Theorem 2.8.13 (Kantorovich's theorem). *Let* $\mathbf{a}_0$ *be a point in* $\mathbb{R}^n$, U *an open neighborhood of* $\mathbf{a}_0$ *in* $\mathbb{R}^n$, *and* $\vec{\mathbf{f}} : U \to \mathbb{R}^n$ *a differentiable mapping, with its derivative* $[\mathbf{D}\vec{\mathbf{f}}(\mathbf{a}_0)]$ *invertible. Define*

$$\vec{\mathbf{h}}_0 = -[\mathbf{D}\vec{\mathbf{f}}(\mathbf{a}_0)]^{-1}\vec{\mathbf{f}}(\mathbf{a}_0), \quad \mathbf{a}_1 = \mathbf{a}_0 + \vec{\mathbf{h}}_0, \quad U_1 = B_{|\vec{\mathbf{h}}_0|}(\mathbf{a}_1). \qquad 2.8.54$$

If $\overline{U_1} \subset U$ *and the derivative* $[\mathbf{D}\vec{\mathbf{f}}(\mathbf{x})]$ *satisfies the Lipschitz condition*

$$\big|[\mathbf{D}\vec{\mathbf{f}}(\mathbf{u}_1)] - [\mathbf{D}\vec{\mathbf{f}}(\mathbf{u}_2)]\big| \le M|\mathbf{u}_1 - \mathbf{u}_2| \quad \textit{for all points } \mathbf{u}_1, \mathbf{u}_2 \in \overline{U_1}, \qquad 2.8.55$$

and if the inequality

$$\big|\vec{\mathbf{f}}(\mathbf{a}_0)\big|\,\big|[\mathbf{D}\vec{\mathbf{f}}(\mathbf{a}_0)]^{-1}\big|^2 M \le \frac{1}{2} \qquad 2.8.56$$

is satisfied, the equation $\vec{\mathbf{f}}(\mathbf{x}) = \vec{\mathbf{0}}$ *has a unique solution in the closed ball* $\overline{U_1}$, *and Newton's method with initial guess* $\mathbf{a}_0$ *converges to it.*

Note that the domain and codomain of the map $\vec{\mathbf{f}}$ have the *same dimension*. Thus, setting $\vec{\mathbf{f}}(\mathbf{x}) = \vec{\mathbf{0}}$, we get the same number of equations as unknowns. This is a reasonable requirement. If we had fewer equations than unknowns, we wouldn't expect them to specify a unique solution, and if we had more equations than unknowns, it would be unlikely that there would be any solutions at all.

In addition, if $n \ne m$, then $[\mathbf{D}\vec{\mathbf{f}}(\mathbf{a}_0)]$ would not be a square matrix, so it would not be invertible.

The basic idea is simple. The first of the three quantities that must be small is the value of the function at $\mathbf{a}_0$. If you are in an airplane flying close to the ground, you are more likely to crash (find a root) than if you are several kilometers up.

The second quantity is the square of the inverse of the derivative of the function at $\mathbf{a}_0$. In one dimension, we can think that the derivative must be big.[19] If your plane is approaching the ground steeply, it is much more likely to crash than if it is flying almost parallel to the ground.

The third quantity is the Lipschitz ratio M, measuring the change in the derivative (i.e., acceleration). If at the last minute the pilot pulls the plane out of a nose dive, flight attendants may be thrown to the floor as the derivative changes sharply, but a crash will be avoided. (Remember that acceleration need not be a change in speed; it can also be a change in direction.)

FIGURE 2.8.6.
Leonid Kantorovich (1912–1986)

Kantorovich was among the first to use linear programming in economics, in a paper published in 1939. He was awarded the Nobel Prize in economics in 1975.

But it is not each quantity individually that must be small: the product must be small. If the airplane starts its nose dive too close to the ground, even a sudden change in derivative may not save it. If it starts its nose dive from an altitude of several kilometers, it will still crash if it falls straight down. And if it loses altitude progressively, rather than plummeting to earth, it will still crash (or at least land) if the derivative never changes.

Remarks.

1. To check whether an equation makes sense, first make sure both sides have the same units. In physics and engineering, this is essential. The right side of inequality 2.8.56 is the unitless number 1/2.

[19]Why the theorem stipulates the *square* of the inverse of the derivative is more subtle. We think of it this way: the theorem should remain true if one changes the scale. Since the "numerator" $\vec{\mathbf{f}}(\mathbf{a}_0)M$ in equation 2.8.56 contains two terms, scaling up will change it by the scale factor squared. So the "denominator" $|[\mathbf{D}\vec{\mathbf{f}}(\mathbf{a}_0)]^{-1}|^2$ must also contain a square.

The left side:

$$\left|\vec{\mathbf{f}}(\mathbf{a}_0)\right| \left|[\mathbf{D}\vec{\mathbf{f}}(\mathbf{a}_0)]^{-1}\right|^2 M \qquad 2.8.57$$

is a complicated mixture of units of domain and codomain, which usually are different. Fortunately, these units cancel. To see this, denote by u the units of the domain, U, and by r the units of the codomain, $\mathbb{R}^n$. The term $|\vec{\mathbf{f}}(\mathbf{a}_0)|$ has units r. A derivative has units codomain/domain (typically, distance divided by time), so the inverse of the derivative has units domain/codomain $= u/r$, and the term $|[\mathbf{D}\vec{\mathbf{f}}(\mathbf{a}_0)]^{-1}|^2$ has units u^2/r^2. The Lipschitz ratio M is the distance between derivatives divided by a distance in the domain, so its units are r/u divided by u. This gives units $r \times \dfrac{u^2}{r^2} \times \dfrac{r}{u^2}$, which cancel out.

2. The Kantorovich theorem does *not* say that if inequality 2.8.56 is not satisfied, the equation has no solutions; it does not even say that if the inequality is not satisfied, there are no solutions in $\overline{U}_1$. In section 2.9 we will see that if we use a different way to measure $[\mathbf{D}\vec{\mathbf{f}}(\mathbf{a}_0)]$, the *norm*, which is harder to compute, then inequality 2.8.56 is easier to satisfy. That version of Kantorovich's theorem thus guarantees convergence for some equations about which this weaker version of the theorem is silent.
3. To see why it is necessary in the final sentence to specify the closed ball $\overline{U_1}$, not U_1, consider example 2.9.1, where Newton's method converges to 1, which is not in U_1 but is in its closure. △

History of the Kantorovich theorem: In one dimension, Cauchy proved that Newton's method converges. As far as we know, no further work on the subject was done until 1932, when Alexander Ostrowski published a proof in the 2-dimensional case. Ostrowski claimed a student proved the general case in his thesis in 1939, but that student was killed in World War II, and we know of no one who has seen this thesis.

Kantorovich approached the problem from a different point of view, that of nonlinear problems in Banach spaces, important in economics (and in many other fields). He proved the general case in a paper in the 1940s; it was included in a book on functional analysis published in 1959.

The proof we give would not need to be modified in order to apply it in the infinite-dimensional setting of Banach spaces.

If inequality 2.8.56 is satisfied, then at each iteration we create a new ball inside the previous ball: U_2 is in U_1, U_3 is in U_2, ..., as shown in the middle of figure 2.8.7. As the radius of the balls goes to zero, the sequence $\mathbf{a}_0, \mathbf{a}_1, \ldots$ converges to $\mathbf{a}$, which we will see is a root.

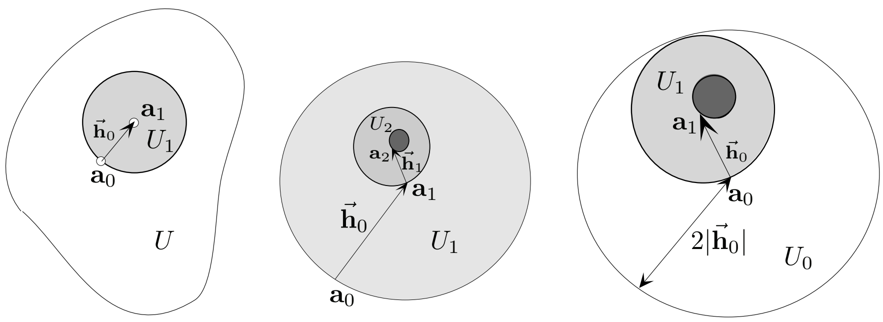

FIGURE 2.8.7. Equation 2.8.54 defines the neighborhood U_1 for which Newton's method is guaranteed to work when the inequality of equation 2.8.56 is satisfied. LEFT: the neighborhood U_1 is the ball of radius $|\vec{\mathbf{h}}_0| = |\mathbf{a}_1 - \mathbf{a}_0|$ around $\mathbf{a}_1$, so $\mathbf{a}_0$ is on the border of U_1. MIDDLE: a blow-up of U_1 (shaded), showing the neighborhoods U_2 (dark) and U_3 (darker). RIGHT: The ball U_0 described in proposition 2.8.14.