(1) Let

$$g(x) = \begin{cases} 2x, x < \frac{1}{2} \\ 2 - 2x, x \geq \frac{1}{2} \end{cases}$$

We want to show

$$g_m(x) = \begin{cases} 2^m(x - \frac{2k}{2^m}), x \in [\frac{2k}{2^m}, \frac{2k+1}{2^m}) \\ 2^m(\frac{2k+2}{2^m} - x), x \in [\frac{2k+1}{2^m}, \frac{2k+2}{2^m}] \end{cases}$$

for $k = 0, ...2^{m-1} - 1$ where g_m denotes "g composes with itself m times."

We shall do it by induction, the base case is clear by plugging in $m = 1$ and $k = 0$ to the g_m formula. For the induction step, suppose the g_m formula is as given. Then $g_{m+1} = g \circ g_m$, which means

$$g_{m+1}(x) = \begin{cases} 2g_m(x), g_m(x) < \frac{1}{2} \\ 2 - 2g_m(x), g_m(x) \geq \frac{1}{2} \end{cases}$$

Consider the interval $[\frac{2k}{2^m}, \frac{2k+2}{2^m}]$, we split it into 4 cases,
$x \in [\frac{2k}{2^m}, \frac{2k+1/2}{2^m})$, $g_m(x) < 1/2$, therefore $g_{m+1}(x) = 2g_m(x) = 2^{m+1}(x - \frac{2(2k)}{2^{m+1}})$
$x \in [\frac{2k}{2^m+1/2}, \frac{2k+1}{2^m}]$, $g_m(x) \geq 1/2$, therefore $g_{m+1}(x) = 2 - 2g_m(x) = 2 - 2^{m+1}(x - \frac{2k}{2^m}) = 2^{m+1}(\frac{2(2k)+2}{2^{m+1}} - x)$
$x \in [\frac{2k}{2^m+1}, \frac{2k+3/2}{2^m}]$, $g_m(x) \geq 1/2$, therefore $g_{m+1}(x) = 2 - 2g_m(x) = 2 - 2^{m+1}(\frac{2k+2}{2^m} - x) = 2^{m+1}(x - \frac{2(2k+1)}{2^{m+1}})$
$x \in (\frac{2k}{2^m+3/2}, \frac{2k+2}{2^m}]$, $g_m(x) < 1/2$, therefore $g_{m+1}(x) = 2g_m(x) = 2^{m+1}(\frac{2(2k+1)+2}{2^{m+1}} - x)$

Now combine the formula in 1st and 3rd case, 2nd and 4th case. Redefine k (in first two cases $k' = 2k$ last two cases $k' = 2k + 1$) we have

$$g_{m+1}(x) = \begin{cases} 2^{m+1}(x - \frac{2k'}{2^{m+1}}), x \in [\frac{2k'}{2^{m+1}}, \frac{2k'+1}{2^{m+1}}) \\ 2^{m+1}(\frac{2k'+2}{2^{m+1}} - x), x \in [\frac{2k'+1}{2^{m+1}}, \frac{2k'+2}{2^{m+1}}] \end{cases}$$

for $k' = 0, 1, ...2^m - 1$. This completes the induction step and the proof.

In order to approximate $f(x) = x^3$ with Span$\{x, g, g_2, ...g_M\}$, we first define $h_m(x) = \frac{1}{2^m}\frac{dg_m(x)}{dx}$. In other words,

$$h_m(x) = \begin{cases} 1, x \in [\frac{2k}{2^m}, \frac{2k+1}{2^m}) \\ -1, x \in [\frac{2k+1}{2^m}, \frac{2k+2}{2^m}] \end{cases}$$

for $k = 0, ...2^{m-1} - 1$.
It's easy to verify $\{1, h, h_2, ...\}$ is an orthonormal basis of $L^2[0, 1]$. Apply the formula for orthonormal projection we have,

$$3x^2 = \sum_{m=0}^{\infty} h_m(x) < h_m(x), 3x^2 >$$

where $<, >$ is the L^2 inner product on $[0, 1]$. Integrate both sides from 0 to x we get

$$x^3 = \sum_{m=0}^{\infty} g_m(x)\frac{< h_m(x), 3x^2 >}{2^m}$$

.

It's left to compute the coefficient,

(1)
$$
\begin{aligned}
&< h_m(x), 3x^2 > \\
&= \int_0^1 3x^2 h_m(x)dx \\
&= \int_0^{1/2^m} 3x^2dx - \int_{1/2^m}^{2/2^m} 3x^2dx + ... - \int_{(2^m-1)/2^m}^{1} 3x^2dx \\
&= \frac{1}{2^{3m}}[(1^3-0^3)-(2^3-1^3)+...-((2^m)^3-(2^m-1)^3)]
\end{aligned}
$$

To simplify further, we need some summation formulas, starting at

$$1^3+2^3+...+N^3 = \frac{N^2(N+1)^2}{4}$$

we can get

$$1^3+2^3+...+(2N)^3 = \frac{(2N)^2(2N+1)^2}{4}$$

and

$$2^3+4^3+...+(2N)^3 = \frac{8N^2(N+1)^2}{4} = 2N^2(N+1)^2$$

, subtract the last two we get

$$1^3+3^3+...+(2N-1)^3 = \frac{4N^2(2N^2-1)}{4} = N^2(2N^2-1)$$

, therefore

(2)
$$
\begin{aligned}
&(1^3-0^3)-(2^3-1^3)+...-((2N)^3-(2N-1)^3) \\
&= 2(1^3+3^3+...+(2N-1)^3) - 2(2^3+4^3+...+(2N)^3) + (2N)^3 \\
&= 2N^2(2N^2-1) - 4N^2(N+1)^2 + (2N)^3 \\
&= -6N^2
\end{aligned}
$$

Plug in $N = 2^{m-1}$, we have $< h_m(x), 3x^2 >= \frac{-6\cdot 2^{2m-2}}{2^{3m}} = \frac{-1.5}{2^m}$ for $m \geq 1$. By direct computation we have $< h_0(x), 3x^2 >=< 1, 3x^2 >= 1$.

Therefore,

$$x^3 = x + \sum_{m=1}^{\infty} \frac{-3}{2^{2m+1}} \cdot g_m(x)$$

.

We can estimate the error term by triangle inequality

$$|e_M(x)| = |\sum_{M+1}^{\infty} \frac{-3}{2^{2m+1}} \cdot g_m(x)| \leq |\sum_{M+1}^{\infty} \frac{-3}{2^{m+1}}| = \frac{3}{2^{M+1}}$$

, and the plot of $f(x)$ vs. approximation is in a separate mathematica notebook.

(2) Define $f_j(x) = ReLU(w_jx+b_j)$ for $x \in \mathbb{R}^{d\times 1}$. We want to show $\{f_j(x)\}_{j=1}^J$ is linearly independent.

Define $Dis_1(f_j) \subset \mathbb{R}^d$ be the set of points where ∇f_j is discontinuous. For the f_j we define, $Dis_1(f_j) = \{x \in \mathbb{R}^d | w_jx+b_j = 0\}$. Notice for a nonzero constant c, we have $Dis_1(f_j) = Dis_1(cf_j)$.

Let $c_j \in \mathbb{R}$ for $1 \leq j \leq J$, suppose we have

$$c_1f_1(x) + ... + c_Jf_J(x) \equiv 0$$

then we must have

$$Dis_1(c_1f_1(x) + ... + c_Jf_J(x)) = Dis_1(0) = \emptyset$$

However, the set of vectors $\{(w_j, b_j)\}_{j=1}^J$ are pairwise independent. That means the hyper-plane $w_j x + b_j = 0$ are all different. In other words $Dis_1(f_j) \neq Dis_1(f_i)$ if $i \neq j$. It's also clear that $Dis_1(f_j) - (\cup_{i \neq j} Dis_1(f_i)) \neq \emptyset$, that is, for each j we can find points in $Dis_1(f_j)$ that is not in every other $Dis_1(f_i)$.

For each j, we find a point $x_j \in Dis_1(f_j)$ that is not in every other $Dis_1(f_i)$. If $c_j \neq 0$ then $\nabla(c_1 f_1(x) + ... + c_J f_J(x))$ is also discontinuous at x_j, that contradict with $Dis_1(c_1 f_1(x) + ... + c_J f_J(x)) = \emptyset$. Therefore c_j must be 0. Since j is arbitrary, we have $c_j = 0$ for $1 \leq j \leq J$. That means the set of functions $\{f_j(x)\}_{j=1}^J$ is linearly independent.