Indeed,

$$\left[\mathbf{D}\widetilde{\mathbf{f}}\begin{pmatrix} x\\ y\\ u_1\\ u_2\\ u_3\end{pmatrix}\right] = \begin{bmatrix} y+u_3 & x & 0 & -1 & x\\ 0 & u_1 & y & 2u_2 & 0\\ -2x & 0 & 1 & 0 & 0\\ 0 & -2y & 0 & 1 & 0\\ 0 & 0 & -2u_1 & 0 & 1\end{bmatrix}. \qquad 2.8.49$$

Thus

$$\left|\left[\mathbf{D}\widetilde{\mathbf{f}}\begin{pmatrix} x\\ y\\ u_1\\ u_2\\ u_3\end{pmatrix}\right] - \left[\mathbf{D}\widetilde{\mathbf{f}}\begin{pmatrix} x'\\ y'\\ u_1'\\ u_2'\\ u_3'\end{pmatrix}\right]\right| = \left|\begin{bmatrix} y-y'+u_3-u_3' & x-x' & 0 & 0 & x-x'\\ 0 & u_1-u_1' & y-y' & 2(u_2-u_2') & 0\\ -2(x-x') & 0 & 0 & 0 & 0\\ 0 & -2(y-y') & 0 & 0 & 0\\ 0 & 0 & -2(u_1-u_1') & 0 & 0\end{bmatrix}\right|$$

$$\leq \Big(6(x-x')^2 + 7(y-y')^2 + 5(u_1-u_1')^2 + 4(u_2-u_2)^2 + 2(u_3-u_3')^2\Big)^{1/2}$$

$$\leq \sqrt{7}\left|\begin{pmatrix} x\\ y\\ u_1\\ u_2\\ u_3\end{pmatrix} - \begin{pmatrix} x'\\ y'\\ u_1'\\ u_2'\\ u_3'\end{pmatrix}\right|. \qquad 2.8.50$$

Going from the first to second line of inequality 2.8.50: Since $(a+b)^2 \leq 2(a^2+b^2)$, we have

$$(y-y'+u_3-u_3')^2 \leq 2(y-y')^2 + 2(u_3-u_3')^2.$$

Thus $\sqrt{7}$ is a Lipschitz ratio for $[\mathbf{D}\widetilde{\mathbf{f}}\,]$. △

For all systems of polynomial equations, a similar trick will work and provide a Lipschitz ratio that can be computed, perhaps laboriously, but requiring no invention.

Kantorovich's theorem

Now we are ready to tackle Kantorovich's theorem. It says that if the product of three quantities is $\leq 1/2$, then the equation $\vec{\mathbf{f}}(\mathbf{x}) = \vec{\mathbf{0}}$ has a unique root in a closed ball $\overline{U_0}$, and if you start with an appropriate initial guess $\mathbf{a}_0$, Newton's method will converge to that root.

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. 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 the domain and codomain of the map $\vec{\mathbf{f}}$ had different dimensions, then $[\mathbf{D}\vec{\mathbf{f}}(\mathbf{a}_0)]$ would not be a square matrix, so it would not be invertible.

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*

$$\vec{\mathbf{h}}_0 \overset{\text{def}}{=} -[\mathbf{D}\vec{\mathbf{f}}(\mathbf{a}_0)]^{-1}\vec{\mathbf{f}}(\mathbf{a}_0), \quad \mathbf{a}_1 \overset{\text{def}}{=} \mathbf{a}_0 + \vec{\mathbf{h}}_0, \quad U_1 \overset{\text{def}}{=} B_{|\vec{\mathbf{h}}_0|}(\mathbf{a}_1). \qquad 2.8.51$$

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

$$\big|[\mathbf{D}\vec{\mathbf{f}}(\mathbf{u}_1)] - [\mathbf{D}\vec{\mathbf{f}}(\mathbf{u}_2)]\big| \leq M|\mathbf{u}_1 - \mathbf{u}_2| \quad \text{for all points } \mathbf{u}_1, \mathbf{u}_2 \in \overline{U_1}, \qquad 2.8.52$$

and if the inequality

$$\big|\vec{\mathbf{f}}(\mathbf{a}_0)\big|\,\big|[\mathbf{D}\vec{\mathbf{f}}(\mathbf{a}_0)]^{-1}\big|^2 M \leq \frac{1}{2} \qquad 2.8.53$$

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.

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.

The theorem is proved in Appendix A5.

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 length of the *inverse* of the derivative 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.)

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.

When discussing Kantorovich's theorem and Newton's method we write $\vec{\mathbf{f}}(\mathbf{x}) = \vec{\mathbf{0}}$, (with arrows) because we think of the codomain of $\vec{\mathbf{f}}$ as a vector space; the definition of $\vec{\mathbf{h}}_0$ only makes sense if $\vec{\mathbf{f}}(\mathbf{a}_0)$ is a vector. Moreover, $\vec{\mathbf{0}}$ plays a distinguished role in Newton's method (as it does in any vector space): we are trying to solve $\vec{\mathbf{f}}(\mathbf{x}) = \vec{\mathbf{0}}$, not $\mathbf{f}(\mathbf{x}) = \mathbf{a}$ for some random $\mathbf{a}$.

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 2.8.53 is the unitless number $1/2$. The left side:
$$\left|\vec{\mathbf{f}}(\mathbf{a}_0)\right| \left|[\mathbf{D}\vec{\mathbf{f}}(\mathbf{a}_0)]^{-1}\right|^2 M \qquad 2.8.54$$
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.
2. The Kantorovich theorem does *not* say that if inequality 2.8.53 is not satisfied, the equation has no solutions; it does not even say that

[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 inequality 2.8.53 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.