connected subsets of $\mathbb{R}$ are the intervals - open, closed, half open, semi or doubly infinite, etc.

- Definition: A subset S of a topological space X is *path connected* if for each $x, y \in S$ there is a *path* from x to y in S: a continuous map $\gamma : [0,1] \to S$ such that $\gamma(0) = x$ and $\gamma(1) = y$.

- Example: The *topologist's sine curve* $S \subset \mathbb{R}^2$,

$$S = \{(x,y) \in \mathbb{R}^2 \mid x = 0,\ -1 \leq y \leq 1\} \cup \left\{\left(x, \sin\frac{1}{x}\right) \in \mathbb{R}^2 \mid x > 0\right\},$$

is connected but not path connected.