High-order open mapping theorems

Héctor J. Sussmann[*]
Department of Mathematics
Rutgers, the State University of New Jersey
Hill Center—Busch Campus
110 Frelinghuysen Road, Piscataway, NJ
08854-8019, USA

E-mail: sussmann@math.rutgers.edu
Web site: http://www.math.rutgers.edu/~sussmann

August 27, 2003

Abstract

The well-known finite-dimensional first-order open mapping theorem says that a continuous map with a finite-dimensional target is open at a point if its differential at that point exists and is surjective. An identical result, due to Graves, is true when the target is infinite-dimensional, if "differentiability" is replaced by "strict differentiability." We prove general theorems in which the linear approximations involved in the usual concept of differentiability is replaced by an approximation by a map which is homogeneous relative to a one-parameter group of dilations, and the error bound in the approximation involves a "homogeneous pseudonorm" or a "homogeneous pseudodistance," rather than the ordinary norm. We outline how these results can be used to derive sufficient conditions for openness involving higher-order derivatives, and carry this out in detail for the second-order case.

---

1 Introduction

Open mapping theorems (abbr. OMTs) are key tools in the derivation of necessary conditions for an optimum in optimization theory, and in particular in optimal control. For example, the usual Lagrange multipliers rule is a trivial corollary of the following OMT.

Theorem 1 *Let X, Y be normed spaces, let Ω be an open subset of X, and let $F : \Omega \mapsto Y$ be a continuous map. Let $x_* \in \Omega$ be such that F is Fréchet differentiable at x_* and the differential $DF(x_*)$ is a surjective linear map from X to Y. Assume that Y is finite-dimensional. Then F is open at x_*, that is, F maps neighborhoods of x_* to neighborhoods of $F(x_*)$.*

(To deduce the Lagrange rule from Theorem 1, let us consider the problem of minimizing the quantity $f_0(x)$ subject to finitely many equality constraints $f_1(x) = \cdots = f_m(x) = 0$, where $f_0, f_1, \ldots, f_m$ are continuous real-valued functions on an open subset Ω of a normed space X, and assume that x_* is a solution. We then define $F : \Omega \mapsto \mathbb{R}^{m+1}$ by letting

$$F(x) = (f_0(x), f_1(x), \ldots, f_m(x)) \ \text{ for } \ x \in \Omega,$$

and observe that, if we let $y_*(\varepsilon) = (f_0(x_*) - \varepsilon, 0, \ldots, 0) \in \mathbb{R}^{m+1}$ for $\varepsilon \in \mathbb{R}$, then $F(x_*) = y_*(0)$, and $y_*(\varepsilon) \notin F(\Omega)$ whenever $\varepsilon > 0$, so F is not open at x_*. If $f_0, f_1, \ldots, f_m$ are Fréchet differentiable at x_*, then F is Fréchet differentiable at x_*, so Theorem 1 implies that $DF(x_*)$ is not surjective. Therefore the $m+1$ vectors $\nabla f_0(x_*), \nabla f_1(x_*), \ldots, \nabla f_m(x_*)$ are not linearly independent. It follows that there exist numbers $\lambda_0, \lambda_1, \ldots, \lambda_m$ that are not all zero and satisfy $\sum_{i=0}^{m} \lambda_i \nabla f_i(x_*) = 0$.)

Theorem 1 is only valid if Y is finite-dimensional. (Precisely: if Y is any normed space, then Y is finite-dimensional if and only if it has the property that whenever $F : Y \mapsto Y$ is a continuous map such that $F(0) = 0$, F is Fréchet differentiable at 0, and the differential $DF(0)$ is the identity map of Y, it follows that F is open at 0.) On the other hand, L.M. Graves proved in [4] (cf. also Dontchev [3]) the following infinite-dimensional result.

Theorem 2 *Let X, Y be Banach spaces, let Ω be an open subset of X, and let $F : \Omega \mapsto Y$ be a continuous map. Let $x_* \in \Omega$ be such that F is strictly differentiable at x_* and the differential $DF(x_*)$ is a surjective linear map from X to Y. Then F is open at x_*.*

(Graves actually proved a stronger result, in which strict differentiability with a surjective differential is replaced by the weaker condition that there exist a surjective bounded linear map $L : X \mapsto Y$ such that

$$\limsup_{x \to x_*, x' \to x_*, x \neq x'} \frac{\|F(x) - F(x') - L(x - x')\|}{\|x - x'\|} < \|L^{-1}\|^{-1} \,, \tag{1}$$

where $\|L^{-1}\|$ is the infimum of the numbers C such that for every $y \in Y$ there exists $x \in X$ such that $L(x) = y$ and $\|x\| \leq C\|y\|$. The definition of "strict differentiability" is as follows: F is strictly differentiable at x_* with differential L if the left-hand side of (1) vanishes.)

The purpose of this note is to present "higher-order" sufficient conditions for openness at a point, generalizing Theorems 1 and 2.

Remark 3 Stronger results can also be proved, in which

1. in the infinite-dimensional case, the analogue of strict differentiability is replaced by the analogue of condition (1);

2. in the finite-dimensional case, the analogue of differentiability is replaced by the analogue of the condition that

$$\limsup_{x \to x_*, x \neq x_*} \frac{\|F(x) - F(x_*) - L(x - x_*)\|}{\|x - x_*\|} < \|L^{-1}\|^{-1} \,;$$

3. F is only required to be defined on a "conic neighborhood" of x_*, i.e., a set S of the form

$$S = S_X(x_*, \varepsilon, C) \stackrel{\text{def}}{=} \{x : \|x - x_*\| < \varepsilon \wedge x - x_* \in C\} \,,$$

where $\varepsilon > 0$ and C is a convex cone in X with nonempty interior such that $0 \in C$;

4. the conclusion says that the map F is "directionally open" at x_* in the direction of a given vector $v_* \in Y$, provided that $v_* \in \operatorname{int} LC$. (Precisely: for every positive ε there exists a positive δ such that the image $F(S_X(x_*, \varepsilon, C))$ contains the set $S_Y(F(x_*), \delta, D)$ for some convex cone D in Y such that $v_* \in \operatorname{int} D$.)

In this paper, however, we will only carry out the simpler task of generalizing the non-directional theorems 1 and 2. $\diamond$

3

Naturally, "higher-order" means "involving higher-oder derivatives." A map F of class C^m has a Taylor approximation

$$F(x_* + h) \sim F(x_*) + P_1(h) + \frac{1}{2}P_2(h, h) + \cdots + \frac{1}{m!}P_m(h, h, \ldots, h),$$

where $P_j = D^j F(x_*)$ for $j = 1, \ldots, m$, so each P_j is a Y-valued, continuous, symmetric, multilinear map defined on the Cartesian product X^j of j copies of X. Openness at x_* follows from the first-order theorems if the linear map P_1 is surjective. When P_1 is not surjective, the high-order results will give openness if the missing directions in the image $P_1 X$ can somehow be realized as image directions using higher-order derivatives. On the other hand, if the "first-order effect" $P_1(h)$ of a particular direction h is nonzero, then this effect will dominate whatever contributions h may make through higher-order terms. This suggests that one should only look at the second-order effects $P_2(h, h)$ of vectors h belonging to the kernel K_2 of P_1. So, if $m = 2$, we take $K_1 = X$, $K_2 = \ker P_1$, and consider the approximation

$$\begin{aligned} F(x_* + h_1 + h_2) \quad \sim \quad & F(x_*) + P_1(h_1) + P_1(h_2) \\ & + \frac{1}{2}P_2(h_1, h_1) + \frac{1}{2}P_2(h_2, h_2) + P_2(h_1, h_2), \end{aligned}$$

where $h_1 \in K_1$, $h_2 \in K_2$. The remainder is clearly $o(\|h_1\|^2 + \|h_2\|^2)$, which is in particular $o(\|h_1\| + \|h_2\|^2)$. Furthermore, the terms $\frac{1}{2}P_2(h_1, h_1)$ and $P_2(h_1, h_2)$ are $o(\|h_1\|)$, so they are $o(\|h_1\| + \|h_2\|^2)$ as well, and can be absorbed into the remainder. Finally, $P_1(h_2)$ vanishes, because $h_2 \in \ker P_1$, and we end up with the approximation

$$\begin{aligned} F(x_* + h_1 + h_2) \quad = \quad & F(x_*) + P_1(h_1) + \frac{1}{2}P_2(h_2, h_2) + o(\|h_1\| + \|h_2\|^2) \\ & \text{as } h_1 \to 0,\ h_2 \to 0,\ h_1 \in K_1,\ h_2 \in K_2. \quad (2) \end{aligned}$$

Such an approximation is valid under more general situations, even if F is not of class C^2 near x_*. (For example, if $F : \mathbb{R}^3 \mapsto \mathbb{R}$ is given by

$$F(x, y, z) = x + |x|^{3/2}(1 + |y|^{1/2}) + y^2 - z^2,$$

then $F(x, y, z) = x + y^2 - z^2 + o(|x| + y^2 + z^2)$, so (2) holds with $K_1 = \mathbb{R}^3$, $K_2 = \{0\} \times \mathbb{R}^2$, and obvious choices of P_1 and P_2.) If we define $\mathcal{X} = K_1 \times K_2$, and write

$$\begin{aligned} \mathcal{F}(h_1, h_2) \quad &= \quad F(x_* + h_1 + h_2) - F(x_*), \\ \mathcal{G}(h_1, h_2) \quad &= \quad P_1(h_1) + \frac{1}{2}P_2(h_2, h_2), \\ \nu(h_1, h_2) \quad &= \quad \|h_1\| + \|h_2\|^2, \end{aligned}$$

4

for $h_1 \in K_1$, $h_2 \in K_2$, then (2) says that

$$\mathcal{F}(\xi) = \mathcal{G}(\xi) + o(\nu(\xi)) \qquad \text{as } \xi \to 0, \quad \xi \in \mathcal{X}. \tag{3}$$

Moreover, if we let

$$\delta_t(h_1, h_2) = (th_1, t^{1/2} h_2) \qquad \text{for} \qquad h_1 \in K_1, \, h_2 \in K_2, \, t > 0,$$

then

(HA.1) $\boldsymbol{\delta} = \{\delta_t\}_{t>0}$ *is a continuous one-parameter group of dilations on* $\mathcal{X}$,

(HA.2) $\mathcal{G}$ *is* $\boldsymbol{\delta}$-*homogeneous, in the sense that* $\mathcal{G}(\delta_t(\xi)) = t\mathcal{G}(\xi)$ *whenever* $\xi \in \mathcal{X}$ *and* $t > 0$,

(HA.3) ν *is* $\boldsymbol{\delta}$-*homogeneous, in the sense that* $\nu(\delta_t(\xi)) = t\nu(\xi)$ *whenever* $\xi \in \mathcal{X}$ *and* $t > 0$,

(HA.4) ν *is a "pseudonorm on* $\mathcal{X}$,*" that is, a continuous nonnegative real-valued function that satisfies the conditions (i)* $\nu(\xi) = 0 \iff \xi = 0$ *and (ii)* $\lim_{\|\xi\| \to +\infty} \nu(\xi) = +\infty$.

So, when we rewrite (2) as (3), we find that we are really dealing with an *approximation of a map* $\mathcal{F}$ *by another map* $\mathcal{G}$ *which is homogeneous with respect to a continuous one-parameter group of dilations.* Furthermore, *the approximation is of exactly the same kind as the approximation by a linear map involved in the usual concept of differentiability, except that the ordinary norm on* $\mathcal{X}$ *is replaced by a dilation-homogeneous pseudonorm.*

Similar considerations apply to higher-order approximations. For example, if $F \in C^3$, and we choose three linear subspaces K_1, K_2, K_3 of X, then

$$
\begin{aligned}
F(x_* + h_1 + h_2 + h_3) = \ & F(x_*) + P_1(h_1) + P_1(h_2) + P_1(h_3) \\
& + \frac{1}{2}\Big(P_2(h_1, h_1) + P_2(h_2, h_2) + P_2(h_3, h_3)\Big) \\
& + P_2(h_1, h_2) + P_2(h_1, h_3) + P_2(h_2, h_3) + P_3(h_1, h_2, h_3) \\
& + \frac{1}{6}\Big(P_3(h_1, h_1, h_1) + P_3(h_2, h_2, h_2) + P_3(h_3, h_3, h_3)\Big) \\
& + \frac{1}{2}\Big(P_3(h_1, h_2, h_2) + P_3(h_1, h_1, h_2) + P_3(h_1, h_3, h_3) \\
& \qquad + P_3(h_1, h_1, h_3) + P_3(h_2, h_3, h_3) + P_3(h_2, h_2, h_3)\Big) \\
& + o(\|h_1\|^3 + \|h_2\|^3 + \|h_3\|^3) \\
& \text{as } h_1 \to 0, \, h_2 \to 0, \, h_3 \to 0, \, h_1 \in K_1, \, h_2 \in K_2, \, h_3 \in K_3.
\end{aligned}
$$

If we choose the K_i such that $K_1 = X$, and K_2 and K_3 are subsets of $\ker P_1$ (so that $P_1(h_2) = P_1(h_3) = 0$), and absorb into the remainder all the terms that are obviously $o(\|h_1\| + \|h_2\|^2 + \|h_3\|^3))$, we find

$$
\begin{aligned}
F(x_* + h_1 + h_2 + h_3) &= F(x_*) + P_1(h_1) + \frac{1}{2}\Big(P_2(h_2, h_2) + P_2(h_3, h_3)\Big) \\
&\quad + P_2(h_2, h_3) + \frac{1}{6}P_3(h_3, h_3, h_3) + o(\|h_1\| + \|h_2\|^2 + \|h_3\|^3).
\end{aligned}
$$

(The term $P_3(h_2, h_3, h_3)$ is eliminated because

$$
\begin{aligned}
\|P_3(h_2, h_3, h_3)\| &\leq \|P_3\|\,\|h_2\|\,\|h_3\|^2 \leq \|P_3\|(\|h_2\|^{7/3} + (\|h_3\|^2)^{7/4}) \\
&= \|P_3\|(\|h_2\|^{7/3} + \|h_3\|^{7/2}) = o(\|h_2\|^2 + \|h_3\|^3),
\end{aligned}
$$

using the fact that the inequality $ab \leq a^p + b^q$ holds if $a \geq 0$, $b \geq 0$, $p > 1$, $q > 1$, and $(1/p) + (1/q) = 1$, and taking $p = 7/3$, $q = 7/4$.)

If we require in addition that "$K_3 \subseteq \ker P_2$," in the precise sense that every $h \in K_3$ must satisfy $P_2(h, h') = 0$ whenever $h' \in K_2$, then we find

$$
\begin{aligned}
F(x_* + h_1 + h_2 + h_3) &= F(x_*) + P_1(h_1) + \frac{1}{2}P_2(h_2, h_2) \\
&\quad + \frac{1}{6}P_3(h_3, h_3, h_3) + o(\|h_1\| + \|h_2\|^2 + \|h_3\|^3) \\
&\text{as } h_1 \to 0,\ h_2 \to 0,\ h_3 \to 0,\ h_1 \in K_1,\ h_2 \in K_2,\ h_3 \in K_3.
\end{aligned}
$$

So, once again, we find that (3) holds, if we define $\mathcal{X} = K_1 \times K_2 \times K_3$, and let

$$
\begin{aligned}
\mathcal{F}(h_1, h_2, h_3) &= F(x_* + h_1 + h_2 + h_3) - F(x_*), \\
\mathcal{G}(h_1, h_2, h_3) &= P_1(h_1) + \frac{1}{2}P_2(h_2, h_2) + \frac{1}{6}P_3(h_3, h_3, h_3), \\
\nu(h_1, h_2, h_3) &= \|h_1\| + \|h_2\|^2 + \|h_3\|^3,
\end{aligned}
$$

for $h_1 \in K_1$, $h_2 \in K_2$, $h_3 \in K_3$. In addition, if we let

$$
\delta_t(h_1, h_2, h_3) = (th_1, t^{1/2}h_2, t^{1/3}h_3) \quad \text{for} \quad h_1 \in K_1,\ h_2 \in K_2,\ h_3 \in K_3,\ t > 0,
$$

then the four homogeneous approximation conditions HA.1-4 hold.

Our higher-order generalization of Theorem 1 involves a condition closely resembling the usual formula $\mathcal{F}(\xi) = \mathcal{L}(\xi) + o(\|\xi\|)$ that characterizes ordinary differentiability of $\mathcal{F}$ at 0. The crucial difference is that (a) the linear map $\mathcal{L}$ is replaced by a map $\mathcal{G}$ which is homogeneous with respect to some suitable group of dilations, and (b) a pseudonorm ν is substituted for the ordinary norm. The requirement that $\mathcal{L}$ be surjective, which occurs in the first-order

theorems, will be replaced by the condition that $\mathcal{G}$ have a surjective differential at some point $\xi_* \in \mathcal{X}$ such that $\mathcal{G}(\xi_*) = 0$. We will show that these conditions imply that $\mathcal{F}$ is open at 0 if the target space $\mathcal{Y}$ of $\mathcal{F}$ is finite-dimensional, generalizing Theorem 1. To get the generalization of Theorem 2, in which $\mathcal{Y}$ is allowed to be infinite-dimensional, we will have use the appropriate "strict" analogues of differentiability: the map $\mathcal{G}$ will have to be assumed to be strictly differentiable at ξ_*, and the approximation formula (3) will have to be replaced by

$$\mathcal{E}(\xi) - \mathcal{E}(\xi') = o(\nu(\xi, \xi')) \qquad \text{as } \xi \to 0, \, \xi' \to 0, \quad \xi \in \mathcal{X}, \, \xi' \in \mathcal{X}, \qquad (4)$$

where the error $\mathcal{E}$ is defined by $\mathcal{E}(\xi) = \mathcal{F}(\xi) - \mathcal{G}(\xi)$, and ν is a "homogeneous Lipschitz-bounded pseudodistance," in a sense to be defined below.

From the general theorems involving homogeneous approximations, it will then be possible to derive high-order open mapping theorems by means of the construction sketched above, using the obvious fact that, when $\mathcal{F}$ is defined in terms of F as we have done in our second- and third-order examples, then if $\mathcal{F}$ is open at 0 it follows that F is open at x_*.

This will be done here for the second-order case in §5, where it will be shown that our sufficient conditions for openness apply in particular when the Hessian of the map has a regular zero, and in the cases considered by Avakov in [1, 2].

2 Preliminaries

Metric spaces, normed spaces, balls, openness at a point. If X is a metric space with distance d, x is a point of X, and $r \geq 0$, we will use $\mathbb{B}_X(x, r)$ to denote the closed r-ball with center x, i.e., the set $\{x' \in X : d(x', x) \leq r\}$.

We use the word "neighborhood" in the usual sense of point set topology: if X is a topological space, and $x_* \in X$, a *neighborhood* of x_* in X is a subset U of X such that x_* is an interior point of U.

Definition 4 If X, Y are topological spaces, $x_* \in X$, and $F : X \mapsto Y$ is a map, then F is said to be *open at x_** if whenever U is a neighborhood of x_* in X, it follows that the image $F(U)$ is a neighborhood of 0 in Y. $\diamondsuit$

All linear spaces will be over $\mathbb{R}$, the field of real numbers. If X is a normed linear space, then we will write $\mathbb{B}_X(r)$ instead of $\mathbb{B}_X(0, r)$. If X, Y are normed linear spaces, then $\mathrm{Lin}(X, Y)$ will denote the space of all bounded linear maps from X to Y. If $A \in \mathrm{Lin}(X, Y)$ and $x \in X$, then we will use interchangeably the expressions Ax, $A \cdot x$ and $A(x)$ to denote the value of A at x. Convergence

in $\mathrm{Lin}(X, Y)$ is uniform convergence, i.e., convergence relative to the operator norm $\mathrm{Lin}(X, Y) \ni A \mapsto \|A\| \overset{\text{def}}{=} \sup\{\|Ax\| : x \in X, \|x\| \le 1\}$.

Differentiability and strict differentiability. The word "differentiable" will only be used to refer to maps between normed spaces, and will mean "Fréchet differentiable." That is, if we assume that

(A) *X and Y are normed spaces, Ω is an open subset of X, $x_* \in \Omega$, $F : \Omega \mapsto Y$ is a map, and $L : X \mapsto Y$ is a bounded linear map,*

then we say that *F is differentiable at x_* with differential L* if

$$\lim_{x \to x_*, x \ne x_*} \frac{\|F(x) - F(x_*) - L.(x - x*)\|}{\|x - x_*\|} = 0 \,. \tag{5}$$

A stronger concept of differentiability is "strict differentiability," defined as follows.

Definition 5 Let X, Y, Ω, x_*, F, L be such that (A) above holds. We say that *F is strictly differentiable at x_* with strict differential L* if the equality

$$\lim_{x \to x_*, x' \to x_*, x \ne x'} \frac{\|F(x) - F(x') - L.(x - x')\|}{\|x - x'\|} = 0 \tag{6}$$

holds. $\diamond$

Clearly, if F is strictly differentiable at x_* with strict differential L, then F is differentiable at x_* with differential L, since we can obtain (5) by taking $x' = x_*$ in (6).

Dilations, pseudonorms, pseudodistances, homogeneous maps.

Definition 6 Assume that $\mathcal{X}$ is a normed real linear space. A *continuous one-parameter group of dilations* on $\mathcal{X}$ is a family $\boldsymbol{\delta} = \{\delta_t\}_{t>0}$ of bounded linear maps $\delta_t : \mathcal{X} \mapsto \mathcal{X}$ such that

D1. δ_1 is the identity map $\mathrm{id}_{\mathcal{X}}$ of $\mathcal{X}$,

D2. $\delta_t \circ \delta_s = \delta_{ts}$ whenever $t > 0$ and $s > 0$,

D3. the map $]0, +\infty[\ni t \mapsto \delta_t \in \mathrm{Lin}(\mathcal{X}, \mathcal{X})$ is continuous with respect to the uniform operator norm on $\mathrm{Lin}(\mathcal{X}, \mathcal{X})$ (that is, $\lim_{t \to s} \|\delta_t - \delta_s\| = 0$ whenever $s > 0$),

D4. $\lim_{t \downarrow 0} \|\delta_t\| = 0 \,.$ $\diamond$

Lemma 7 *If $\mathcal{X}$ is a normed linear space and $\boldsymbol{\delta} = \{\delta_t\}_{t>0}$ is a continuous one-parameter group of dilations on $\mathcal{X}$, then*

$$\lim_{t\to+\infty} \|\delta_t(x)\| = +\infty \qquad \text{for all} \qquad x \in \mathcal{X}\backslash\{0\}\,, \tag{7}$$

and

$$\text{the map }]0, +\infty[\,\ni t \mapsto \delta_t(x) \in \mathcal{X} \text{ is one-to-one whenever } x \in X\backslash\{0\}\,. \tag{8}$$

Proof. If $x \neq 0$ and it is not true that $\lim_{t\to+\infty} \|\delta_t(x)\| = +\infty$, then there exists a sequence $\{t_j\}_{j=1}^\infty$ such that $\lim_{j\to\infty} t_j = +\infty$ while, on the other hand, the set $S = \{\delta_{t_j}(x) : j \in \mathbb{N}\}$ is bounded. Then, if we let $\xi_j = \delta_{t_j}(x)$, we can pick a constant C such that $\|\xi_j\| \leq C$ for all j, and conclude that $x = \delta_{t_j^{-1}}(\xi_j)$, so $\|x\| \leq C\|\delta_{t_j^{-1}}\|$. Since $\|\delta_{t_j^{-1}}\| \to 0$—because $t_j^{-1} \to 0$—it follows that $x = 0$. This completes the proof of (7).

To prove (8), we observe that if $x \neq 0$, $\delta_t(x) = \delta_s(x)$, and $0 < t < s$, then $\delta_\tau(x) = x$, if $\tau = s/t$. Then $\tau > 1$, and $\delta_{\tau^k}(x) = x$ for $k = 1, 2, \ldots$, so the norm $\|\delta_{\tau^k}(x)\|$ does not go to $+\infty$ as $k \to \infty$ but $\tau^k \to +\infty$ as $k \to \infty$, contradicting (7). $\diamond$

Definition 8 Assume that $\mathcal{X}$, $\boldsymbol{\delta}$ are as in Definition 6. A $\boldsymbol{\delta}$-*pseudonorm* is a continuous map $\nu : \mathcal{X} \mapsto \mathbb{R}$ such that (1) $\nu(x) \geq 0$ for all $x \in \mathcal{X}$, (2) $\nu(x) = 0 \iff x = 0$, (3) $\lim_{\|x\|\to+\infty} \nu(x) = +\infty$, and (4) $\nu(\delta_t(x)) = t\nu(x)$ whenever $x \in \mathcal{X}$, $t \in \mathbb{R}$, and $t > 0$. $\diamond$

Definition 9 Assume that the space $\mathcal{X}$ and the dilation group $\boldsymbol{\delta}$ are as in Definition 6. A $\boldsymbol{\delta}$-*pseudodistance* is a continuous map $\nu : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}_+$ such that (1) $\nu(x, x') = \nu(x', x) \geq 0$ for all $x, x' \in \mathcal{X}$, (2) $\nu(x, x') = 0 \iff x = x'$, (3) $\lim_{\|x\|\to+\infty} \nu(x, 0) = +\infty$, and (4) $\nu(\delta_t(x), \delta_t(x')) = t\nu(x, x')$ whenever $x, x' \in \mathcal{X}$, $t \in \mathbb{R}$, and $t > 0$. $\diamond$

If $\nu : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ is a $\boldsymbol{\delta}$-pseudodistance on $\mathcal{X}$, we define, for each positive number R,

$$\kappa_\nu(R) \stackrel{\text{def}}{=} \sup\left\{ \frac{\nu(x, x')}{\|x - x'\|} : x, x' \in \mathbb{B}_\mathcal{X}(R)\,, x \neq x' \right\}, \tag{9}$$

Then

$$\kappa_\nu(R) < \infty \text{ for some positive } R \iff \kappa_\nu(R) < \infty \text{ whenever } R > 0. \tag{10}$$

9

(Indeed, if R_0 is such that $R_0 > 0$ and $\kappa_\nu(R_0) < \infty$, then there exists a positive $\bar{t}$ such that $R_0\|\delta_{\bar{t}}\| < R$, since $\lim_{t\downarrow 0}\|\delta_t\| = 0$. Then $\delta_{\bar{t}}(\mathbb{B}_{\mathcal{X}}(R)) \subseteq \mathbb{B}_{\mathcal{X}}(R_0)$. If $x, x' \in \mathbb{B}_{\mathcal{X}}(R)$, then

$$
\begin{aligned}
\nu(x, x') &= \bar{t}^{-1}\nu(\delta_{\bar{t}}(x), \delta_{\bar{t}}(x')) \\
&\leq \bar{t}^{-1}\kappa_\nu(R_0)\|\delta_{\bar{t}}(x) - \delta_{\bar{t}}(x')\| \\
&\leq \bar{t}^{-1}\kappa_\nu(R_0)\|\delta_{\bar{t}}\|\,\|x - x'\|\,,
\end{aligned}
$$

so $\kappa_\nu(R) \leq \bar{t}^{-1}\kappa_\nu(R_0)\|\delta_{\bar{t}}\| < +\infty$.)

Definition 10 A pseudodistance $\nu : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ is *Lipschitz-bounded* if $\kappa_\nu(R) < \infty$ for some (and hence every) positive number R. $\diamond$

Definition 11 Assume that $\mathcal{X}$, $\boldsymbol{\delta}$ are as in Definition 6, $\mathcal{Y}$ is a normed linear space, and $\mathcal{G} : \mathcal{X} \mapsto \mathcal{Y}$ is a map. We say that $\mathcal{G}$ is $\boldsymbol{\delta}$-*homogeneous* if $\mathcal{G}(\delta_t(x)) = t\,\mathcal{G}(x)$ whenever $x \in \mathcal{X}$ and $t \geq 0$. $\diamond$

Regular zeros. If X, Y are normed spaces, Ω is open in X, and F is a map from Ω to Y, then a *regular zero* (resp. a *strictly regular zero*) of F is a point $x_* \in \Omega$ such that $F(x_*) = 0$, F is Fréchet differentiable (resp. strictly Fréchet differentiable) at x_*, and the differential $DF(x_*) : X \mapsto Y$ is surjective.

3   The finite-dimensional theorem

Theorem 12 *Assume that $\mathcal{X}$ and $\mathcal{Y}$ are real linear normed spaces, $\mathcal{Y}$ is finite-dimensional, $\boldsymbol{\delta} = \{\delta_t\}_{t>0}$ is a continuous one-parameter group of dilations of $\mathcal{X}$, $\nu : \mathcal{X} \mapsto \mathbb{R}_+$ is a $\boldsymbol{\delta}$-pseudonorm, and $\mathcal{G} : \mathcal{X} \mapsto \mathcal{Y}$ is a continuous $\boldsymbol{\delta}$-homogeneous map. Let Ω be an open subset of $\mathcal{X}$ such that $0 \in \Omega$, and let $\mathcal{F} : \Omega \mapsto \mathcal{Y}$ be a continuous map such that $\mathcal{F}(0) = 0$ and*

$$
\lim_{\xi \to 0} \frac{\|\mathcal{F}(\xi) - \mathcal{G}(\xi)\|}{\nu(\xi)} = 0\,. \tag{11}
$$

Assume that $\mathcal{G}$ has a regular zero. Then $\mathcal{F}$ is open at 0.

Proof. Let V be a neighborhood of 0 in $\mathcal{X}$ such that $V \subseteq \Omega$. Let R be such that $R > 0$ and $\mathbb{B}_{\mathcal{X}}(2R) \subseteq V$. Let ξ_* be a regular zero of $\mathcal{G}$. For $t > 0$, let $\xi_{*,t} = \delta_t(\xi_*)$. Let $\bar{t}$ be such that $\bar{t} > 0$ and the inequalities

$$
\|\delta_t\| \leq 1\,, \qquad \|\xi_{*,t}\| \leq R\,,
$$

hold whenever $0 < t \leq \bar{t}$. Let $\mathcal{L} = D\mathcal{G}(\xi_*)$, so $\mathcal{L}$ is surjective. Let $\mathcal{M} : \mathcal{Y} \mapsto \mathcal{X}$ be a linear map such that $\mathcal{L} \circ \mathcal{M} = \mathrm{id}_{\mathcal{Y}}$. Then $\mathcal{M}$ is bounded, because $\mathcal{Y}$ is finite-dimensional. Let r be such that $0 < r < R$ and

$$\|\mathcal{G}(\xi) - \mathcal{G}(\xi_*) - \mathcal{L}(\xi - \xi_*)\| \leq \frac{\|\xi - \xi_*\|}{4\|\mathcal{M}\|} \qquad \text{whenever} \qquad \|\xi - \xi_*\| \leq r\,.$$

Let s be such that $s > 0$, $s\|\mathcal{M}\| \leq r$, and $\bar{\nu} < \infty$, where

$$\bar{\nu} = \sup\Big\{\nu(\xi) : \xi \in \mathbb{B}_{\mathcal{X}}(\xi_*; s\|\mathcal{M}\|)\Big\}\,.$$

Let Ψ be the map from $\mathbb{B}_{\mathcal{Y}}(s)$ to $\mathcal{Y}$ given by $\Psi(y) = \mathcal{G}(\xi_* + \mathcal{M} \cdot y)$. Then, if $y \in \mathbb{B}_{\mathcal{Y}}(s)$, $\mathcal{M}(y)$ belongs to $\mathbb{B}_{\mathcal{X}}(r)$, so

$$
\begin{aligned}
\|\Psi(y) - y\| &= \|\mathcal{G}(\xi_* + \mathcal{M} \cdot y) - \mathcal{L}(\mathcal{M}(y))\| \\
&= \|\mathcal{G}(\xi_* + \mathcal{M} \cdot y) - \mathcal{G}(\xi_*) - \mathcal{L}(\mathcal{M}(y))\| \\
&\leq \frac{1}{4\|\mathcal{M}\|}\|\mathcal{M}\| \cdot \|y\| \\
&\leq \frac{s}{4}\,,
\end{aligned}
$$

using the fact that $\mathcal{G}(\xi_*) = 0$. Next, we define $W_t = \delta_t\Big(\mathbb{B}_{\mathcal{X}}(\xi_*, s\|\mathcal{M}\|)\Big)$, and observe that $W_t \subseteq \mathbb{B}_{\mathcal{X}}(\|\xi_{*,t}\| + s\|\delta_t\|\,\|\mathcal{M}\|)$. In particular,

$$\lim_{t\downarrow 0}\Big(\sup\{\|\xi\| : \xi \in W_t\}\Big) = 0\,,$$

$$0 < t \leq \bar{t} \implies W_t \subseteq \mathbb{B}_{\mathcal{X}}(2R)\,,$$

and

$$\lim_{t\downarrow 0}\lambda_t = 0\,,$$

where

$$\lambda_t \overset{\mathrm{def}}{=} \sup\Big\{\frac{\|\mathcal{F}(\xi) - \mathcal{G}(\xi)\|}{\nu(\xi)} : \xi \in W_t\Big\}\,.$$

Define

$$\Phi_t(y) = t^{-1}(\mathcal{F} \circ \delta_t)(\xi_* + \mathcal{M} \cdot y) \qquad \text{for} \qquad y \in \mathbb{B}_{\mathcal{Y}}(s)\,, \quad 0 < t \leq \bar{t}\,.$$

(The definition is possible because, if $y \in \mathbb{B}_{\mathcal{Y}}(s)$ and $0 < t \leq \bar{t}$, then

$$\|\delta_t(\xi_* + \mathcal{M} \cdot y)\| = \|\xi_{*,t} + \delta_t(\mathcal{M} \cdot y)\| \leq \|\xi_{*,t}\| + s\|\delta_t\|\,\|\mathcal{M}\| \leq R + r < 2R\,,$$

11

so $\delta_t(\xi_* + \mathcal{M} \cdot y) \in \mathbb{B}_{\mathcal{X}}(2R)$ and then $(\mathcal{F} \circ \delta_t)(\xi_* + \mathcal{M} \cdot y)$ is well defined.)

Then, if we let

$$
\begin{aligned}
\mu(t, y) &= \left\| (\mathcal{F} \circ \delta_t)(\xi_* + \mathcal{M} \cdot y) - (\mathcal{G} \circ \delta_t)(\xi_* + \mathcal{M} \cdot y) \right\|, \\
\tilde{\mu}(t, y) &= \frac{\mu(t, y)}{\nu(\delta_t(\xi_* + \mathcal{M} \cdot y))}, \\
\mu_t &= \sup \left\{ \mu(t, y) : y \in \mathbb{B}_{\mathcal{Y}}(s) \right\},
\end{aligned}
$$

we have

$$
\tilde{\mu}(t, y) \leq \sup \left\{ \frac{\|\mathcal{F}(\xi) - \mathcal{G}(\xi)\|}{\nu(\xi)} : \xi \in W_t \right\} = \lambda_t,
$$

so that

$$
\begin{aligned}
\mu_t &= \sup \left\{ \tilde{\mu}(t, y) \cdot \nu(\delta_t(\xi_* + \mathcal{M} \cdot y)) : y \in \mathbb{B}_{\mathcal{Y}}(s) \right\} \\
&\leq \lambda_t \sup \left\{ \nu(\delta_t(\xi)) : \xi \in \mathbb{B}_{\mathcal{X}}(\xi_*; s\|\mathcal{M}\|) \right\} \\
&= \lambda_t \sup \left\{ t\nu(\xi) : \xi \in \mathbb{B}_{\mathcal{X}}(\xi_*; s\|\mathcal{M}\|) \right\} \\
&= t \lambda_t \sup \left\{ \nu(\xi) : \xi \in \mathbb{B}_{\mathcal{X}}(\xi_*; s\|\mathcal{M}\|) \right\} \\
&= t \lambda_t \bar{\nu} .
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
\|\Phi_t(y) - y\| &\leq t^{-1}\mu(t, y) + \|t^{-1}\mathcal{G}(\delta_t(\xi_* + \mathcal{M} \cdot y)) - y\| \\
&\leq t^{-1}\mu(t, y) + \|\mathcal{G}(\xi_* + \mathcal{M} \cdot y) - y\| \\
&\leq t^{-1}\mu_t + \|\Psi(y) - y\| \\
&\leq \lambda_t \bar{\nu} + \frac{s}{4} .
\end{aligned}
$$

Now choose t such that $\bar{\nu}\lambda_t \leq \frac{s}{4}$. Then $\|\Phi_t(y) - y\| \leq \frac{s}{2}$. Let $B = \mathbb{B}_{\mathcal{Y}}(\frac{s}{2})$, $\tilde{B} = \mathbb{B}_{\mathcal{Y}}(s)$. If $y \in B$, define a map $\zeta_y : \tilde{B} \mapsto \mathcal{Y}$ by letting $\zeta_y(y') = y' - \Phi_t(y') + y$ if $y' \in \tilde{B}$. Then, if $y' \in \tilde{B}$, the inequalities

$$
\|\zeta_y(y')\| \leq \|y' - \Phi_t(y')\| + \|y\| \leq \frac{s}{2} + \frac{s}{2} = s
$$

imply that $\zeta_y(y') \in \tilde{B}$ as well. Hence ζ_y is a continuous map from $\tilde{B}$ to $\tilde{B}$, so ζ_y has a fixed point by Brouwer's theorem. If $\bar{y}'$ is a fixed point of ζ_y, then $\bar{y}' - \Phi_t(\bar{y}') + y = \zeta_y(\bar{y}') = \bar{y}'$, so $\Phi_t(\bar{y}') = y$. Then

$$
t^{-1}\mathcal{F}\left(\delta_t(\xi_* + \mathcal{M} \cdot \bar{y}') \right) = y,
$$

12

and $\delta_t(\xi_* + \mathcal{M} \cdot \bar{y}') \in W_t$. So, if we let $z = \delta_t(\xi_* + \mathcal{M} \cdot \bar{y}')$, it follows that $z \in V$ (because $W_t \subseteq \mathbb{B}_{\mathcal{X}}(2R) \subseteq V$) and $\mathcal{F}(z) = ty$. Since y is an arbitrary member of B, we have shown that

$$tB \subseteq \mathcal{F}(V) \,.$$

Therefore tB is a neighborhood of 0 in $\mathcal{Y}$, and $tB \subseteq \mathcal{F}(V)$. This completes our proof. $\diamond$

4 The infinite-dimensional theorem

Theorem 13 *Assume that $\mathcal{X}$ and $\mathcal{Y}$ are real Banach spaces, $\boldsymbol{\delta} = \{\delta_t\}_{t>0}$ is a continuous one-parameter group of dilations of $\mathcal{X}$, $\nu : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}_+$ is a Lipschitz-bounded $\boldsymbol{\delta}$-pseudodistance, and $\mathcal{G} : \mathcal{X} \mapsto \mathcal{Y}$ is a continuous $\boldsymbol{\delta}$-homogeneous map. Let Ω be an open subset of $\mathcal{X}$ such that $0 \in \Omega$, and let $\mathcal{F} : \Omega \mapsto \mathcal{Y}$ be a continuous map such that $\mathcal{F}(0) = 0$ and*

$$\lim_{\xi \to 0, \xi' \to 0, \xi \neq \xi'} \frac{\left\| \Big(\mathcal{F}(\xi) - \mathcal{G}(\xi) \Big) - \Big(\mathcal{F}(\xi') - \mathcal{G}(\xi') \Big) \right\|}{\nu(\xi, \xi')} = 0 \,. \tag{12}$$

Assume that $\mathcal{G}$ has a strictly regular zero. Then $\mathcal{F}$ is open at 0.

Proof. We define the error functions $\mathcal{E} : \Omega \mapsto \mathcal{Y}$, $\tilde{\mathcal{E}} : \mathcal{X} \times \mathcal{X} \mapsto \mathcal{Y}$, by letting

$$\mathcal{E}(\xi) = \mathcal{F}(\xi) - \mathcal{G}(\xi) \ \text{ for } \ \xi \in \Omega \,,$$
$$\tilde{\mathcal{E}}(\xi, \xi') = \mathcal{G}(\xi) - \mathcal{G}(\xi') - \mathcal{L}(\xi - \xi') \ \text{ for } \ \xi, \xi' \in \mathcal{X} \,.$$

Let V be a neighborhood of 0 in $\mathcal{X}$ such that $V \subseteq \Omega$. Let R be such that $R > 0$ and $\mathbb{B}_{\mathcal{X}}(2R) \subseteq V$. Let ξ_* be a strictly regular zero of $\mathcal{G}$. For $t > 0$, let $\xi_{*,t} = \delta_t(\xi_*)$. Let $\bar{t}$ be such that $\bar{t} > 0$ and the inequalities

$$\|\delta_t\| \leq 1 \,, \qquad \|\xi_{*,t}\| \leq R \,,$$

hold whenever $0 < t \leq \bar{t}$. Let $\mathcal{L} = D\mathcal{G}(\xi_*)$, so $\mathcal{L}$ is a bounded, surjective linear map from $\mathcal{X}$ to $\mathcal{Y}$. Then the Banach open mapping theorem implies that there is a positive constant C such that

(#) *for every $y \in \mathcal{X}$ there exists a $\xi \in \mathcal{X}$ such that $\mathcal{L}(\xi) = y$ and $\|\xi\| \leq C\|y\|$.*

Using the strict differentiability of $\mathcal{G}$ at ξ_*, we choose r such that $0 < r < R$ and

$$\|\tilde{\mathcal{E}}(\xi, \xi')\| \leq \frac{\|\xi - \xi'\|}{4C} \text{ whenever } \|\xi - \xi_*\| \leq r \text{ and } \|\xi' - \xi_*\| \leq r \,.$$

Let s be such that $s > 0$, $sC \leq r$, and $\bar{\nu} < \infty$, where

$$\bar{\nu} = \sup \left\{ \nu(\xi, 0) : \xi \in \mathbb{B}_{\mathcal{X}}(\xi_*; sC) \right\}.$$

Define

$$W_t = \delta_t \Big(\mathbb{B}_{\mathcal{X}}(\xi_*, sC) \Big) \,,$$

$$\omega_t = \sup \left\{ \|\xi\| : \xi \in W_t \right\},$$

and observe that $\omega_t \leq \|\xi_{*,t}\| + s\|\delta_t\| \, C$ for every t. In particular,

$$\lim_{t \downarrow 0} \omega_t = 0 \,,$$

$$0 < t \leq \bar{t} \implies W_t \subseteq \mathbb{B}_{\mathcal{X}}(\omega_t) \subseteq \mathbb{B}_{\mathcal{X}}(2R) \,,$$

and

$$\lim_{t \downarrow 0} \lambda_t = 0 \,,$$

where

$$\lambda_t \stackrel{\text{def}}{=} \sup \left\{ \frac{\left\| \mathcal{E}(\xi) - \mathcal{E}(\xi') \right\|}{\nu(\xi, \xi')} : \xi \in \mathbb{B}_{\mathcal{X}}(\omega_t), \, \xi' \in \mathbb{B}_{\mathcal{X}}(\omega_t), \, \xi \neq \xi' \right\}.$$

The Lipschitz-boundedness of ν implies that the constant $\bar{\kappa} \stackrel{\text{def}}{=} \kappa_\nu(R + \|\xi_*\|)$ is finite. For each t, we let

$$\alpha_t \stackrel{\text{def}}{=} \frac{1}{4} + C\lambda_t \bar{\kappa} \,.$$

We now fix a t such that

$$0 < t \leq \bar{t} \,, \qquad 4\lambda_t \bar{\nu} \leq s \,, \qquad \text{and} \qquad 2\alpha_t \leq 1 \,, \tag{13}$$

write $\tau = 1/t$, and prove that the ball $\mathbb{B}_{\mathcal{Y}}(t\lambda_t \bar{\nu})$ is contained in $\mathcal{F}(W_t)$. For this purpose, we fix a $y \in \mathcal{Y}$ such that $\|y\| \leq t\lambda_t \bar{\nu}$, and construct a $\xi \in W_t$ such that $\mathcal{F}(\xi) = y$. This ξ will be the limit of a sequence $\{\xi^j\}_{j=0}^\infty$ of points of W_t.

14

To begin with, we let $\xi^0 = \xi_{*,t}$, and observe that $\xi^0 \in W_t$, and the error $e^0 \overset{\text{def}}{=} \mathcal{F}(\xi^0) - y$ satisfies

$$e^0 = \mathcal{F}(\xi_{*,t}) - \mathcal{G}(\xi_{*,t}) - (\mathcal{F}(0) - \mathcal{G}(0)) - y \,,$$

since $\mathcal{G}(\xi_{*,t}) = \mathcal{G}(0) = \mathcal{F}(0) = 0$. So

$$
\begin{aligned}
\|e^0\| &\leq \|\mathcal{F}(\xi_{*,t}) - \mathcal{G}(\xi_{*,t}) - (\mathcal{F}(0) - \mathcal{G}(0))\| + \|y\| \\
&\leq \lambda_t \nu(\xi_{*,t}, 0) + \|y\| \\
&= \lambda_t \nu(\delta_t(\xi_*), \delta_t(0)) + \|y\| \\
&= t\lambda_t \nu(\xi_*, 0) + \|y\| \\
&\leq t\lambda_t \bar{\nu} + \|y\| \\
&= 2t\lambda_t \bar{\nu} \,.
\end{aligned}
$$

We then choose $\zeta^1 \in \mathcal{X}$ such that $\mathcal{L}(\zeta^1) = -t^{-1}e^0$ and $\|\zeta^1\| \leq t^{-1}C\|e^0\|$, and define $\eta^1 = \delta_t(\zeta^1)$, $\xi^1 = \xi^0 + \eta^1 = \delta_t(\xi_* + \zeta^1)$. Then $t\mathcal{L}(\delta_\tau(\eta^1)) = -e^0$, and $\|\zeta^1\| \leq 2C\lambda_t\bar{\nu}$. Therefore $\|\delta_\tau(\xi^1) - \xi_*\| = \|\zeta^1\| \leq 2C\lambda_t\bar{\nu} \leq 4C\lambda_t\bar{\nu} \leq sC$, from which it follows that $\delta_\tau(\xi^1) \in \mathbb{B}_\mathcal{X}(\xi_*, sC)$, so that $\xi^1 \in W_t$.

We then let $e^1 = \mathcal{F}(\xi^1) - y$. It follows that

$$
\begin{aligned}
e^1 &= \mathcal{E}(\xi^1) + \Big(\mathcal{G}(\xi^1) - y\Big) \\
&= \Big(\mathcal{E}(\xi^1) - \mathcal{E}(\xi^0)\Big) + \Big(\mathcal{F}(\xi^0) - \mathcal{G}(\xi^0)\Big) + \Big(\mathcal{G}(\xi^1) - y\Big) \\
&= \mathcal{E}(\xi^1) - \mathcal{E}(\xi^0) + \Big(\mathcal{F}(\xi^0) - y\Big) + \Big(\mathcal{G}(\xi^1) - \mathcal{G}(\xi^0)\Big) \\
&= \mathcal{E}(\xi^1) - \mathcal{E}(\xi^0) + e^0 + t\Big(\mathcal{G}(\delta_\tau(\xi^1)) - \mathcal{G}(\delta_\tau(\xi^0))\Big) \\
&= \mathcal{E}(\xi^1) - \mathcal{E}(\xi^0) + e^0 + t\tilde{\mathcal{E}}(\delta_\tau(\xi^1), \delta_\tau(\xi^0)) + t\mathcal{L}(\delta_\tau(\eta^1)) \\
&= \mathcal{E}(\xi^1) - \mathcal{E}(\xi^0) + t\tilde{\mathcal{E}}(\delta_\tau(\xi^1), \delta_\tau(\xi^0)) \,.
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
\|\mathcal{E}(\xi^1) - \mathcal{E}(\xi^0)\| &\leq \lambda_t \nu(\xi^1, \xi^0) \\
&= t\lambda_t \nu(\delta_\tau(\xi^1), \delta_\tau(\xi^0)) \\
&\leq t\lambda_t \bar{\kappa} \|\delta_\tau(\xi^1) - \delta_\tau(\xi^0))\| \\
&= t\lambda_t \bar{\kappa} \|\delta_\tau(\eta^1)\| \\
&= t\lambda_t \bar{\kappa} \|\zeta^1\| \\
&\leq 2Ct\lambda_t^2 \bar{\kappa}\bar{\nu} \,,
\end{aligned}
$$

15

and

$$\|\tilde{\mathcal{E}}(\delta_\tau(\xi^1), \delta_\tau(\xi^0))\| \leq \frac{\|\delta_\tau(\xi^1) - \delta_\tau(\xi^0)\|}{4C} = \frac{\|\zeta^1\|}{4C} \leq \frac{2C\lambda_t\bar{\nu}}{4C} = \frac{1}{2}\lambda_t\bar{\nu}\,.$$

Therefore

$$\|e^1\| \leq 2Ct\lambda_t^2\bar{\kappa}\bar{\nu} + \frac{t}{2}\lambda_t\bar{\nu} = t\lambda_t\bar{\nu}\Big(2C\lambda_t\bar{\kappa} + \frac{1}{2}\Big) = 2t\lambda_t\bar{\nu}\alpha_t\,.$$

Next, we choose $\zeta^2 \in \mathcal{X}$ such that $\mathcal{L}(\zeta^2) = -t^{-1}e^1$ and $\|\zeta^2\| \leq t^{-1}C\|e^1\|$, and define $\eta^2 = \delta_t(\zeta^2)$, $\xi^2 = \xi^1 + \eta^2 = \xi^1 + \delta_t(\zeta^2) = \delta_t(\xi_* + \zeta^1 + \zeta^2)$. Then $t\mathcal{L}(\delta_\tau(\eta^2)) = -e^1$, and $\|\zeta^2\| \leq 2C\lambda_t\bar{\nu}\alpha_t$. Therefore

$$\|\delta_\tau(\xi^2) - \xi_*\| = \|\zeta^1 + \zeta^2\| \leq 2C\lambda_t\bar{\nu}(1 + \alpha_t) \leq 4C\lambda_t\bar{\nu} \leq sC\,,$$

from which it follows that $\delta_\tau(\xi^2) \in \mathbb{B}_{\mathcal{X}}(\xi_*, sC)$, so that $\xi^2 \in W_t$.

We then let $e^2 = \mathcal{F}(\xi^2) - y$. It follows that

$$
\begin{aligned}
e^2 &= \mathcal{E}(\xi^2) + \Big(\mathcal{G}(\xi^2) - y\Big) \\
&= \Big(\mathcal{E}(\xi^2) - \mathcal{E}(\xi^1)\Big) + \Big(\mathcal{F}(\xi^1) - \mathcal{G}(\xi^1)\Big) + \Big(\mathcal{G}(\xi^2) - y\Big) \\
&= \mathcal{E}(\xi^2) - \mathcal{E}(\xi^1) + \Big(\mathcal{F}(\xi^1) - y\Big) + \Big(\mathcal{G}(\xi^2) - \mathcal{G}(\xi^1)\Big) \\
&= \mathcal{E}(\xi^2) - \mathcal{E}(\xi^1) + e^1 + t\Big(\mathcal{G}(\delta_\tau(\xi^2)) - \mathcal{G}(\delta_\tau(\xi^1))\Big) \\
&= \mathcal{E}(\xi^2) - \mathcal{E}(\xi^1) + e^1 + t\tilde{\mathcal{E}}(\delta_\tau(\xi^2), \delta_\tau(\xi^1)) + t\mathcal{L}(\delta_\tau(\eta^2)) \\
&= \mathcal{E}(\xi^2) - \mathcal{E}(\xi^1) + t\tilde{\mathcal{E}}(\delta_\tau(\xi^2), \delta_\tau(\xi^1))\,.
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
\|\mathcal{E}(\xi^2) - \mathcal{E}(\xi^1)\| &\leq \lambda_t\nu(\xi^2, \xi^1) \\
&= t\lambda_t\nu(\delta_\tau(\xi^2), \delta_\tau(\xi^1)) \\
&\leq t\lambda_t\bar{\kappa}\|\delta_\tau(\xi^2) - \delta_\tau(\xi^1))\| \\
&= t\lambda_t\bar{\kappa}\|\delta_\tau(\eta^2)\| \\
&= t\lambda_t\bar{\kappa}\|\zeta^2\| \\
&\leq t\lambda_t\bar{\kappa}(2C\lambda_t\bar{\nu}\alpha_t) \\
&\leq 2Ct\lambda_t^2\bar{\kappa}\bar{\nu}\alpha_t\,,
\end{aligned}
$$

and

$$\|\tilde{\mathcal{E}}(\delta_\tau(\xi^2), \delta_\tau(\xi^1))\| \leq \frac{\|\delta_\tau(\xi^2) - \delta_\tau(\xi^1)\|}{4C} = \frac{\|\zeta^2\|}{4C} \leq \frac{2C\lambda_t\bar{\nu}\alpha_t}{4C} = \frac{1}{2}\lambda_t\bar{\nu}\alpha_t\,.$$

16

Therefore

$$\|e^2\| \le 2Ct\lambda_t^2\bar{\kappa}\bar{\nu}\alpha_t + \frac{t}{2}\lambda_t\bar{\nu}\alpha_t = 2t\lambda_t\bar{\nu}\Big(C\lambda_t\bar{\kappa} + \frac{1}{4}\Big)\alpha_t = 2t\lambda_t\bar{\nu}\alpha_t^2\,.$$

We continue this construction inductively. Suppose we have defined

$$\xi^0,\ldots,\xi^k \in W_t\,,\quad e^0,\ldots,e^k \in \mathcal{Y}\,,\quad \zeta^1,\ldots,\zeta^k \in \mathcal{X}\,,\quad \eta^1,\ldots,\eta^k \in \mathcal{Y}\,,$$

such that

$$
\begin{aligned}
\xi^j &= \xi^{j-1} + \eta^j & &\text{for } j = 1,\ldots,k\,,\\
e^j &= \mathcal{F}(\xi^j) - y & &\text{for } j = 0,\ldots,k\,,\\
t\mathcal{L}(\zeta^j) &= -e^{j-1} & &\text{for } j = 1,\ldots,k\,,\\
\eta^j &= \delta_t(\zeta_j) & &\text{for } j = 1,\ldots,k\,,\\
\|\zeta^j\| &\le 2C\lambda_t\bar{\nu}\alpha_t^{j-1} & &\text{for } j = 1,\ldots,k\,,\\
\|e^j\| &\le 2t\lambda_t\bar{\nu}\alpha_t^j & &\text{for } j = 0,\ldots,k\,.
\end{aligned}
$$

We then choose $\zeta^{k+1} \in \mathcal{X}$ such that

$$\mathcal{L}(\zeta^{k+1}) = -t^{-1}e^k \text{ and } \|\zeta^{k+1}\| \le t^{-1}C\|e^k\|\,,$$

and define

$$
\begin{aligned}
\eta^{k+1} &= \delta_t(\zeta^{k+1})\,,\\
\xi^{k+1} &= \xi^k + \eta^{k+1} = \xi^k + \delta_t(\zeta^{k+1}) = \delta_t(\xi_* + \zeta^1 + \ldots + \zeta^{k+1})\,.
\end{aligned}
$$

Then $t\mathcal{L}(\delta_\tau(\eta^{k+1})) = -e^k$, and

$$\|\zeta^{k+1}\| \le t^{-1}C\|e^k\| \le 2C\lambda_t\bar{\nu}\alpha_t^k\,.$$

Therefore

$$\|\delta_\tau(\xi^{k+1}) - \xi_*\| = \|\zeta^1 + \ldots + \zeta^{k+1}\| \le 2C\lambda_t\bar{\nu}\sum_{j=0}^k \alpha_t^j = \frac{2C\lambda_t\bar{\nu}}{1 - \alpha_t} \le 4C\lambda_t\bar{\nu}\,,$$

so $\|\delta_\tau(\xi^{k+1}) - \xi_*\| \le sC$, from which it follows that $\delta_\tau(\xi^{k+1}) \in \mathbb{B}_\mathcal{X}(\xi_*, sC)$, so that $\xi^{k+1} \in W_t$.

17

We then let $e^{k+1} = \mathcal{F}(\xi^{k+1}) - y$. It follows that

$$
\begin{aligned}
e^{k+1} &= \mathcal{E}(\xi^{k+1}) + \Big(\mathcal{G}(\xi^{k+1}) - y\Big) \\
&= \Big(\mathcal{E}(\xi^{k+1}) - \mathcal{E}(\xi^k)\Big) + \Big(\mathcal{F}(\xi^k) - \mathcal{G}(\xi^k)\Big) + \Big(\mathcal{G}(\xi^{k+1}) - y\Big) \\
&= \mathcal{E}(\xi^{k+1}) - \mathcal{E}(\xi^k) + \Big(\mathcal{F}(\xi^k) - y\Big) + \Big(\mathcal{G}(\xi^{k+1}) - \mathcal{G}(\xi^k)\Big) \\
&= \mathcal{E}(\xi^{k+1}) - \mathcal{E}(\xi^k) + e^k + t\Big(\mathcal{G}(\delta_\tau(\xi^{k+1})) - \mathcal{G}(\delta_\tau(\xi^k))\Big) \\
&= \mathcal{E}(\xi^{k+1}) - \mathcal{E}(\xi^k) + e^1 + t\tilde{\mathcal{E}}(\delta_\tau(\xi^{k+1}), \delta_\tau(\xi^k)) + t\mathcal{L}(\delta_\tau(\eta^{k+1})) \\
&= \mathcal{E}(\xi^{k+1}) - \mathcal{E}(\xi^k) + t\tilde{\mathcal{E}}(\delta_\tau(\xi^{k+1}), \delta_\tau(\xi^k)).
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
\|\mathcal{E}(\xi^{k+1}) - \mathcal{E}(\xi^k)\| &\leq \lambda_t \nu(\xi^{k+1}, \xi^k) \\
&= t\lambda_t \nu(\delta_\tau(\xi^{k+1}), \delta_\tau(\xi^k)) \\
&\leq t\lambda_t \bar{\kappa} \|\delta_\tau(\xi^{k+1}) - \delta_\tau(\xi^k))\| \\
&= t\lambda_t \bar{\kappa} \|\delta_\tau(\eta^{k+1})\| \\
&= t\lambda_t \bar{\kappa} \|\zeta^{k+1}\| \\
&\leq t\lambda_t \bar{\kappa}(2C\lambda_t \bar{\nu}\alpha_t) \\
&\leq 2Ct\lambda_t^2 \bar{\kappa}\bar{\nu}\alpha_t^k,
\end{aligned}
$$

and

$$
\|\tilde{\mathcal{E}}(\delta_\tau(\xi^{k+1}), \delta_\tau(\xi^k))\| \leq \frac{\|\delta_\tau(\xi^{k+1}) - \delta_\tau(\xi^k)\|}{4C} = \frac{\|\zeta^{k+1}\|}{4C} \leq \frac{1}{2}\lambda_t \bar{\nu}\alpha_t^k.
$$

Therefore

$$
\|e^{k+1}\| \leq 2Ct\lambda_t^2 \bar{\kappa}\bar{\nu}\alpha_t^k + \frac{t}{2}\lambda_t \bar{\nu}\alpha_t^k = 2t\lambda_t \bar{\nu}\Big(C\lambda_t \bar{\kappa} + \frac{1}{4}\Big)\alpha_t^k = 2t\lambda_t \bar{\nu}\alpha_t^{k+1},
$$

and our inductive construction is complete.

The bound $\|\zeta^j\| \leq 2C\lambda_t \bar{\nu}\alpha_t^{j-1}$ implies—since $2\alpha_t \leq 1$—that the series $\sum_{j=1}^\infty \zeta^j$ converges, and the sum ζ of the series satisfies $\|\zeta\| \leq 4C\lambda_t \bar{\nu} \leq sC$. Therefore the limit $\Xi = \lim_{j\to\infty} \delta_\tau(\xi^j)$ exists and satisfies $\Xi \in \mathbb{B}_{\mathcal{X}}(\xi_*, sC)$. So the limit $\xi = \delta_t(\Xi) = \lim_{j\to\infty} \xi^j$ exists and belongs to W_t. Furthermore, $\mathcal{F}(\xi) - y = \lim_{j\to\infty}(\mathcal{F}(\xi^j) - y) = \lim_{j\to\infty} e^j = 0$, because of the bound $\|e^j\| \leq 2t\lambda_t \bar{\nu}\alpha_t^j$. Therefore $\mathcal{F}(\xi) = y$, and our proof is complete. $\diamond$

5 Second-order open mapping theorems

If X and Y are real linear spaces, and $X \times X \ni (x, x') \mapsto B(x, x') \in Y$ is a symmetric bilinear map, we write Q_B to denote the quadratic map associated with B, i.e., the map $X \ni x \mapsto B(x, x) \overset{\text{def}}{=} Q_B(x) \in Y$. It is well known that B is completely determined by Q_B, since

$$B(x, y) = \frac{1}{4}\Big(Q_B(x + y) - Q_B(x - y)\Big). \tag{14}$$

A *quadratic map* from X to Y is a map Q such that $Q = Q_B$ for some (unique) bilinear symmetric map $B : X \times X \mapsto Y$. If Q is a quadratic map, then we will use B^Q to denote the corresponding symmetric bilinear map.

If X, Y are normed, then a bilinear map $B : X \times X \mapsto Y$ is continuous if and only if it is bounded, in the sense that there exists a constant C such that $\|B(x, x')\| \leq C\|x\| \, \|x'\|$ for all $x, x' \in X$. It follows from (14) that a quadratic map $Q : X \to Y$ is continuous if and only if the bilinear map B^Q is continuous.

Definition 14 If X, Y are normed spaces, Ω is open in X, $x_* \in \Omega$, and F is a map from Ω to Y, then a *linear-quadratic approximation* (abbreviated LQA) of F at x_* is a triple $A = (L, K, Q)$ such that

LQA1. L is a bounded linear map from X to Y,

LQA2. K is a closed linear subspace of X,

LQA3. Q is a continuous quadratic map from K to Y.

LQA4. $F(x_* + x + k) = F(x_*) + Lx + \frac{1}{2}Q(k) + o(\|k\|^2 + \|x\|)$ as (x, k) goes to zero via values in $X \times K$. $\diamond$

If (L, K, Q) is a LQA of F at x_*, then it is clear that F is Fréchet differentiable at x_* and $DF(x_*) = L$, $K \subseteq \ker L$, and the quadratic map Q is completely determined by K, by means of the formula.

$$Q(k) = 2 \lim_{t \downarrow 0} t^{-2}(F(x_* + tk) - F(x_*)) \qquad \text{for} \qquad x \in K.$$

The error bound of LQA4 is important because it gives an estimate of the the error in terms of the function $X \times K \ni (x, k) \mapsto \|x\| + \|k\|^2$, which is positively homogeneous of degree 1 relative to an appropriate group of dilations on $X \times K$. We make this observation precise in the following statement, whose proof is trivial.

Lemma 15 *Assume that (a) X, Y, Ω, x_*, F are as in Definition 14, and (b) $A = (L, K, Q)$ is a linear-quadratic approximation of F at x_*. Let*

$$\mathcal{X} \stackrel{\text{def}}{=} X \times K, \qquad \tilde{\Omega} = \left\{ (x, k) \in \mathcal{X} : x_* + x + k \in \Omega \right\},$$

and define maps $\nu : \mathcal{X} \mapsto \mathbb{R}$, $\delta_t : \mathcal{X} \mapsto \mathcal{X}$ for $t > 0$, $\mathcal{F} : \tilde{\Omega} \mapsto Y$, $\mathcal{G} : \mathcal{X} \mapsto \mathcal{Y}$, $\mathcal{E} : \tilde{\Omega} \mapsto Y$, by

$$
\begin{aligned}
\nu(x, k) &= \|x\| + \|k\|^2 && \text{for} \quad (x, k) \in \mathcal{X}, \\
\delta_t(x, k) &= (tx, \sqrt{t}k) && \text{for} \quad (x, k) \in \mathcal{X}, \ t > 0, \\
\mathcal{F}(x, k) &= F(x_* + x + k) - F(x_*) && \text{for} \quad (x, k) \in \tilde{\Omega}, \\
\mathcal{G}_A(x, k) &= Lx + \frac{1}{2}Q(k) && \text{for} \quad (x, k) \in \mathcal{X}, \\
\mathcal{E}(x, k) &= F(x_* + x + k) - F(x_*) - \mathcal{G}_A(x, k) \\
&= \mathcal{F}(x, k) - \mathcal{G}_A(x, k) && \text{for} \quad (x, k) \in \tilde{\Omega}.
\end{aligned}
$$

Then $\boldsymbol{\delta} = \{\delta_t\}_{t>0}$ is a continuous one-parameter group of dilations of $\mathcal{X}$, $\mathcal{G}_A$ is $\boldsymbol{\delta}$-homogeneous, ν is a $\boldsymbol{\delta}$-pseudonorm, and the error bound

$$\lim_{x \to 0, \, k \to 0} \frac{\|\mathcal{E}(x, k)\|}{\nu(x, k)} = 0 \tag{15}$$

is satisfied. $\diamond$

Definition 16 Assume that X, Y, Ω, x_*, F are as in Definition 14. A *strict linear-quadratic approximation* (abbreviated SLQA) of F at x_* is a triple $A = (L, K, Q)$ that satisfies conditions LQA1,2,3 of Definition 14 and is such that

$$\text{(SLQA)} \ \left(F(x_* + x + k) - Lx - \tfrac{1}{2}Q(k) \right) - \left(F(x_* + x' + k') - Lx' - \tfrac{1}{2}Q(k') \right)$$
$$= o\left(\|x - x'\| + (\sqrt{\|x\| + \|x'\|} + \|k\| + \|k'\|).\|k - k'\| \right) \text{ as } (x, k, x', k')$$
$$\text{goes to zero via values in } X \times K \times X \times K. \qquad \diamond$$

If $A = (L, K, Q)$ is a SLQA of F at x_*, then A is a LQA of F at x_*, F is strictly Fréchet differentiable at x_*, and $DF(x_*) = L$.

As in the case of (non-strict) LQAs, the error bound of condition (SLQA) is important because it estimates the error in terms of a function which is positively homogeneous of degree 1 relative to an appropriate group of dilations. We make this precise in the following statement, whose proof is trivial.

Lemma 17 *Let X, Y, Ω, x_*, F, A, L, K, Q, $\mathcal{X}$, $\tilde{\Omega}$, $\mathcal{F}$, $\mathcal{G}_A$, $\boldsymbol{\delta}$, $\mathcal{E}$ be as in the statement of Lemma 15. Assume that A is a strict linear-quadratic approximation of F at x_*. Define a map $\nu : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ by letting*

$$\nu(x, k, x', k') = \|x - x'\| + \left(\sqrt{\|x\| + \|x'\|} + \|k\| + \|k'\| \right) \cdot \|k - k'\| \qquad (16)$$

for $(x, k, x', k') \in \mathcal{X} \times \mathcal{X}$. Then

1. *$\boldsymbol{\delta}$ is a continuous one-parameter group of dilations of $\mathcal{X}$,*
2. *$\mathcal{G}_A$ is $\boldsymbol{\delta}$-homogeneous,*
3. *ν is a Lipschitz-bounded $\boldsymbol{\delta}$-pseudodistance,*
4. *the error bound*

$$\lim_{x \to 0, x' \to 0, x' \neq x} \frac{\|\mathcal{E}(x, k) - \mathcal{E}(x', k')\|}{\nu(x, k, x', k')} = 0 \qquad (17)$$

is satisfied. ◇

An important class of maps that necessarily admit strict linear-quadratic approximations consists of the *maps of class C^1 with a differentiable derivative*. Precisely, let us assume that

(#) *X, Y are normed spaces, Ω is an open subset of X, $F : \Omega \mapsto Y$ is a map of class C^1, $x_* \in \Omega$, and the map $\Omega \ni x \mapsto DF(x) \in \mathrm{Lin}(X, Y)$ is differentiable at x_*.*

Then DF is a continuous map from Ω to the space $\mathrm{Lin}(X, Y)$ of bounded linear maps from X to Y, and the second derivative $D(DF)(x_*) = D^2F(x_*)$ is a bounded linear map from X to $\mathrm{Lin}(X, \mathrm{Lin}(X, Y))$. Furthermore, *this map is symmetric*, i.e., $D^2F(x_*)(x) \cdot x' = D^2F(x_*)(x') \cdot x$ for all $x, x' \in X$. (This is true because of the identity

$$D^2F(x_*)(x) \cdot x' =$$
$$\lim_{\alpha \downarrow 0} \alpha^{-2} \Big(F(x_* + \alpha x + \alpha x') - F(x_* + \alpha x) - F(x_* + \alpha x') + F(x_*) \Big),$$

whose right-hand side is clearly symmetric under the interchange of x and x'.)

It follows that we can regard $D^2F(x_*)$ as a bounded symmetric bilinear map $B : X \times X \mapsto Y$, given by $B(x, x') = D^2F(x_*)(x) \cdot x'$. Then $B = B^{\mathcal{Q}}$, where $\mathcal{Q} : X \mapsto Y$ is the quadratic map given by $\mathcal{Q}(x) = B(x, x)$, so that $\mathcal{Q} = Q_B$ and $B = B^{\mathcal{Q}}$. Even more important for us will be the restriction Q of $\mathcal{Q}$ to the kernel $K = \ker L$, where $L = DF(x_*)$.

It turns out that the triple $A = (L, K, \mathcal{Q})$ is a strict linear-quadratic approximation of F at x_*, as we now show.

First, we write

$$
\begin{aligned}
M(x) &= L \cdot x + \frac{1}{2}\mathcal{Q}(x) \ , \\
\mathcal{G}_A(x, k) &= L \cdot x + \frac{1}{2}Q(k) \ , \\
E(x) &= F(x_* + x) - F(x_*) - M(x) \ , \\
\mathcal{E}(x, k) &= F(x_* + x + k) - F(x_*) - \mathcal{G}_A(x, k) \ ,
\end{aligned}
$$

for $x \in X$, $k \in K$.

Lemma 18 *Let X, Y, Ω, F, L, K, $\mathcal{Q}$, Q, M, $\mathcal{G}_A$, A, E, $\mathcal{E}$, be as above. Then*

$$
\lim_{x \to 0, x' \to 0, x' \neq x} \frac{\|E(x) - E(x')\|}{(\|x\| + \|x'\|)\|x - x'\|} = 0 \ , \tag{18}
$$

and A is a strict linear-quadratic approximation of F at x_*.

Proof. Without loss of generality, we assume that $x_* = 0$ and $F(x_*) = 0$. We fix a positive R such that $\mathbb{B}_X(R) \subseteq \Omega$. For $0 < r \leq R$, let

$$
\theta(r) = \sup\left\{ \frac{\|DF(x) - \mathcal{L} - B^{\mathcal{Q}}(x, \cdot)\|}{\|x\|} : 0 < \|x\| \leq r \right\} .
$$

Then θ is monotonically nondecreasing, and the differentiability assumption implies that $\lim_{r \downarrow 0} \theta(r) = 0$. Clearly, the bound

$$
\|DF(\xi).v - \mathcal{L}.v - B^{\mathcal{Q}}(\xi, v)\| \leq \theta(r)\,\|\xi\|\,\|v\| \tag{19}
$$

is satisfied whenever $0 < r \leq R$, $\xi \in \mathbb{B}_X(r)$, and $v \in X$.

22

Then, if $x, x' \in \mathbb{B}_X(R)$, and we write $v = x - x'$ $\xi_s = x' + sv$, we have

$$
\begin{aligned}
E(x) - E(x') &= F(x) - F(x') - \mathcal{L} \cdot v - \frac{1}{2}\Big(\mathcal{Q}(x) - \mathcal{Q}(x')\Big) \\
&= \Big(\int_0^1 DF(\xi_s)\,ds\Big) \cdot v - \mathcal{L} \cdot v - \frac{1}{2}\Big(\mathcal{Q}(x) - \mathcal{Q}(x')\Big) \\
&= \Big(\int_0^1 (DF(\xi_s) - \mathcal{L})\,ds\Big) \cdot v - \frac{1}{2}\Big(\mathcal{Q}(x) - \mathcal{Q}(x')\Big) \\
&= \int_0^1 \Big((DF(\xi_s) - \mathcal{L}).v - B^{\mathcal{Q}}(\xi_s, v)\Big)\,ds \\
&\quad + \int_0^1 B^{\mathcal{Q}}(\xi_s, v)\,ds - \frac{1}{2}\Big(B^{\mathcal{Q}}(x, x) - B^{\mathcal{Q}}(x', x')\Big) \\
&= \int_0^1 \Big((DF(\xi_s) - \mathcal{L}).v - B^{\mathcal{Q}}(\xi_s, v)\Big)\,ds \\
&\quad + B^{\mathcal{Q}}(x', v) + \frac{1}{2}B^{\mathcal{Q}}(v, v) - \frac{1}{2}\Big(B^{\mathcal{Q}}(x, x) - B^{\mathcal{Q}}(x', x')\Big) \\
&= \int_0^1 \Big((DF(\xi_s) - \mathcal{L}).v - B^{\mathcal{Q}}(\xi_s, v)\Big)\,ds,
\end{aligned}
$$

using the identities $B^{\mathcal{Q}}(x', v) + \frac{1}{2}B^{\mathcal{Q}}(v, v) - \frac{1}{2}\Big(B^{\mathcal{Q}}(x, x) - B^{\mathcal{Q}}(x', x')\Big) = 0$ and $\int_0^1 B^{\mathcal{Q}}(\xi_s, v)\,ds = B^{\mathcal{Q}}(x', v) + \frac{1}{2}B^{\mathcal{Q}}(v, v)$. Then (19), with $\xi = \xi_s$, yields

$$
\|(DF(\xi_s) - \mathcal{L}).v - B^{\mathcal{Q}}(\xi_s, v)\| \le \theta(\|x\| + \|x'\|) \cdot (\|x\| + \|x'\|) \cdot \|x - x'\|,
$$

since $\|\xi_s\| \le \|x\| + \|x'\|$ whenever $0 \le s \le 1$. Integrating this inequality, we get the bound

$$
\|E(x) - E(x')\| \le \theta(\|x\| + \|x'\|) \cdot (\|x\| + \|x'\|) \cdot \|x - x'\|, \tag{20}
$$

and (18) follows.

To prove that A is a strict linear-quadratic approximation of F at 0, we have to show that

$$
\mathcal{E}(x, k) - \mathcal{E}(x', k') = o\Big(\|x - x'\| + (\sqrt{\|x\| + \|x'\|} + \|k\| + \|k'\|).\|k - k'\|\Big)
$$

23

as $(x, x', k, k') \rightarrow (0, 0, 0, 0)$. But

$$
\begin{aligned}
\mathcal{E}(x, k) - \mathcal{E}(x', k') &= (F(x+k) - \mathcal{G}_A(x,k)) - (F(x'+k') - \mathcal{G}_A(x',k')) \\
&= \left(F(x+k) - L \cdot x - \frac{1}{2}Q(k) \right) - \left(F(x'+k') - L \cdot x' - \frac{1}{2}Q(k') \right) \\
&= \left(F(x+k) - L \cdot (x+k) - \frac{1}{2}\mathcal{Q}(x+k) \right) - \left(F(x'+k') - L \cdot (x'+k') \right. \\
&\quad \left. - \frac{1}{2}\mathcal{Q}(x'+k') \right) + \frac{1}{2} \left(\mathcal{Q}(x+k) - \mathcal{Q}(x'+k') - \mathcal{Q}(k) + \mathcal{Q}(k') \right) \\
&= E(x+k) - E(x'+k') + \frac{1}{2} \left(\mathcal{Q}(x+k) - \mathcal{Q}(x'+k') - \mathcal{Q}(k) + \mathcal{Q}(k') \right) \\
&= E(x+k) - E(x'+k') + \frac{1}{2} \left(\mathcal{Q}(x) - \mathcal{Q}(x') \right) + B^{\mathcal{Q}}(x,k) + B^{\mathcal{Q}}(x',k') \\
&= E(x+k) - E(x'+k') \\
&\quad + \frac{1}{2}\mathcal{Q}(x-x') + B^{\mathcal{Q}}(x', x'-x) + B^{\mathcal{Q}}(x-x', k) + B^{\mathcal{Q}}(x', k-k') \,.
\end{aligned}
$$

Therefore (20) implies, if we write $\kappa = \|B^{\mathcal{Q}}\|$, $\Theta(s) = s\theta(s)$, that

$$
\begin{aligned}
\|\mathcal{E}(x, k) - \mathcal{E}(x', k')\| &\leq \Theta \left(\|x\| + \|x'\| + \|k\| + \|k'\| \right) \cdot \left(\|x - x'\| + \|k - k'\| \right) \\
&\quad + \frac{\kappa}{2}\|x - x'\|^2 + \kappa\|x'\|.\|x - x'\| + \kappa\|x - x'\|.\|k\| + \kappa\|x'\|.\|k - k'\| \,,
\end{aligned}
$$

which is clearly $o\left(\|x - x'\| + (\sqrt{\|x\| + \|x'\|} + \|k\| + \|k'\|).\|k - k'\| \right)$. $\diamondsuit$

We can then apply Theorems 12 and 13 and obtain a number of open mapping theorems. The crucial condition in all of them is, of course, the existence of a regular zero (ξ_*, k_*) of the map $\mathcal{G}_A$ associated to the LQA $A = (L, K, Q)$. (Such a point will then automatically be a strictly regular zero, because $\mathcal{G}_A$ is a polynomial map.) It turns out that this condition is equivalent to a statement in terms of the "Hessian" H_A of A, as we now explain.

Let Z be the quotient space Y/LX, and let π be the canonical projection from Y onto Z. The *Hessian H_A* is the quadratic map

$$
K \ni k \mapsto \pi(Q(k, k)) \overset{\text{def}}{=} H_A(k) \in Z \,.
$$

Definition 19 A *regular zero* of H_A is a $k_* \in K$ such that $H_A(k_*, k_*) = 0$ and the map $K \ni k \mapsto \pi(B^Q(k_*, k)) \in Z$ is surjective. $\diamondsuit$

Remark 20 If LX is a closed subspace of Y then Z is a normed space and the quadratic map $H_A : K \mapsto Z$ is continuous, and hence smooth. In that case, k_* is a regular zero of H_A in the sense defined above if and only if it is a regular zero in the sense defined earlier in §2. Here, however, *we are not requiring LX to be closed*, and the concept of a "regular zero" of H_A has to be understood in the purely algebraic sense of Definition 19. $\diamond$

Lemma 21 *If $A = (L, K, Q)$ is a LQA of a map F at a point x_*, and $k_* \in K$, then k_* is a regular zero of H_A if and only if there exists $\xi_* \in X$ such that (ξ_*, k_*) is a regular zero of $\mathcal{G}_A$. In particular, H_A has a regular zero if and only if $\mathcal{G}_A$ has a regular zero.* $\diamond$

Proof. Recall that a regular zero of $\mathcal{G}_A$ is a pair $(\xi_*, k_*) \in X \times K$ such that $L\xi_* + \frac{1}{2}Q(k_*) = 0$ and the linear map $X \times K \ni (x, k) \mapsto L\,x + B^Q(k_*, k) \in Y$ is surjective.

If k_* is a regular zero of the Hessian H_A, then $\pi(Q(k_*)) = 0$, so $Q(k_*)$ belongs to LX, and then there exists $\xi_* \in X$ such that $\frac{1}{2}Q(k_*) + L \cdot \xi_* = 0$. Therefore $\mathcal{G}_A(\xi_*, k_*) = 0$, and $D\mathcal{G}_A(\xi_*, k_*)(x, k) = L \cdot x + B^Q(k_*, k)$ if $x \in X$, $k \in K$. If $y \in Y$, then the surjectivity of the map $K \ni k \mapsto \pi(B^Q(k_*, k)) \in Z$ implies that there exists a $k \in K$ such that $\pi(B^Q(k_*, k)) = \pi(y)$, i.e., that $y - B^Q(k_*, k)$ belongs to LX. It follows that there exists $x \in X$ such that $y = B^Q(k_*, k) + Lx = D\mathcal{G}_A(h_*, k_*)(x, k)$. Since y is an arbitrary member of Y, we have shown that the linear map $D\mathcal{G}_A(\xi_*, k_*) : X \times K \mapsto Y$ is surjective, so (ξ_*, k_*) is a regular zero of $\mathcal{G}_A$.

Conversely, if (ξ_*, k_*) is a regular zero of $\mathcal{G}_A$, then $\frac{1}{2}Q(k_*) + L \cdot \xi_* = 0$, so $Q(k_*) \in LX$, and then $H_A(k_*) = 0$. Furthermore, if $v \in Z$, and $y \in Y$ is such that $\pi(y) = v$, then we can write $y = D\mathcal{G}_A(\xi_*, k_*)(x, k) = Lx + B^Q(k_*, k)$ for some $x \in X$, $k \in K$. But then $v = \pi(y) = \pi(B^Q(k_*, k))$. Therefore the map $K \ni k \mapsto \pi(B^Q(k_*, k)) \in Z$ is surjective, and we have shown that k_* is a regular zero of H_A. $\diamond$

Theorem 22 *Assume that X, Y are normed spaces, Y is finite-dimensional, Ω is open in X, $x_* \in \Omega$, $F : \Omega \mapsto Y$ is a continuous map, and $A = (L, K, Q)$ is a linear-quadratic approximation of F at x_* such that the Hessian H_A has a regular zero. Then F is open at x_*.*

Theorem 23 *Assume that X, Y are Banach spaces, Ω is open in X, $x_* \in \Omega$, $F : \Omega \mapsto Y$ is a continuous map and $A = (L, K, Q)$ is a strict linear-quadratic approximation of F at x_* such that the Hessian H_A has a regular zero. Then F is open at x_*.*

In particular, we can take F to be a map of the kind considered in Lemma 18. In that case, the *Hessian of F at x_** is the Hessian of the strict LQA (L, K, Q), where $L = DF(x_*)$, $K = \ker L$, and Q is the quadratic map $K \ni k \mapsto D^2 F(x_*)(k, k)$.

Theorem 24 *Assume that X, Y are Banach spaces, Ω is open in X, $x_* \in \Omega$, $F : \Omega \mapsto Y$ is a map of class C^1 such that the derivative DF is differentiable at x_* and the Hessian H of F at x_* has a regular zero. Then F is open at x_*.*

Remark 25 Theorem 24 is very similar to the result proved by Avakov (cf. [1, 2]). Avakov's sufficient condition for openness—called "2-regularity" by some authors, e.g., Ledzewicz and Schättler, cf. [5, 6]—is fomulated in slightly different terms, but is easily seen to be equivalent to the existence of a regular zero of the Hessian.

Precisely, the algebraic part of Avakov's condition says—using

$$L = DF(x_*)(k_*, k_*), \quad K = \ker L,$$

and writing $\mathcal{Q}$ for the map $X \ni x \mapsto D^2 F(x_*)(x, x) \in Y$, and Q for the restriction of $\mathcal{Q}$ to K—that

Av1. $Lk_* = 0$,

Av2. $\mathcal{Q}(k_*) \in LX$,

and

Av3. *the map $X \ni x \mapsto \mathcal{A}(x) \overset{\text{def}}{=} (Lx, \pi(B^{\mathcal{Q}}(k_*, x))) \in LX \times Z$ is surjective.*

Lemma 26 *The algebraic part of Avakov's condition holds if and only if k_* is a regular zero of the Hessian.*

Proof. If Avakov's condition holds, then of course $k_* \in K$, so $Q(k_*) \in LX$, and then $H(k_*) = 0$. Furthermore, the surjectivity of $\mathcal{A}$ implies, in particular, that given any $z \in Z$ there exists a $k \in X$ such that $(Lk, \pi(B^{\mathcal{Q}}(k_*, k))) = (0, z)$. But then $Lk = 0$, so $k \in K$, and $\pi(B^Q(k_*, k)) = z$. This shows that the map $K \ni k \mapsto \pi(B^Q(k_*, k)) \in Z$ is surjective, so k_* is a regular zero of H.

Conversely, suppose that k_* is a regular zero of H. Then $\pi(Q(k_*)) = 0$, so $\mathcal{Q}(k_*) \in LX$. We now show that $\mathcal{A}$ is surjective. Pick $(y, z) \in LX \times Z$. Write $y = Lx$, $x \in X$. Let $v' = B^{\mathcal{Q}}(k_*, x)$, $z' = \pi(v')$. Then the fact that the map $K \ni k \mapsto \pi(B^Q(k_*, k)) \in Z$ is surjective implies that there exists a $k \in K$ such that $\pi(B^Q(k_*, k)) = z - z'$. Then $\pi(B^{\mathcal{Q}}(k_*, x + k)) = z' + (z - z') = z$, and $L(x + k) = Lx = y$, since $Lk = 0$. So $\mathcal{A}(x + k) = (y, z)$. Hence $\mathcal{A}$ is surjective, and our proof is complete. $\diamond$

We point out, however, that in the work of Avakov and Ledzewicz-Schättler it also required, in addition to the algebraic condition described above, that the space LX be closed, whereas we do not need to make that extra requirement, because in our framework the purely algebraic condition on the Hessian suffices to obtain the openness theorem. ◇

References

[1] Avakov, E.R., "Extremum conditions for smooth problems with equality-type constraints." (Russian) *Zh. Vychisl. Mat. Mat. Fiz.* **25**, No. 5, 1985, pp. 680-693. Eng. translation in *U.S.S.R. Comput. Maths. Math. Phys.* **25**, No. 3 (Pergamon Press), 1985, pp. 24-32.

[2] Avakov, E.R., "Necessary conditions for an extremum for smooth abnormal problems with constraints of equality and inequality type." (Russian) *Matematicheskie Zametki* **45**, No. 6, 1989, pp. 3-11. Eng. translation in *Math. Notes* **47**, no. 5-6, 1990, pp. 425-432.

[3] Dontchev, A.L., *The Graves theorem revisited.* J. Convex Analysis **3**, 1996, pp. 45-53.

[4] Graves, L. M., *Some mapping theorems.* Duke Math. J. **17**, 1950, pp. 111–114.

[5] Ledzewicz, U., and H. Schättler, "An extended maximum principle." *Nonlinear Analysis, Theory, Methods and Applications* **29**, 1997, pp. 159-183.

[6] Ledzewicz, U., and H. Schättler, "A high-order generalized local maximum principle." *SIAM J. Control Optim.* **38**, 2000, pp. 823-854.