Weierstraß proved that polynomials are dense in $C^0(\mathbb{R}^n)$. Decades later, Stone greatly abstracted this. Prior to Stone, S. Bernstein gave a memorable argument for Weierstraß’ concrete case, with the additional virtue of suggesting similarly intuitive arguments for various function spaces on topological spaces with transitive group actions.

[0.0.1] Theorem: Given $f \in C^0(U)$ for $U$ open in $\mathbb{R}^n$, for compact $C \subset U$ and $\varepsilon > 0$, there is a polynomial $P$ such that

$$\sup_{x \in C} |f(x) - P(x)| < \varepsilon$$

Proof: The idea is to create a sequence $P_\ell$ of polynomials on $\mathbb{R}^n$ whose restrictions $\varphi_\ell$ to a fixed compact, such as the cube $C = \{x = (x_1, \ldots, x_n) : |x_i| \leq 1, \text{ for all } i\}$ form an approximate identity, in the sense that their masses bunch up at 0.

More precisely, we want the restrictions $\varphi_\ell$ to $C$ to be non-negative, to have integrals 1 on the smaller cube $\frac{1}{2}C$, and to satisfy

$$\lim_{\ell \to \infty} \frac{\int_{\delta C} \varphi_\ell}{\int_{\frac{1}{2}C} \varphi_\ell} = 1 \quad (\text{for each fixed positive } \delta \leq \frac{1}{2})$$

Granting all that, we can show that the mollifications of $f$ by the $\varphi_\ell$ approach $f$ in the sup norm:

$$\lim_{\ell \to \infty} \sup_{x \in \frac{1}{2}C} \left| f(x) - \int_{\frac{1}{2}C} f(x + y) \varphi_\ell(y) \, dy \right| = 0 \quad (\text{as } \ell \to \infty)$$

Indeed,

$$\int_{\frac{1}{2}C} f(x + y) \varphi_\ell(y) \, dy = \int_{\frac{1}{2}C} f(x + y) \varphi_\ell(y) \, dy + \int_{\frac{1}{2}C - \delta C} f(x + y) \varphi_\ell(y) \, dy = f(x) \int_{\frac{1}{2}C} \varphi_\ell(y) \, dy + \int_{\frac{1}{2}C - \delta C} (f(x + y) - f(x)) \varphi_\ell(y) \, dy + \int_{\frac{1}{2}C - \delta C} f(x + y) \varphi_\ell(y) \, dy$$

The first integral goes to $1$ as $\ell \to \infty$, for fixed $\delta > 0$, by the bunching-up property. The second integral goes to 0 uniformly in $x$ as $\delta \to 0$, by the uniform continuity of $f$ on $C$. The third integral goes to 0 as $\ell \to \infty$, since the masses of the $\varphi_\ell$ bunch up inside $\delta C$. Thus, assuming we have such polynomials,

$$\lim_{\ell \to \infty} \int_{\frac{1}{2}C} f(x + y) \varphi_\ell(y) \, dy = f(x) \quad (\text{uniformly in } x \in \frac{1}{2}C)$$

At the same time,

$$\int_{\frac{1}{2}C} f(x + y) \varphi_\ell(y) \, dy = \int_{x + \frac{1}{2}C} f(y) \varphi_\ell(-x + y) \, dy$$

is a superposition of polynomials of degrees at most that of $\varphi_\ell$. The space $V$ of such polynomials is finite-dimensional. Thus, this integral of a compactly-supported continuous $V$-valued function lies in $V$. That is, this integral is equal to a polynomial, as a function. This would prove the theorem.
Paul Garrett: S. Bernstein’s proof of Weierstraß’ approximation theorem (February 28, 2011)

To make suitable polynomials \( P_\ell \), it suffices to treat the single-variable case. Let

\[ P_\ell(x) = (1 - x^2)^\ell \quad \text{(for } x \in \mathbb{R}) \]

First, determine where the second derivative vanishes: solve

\[
0 = \frac{d}{dx} \left(-2\ell x(1 - x^2)^{\ell - 1}\right) = 4\ell(\ell - 1)x^2(1 - x^2)^{\ell - 2} - 2\ell(1 - x^2)^{\ell - 1}
\]

\[
= 2\ell \cdot \left((\ell - 1)x^2 - (1 - x^2)\right) \cdot (1 - x^2)^{\ell - 2}
\]

Thus, in the interior of \([-1, 1]\), the second derivative vanishes at \( \pm 1/\sqrt{\ell} \), so the curve bends downward in \([-1/\sqrt{\ell}, 1/\sqrt{\ell}]\), and bends upward outside that interval. In particular, the line segments from the points \((\pm 1/\sqrt{\ell}, 0)\) to \((0, 1)\) are below the graph of \( P_\ell \), so

\[
\int_{|x| \leq 1/\sqrt{\ell}} P_\ell(x) \, dx \geq \frac{2}{\sqrt{\ell}}
\]

On the other hand, the standard fact that

\[
\lim_{\ell \to \infty} (1 - x/\ell)^\ell = e^{-x}
\]

suggests a certain approach. For example,

\[
P_\ell \left( \frac{\sqrt{\log \ell}}{\sqrt{\ell}} \right) = \left(1 - \frac{\log \ell}{\ell}\right)^\ell
\]

Since \( \log(1 - x) \leq -x \) for \( x \geq 0 \),

\[
\log P_\ell \left( \frac{\sqrt{\log \ell}}{\sqrt{\ell}} \right) \leq -\ell
\]

Thus,

\[
P_\ell \left( \frac{\sqrt{\log \ell}}{\sqrt{\ell}} \right) \leq \frac{1}{\ell}
\]

Thus we have a sufficient bunching-up result: obviously \( 1/\sqrt{\ell} < \frac{\log \ell}{\sqrt{\ell}} \), so

\[
\int_{|x| < \frac{\log \ell}{\sqrt{\ell}}} P_\ell(x) \, dx \geq \frac{2}{\sqrt{\ell}}
\]

while

\[
\int_{\frac{\log \ell}{\sqrt{\ell}} < |x| < 1} P_\ell(x) \, dx \leq \frac{2}{\ell}
\]

That is, letting

\[
\varphi_\ell(x) = \frac{P_\ell(x)}{\int_{|x| \leq 1/2} P_\ell(x) \, dx}
\]

gives the single-variable approximate identity desired. The product

\[
\varphi_\ell(x_1) \ldots \varphi_\ell(x_n)
\]

is the desired collection for \( \mathbb{R}^n \).

///

[0.0.2] Remark: Although it is unnecessary for the above, it is interesting to determine the integral of the single-variable \( P_\ell \) over \([-1, 1]\). Integrating by parts repeatedly, it is

\[
\int_{-1}^{1} (1 - x)^\ell \cdot (1 + x)^\ell \, dx = \frac{\ell}{\ell + 1} \int_{-1}^{1} (1 - x)^{\ell - 1} \cdot (1 + x)^{\ell + 1} \, dx = \frac{\ell(\ell - 1)}{(\ell + 1)(\ell + 2)} \int_{-1}^{1} (1 - x)^{\ell - 2} \cdot (1 + x)^{\ell + 2} \, dx
\]

\[
= \ldots = \frac{\ell! \ell!}{(2\ell)!} \int_{-1}^{1} (1 + x)^{2\ell} \, dx = \frac{\ell! \ell!}{(2\ell)!} \frac{2^{2\ell + 1}}{2\ell + 1}
\]

2