

# The Contraction Mapping Theorem

**Theorem (The Contraction Mapping Theorem)** Let  $B_a = \{ \vec{x} \in \mathbb{R}^d \mid \|\vec{x}\| < a \}$  denote the open ball of radius  $a$  centred on the origin in  $\mathbb{R}^d$ . If the function

$$\vec{g}: B_a \rightarrow \mathbb{R}^d$$

obeys

(H1) there is a constant  $G < 1$  such that  $\|\vec{g}(\vec{x}) - \vec{g}(\vec{y})\| \leq G \|\vec{x} - \vec{y}\|$  for all  $\vec{x}, \vec{y} \in B_a$

(H2)  $\|\vec{g}(\vec{0})\| < (1 - G)a$

then the equation

$$\vec{x} = \vec{g}(\vec{x})$$

has exactly one solution.

**Discussion of hypothesis (H1):** Hypothesis (H1) is responsible for the word ‘‘Contraction’’ in the name of the theorem. Because  $G < 1$  (and it is crucial that  $G < 1$ ) the distance between the images  $\vec{g}(\vec{x})$  and  $\vec{g}(\vec{y})$  of  $\vec{x}$  and  $\vec{y}$  is smaller than the original distance between  $\vec{x}$  and  $\vec{y}$ . Thus the function  $g$  contracts distances. Note that, when the dimension  $d = 1$ ,

$$|g(x) - g(y)| = \left| \int_x^y g'(t) dt \right| \leq \left| \int_x^y |g'(t)| dt \right| \leq \left| \int_x^y \sup_{t' \in B_a} |g'(t')| dt \right| = |x - y| \sup_{t' \in B_a} |g'(t')|$$

For a once continuously differentiable function, the smallest  $G$  that one can pick and still have  $|g(x) - g(y)| \leq G|x - y|$  for all  $x, y$  is  $G = \sup_{t' \in B_a} |g'(t')|$ . In this case (H1) comes down to the requirement that there exist a constant  $G < 1$  such that  $|g'(t)| \leq G < 1$  for all  $t' \in B_a$ . For dimensions  $d > 1$ , one has a whole matrix  $\mathcal{G}(\vec{x}) = \left[ \frac{\partial g_i}{\partial x_j}(\vec{x}) \right]_{1 \leq i, j \leq d}$  of first derivatives. There is a measure of the size of this matrix, called the norm of the matrix and denoted  $\|\mathcal{G}(\vec{x})\|$  such that

$$\|\vec{g}(\vec{x}) - \vec{g}(\vec{y})\| \leq \|\vec{x} - \vec{y}\| \sup_{\vec{t} \in B_a} \|\mathcal{G}(\vec{t})\|$$

Once again (H1) comes down to  $\|\mathcal{G}(\vec{t})\| \leq G < 1$  for all  $\vec{t} \in B_a$ . Roughly speaking, (H1) forces the derivative of  $\vec{g}$  to be sufficiently small, which forces the derivative of  $\vec{x} - \vec{g}(\vec{x})$  to be bounded away from zero.

If we were to relax (H1) to  $G \leq 1$ , the theorem would fail. For example,  $g(x) = x$  obeys  $|g(x) - g(y)| = |x - y|$  for all  $x$  and  $y$ . So  $G$  would be one in this case. But every  $x$  obeys  $g(x) = x$ , so the solution is certainly not unique.

**Discussion of hypothesis (H2):** If  $\vec{g}$  only takes values that are outside of  $B_a$ , then  $\vec{x} = \vec{g}(\vec{x})$  cannot possibly have any solutions. So there has to be a requirement that  $\vec{g}(\vec{x})$  lies in  $B_a$  for at least some values of  $\vec{x} \in B_a$ . Our hypotheses are actually somewhat stronger than this:

$$\|\vec{g}(\vec{x})\| = \|\vec{g}(\vec{x}) - \vec{g}(\vec{0}) + \vec{g}(\vec{0})\| \leq \|\vec{g}(\vec{x}) - \vec{g}(\vec{0})\| + \|\vec{g}(\vec{0})\| \leq G\|\vec{x} - \vec{0}\| + (1 - G)a$$

by (H1) and (H2). So, for all  $\vec{x}$  in  $B_a$ , that is, all  $\vec{x}$  with  $\|\vec{x}\| < a$ ,  $\|\vec{g}(\vec{x})\| < Ga + (1 - G)a = a$ . With our hypotheses  $\vec{g} : B_a \rightarrow B_a$ . Roughly speaking, (H2) requires that  $\vec{g}(\vec{x})$  be sufficiently small for at least one  $\vec{x}$ .

If we were to relax (H2) to  $\|\vec{g}(\vec{0})\| \leq (1 - G)a$ , the theorem would fail. For example, pick any  $a > 0$ ,  $0 < G < 1$  and define  $g : B_a \rightarrow \mathbb{R}$  by  $g(x) = a(1 - G) + Gx$ . Then  $g'(x) = G$  for all  $x$  and  $g(0) = a(1 - G)$ . For this  $g$ ,

$$g(x) = x \iff a(1 - G) + Gx = x \iff a(1 - G) = (1 - G)x \iff x = a$$

As  $x = a$  is not in the domain of definition of  $g$ , there is no solution.

**Proof that there is at most one solution:** Suppose that  $\vec{x}^*$  and  $\vec{y}^*$  are two solutions. Then

$$\begin{aligned} \vec{x}^* = \vec{g}(\vec{x}^*), \vec{y}^* = \vec{g}(\vec{y}^*) &\implies \|\vec{x}^* - \vec{y}^*\| = \|\vec{g}(\vec{x}^*) - \vec{g}(\vec{y}^*)\| \\ &\stackrel{(H1)}{\implies} \|\vec{x}^* - \vec{y}^*\| \leq G\|\vec{x}^* - \vec{y}^*\| \\ &\implies (1 - G)\|\vec{x}^* - \vec{y}^*\| = 0 \end{aligned}$$

As  $G < 1$ ,  $1 - G$  is nonzero and  $\|\vec{x}^* - \vec{y}^*\|$  must be zero. That is,  $\vec{x}^*$  and  $\vec{y}^*$  must be the same.

**Proof that there is at least one solution:** Set

$$\vec{x}_0 = 0 \quad \vec{x}_1 = \vec{g}(\vec{x}_0) \quad \vec{x}_2 = \vec{g}(\vec{x}_1) \quad \cdots \quad \vec{x}_n = \vec{g}(\vec{x}_{n-1}) \quad \cdots$$

We showed in ‘‘Significance of hypothesis (H2)’’ that  $\vec{g}(\vec{x})$  is in  $B_a$  for all  $\vec{x}$  in  $B_a$ . So  $\vec{x}_0, \vec{x}_1, \vec{x}_2, \dots$  are all in  $B_a$ . So the definition  $\vec{x}_n = \vec{g}(\vec{x}_{n-1})$  is legitimate. We shall show that the sequence  $\vec{x}_0, \vec{x}_1, \vec{x}_2, \dots$  converges to some vector  $\vec{x}^*$ . Since  $\vec{g}$  is continuous, this vector will obey

$$\vec{x}^* = \lim_{n \rightarrow \infty} \vec{x}_n = \lim_{n \rightarrow \infty} \vec{g}(\vec{x}_{n-1}) = \vec{g}\left(\lim_{n \rightarrow \infty} \vec{x}_{n-1}\right) = \vec{g}(\vec{x}^*)$$

In other words,  $\vec{x}^*$  is a solution of  $\vec{x} = \vec{g}(\vec{x})$ .

To prove that the sequence converges, we first observe that, applying (H1) numerous times,

$$\begin{aligned} \|\vec{x}_m - \vec{x}_{m-1}\| &= \|\vec{g}(\vec{x}_{m-1}) - \vec{g}(\vec{x}_{m-2})\| \\ &\leq G\|\vec{x}_{m-1} - \vec{x}_{m-2}\| = G\|\vec{g}(\vec{x}_{m-2}) - \vec{g}(\vec{x}_{m-3})\| \\ &\leq G^2\|\vec{x}_{m-2} - \vec{x}_{m-3}\| = G^2\|\vec{g}(\vec{x}_{m-3}) - \vec{g}(\vec{x}_{m-4})\| \\ &\vdots \\ &\leq G^{m-1}\|\vec{x}_1 - \vec{x}_0\| = G^{m-1}\|\vec{g}(\vec{0})\| \end{aligned}$$

Remember that  $G < 1$ . So the distance  $\|\vec{x}_m - \vec{x}_{m-1}\|$  between the  $(m-1)^{\text{st}}$  and  $m^{\text{th}}$  entries in the sequence gets really small for  $m$  as large. As

$$\vec{x}_n = \vec{x}_0 + (\vec{x}_1 - \vec{x}_0) + (\vec{x}_2 - \vec{x}_1) + \cdots + (\vec{x}_n - \vec{x}_{n-1}) = \sum_{m=1}^n (\vec{x}_m - \vec{x}_{m-1})$$

(recall that  $\vec{x}_0 = \vec{0}$ ) it suffices to prove that  $\sum_{m=1}^n (\vec{x}_m - \vec{x}_{m-1})$  converges as  $n \rightarrow \infty$ . To do so it suffices to prove that  $\sum_{m=1}^n \|\vec{x}_m - \vec{x}_{m-1}\|$  converges as  $n \rightarrow \infty$ , which we do now.

$$\sum_{m=1}^n \|\vec{x}_m - \vec{x}_{m-1}\| \leq \sum_{m=1}^n G^{m-1} \|\vec{g}(\vec{0})\| = \frac{1 - G^n}{1 - G} \|\vec{g}(\vec{0})\|$$

As  $n$  tends to  $\infty$ ,  $G^n$  converges to zero (because  $0 \leq G < 1$ ) and  $\frac{1 - G^n}{1 - G} \|\vec{g}(\vec{0})\|$  converges to  $\frac{1}{1 - G} \|\vec{g}(\vec{0})\|$ . ■

Now consider some function  $\vec{f}(\vec{x}, \vec{y})$  with  $\vec{x}$  running over  $\mathbb{R}^d$ ,  $\vec{y}$  running over  $\mathbb{R}^n$  and  $\vec{f}$  taking values in  $\mathbb{R}^d$ . Suppose that we have one point  $(\vec{x}_0, \vec{y}_0)$  on the surface  $\vec{f}(\vec{x}, \vec{y}) = 0$ . In other words, suppose that  $\vec{f}(\vec{x}_0, \vec{y}_0) = 0$ . And suppose that we wish to solve  $\vec{f}(\vec{x}, \vec{y}) = 0$  for  $\vec{x} = \vec{x}(\vec{y})$  near  $(\vec{x}_0, \vec{y}_0)$ . First observe that for each fixed  $\vec{y}$ ,  $\vec{f}(\vec{x}, \vec{y}) = 0$  is a system of  $d$  equations in  $d$  unknowns. So at least the number of unknowns matches the number of equations. Denote by  $A$  the  $d \times d$  matrix  $\left[ \frac{\partial f_i}{\partial x_j}(\vec{x}_0, \vec{y}_0) \right]_{1 \leq i, j \leq d}$  of first partial  $\vec{x}$  derivatives at  $(\vec{x}_0, \vec{y}_0)$ . Suppose that this matrix has an inverse. When  $d = 1$ , the invertibility of  $A$  just means that  $\frac{\partial f}{\partial x}(\vec{x}_0, \vec{y}_0) \neq 0$ . For  $d > 1$ , it just means that 0 is not an eigenvalue of  $A$ . Then

$$\vec{f}(\vec{x}, \vec{y}) = 0 \iff A^{-1} \vec{f}(\vec{x}, \vec{y}) = 0 \iff \vec{x} - \vec{x}_0 - A^{-1} \vec{f}(\vec{x}, \vec{y}) = \vec{x} - \vec{x}_0$$

Rename  $\vec{x} - \vec{x}_0 = \vec{z}$  and define  $\vec{g}(\vec{z}, \vec{y}) = \vec{z} - A^{-1} \vec{f}(\vec{z} + \vec{x}_0, \vec{y})$ . Then

$$\vec{f}(\vec{x}, \vec{y}) = 0 \iff \vec{x} = \vec{x}_0 + \vec{z} \text{ and } \vec{g}(\vec{z}, \vec{y}) = \vec{z}$$

Now apply the Contraction Mapping Theorem with  $\vec{y}$  viewed as a parameter. That is, fix any  $\vec{y}$  sufficiently near  $\vec{y}_0$ . Then  $\vec{g}(\vec{z}, \vec{y})$  is a function of  $\vec{z}$  only and one may apply the Contraction Mapping Theorem to it.

We must of course check that the hypotheses are satisfied. Observe first, that when  $\vec{z} = \vec{0}$  and  $\vec{y} = \vec{y}_0$ , the matrix  $\left[ \frac{\partial g_i}{\partial z_j}(\vec{0}, \vec{y}_0) \right]_{1 \leq i, j \leq d}$  of first derivatives of  $\vec{g}$  is exactly  $\mathbb{1} - A^{-1}A$ . The identity matrix  $\mathbb{1}$  arises from differentiating the term  $\vec{z}$  of  $\vec{g}$  and  $-A^{-1}A$  arises from differentiating  $-A^{-1} \vec{f}(\vec{z} + \vec{x}_0, \vec{y})$ . So  $\left[ \frac{\partial g_i}{\partial z_j}(\vec{0}, \vec{y}_0) \right]_{1 \leq i, j \leq d}$  is exactly the zero matrix. For  $(\vec{z}, \vec{y})$  sufficiently close to  $(\vec{0}, \vec{y}_0)$ , the matrix  $\left[ \frac{\partial g_i}{\partial z_j}(\vec{z}, \vec{y}) \right]_{1 \leq i, j \leq d}$  will, by continuity, be small enough that (H1) is satisfied. Also observe that  $\vec{g}(\vec{0}, \vec{y}_0) = -A^{-1} \vec{f}(\vec{x}_0, \vec{y}_0) = \vec{0}$ . So, once again, by continuity, if  $\vec{y}$  is sufficiently close to  $\vec{y}_0$ ,  $\vec{g}(\vec{0}, \vec{y})$  will be small enough that (H2) is satisfied.

We conclude from the Contraction Mapping Theorem that, assuming  $A$  is invertible,  $\vec{f}(\vec{x}, \vec{y}) = 0$  has exactly one solution,  $\vec{x}(\vec{y})$ , near  $\vec{x}_0$  for each  $\vec{y}$  sufficiently near  $\vec{y}_0$ .