Notes on Brouwer and Kakutani fixed point theorems

Marcelo Gallardo

August 2026

These notes present the proofs of two papers together with the material each one needs. Milnor [8] deduces Brouwer's fixed point theorem from the hairy ball theorem, which he proves by a volume computation (Section 1). Section 2 records the reduction of Brouwer's theorem to its smooth case, the route taken by the proofs that begin with a differentiable map. Cellina [2, 3] approximates an upper hemicontinuous correspondence with convex values by continuous functions whose graphs lie in a prescribed neighborhood of its graph; Kakutani's theorem follows from this together with Brouwer's (Section 3).

1 Brouwer's theorem, following Milnor

Notation: $S^{n-1} = \{\mathbf{u} \in \mathbb{R}^n : \|\mathbf{u}\| = 1\}$ and $D^n = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leq 1\}$, where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^n$ and $\mathbf{x} \cdot \mathbf{y}$ the Euclidean inner product. For maps $\mathbf{F}, \mathbf{G} : D^n \to \mathbb{R}^n$ we write

$$\|\mathbf{F} - \mathbf{G}\|_\infty = \sup_{\mathbf{x} \in D^n} \|\mathbf{F}(\mathbf{x}) - \mathbf{G}(\mathbf{x})\|.$$

A vector $\mathbf{v}(\mathbf{u}) \in \mathbb{R}^n$ is tangent to S^{n-1} at $\mathbf{u}$ if $\mathbf{u} \cdot \mathbf{v}(\mathbf{u}) = 0$.

The structure of Milnor's proof [8] is

$$\text{Lemma 1.1} \implies \text{Theorem 1.3} \implies \text{Theorem 1.5} \implies \text{Theorem 1.6}.$$

1.1 The volume lemma

Let $A \subset \mathbb{R}^n$ be compact and let $\mathbf{v} : U \to \mathbb{R}^n$ be a C^1 vector field on a neighborhood U of A. For $t \in \mathbb{R}$ set

$$\boldsymbol{\varphi}_t(\mathbf{x}) = \mathbf{x} + t\,\mathbf{v}(\mathbf{x}), \qquad \mathbf{x} \in A.$$

Lemma 1.1. *For $|t|$ sufficiently small, $\boldsymbol{\varphi}_t$ is one-to-one on A and*

$$\operatorname{vol} \boldsymbol{\varphi}_t(A) = a_0 + a_1 t + \cdots + a_n t^n,$$

a polynomial in t.[1]

[1] Here vol is n-dimensional Lebesgue measure.

Proof. Step 1: $\mathbf{v}$ *is Lipschitz on* A*.* If A is a cube with edges parallel to the axes, moving from $\mathbf{x}$ to $\mathbf{y}$ one coordinate at a time and applying the mean value theorem in each variable gives $|v_i(\mathbf{x}) - v_i(\mathbf{y})| \le \sum_j c_{ij}|x_j - y_j|$ with $c_{ij} = \sup_A |\partial v_i/\partial x_j| < \infty$; summing over i yields a Lipschitz constant. For general compact A, cover A by finitely many open cubes $I_\alpha \subset U$ so that the cube estimate applies whenever $\mathbf{x}, \mathbf{y}$ lie in a common cube. If $\mathbf{x}, \mathbf{y} \in A$ share no cube, then $\|\mathbf{x} - \mathbf{y}\|$ is bounded below by a positive constant, since $(\mathbf{x}, \mathbf{y}) \mapsto \|\mathbf{x} - \mathbf{y}\|$ is continuous and nowhere zero on the compact set $A \times A \setminus \bigcup_\alpha I_\alpha \times I_\alpha$, while $\|\mathbf{v}(\mathbf{x}) - \mathbf{v}(\mathbf{y})\|$ is bounded above. A single constant c then works uniformly:

$$\|\mathbf{v}(\mathbf{x}) - \mathbf{v}(\mathbf{y})\| \le c\,\|\mathbf{x} - \mathbf{y}\|, \qquad \mathbf{x}, \mathbf{y} \in A. \tag{$*$}$$

Step 2: injectivity for $|t| < c^{-1}$*.* If $\boldsymbol{\varphi}_t(\mathbf{x}) = \boldsymbol{\varphi}_t(\mathbf{y})$, then $\mathbf{x} - \mathbf{y} = t(\mathbf{v}(\mathbf{y}) - \mathbf{v}(\mathbf{x}))$, so $\|\mathbf{x} - \mathbf{y}\| \le |t|c\|\mathbf{x} - \mathbf{y}\|$ by $(*)$, forcing $\mathbf{x} = \mathbf{y}$.

Step 3: $\det D\boldsymbol{\varphi}_t$ *is polynomial in* t*.* We have $D\boldsymbol{\varphi}_t(\mathbf{x}) = I + t\,[\partial v_i/\partial x_j(\mathbf{x})]$, so each entry of $D\boldsymbol{\varphi}_t(\mathbf{x})$ is affine in t, and

$$\det D\boldsymbol{\varphi}_t(\mathbf{x}) = 1 + t\,\sigma_1(\mathbf{x}) + \cdots + t^n \sigma_n(\mathbf{x}),$$

where the σ_k, sums of minors of $D\mathbf{v}$, are continuous on A. Since $\det D\boldsymbol{\varphi}_t(\mathbf{x})$ is continuous in $(t, \mathbf{x})$ and equals 1 at $t = 0$, it is strictly positive on A for $|t|$ small.

Step 4: change of variables.[2] For such t, $\boldsymbol{\varphi}_t$ is injective and C^1 with positive Jacobian, so

$$\operatorname{vol} \boldsymbol{\varphi}_t(A) = \int_A \det D\boldsymbol{\varphi}_t(\mathbf{x})\, d\mathbf{x} = \sum_{k=0}^{n} \left(\int_A \sigma_k(\mathbf{x})\, d\mathbf{x} \right) t^k, \qquad \sigma_0 \equiv 1. \qquad \square$$

1.2 The sphere-to-sphere lemma

Suppose S^{n-1} carries a C^1 field $\mathbf{u} \mapsto \mathbf{v}(\mathbf{u})$ of unit tangent vectors: $\|\mathbf{v}(\mathbf{u})\| = 1$ and $\mathbf{u} \cdot \mathbf{v}(\mathbf{u}) = 0$. Then

$$\|\mathbf{u} + t\,\mathbf{v}(\mathbf{u})\|^2 = \|\mathbf{u}\|^2 + t^2 \|\mathbf{v}(\mathbf{u})\|^2 = 1 + t^2,$$

so $\boldsymbol{\varphi}_t(\mathbf{u}) = \mathbf{u} + t\mathbf{v}(\mathbf{u})$ maps S^{n-1} into the sphere of radius $\sqrt{1 + t^2}$.

Lemma 1.2. *For* $|t|$ *sufficiently small,* $\boldsymbol{\varphi}_t$ *maps* S^{n-1} *onto the sphere of radius* $\sqrt{1 + t^2}$*.*

First proof (Banach fixed point theorem). Extend $\mathbf{v}$ to the annulus $A = \{\mathbf{x} : \frac{1}{2} \le \|\mathbf{x}\| \le \frac{3}{2}\}$ by $\mathbf{v}(r\mathbf{u}) = r\,\mathbf{v}(\mathbf{u})$; the extension is C^1, with Lipschitz constant c on A. Fix $\mathbf{u}_0 \in S^{n-1}$, take $|t| < \min\{1/3, c^{-1}\}$, and define

$$\boldsymbol{\Phi}(\mathbf{x}) = \mathbf{u}_0 - t\,\mathbf{v}(\mathbf{x}).$$

Since $\|t\,\mathbf{v}(\mathbf{x})\| \le |t| \cdot \frac{3}{2} < \frac{1}{2}$, the map $\boldsymbol{\Phi}$ sends the complete metric space A into itself, and

$$\|\boldsymbol{\Phi}(\mathbf{x}) - \boldsymbol{\Phi}(\mathbf{y})\| = |t|\|\mathbf{v}(\mathbf{x}) - \mathbf{v}(\mathbf{y})\| \le |t|c\|\mathbf{x} - \mathbf{y}\|, \qquad |t|c < 1,$$

[2] If $\mathbf{g} : A \to \mathbf{g}(A)$ is injective, C^1, and $\det D\mathbf{g}$ vanishes nowhere, then $\operatorname{vol} \mathbf{g}(A) = \int_A |\det D\mathbf{g}(\mathbf{x})|\, d\mathbf{x}$. The number $\det D\mathbf{g}(\mathbf{x})$ is the factor by which $\mathbf{g}$ rescales volume near $\mathbf{x}$: the derivative $D\mathbf{g}(\mathbf{x})$ sends the unit cube to a parallelepiped of volume $|\det D\mathbf{g}(\mathbf{x})|$, and integrating this local factor over A gives the volume of the image. Here $\det D\boldsymbol{\varphi}_t > 0$ on A, so the absolute value can be dropped.

so $\mathbf{\Phi}$ is a contraction. Its unique fixed point $\bar{\mathbf{x}}$ satisfies $\boldsymbol{\varphi}_t(\bar{\mathbf{x}}) = \mathbf{u}_0$. Thus every $\mathbf{u}_0 \in S^{n-1}$ lies in $\boldsymbol{\varphi}_t(A)$; since $\boldsymbol{\varphi}_t(r\mathbf{u}) = r\boldsymbol{\varphi}_t(\mathbf{u})$, multiplying by $\sqrt{1+t^2}$ shows every point of the sphere of radius $\sqrt{1+t^2}$ equals $\boldsymbol{\varphi}_t(\mathbf{x})$ for some $\mathbf{x} \in S^{n-1}$. ☐

Second proof (inverse function theorem). Assume $n \geq 2$. For $|t|$ small, $D\boldsymbol{\varphi}_t$ is nonsingular on A by Step 3 above, so $\boldsymbol{\varphi}_t$ maps open subsets of the interior of A to open sets;[3] hence $\boldsymbol{\varphi}_t(S^{n-1})$ is relatively open in the sphere of radius $\sqrt{1+t^2}$. It is also compact, hence closed. The sphere of radius $\sqrt{1+t^2}$ is connected, so $\boldsymbol{\varphi}_t(S^{n-1})$, nonempty, open and closed, is the whole sphere. ☐

1.3 Hairy ball theorem, C^1 case

Theorem 1.3. *An even-dimensional sphere admits no C^1 field of unit tangent vectors.*

Proof. The sphere $S^{n-1} \subset \mathbb{R}^n$ has dimension $n-1$; even-dimensional means n odd.

Suppose such a field $\mathbf{v}$ exists on S^{n-1}, n odd. Extend it to $A = \{\mathbf{x} : a \leq \|\mathbf{x}\| \leq b\}$ by $\mathbf{v}(r\mathbf{u}) = r\,\mathbf{v}(\mathbf{u})$. By Lemma 1.1, for $|t|$ small,

$$\operatorname{vol}\boldsymbol{\varphi}_t(A) \text{ is a polynomial in } t. \tag{i}$$

By homogeneity, $\boldsymbol{\varphi}_t(r\mathbf{u}) = r\boldsymbol{\varphi}_t(\mathbf{u})$, so Lemma 1.2 rescaled gives: $\boldsymbol{\varphi}_t$ maps the sphere of radius r onto the sphere of radius $r\sqrt{1+t^2}$, for every $a \leq r \leq b$. Hence $\boldsymbol{\varphi}_t(A) = \{\mathbf{y} : a\sqrt{1+t^2} \leq \|\mathbf{y}\| \leq b\sqrt{1+t^2}\}$, the image of A under scaling by $\sqrt{1+t^2}$, and

$$\operatorname{vol}\boldsymbol{\varphi}_t(A) = (1+t^2)^{n/2}\operatorname{vol}(A). \tag{ii}$$

For n odd, $(1+t^2)^{n/2}$ is not a polynomial in t: if $p(t) = (1+t^2)^{n/2}$ with p polynomial, then $p(t)^2 = (1+t^2)^n$, so p has degree n, while $p(t) = (1+t^2)^{(n-1)/2}\sqrt{1+t^2}$ forces $\sqrt{1+t^2}$ to be a rational function, hence a polynomial of degree 1; but $(a+bt)^2 = 1+t^2$ has no solution. Then (i) and (ii) are incompatible. ☐

Remark 1.4. The parity hypothesis is sharp: for $n = 2m$,

$$\mathbf{v}(u_1, \ldots, u_{2m}) = (u_2, -u_1, u_4, -u_3, \ldots, u_{2m}, -u_{2m-1})$$

is a linear field of unit tangent vectors on S^{2m-1}.

1.4 Hairy ball theorem, continuous case

Theorem 1.5. *An even-dimensional sphere admits no continuous field of nowhere-zero tangent vectors.*

Proof. Let $\mathbf{v} : S^{n-1} \to \mathbb{R}^n$ be continuous, tangent, nowhere zero, with $n-1$ even, and let

$$m = \min_{\mathbf{u} \in S^{n-1}} \|\mathbf{v}(\mathbf{u})\| > 0.$$

[3] Inverse function theorem: if $\mathbf{g}$ is C^1 near $\mathbf{x}_0$ and $D\mathbf{g}(\mathbf{x}_0)$ is invertible, then $\mathbf{g}$ restricts to a bijection from some open neighborhood of $\mathbf{x}_0$ onto an open neighborhood of $\mathbf{g}(\mathbf{x}_0)$, with C^1 inverse. Such a bijection is called a diffeomorphism; the term reappears in the proof of Theorem 1.6. Openness of the image follows by applying this at each point.

By the Weierstrass approximation theorem on the compact set $S^{n-1} \subset \mathbb{R}^n$, applied componentwise (Section 2.1), there is a polynomial map $\mathbf{p} : \mathbb{R}^n \to \mathbb{R}^n$ with

$$\|\mathbf{p}(\mathbf{u}) - \mathbf{v}(\mathbf{u})\| < \frac{m}{2}, \qquad \mathbf{u} \in S^{n-1}.$$

Project $\mathbf{p}$ onto the tangent spaces:

$$\mathbf{w}(\mathbf{u}) = \mathbf{p}(\mathbf{u}) - (\mathbf{p}(\mathbf{u}) \cdot \mathbf{u})\, \mathbf{u}.$$

Then $\mathbf{w}(\mathbf{u}) \cdot \mathbf{u} = \mathbf{p}(\mathbf{u}) \cdot \mathbf{u} - (\mathbf{p}(\mathbf{u}) \cdot \mathbf{u})\|\mathbf{u}\|^2 = 0$, so $\mathbf{w}$ is tangent. Moreover, since $\mathbf{v}(\mathbf{u}) \cdot \mathbf{u} = 0$ and $\|\mathbf{u}\| = 1$,

$$|\mathbf{p}(\mathbf{u}) \cdot \mathbf{u}| = |(\mathbf{p}(\mathbf{u}) - \mathbf{v}(\mathbf{u})) \cdot \mathbf{u}| \le \|\mathbf{p}(\mathbf{u}) - \mathbf{v}(\mathbf{u})\| < \frac{m}{2},$$

so $\|\mathbf{w}(\mathbf{u}) - \mathbf{p}(\mathbf{u})\| = |\mathbf{p}(\mathbf{u}) \cdot \mathbf{u}| < m/2$, and

$$\|\mathbf{w}(\mathbf{u})\| \ge \|\mathbf{v}(\mathbf{u})\| - \|\mathbf{p}(\mathbf{u}) - \mathbf{v}(\mathbf{u})\| - \|\mathbf{w}(\mathbf{u}) - \mathbf{p}(\mathbf{u})\| > m - \frac{m}{2} - \frac{m}{2} = 0.$$

Thus $\mathbf{w}$ is a polynomial expression in $\mathbf{u}$, hence C^∞, tangent, and nowhere zero, and

$$\mathbf{u} \longmapsto \frac{\mathbf{w}(\mathbf{u})}{\|\mathbf{w}(\mathbf{u})\|}$$

is a C^∞ field of unit tangent vectors on S^{n-1}, contradicting Theorem 1.3. □

1.5 Brouwer's theorem

Theorem 1.6. *Every continuous map $\mathbf{f} : D^n \to D^n$ has a fixed point.*

Proof. Suppose $\mathbf{f}(\mathbf{x}) \neq \mathbf{x}$ for all $\mathbf{x} \in D^n$.

Step 1: a nowhere-zero field, radial on the boundary. Set $\mathbf{y} = \mathbf{f}(\mathbf{x})$ and

$$\mathbf{w}(\mathbf{x}) = \mathbf{x} - \mathbf{y}\, \frac{1 - \mathbf{x} \cdot \mathbf{x}}{1 - \mathbf{x} \cdot \mathbf{y}}.$$

The denominator never vanishes: $\mathbf{x} \cdot \mathbf{y} = 1$ with $\|\mathbf{x}\|, \|\mathbf{y}\| \le 1$ forces, by Cauchy–Schwarz with equality, $\mathbf{x} = \mathbf{y}$, that is, a fixed point. So $\mathbf{w}$ is continuous on D^n. If $\mathbf{x} \in S^{n-1}$, then $\mathbf{x} \cdot \mathbf{x} = 1$ and $\mathbf{w}(\mathbf{x}) = \mathbf{x}$. And $\mathbf{w}(\mathbf{x}) \neq \mathbf{0}$ for all $\mathbf{x}$: if $\mathbf{x}, \mathbf{y}$ are linearly independent, $\mathbf{w}(\mathbf{x})$ is a combination of $\mathbf{x}, \mathbf{y}$ with the coefficient of $\mathbf{x}$ equal to 1; if $\mathbf{y} = \lambda\mathbf{x}$, the identity $(\mathbf{x} \cdot \mathbf{x})\mathbf{y} = (\mathbf{x} \cdot \mathbf{y})\mathbf{x}$ gives

$$\mathbf{w}(\mathbf{x}) = \frac{\mathbf{x} - \mathbf{y}}{1 - \mathbf{x} \cdot \mathbf{y}} \neq \mathbf{0},$$

since $\mathbf{x} \neq \mathbf{y}$.

Step 2: transplant to S^n. Assume n even, so $S^n \subset \mathbb{R}^{n+1}$ is even-dimensional. Identify $\mathbb{R}^n$ with the hyperplane $x_{n+1} = 0$ in $\mathbb{R}^{n+1}$. Stereographic projection from the north pole,[4]

$$\mathbf{s}(\mathbf{x}) = \frac{(2x_1, \ldots, 2x_n,\ \mathbf{x}\cdot\mathbf{x} - 1)}{\mathbf{x}\cdot\mathbf{x} + 1},$$

maps D^n diffeomorphically onto the closed southern hemisphere of S^n. At $\mathbf{u} = \mathbf{s}(\mathbf{x})$ define

$$\mathbf{W}(\mathbf{u}) = \left.\frac{d}{dt}\right|_{t=0} \mathbf{s}\big(\mathbf{x} + t\,\mathbf{w}(\mathbf{x})\big),$$

the pushforward of $\mathbf{w}$ by $D\mathbf{s}$.[5] Then $\mathbf{W}(\mathbf{u})$ is tangent to S^n at $\mathbf{u}$ and nowhere zero on the southern hemisphere, since $\mathbf{s}$ is a diffeomorphism and $\mathbf{w} \neq \mathbf{0}$. At an equator point $\mathbf{u} = \mathbf{s}(\mathbf{x})$ with $\mathbf{x} \in S^{n-1}$, because $\mathbf{w}(\mathbf{x}) = \mathbf{x}$ is radial, a direct computation gives

$$\mathbf{W}(\mathbf{u}) = (0, \ldots, 0, 1).$$

Step 3: glue. Apply stereographic projection from the south pole to the field $-\mathbf{w}$: this yields a nowhere-zero tangent field on the closed northern hemisphere, equal to $(0, \ldots, 0, 1)$ at every equator point. Both fields agree on the equator, so together they define a continuous nowhere-zero tangent field on S^n, contradicting Theorem 1.5 for n even. Hence $\mathbf{f}$ has a fixed point when n is even.

Step 4: odd n. If $n = 2k-1$ and $\mathbf{f} : D^{2k-1} \to D^{2k-1}$ had no fixed point, then

$$\mathbf{F}(x_1, \ldots, x_{2k}) = \big(\mathbf{f}(x_1, \ldots, x_{2k-1}),\, 0\big)$$

is a continuous map $D^{2k} \to D^{2k}$ with no fixed point, since a fixed point of $\mathbf{F}$ has $x_{2k} = 0$ and first coordinates fixed by $\mathbf{f}$. This contradicts the even case. □

2 Brouwer's theorem for differentiable maps, and Stone–Weierstrass

Several proofs of Brouwer's theorem give the smooth case first and then pass to the continuous case by approximation. We record that passage here.

Hypothesis (smooth Brouwer). Every smooth map $\mathbf{g} : D^n \to D^n$ has a fixed point [9, 7, 10].

Theorem 2.1. *Assume the Hypothesis. Then every continuous map $\mathbf{f} : D^n \to D^n$ has a fixed point.*

2.1 Weierstrass approximation on compact subsets of $\mathbb{R}^n$

The classical Weierstrass theorem concerns a compact interval: every $g \in C([a,b], \mathbb{R})$ is a uniform limit of polynomials. Two standard arguments extend it to $D^n \subset \mathbb{R}^n$.

[4]Put D^n in the equatorial hyperplane of $S^n \subset \mathbb{R}^{n+1}$. For $\mathbf{x}$ in that hyperplane, draw the line through the north pole $(0, \ldots, 0, 1)$ and $\mathbf{x}$; the point $\mathbf{s}(\mathbf{x})$ is the second point where this line meets S^n. The displayed formula is the closed form of that construction. It sends $\mathbf{0}$ to the south pole, the boundary S^{n-1} of D^n to the equator of S^n, and D^n diffeomorphically onto the closed southern hemisphere.

[5]To push a tangent vector forward by a differentiable map, take a curve through the point with that velocity, here the straight line $t \mapsto \mathbf{x} + t\,\mathbf{w}(\mathbf{x})$, and differentiate its image at $t = 0$. By the chain rule this equals $D\mathbf{s}(\mathbf{x})\mathbf{w}(\mathbf{x})$ and does not depend on the curve chosen. The resulting vector is tangent to S^n at $\mathbf{s}(\mathbf{x})$, being the velocity of a curve that lies on S^n.

Via Stone–Weierstrass. If K is a compact Hausdorff space and $\mathcal{A} \subseteq C(K, \mathbb{R})$ is a subalgebra containing the constants and separating points of K, meaning that for $x \neq y$ in K there is $g \in \mathcal{A}$ with $g(x) \neq g(y)$, then $\mathcal{A}$ is uniformly dense in $C(K, \mathbb{R})$. Take $K = D^n$ and $\mathcal{A}$ the restrictions to D^n of real polynomials in n variables: $\mathcal{A}$ is an algebra, contains the constants, and separates points, since the coordinate functions $\mathbf{x} \mapsto x_i$ already do. Hence every $g \in C(D^n, \mathbb{R})$ is a uniform limit of polynomials in n variables. The same applies with $K = S^{n-1}$, as used in Theorem 1.5.

Via multivariate Bernstein polynomials. Enclose D^n in $[-1, 1]^n$, extend g continuously to the cube,[6] and run the n-variable Bernstein construction,[7] the product-form generalization of the one-variable proof.

Vector-valued maps. Given $\mathbf{f} = (f_1, \ldots, f_n) : D^n \to \mathbb{R}^n$ continuous and $\delta > 0$, choose polynomials p_i with $\sup_{D^n} |p_i - f_i| < \delta/\sqrt{n}$ and set $\mathbf{p} = (p_1, \ldots, p_n)$. Then

$$\|\mathbf{p} - \mathbf{f}\|_\infty = \sup_{\mathbf{x} \in D^n} \left(\sum_{i=1}^{n} |p_i(\mathbf{x}) - f_i(\mathbf{x})|^2 \right)^{1/2} < \delta,$$

and $\mathbf{p}$ is C^∞ on $\mathbb{R}^n$.

2.2 Proof of Theorem 2.1

Proof. Suppose $\mathbf{f} : D^n \to D^n$ is continuous with $\mathbf{f}(\mathbf{x}) \neq \mathbf{x}$ for every $\mathbf{x} \in D^n$.

Step 1. The function $\mathbf{x} \mapsto \|\mathbf{f}(\mathbf{x}) - \mathbf{x}\|$ is continuous and strictly positive on the compact set D^n, so it attains a positive minimum: there is $\varepsilon > 0$ with

$$\|\mathbf{f}(\mathbf{x}) - \mathbf{x}\| \geq \varepsilon, \qquad \mathbf{x} \in D^n.$$

Step 2. Choose a polynomial map $\mathbf{p} : D^n \to \mathbb{R}^n$ with $\|\mathbf{p} - \mathbf{f}\|_\infty < \varepsilon/2$.

Step 3. Since $\|\mathbf{f}(\mathbf{x})\| \leq 1$,

$$\|\mathbf{p}(\mathbf{x})\| \leq \|\mathbf{f}(\mathbf{x})\| + \|\mathbf{p}(\mathbf{x}) - \mathbf{f}(\mathbf{x})\| < 1 + \frac{\varepsilon}{2},$$

so $\mathbf{p}$ need not map D^n into D^n. Set

$$\mathbf{q}(\mathbf{x}) = \frac{\mathbf{p}(\mathbf{x})}{1 + \varepsilon/2}.$$

---

[6] Tietze extension theorem: if X is a normal topological space, in particular any metric space, $A \subseteq X$ is closed, and $g : A \to \mathbb{R}$ is continuous, then there is a continuous $\tilde{g} : X \to \mathbb{R}$ with $\tilde{g}|_A = g$; if $|g| \leq M$ on A, the extension can be chosen with $|\tilde{g}| \leq M$. In the present case no general theorem is needed: $\tilde{g}(\mathbf{x}) = g\big(\mathbf{x}/\max\{1, \|\mathbf{x}\|\}\big)$ extends g from D^n to all of $\mathbb{R}^n$ continuously, by composing with the radial retraction.

[7] For $g \in C([0,1]^n, \mathbb{R})$, the n-variate Bernstein polynomial of order k is

$$B_k g(x_1, \ldots, x_n) = \sum_{j_1=0}^{k} \cdots \sum_{j_n=0}^{k} g\left(\frac{j_1}{k}, \ldots, \frac{j_n}{k}\right) \prod_{i=1}^{n} \binom{k}{j_i} x_i^{j_i} (1 - x_i)^{k - j_i},$$

and $B_k g \to g$ uniformly on $[0,1]^n$ as $k \to \infty$. The proof is the probabilistic argument of the one-variable case, using the product structure of the weights: each factor is the probability mass function of a binomial distribution. The cube $[-1,1]^n$ is handled by an affine change of variables.

Then $\|\mathbf{q}(\mathbf{x})\| < 1$, so $\mathbf{q} : D^n \to D^n$, and $\mathbf{q}$ is polynomial, hence smooth. Moreover

$$\|\mathbf{q}(\mathbf{x}) - \mathbf{p}(\mathbf{x})\| = \|\mathbf{p}(\mathbf{x})\| \left(1 - \frac{1}{1+\varepsilon/2}\right) = \|\mathbf{p}(\mathbf{x})\| \frac{\varepsilon/2}{1+\varepsilon/2} \le \frac{\varepsilon}{2},$$

using $\|\mathbf{p}(\mathbf{x})\| \le 1 + \varepsilon/2$.

Step 4. By the triangle inequality, for every $\mathbf{x} \in D^n$,

$$\|\mathbf{q}(\mathbf{x}) - \mathbf{f}(\mathbf{x})\| \le \|\mathbf{q}(\mathbf{x}) - \mathbf{p}(\mathbf{x})\| + \|\mathbf{p}(\mathbf{x}) - \mathbf{f}(\mathbf{x})\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Step 5. For every $\mathbf{x} \in D^n$,

$$\|\mathbf{q}(\mathbf{x}) - \mathbf{x}\| \ge \|\mathbf{f}(\mathbf{x}) - \mathbf{x}\| - \|\mathbf{q}(\mathbf{x}) - \mathbf{f}(\mathbf{x})\| > \varepsilon - \varepsilon = 0,$$

so $\mathbf{q}$ is a smooth self-map of D^n with no fixed point, contradicting the Hypothesis. □

3 Cellina's approximation theorem

Throughout this section $E \subseteq \mathbb{R}^m$ is compact, $F \subseteq \mathbb{R}^k$ is convex, and $\Gamma : E \to 2^F$ is upper hemicontinuous (uhc) with nonempty convex values. For $\delta > 0$ and $\mathbf{x} \in E$, define

$$\Gamma^\delta(\mathbf{x}) = \operatorname{co} \bigcup_{\mathbf{z} \in N_\delta(\mathbf{x}) \cap E} \Gamma(\mathbf{z}),$$

where $N_\delta(\mathbf{x})$ is the open δ-ball about $\mathbf{x}$ and co denotes the convex hull.[8] Write $\operatorname{Graph}\Gamma = \{(\mathbf{x}, \mathbf{y}) : \mathbf{x} \in E,\ \mathbf{y} \in \Gamma(\mathbf{x})\} \subseteq \mathbb{R}^m \times \mathbb{R}^k$. For $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^m \times \mathbb{R}^k$ set

$$\operatorname{dist}\big((\mathbf{x}, \mathbf{y}), \operatorname{Graph}\Gamma\big) = \inf\big\{\|(\mathbf{x}, \mathbf{y}) - (\mathbf{x}', \mathbf{y}')\| : (\mathbf{x}', \mathbf{y}') \in \operatorname{Graph}\Gamma\big\},$$

the distance with respect to the Euclidean norm on $\mathbb{R}^m \times \mathbb{R}^k$, and define

$$N_\varepsilon(\operatorname{Graph}\Gamma) = \big\{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^m \times \mathbb{R}^k : \operatorname{dist}\big((\mathbf{x}, \mathbf{y}), \operatorname{Graph}\Gamma\big) < \varepsilon\big\}.$$

Lemma 3.1. *Let $\Gamma : E \to 2^F$ be uhc with nonempty values.*

(i) If $\Gamma(\mathbf{x})$ is compact for every $\mathbf{x} \in E$, then $\Gamma(E) = \bigcup_{\mathbf{x} \in E} \Gamma(\mathbf{x})$ is compact.

(ii) If $\Gamma(\mathbf{x})$ is closed for every $\mathbf{x} \in E$, then Graph Γ *is closed.*

Proof. See de la Fuente [4], Chapter 2, and Border [1], Chapter 11. □

[8] For $A \subseteq \mathbb{R}^k$, co A is the set of all finite convex combinations of points of A, equivalently the smallest convex set containing A.

3.1 The compact-valued version

Lemma 3.2 (Cellina; Border, Lemma 13.1)**.** *Assume* $\Gamma(\mathbf{x})$ *is compact for every* $\mathbf{x} \in E$*. Then for every* $\varepsilon > 0$ *there exists* $\delta > 0$ *such that* $\operatorname{Graph}\Gamma^{\delta} \subseteq N_{\varepsilon}(\operatorname{Graph}\Gamma)$*.*

Proof (following Border, 13.2). Suppose not. Then there exist $\varepsilon > 0$ and, for each $n \in \mathbb{N}$, a pair $(\mathbf{x}_n, \mathbf{y}_n) \in \operatorname{Graph}\Gamma^{1/n}$ with $\operatorname{dist}\big((\mathbf{x}_n, \mathbf{y}_n), \operatorname{Graph}\Gamma\big) \geq \varepsilon$. Since $\mathbf{y}_n \in \operatorname{co}\bigcup_{\mathbf{z}\in N_{1/n}(\mathbf{x}_n)\cap E}\Gamma(\mathbf{z}) \subseteq \mathbb{R}^k$, Carathéodory's theorem[9] lets us write

$$\mathbf{y}_n = \sum_{i=0}^{k} \lambda_{i,n}\, \mathbf{w}_{i,n}, \qquad (\lambda_{0,n}, \ldots, \lambda_{k,n}) \in \Delta^k, \quad \mathbf{w}_{i,n} \in \Gamma(\mathbf{z}_{i,n}), \quad \|\mathbf{z}_{i,n} - \mathbf{x}_n\| < \frac{1}{n},$$

with Δ^k the standard k-simplex. All sequences live in compact sets: $\mathbf{x}_n, \mathbf{z}_{i,n} \in E$, the coefficient vectors lie in Δ^k, and $\mathbf{w}_{i,n} \in \Gamma(E)$, compact by Lemma 3.1(i). Passing to a common subsequence,

$$\mathbf{x}_n \to \bar{\mathbf{x}} \in E, \qquad \lambda_{i,n} \to \bar{\lambda}_i, \qquad \mathbf{w}_{i,n} \to \bar{\mathbf{w}}_i, \qquad \mathbf{z}_{i,n} \to \bar{\mathbf{x}},$$

with $(\bar{\lambda}_0, \ldots, \bar{\lambda}_k) \in \Delta^k$, the last limit because $\|\mathbf{z}_{i,n} - \mathbf{x}_n\| < 1/n$. By Lemma 3.1(ii), $\operatorname{Graph}\Gamma$ is closed, and $(\mathbf{z}_{i,n}, \mathbf{w}_{i,n}) \in \operatorname{Graph}\Gamma$, so $\bar{\mathbf{w}}_i \in \Gamma(\bar{\mathbf{x}})$ for each i; by convexity of $\Gamma(\bar{\mathbf{x}})$,

$$\mathbf{y}_n \to \bar{\mathbf{y}} = \sum_{i=0}^{k} \bar{\lambda}_i \bar{\mathbf{w}}_i \in \Gamma(\bar{\mathbf{x}}).$$

Hence $(\mathbf{x}_n, \mathbf{y}_n) \to (\bar{\mathbf{x}}, \bar{\mathbf{y}}) \in \operatorname{Graph}\Gamma$ along the subsequence, so $\operatorname{dist}((\mathbf{x}_n, \mathbf{y}_n), \operatorname{Graph}\Gamma) \to 0$, a contradiction. $\square$

Theorem 3.3 (Approximation theorem; von Neumann, Cellina)**.** *Assume* $\Gamma(\mathbf{x})$ *is compact for every* $\mathbf{x} \in E$*. Then for every* $\varepsilon > 0$ *there exists a continuous* $\mathbf{f}_{\varepsilon} : E \to F$ *with* $\operatorname{Graph}\mathbf{f}_{\varepsilon} \subseteq N_{\varepsilon}(\operatorname{Graph}\Gamma)$ *and* $\mathbf{f}_{\varepsilon}(E) \subseteq \operatorname{co}\Gamma(E)$*.*

Proof (following Border, 13.4). Take $\delta > 0$ as in Lemma 3.2 and cover E with finitely many balls $N_{\delta/2}(\mathbf{x}_i)$, $i = 1, \ldots, p$, with $\mathbf{x}_i \in E$. Set $\mu_i(\mathbf{x}) = \max\{\delta/2 - \|\mathbf{x} - \mathbf{x}_i\|, 0\}$ and $\lambda_i = \mu_i / \sum_j \mu_j$, a continuous partition of unity subordinate to the cover.[10] Choose $\mathbf{y}_i \in \Gamma(\mathbf{x}_i)$ and define $\mathbf{f}_{\varepsilon}(\mathbf{x}) = \sum_{i=1}^{p} \lambda_i(\mathbf{x})\, \mathbf{y}_i$, a continuous function. If $\lambda_i(\mathbf{x}) > 0$ then $\|\mathbf{x} - \mathbf{x}_i\| < \delta/2 < \delta$, so $\mathbf{y}_i \in \bigcup_{\mathbf{z}\in N_{\delta}(\mathbf{x})\cap E}\Gamma(\mathbf{z})$; hence $\mathbf{f}_{\varepsilon}(\mathbf{x})$, being a convex combination of such points, lies in $\Gamma^{\delta}(\mathbf{x})$, and Lemma 3.2 gives $(\mathbf{x}, \mathbf{f}_{\varepsilon}(\mathbf{x})) \in N_{\varepsilon}(\operatorname{Graph}\Gamma)$. Finally $\mathbf{f}_{\varepsilon}(\mathbf{x}) \in \operatorname{co}\Gamma(E) \subseteq F$ by convexity of F. $\square$

3.2 The version without compactness of the values

The statement below is that of Cellina [2]; the proof given is the direct estimate of Hildenbrand–Kirman, Lemma AIV.1, not a transcription of Cellina's argument.

[9] If $A \subseteq \mathbb{R}^k$ and $\mathbf{y} \in \operatorname{co} A$, then $\mathbf{y}$ is a convex combination of at most $k+1$ points of A.

[10] That is, each $\lambda_i : E \to [0,1]$ is continuous, $\sum_{i=1}^{p} \lambda_i \equiv 1$ on E, and λ_i vanishes outside $N_{\delta/2}(\mathbf{x}_i)$. It is well defined because the balls cover E, so the denominator is positive.

Theorem 3.4 (Cellina)**.** *Let $E \subseteq \mathbb{R}^m$ be compact, $F \subseteq \mathbb{R}^k$ convex, and $\Gamma : E \to 2^F$ uhc with nonempty convex values. Then for every $\varepsilon > 0$ there exists a continuous $\mathbf{f}_\varepsilon : E \to F$ with $\operatorname{Graph} \mathbf{f}_\varepsilon \subseteq N_\varepsilon(\operatorname{Graph}\Gamma)$ and $\mathbf{f}_\varepsilon(E) \subseteq \operatorname{co}\Gamma(E)$.*

Proof. Fix $\varepsilon > 0$. Since Γ is uhc, for each $\mathbf{x} \in E$ there is $\delta(\mathbf{x}) \in (0, \varepsilon/2)$ such that

$$\Gamma\big(N_{\delta(\mathbf{x})}(\mathbf{x}) \cap E\big) \subseteq B_{\varepsilon/2}\big(\Gamma(\mathbf{x})\big).$$

The balls $N_{\delta(\mathbf{x})/4}(\mathbf{x})$ cover E; extract a finite subcover $N_{\delta_1/4}(\mathbf{x}_1), \dots, N_{\delta_p/4}(\mathbf{x}_p)$ with $\delta_i = \delta(\mathbf{x}_i)$, take a continuous partition of unity $\{\lambda_i\}$ subordinate to it,[11] choose $\mathbf{y}_i \in \Gamma(\mathbf{x}_i)$, and set $\mathbf{f}_\varepsilon(\mathbf{z}) = \sum_{i=1}^p \lambda_i(\mathbf{z})\,\mathbf{y}_i$, continuous with values in $\operatorname{co}\Gamma(E) \subseteq F$.

Fix $\mathbf{z} \in E$ and let $I(\mathbf{z}) = \{i : \lambda_i(\mathbf{z}) > 0\}$, so $\|\mathbf{z} - \mathbf{x}_i\| < \delta_i/4$ for $i \in I(\mathbf{z})$. Choose $j \in I(\mathbf{z})$ with δ_j maximal. For every $i \in I(\mathbf{z})$,

$$\|\mathbf{x}_i - \mathbf{x}_j\| \le \|\mathbf{x}_i - \mathbf{z}\| + \|\mathbf{z} - \mathbf{x}_j\| < \frac{\delta_i}{4} + \frac{\delta_j}{4} \le \frac{\delta_j}{2} < \delta_j,$$

so $\mathbf{x}_i \in N_{\delta_j}(\mathbf{x}_j)$ and hence $\mathbf{y}_i \in \Gamma(\mathbf{x}_i) \subseteq B_{\varepsilon/2}(\Gamma(\mathbf{x}_j))$. The $\varepsilon/2$-neighborhood of the convex set $\Gamma(\mathbf{x}_j)$ is convex, so the convex combination satisfies $\mathbf{f}_\varepsilon(\mathbf{z}) \in B_{\varepsilon/2}(\Gamma(\mathbf{x}_j))$: there is $\mathbf{w} \in \Gamma(\mathbf{x}_j)$ with $\|\mathbf{f}_\varepsilon(\mathbf{z}) - \mathbf{w}\| < \varepsilon/2$. Since also $\|\mathbf{z} - \mathbf{x}_j\| < \delta_j/4 < \varepsilon/2$, the point $(\mathbf{x}_j, \mathbf{w}) \in \operatorname{Graph}\Gamma$ satisfies

$$\big\|(\mathbf{z}, \mathbf{f}_\varepsilon(\mathbf{z})) - (\mathbf{x}_j, \mathbf{w})\big\| = \sqrt{\|\mathbf{z} - \mathbf{x}_j\|^2 + \|\mathbf{f}_\varepsilon(\mathbf{z}) - \mathbf{w}\|^2} < \sqrt{\frac{\varepsilon^2}{4} + \frac{\varepsilon^2}{4}} < \varepsilon,$$

so $(\mathbf{z}, \mathbf{f}_\varepsilon(\mathbf{z})) \in N_\varepsilon(\operatorname{Graph}\Gamma)$. ∎

3.3 Comparison of the two versions

Theorems 3.3 and 3.4 have the same conclusion under different hypotheses.

Theorem 3.3 assumes $\Gamma(\mathbf{x})$ compact for every $\mathbf{x} \in E$, and its proof runs through Lemma 3.2, an argument by contradiction and subsequence extraction. Compactness of the values enters there twice. It makes $\Gamma(E)$ compact by Lemma 3.1(i), which is what allows a convergent subsequence of the $\mathbf{w}_{i,n}$ to be extracted; and it makes $\operatorname{Graph}\Gamma$ closed by Lemma 3.1(ii), which is what puts the limit $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ back in the graph. Neither step survives if the values are unbounded.

Theorem 3.4 is Cellina's 1969 statement and assumes only that Γ is uhc with nonempty convex values. Its proof is a direct estimate from the definition of upper hemicontinuity: there is no subsequence, no contradiction, and no appeal to compactness of $\Gamma(E)$ or closedness of the graph. The values may then be closed and unbounded, or convex and not closed, and the approximation still goes through.

The result goes back to von Neumann's 1937 paper on the expanding economy model [11]. Border states it as "von Neumann's Approximation Lemma" (13.3) and proves it through Lemma 3.2, that is, by the compact-valued route. Cellina rediscovered and generalized it in two 1969 papers. The first, in the *Annali*

[11] Set $\mu_i(\mathbf{z}) = \max\{\delta_i/4 - \|\mathbf{z} - \mathbf{x}_i\|, 0\}$ and $\lambda_i = \mu_i / \sum_{j=1}^p \mu_j$. By construction $\lambda_i(\mathbf{z}) > 0$ if and only if $\|\mathbf{z} - \mathbf{x}_i\| < \delta_i/4$, which is the implication used below.

di Matematica [2], is Theorem 3.4 above and derives Kakutani's theorem as an application; it is the version reproduced as Lemma AIV.1 in Hildenbrand–Kirman [5]. The second, the Lincei note [3], holds over an arbitrary metric domain with values in a metric locally convex space, trading compactness of the domain for total boundedness of the range $R(\Gamma)$ and closedness of the values; as Cellina notes in his introduction, neither theorem contains the other. Incidentally, the Hildenbrand–Kirman footnote credits "Celina (1969)" with a single ell.

3.4 Kakutani's theorem via Cellina

Theorem 3.5 (Kakutani)**.** *Let $X \subset \mathbb{R}^n$ be nonempty, compact, and convex, and let $\Psi : X \to 2^X$ be uhc with nonempty, compact, convex values. Then Ψ has a fixed point: there exists $\hat{\mathbf{x}} \in X$ with $\hat{\mathbf{x}} \in \Psi(\hat{\mathbf{x}})$.*

Proof. Apply Theorem 3.4 with $E = F = X$ and $\varepsilon = 1/n$: there are continuous functions $\mathbf{f}_n : X \to X$ with

$$\operatorname{Graph} \mathbf{f}_n \subseteq B_{1/n}(\operatorname{Graph} \Psi), \qquad n \in \mathbb{N},$$

using here only that Ψ is uhc with convex values, together with $\mathbf{f}_n(X) \subseteq \operatorname{co} \Psi(X) \subseteq X$, which holds by convexity of X. By Brouwer's fixed point theorem (Theorem 1.6), applicable since X is nonempty, compact, and convex, each $\mathbf{f}_n$ has a fixed point $\hat{\mathbf{x}}_n$, and

$$(\hat{\mathbf{x}}_n, \hat{\mathbf{x}}_n) = (\hat{\mathbf{x}}_n, \mathbf{f}_n(\hat{\mathbf{x}}_n)) \in \operatorname{Graph} \mathbf{f}_n \subseteq B_{1/n}(\operatorname{Graph} \Psi).$$

Hence for each n there exists $(\mathbf{x}_n, \mathbf{y}_n) \in \operatorname{Graph} \Psi$ with $\|\hat{\mathbf{x}}_n - \mathbf{x}_n\| < 1/n$ and $\|\hat{\mathbf{x}}_n - \mathbf{y}_n\| < 1/n$. By compactness of X, a subsequence $\hat{\mathbf{x}}_{n_k} \to \hat{\mathbf{x}} \in X$; then $\mathbf{x}_{n_k} \to \hat{\mathbf{x}}$ and $\mathbf{y}_{n_k} \to \hat{\mathbf{x}}$ as well. Since Ψ is uhc with closed values, $\operatorname{Graph} \Psi$ is closed by Lemma 3.1(ii), so $(\hat{\mathbf{x}}, \hat{\mathbf{x}}) \in \operatorname{Graph} \Psi$, that is, $\hat{\mathbf{x}} \in \Psi(\hat{\mathbf{x}})$. □

Remark 3.6. With X compact, the following three hypotheses on $\Psi : X \to 2^X$ with nonempty values are equivalent: (a) $\operatorname{Graph} \Psi$ is closed; (b) Ψ is uhc with closed values; (c) Ψ is uhc with compact values. Kakutani's original statement [6] and Hildenbrand–Kirman use (a). They are the same theorem.

The approximation step needs no compactness of the values, but Kakutani's theorem assumes it. The hypothesis is used at one point only: it makes $\operatorname{Graph} \Psi$ closed, and so puts the limit of the approximate fixed points back in the graph. The rest is carried by convexity, of the values in Theorem 3.4 and of X in Brouwer's theorem and in $\mathbf{f}_n(X) \subseteq \operatorname{co} \Psi(X) \subseteq X$, and by compactness of X, which yields the finite subcover in the approximation and the convergent subsequence of fixed points.

References

[1] K. C. Border, *Fixed Point Theorems with Applications to Economics and Game Theory*, Cambridge University Press, 1985.

[2] A. Cellina, *Approximation of set valued functions and fixed point theorems*, Annali di Matematica Pura ed Applicata **82** (1969), 17–24.

[3] A. Cellina, *A theorem on the approximation of compact multi-valued mappings*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. **47** (1969), 429–433.

[4] A. de la Fuente, *Mathematical Methods and Models for Economists*, Cambridge University Press, 2000.

[5] W. Hildenbrand and A. P. Kirman, *Introduction to Equilibrium Analysis*, North-Holland, 1976, Mathematical Appendix IV.

[6] S. Kakutani, *A generalization of Brouwer's fixed point theorem*, Duke Mathematical Journal **8** (1941), 457–459.

[7] M. Laczkovich and V. T. Sós, *Real Analysis: Series, Functions of Several Variables, and Applications*, Springer, 2017.

[8] J. Milnor, *Analytic proofs of the hairy ball theorem and the Brouwer fixed point theorem*, American Mathematical Monthly **85** (1978), 521–524.

[9] C. A. Rogers, *A less strange version of Milnor's proof of Brouwer's fixed-point theorem*, American Mathematical Monthly **87** (1980), 525–527.

[10] P. V. Subrahmanyam, *Elementary Fixed Point Theorems*, Forum for Interdisciplinary Mathematics, Springer, Singapore, 2018.

[11] J. von Neumann, *Über ein ökonomisches Gleichungssystem und eine Verallgemeinerung des Brouwerschen Fixpunktsatzes*, Ergebnisse eines Mathematischen Kolloquiums **8** (1937), 73–83.