Supplement 1: Manifolds

Mladen Victor Wickerhauser

PMF 2022: Dimensionality Reduction and Manifold Estimation

January 10, 2022

1 Inverse Function Theorem proof.

1.0.1 K is a contraction map.

For each $\mathbf{y}' \in \mathbf{E}^d$, define a function $K : \mathbf{E}^d \to \mathbf{E}^d$ by

$$K(\mathbf{u}) \stackrel{\text{def}}{=} \mathbf{u} - D\mathbf{f}(\mathbf{x})^{-1}[\mathbf{f}(\mathbf{u}) - \mathbf{y}'],$$

relying on the hypothesis that $D\mathbf{f}(\mathbf{x})$ is nonsingular, hence invertible.

Use K to define a sequence $\{\mathbf{x}_n\}$ recursively:

$$\mathbf{x}_0 \stackrel{\text{def}}{=} \mathbf{x}; \quad \mathbf{x}_{n+1} = K(\mathbf{x}_n) = \mathbf{x}_n - D\mathbf{f}(\mathbf{x})^{-1}[\mathbf{f}(\mathbf{x}_n) - \mathbf{y}'], \qquad n = 0, 1, 2, \ldots.$$

From the definition of differentiability for functions $\mathbf{f} : \mathbf{E}^n \to \mathbf{E}^m$,

$$\mathbf{f}(\mathbf{u}) = \mathbf{f}(\mathbf{v}) + D\mathbf{f}(\mathbf{v})(\mathbf{u} - \mathbf{v}) + E(\mathbf{u}, \mathbf{v}),$$

where the error satisfies $\|E(\mathbf{u}, \mathbf{v})\| = o(\|\mathbf{u} - \mathbf{v}\|)$ as $\mathbf{u} \to \mathbf{v}$. Substitute this estimate into $K(\mathbf{u})$ and $K(\mathbf{v})$ to show that K is a contraction map near $\mathbf{x}$:

$$\begin{aligned} K(\mathbf{u}) - K(\mathbf{v}) &= \mathbf{u} - \mathbf{v} - D\mathbf{f}(\mathbf{x})^{-1}[\mathbf{f}(\mathbf{u}) - \mathbf{f}(\mathbf{v})] \\ &= \mathbf{u} - \mathbf{v} - D\mathbf{f}(\mathbf{x})^{-1}\Big[D\mathbf{f}(\mathbf{v})(\mathbf{u} - \mathbf{v}) + E(\mathbf{u}, \mathbf{v})\Big] \\ &= \Big[\mathrm{I} - D\mathbf{f}(\mathbf{x})^{-1}D\mathbf{f}(\mathbf{v})\Big](\mathbf{u} - \mathbf{v}) + D\mathbf{f}(\mathbf{x})^{-1}E(\mathbf{u}, \mathbf{v}). \end{aligned}$$

Now $\mathbf{f}$ is continuously differentiable near $\mathbf{x}$, so $D\mathbf{f}(\mathbf{v}) \to D\mathbf{f}(\mathbf{x})$ as $\mathbf{v} \to \mathbf{x}$, so $\mathrm{I} - D\mathbf{f}(\mathbf{x})^{-1}D\mathbf{f}(\mathbf{v}) \to 0$. (Every coefficient of the matrix tends to 0.) Thus

$$(\exists \delta_1 > 0) \quad \mathbf{v} \in B(\mathbf{x}, \delta_1) \implies \left\|\mathrm{I} - D\mathbf{f}(\mathbf{x})^{-1}D\mathbf{f}(\mathbf{v})\right\|_{\text{op}} < \frac{1}{4} \tag{1}$$

Here $\|\cdot\|_{\text{op}}$ is the "operator norm." For $d \times d$ matrix A, $\|A\|_{\text{op}} \stackrel{\text{def}}{=} \sup\{\|A\mathbf{x}\|/\|\mathbf{x}\| : \mathbf{x} \in \mathbf{E}^d,\ \mathbf{x} \neq \mathbf{0}\}$. It is a continuous function of the coefficients of A and thus $\|A\|_{\text{op}} \to 0$ as $A \to 0$. It satisfies $\|A\mathbf{x}\| \leq \|A\|_{\text{op}}\|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbf{E}^d$.

Likewise, since $D\mathbf{f}(\mathbf{x})^{-1}$ is bounded and $\|E(\mathbf{u},\mathbf{v})\| = o(\|\mathbf{u}-\mathbf{v}\|)$,

$$(\exists \delta_2 > 0) \quad \mathbf{u},\mathbf{v} \in B(\mathbf{x},\delta_2) \implies \|D\mathbf{f}(\mathbf{x})^{-1}E(\mathbf{u},\mathbf{v})\| \le \frac{1}{4}\|\mathbf{u}-\mathbf{v}\|. \qquad (2)$$

Put $\delta = \min\{\delta_1, \delta_2\}$ to conclude that if $\mathbf{u},\mathbf{v} \in B(\mathbf{x},\delta)$, then

$$\begin{aligned} \|K(\mathbf{u}) - K(\mathbf{v})\| &= \left\| \left[\mathrm{I} - D\mathbf{f}(\mathbf{x})^{-1}D\mathbf{f}(\mathbf{v})\right](\mathbf{u}-\mathbf{v}) + D\mathbf{f}(\mathbf{x})^{-1}E(\mathbf{u},\mathbf{v}) \right\| \\ &\le \left\|\mathrm{I} - D\mathbf{f}(\mathbf{x})^{-1}D\mathbf{f}(\mathbf{v})\right\| \|\mathbf{u}-\mathbf{v}\| + \left\|D\mathbf{f}(\mathbf{x})^{-1}E(\mathbf{u},\mathbf{v})\right\| \\ &\le \frac{1}{4}\|\mathbf{u}-\mathbf{v}\| + \frac{1}{4}\|\mathbf{u}-\mathbf{v}\| \;=\; \frac{1}{2}\|\mathbf{u}-\mathbf{v}\|. \end{aligned}$$

If $\{\mathbf{x}_n\}$ stays in this ball $B(\mathbf{x},\delta)$, then by the contraction mapping theorem it will converge to a unique fixed point.

1.0.2 K preserves a neighborhood of $\mathbf{x}$.

To show that $\{\mathbf{x}_n\}$ stays sufficiently near $\mathbf{x}$ if $\mathbf{y}'$ is sufficiently near $\mathbf{y}$, use

$$\mathbf{f}(\mathbf{u}) = \mathbf{y} + D\mathbf{f}(\mathbf{x})(\mathbf{u}-\mathbf{x}) + E(\mathbf{u},\mathbf{x}),$$

where $\|E(\mathbf{u},\mathbf{x})\| = o(\|\mathbf{u}-\mathbf{x}\|)$ as $\mathbf{u} \to \mathbf{x}$, to estimate

$$\begin{aligned} K(\mathbf{u}) - \mathbf{x} &= \mathbf{u} - D\mathbf{f}(\mathbf{x})^{-1}\left[\mathbf{f}(\mathbf{u}) - \mathbf{y}'\right] - \mathbf{x} \\ &= \mathbf{u}-\mathbf{x} - D\mathbf{f}(\mathbf{x})^{-1}\Big[\mathbf{y}-\mathbf{y}' + D\mathbf{f}(\mathbf{x})(\mathbf{u}-\mathbf{x}) + E(\mathbf{u},\mathbf{x})\Big] \\ &= -D\mathbf{f}(\mathbf{x})^{-1}(\mathbf{y}-\mathbf{y}') - D\mathbf{f}(\mathbf{x})^{-1}E(\mathbf{u},\mathbf{x}). \end{aligned}$$

But $D\mathbf{f}(\mathbf{x})^{-1}$ is bounded, so there exists $\epsilon > 0$ such that

$$\mathbf{y}' \in B(\mathbf{y},\epsilon) \implies \|D\mathbf{f}(\mathbf{x})^{-1}(\mathbf{y}-\mathbf{y}')\| < \frac{1}{2}\delta,$$

where $\delta = \min\{\delta_1,\delta_2\}$ from Eq.1 and Eq.2. Then Eq.2 implies

$$\mathbf{u} \in B(\mathbf{x},\delta) \implies \|D\mathbf{f}(\mathbf{x})^{-1}E(\mathbf{u},\mathbf{x})\| \le \frac{1}{4}\|\mathbf{u}-\mathbf{x}\| \le \frac{1}{4}\delta.$$

Hence, if $\mathbf{y}' \in B(\mathbf{y},\epsilon)$, then

$$\mathbf{u} \in B(\mathbf{x},\delta) \implies \|K(\mathbf{u}) - \mathbf{x}\| \le \frac{1}{2}\delta + \frac{1}{4}\delta < \delta.$$

Conclude that $\mathbf{x}_n \in B(\mathbf{x},\delta) \implies \mathbf{x}_{n+1} = K(\mathbf{x}_n) \in B(\mathbf{x},\delta)$, so that for all $\mathbf{y}' \in B(\mathbf{y},\epsilon)$,

$$\mathbf{x}_0 \in B(\mathbf{x},\delta) \implies \{\mathbf{x}_n\} \subset B(\mathbf{x},\delta).$$

1.0.3 Sequence $\{\mathbf{x}_n\}$ converges to a unique fixed point.

Suppose that $\mathbf{y}' \in B(\mathbf{y}, \epsilon)$ so that by the previous results, $K : B(\mathbf{x}, \delta) \to B(\mathbf{x}, \delta)$ is a contraction map satisfying

$$\mathbf{u}, \mathbf{v} \in B(\mathbf{x}, \delta) \implies \|K(\mathbf{u}) - K(\mathbf{v})\| \leq \frac{1}{2}\|\mathbf{u} - \mathbf{v}\|. \tag{3}$$

Then for any $\mathbf{x}_0 \in B(\mathbf{x}, \delta)$, the sequence defined by $\mathbf{x}_{n+1} = K(\mathbf{x}_n)$ will satisfy

$$\|\mathbf{x}_{N+1} - \mathbf{x}_N\| = \|K(\mathbf{x}_N) - K(\mathbf{x}_{N-1}\| \leq \frac{1}{2}\left\|\mathbf{x}_N - \mathbf{x}_{N-1}\right\|,$$

for $N = 0, 1, 2, \ldots$. Repeat this estimate N times to get the inequality

$$\|\mathbf{x}_{N+1} - \mathbf{x}_N\| \leq \left(\frac{1}{2}\right)^N \|\mathbf{x}_1 - \mathbf{x}_0\| < 2\delta\left(\frac{1}{2}\right)^N,$$

since $\mathbf{x}_1, \mathbf{x}_0 \in B(\mathbf{x}, \delta)$. But this implies that $\{\mathbf{x}_n\}$ is a Cauchy sequence:

$$\|\mathbf{x}_{N+n} - \mathbf{x}_N\| \leq \sum_{i=0}^{n-1} \|\mathbf{x}_{N+i+1} - \mathbf{x}_{N+i}\| < \sum_{i=0}^{n-1} 2\delta\left(\frac{1}{2}\right)^{N+i} < \frac{4\delta}{2^N}.$$

Metric space $\mathbf{E}^d$ is complete, so $\{\mathbf{x}_n\}$ converges to a point $\mathbf{x}' \in \mathbf{E}^d$ that satisfies $K(\mathbf{x}') = \mathbf{x}'$. This limit is unique in $B(\mathbf{x}, \delta)$, since for any other point $\mathbf{x}'' \in B(\mathbf{x}, \delta)$ with $\mathbf{x}'' = K(\mathbf{x}'')$, the contraction property implies

$$\|\mathbf{x}'' - \mathbf{x}'\| = \|K(\mathbf{x}'') - K(\mathbf{x}')\| \leq \frac{1}{2}\|\mathbf{x}'' - \mathbf{x}'\|,$$

which forces $\|\mathbf{x}'' - \mathbf{x}'\| = 0$ and thus $\mathbf{x}'' = \mathbf{x}'$.

Since $\mathbf{x}' = K(\mathbf{x}') = \mathbf{x}' - D\mathbf{f}(\mathbf{x})^{-1}[\mathbf{f}(\mathbf{x}') - \mathbf{y}']$, conclude that this unique fixed point satisfies $\mathbf{f}(\mathbf{x}') = \mathbf{y}'$. Thus it defines a function

$$\mathbf{g}(\mathbf{y}') = \mathbf{x}' \in B(\mathbf{x}, \delta), \qquad \mathbf{y}' \in B(\mathbf{y}, \epsilon).$$

This $\mathbf{g}$ is the inverse function for $\mathbf{f}$.

1.0.4 The inverse map is differentiable.

Since $\mathbf{y}' = \mathbf{f} \circ \mathbf{g}(\mathbf{y}')$ at all $\mathbf{y}'$ in an open neighborhood of $\mathbf{y} = \mathbf{f}(\mathbf{x})$, apply the chain rule to compute

$$\mathrm{I} = D[\mathbf{f} \circ \mathbf{g}](\mathbf{y}') = D\mathbf{f}(\mathbf{g}(\mathbf{y}'))D\mathbf{g}(\mathbf{y}') = D\mathbf{g}(\mathbf{y}')D\mathbf{f}(\mathbf{x}'),$$

for $\mathbf{x}' = \mathbf{g}(\mathbf{y}')$. Neither $d \times d$ factor matrix is singular since their product is the nonsingular identity matrix. It follows that $D\mathbf{f}(\mathbf{x}')$ is invertible with

$$D\mathbf{g}(\mathbf{y}') = D\mathbf{f}(\mathbf{x}')^{-1}.$$

This completes the proof of the Inverse Function Theorem. □