MINIMAL PERTURBATIONS TO ROOTS OF PARAMETERIZED EQUATIONS

JOSEPH F. GR CAR∗

Abstract. The size of minimal perturbations to roots of parameterized equations can be estimated reliably from linearizations of the equations.

Key words. parametric optimization, perturbation analysis

AMS subject classifications. 90C31, 26B05, 41A29, 46N10

1. Introduction. This paper offers a systematic way to answer the question: how much change must occur in a solution of equations to compensate for perturbations to the equations? Short of finding all the nearby roots of the new equations, the minimal change can be determined in an asymptotic sense by linearizing the equations and considering the dual problems. This conclusion is exhaustive because all nearby roots are considered, and strong because the asymptotics imply differential approximations.

The asymptotic relationship is proved here. Companion papers make applications to differentiability of best approximations and to numerical analysis.

2. Approach.

2.1. Introduction. Let the equations be $F(y, x) = 0$ with the specific root $(y_0, x_0)$. The variable $x$ is regarded as the parameter so that $y$ depends on $x$ constrained by $F(y, x) = 0$. The equations for $y$ may be underdetermined so $y$ may not be a function of $x$. Nevertheless, the size of minimal perturbations to $y_0$ is a function,

$$\mu_F(x) = \min_{y : F(y, x) = 0} \|y - y_0\|.$$ (2.1)

The idea is to study the value of this optimization problem by linearizing the equations. There are two requirements for the altered problems:

1. The values of the simplified problems should mimic how $\mu_F(x)$ varies with $x$.
2. Since $\mu_F(x)$ is of interest when $x \approx x_0$, good mimicry is needed near $x_0$.

The novelty of the present approach is to formalize these requirements by equivalence relations, $\equiv$, among functions of $x$; two equivalences are chosen in section 2.2. Problem (2.1) is then altered by linearizing $F$; three linearizations, $F^{(i)}$, are constructed in section 2.3. The bulk of the paper establishes equivalences $\mu_F \equiv \mu_{F^{(i)}}$. For simplicity, the values of the altered problems are written $\mu_{F^{(i)}} = \mu_i$.

2.2. Equivalence Relations. The following equivalence relation is appropriate when differentiability at $x_0$ is the object of study.

Definition 2.1 (Differential equivalence). The functions $f$ and $g$ defined on a neighborhood of $x_0 \in \mathbb{R}^n$ with values in $\mathbb{R}^p$ are differentially equivalent at $x_0$ provided $f - g$ has a Fréchet derivative of 0 at $x_0$, equivalently,

$$f \equiv_{x_0} g \iff \lim_{x \to x_0} \frac{\|f(x) - g(x)\|}{\|x - x_0\|} = 0.$$ (2.2)

Lemma 2.2. Differential equivalence is an equivalence relation. (This lemma is clear and not proved.)

∗6059 Castlebrook Drive, Castro Valley, CA 94552 USA (jfgrcar@comcast.net).
If \( g \) is an affine function, then equation (2.2) becomes the definition for the Fréchet derivative of \( f \) at \( x_0 \). In this way the differential properties of \( f \) at \( x_0 \) are determined by the differential equivalence class.

A simpler but stronger equivalence relation is that real-valued functions should be relatively closer as \( x \) approaches \( x_0 \).

**Definition 2.3 (Asymptotic equality).** The real-valued functions \( f \) and \( g \) defined on a neighborhood of \( x_0 \in \mathbb{R}^n \) are asymptotically equal at \( x_0 \) provided for every \( \epsilon > 0 \) there is a neighborhood \( N(\epsilon) \) of \( x_0 \) such that \( x \in N(\epsilon) \) implies

\[
(2.3) \quad f \equiv^\epsilon x_0 \ g \iff (1 - \epsilon)g(x) \leq f(x) \leq (1 + \epsilon)g(x).
\]

**Lemma 2.4.** Asymptotic equality is an equivalence relation. (This lemma is clear and not proved.)

Asymptotic equality is stronger than differential equivalence. For example, all functions with vanishing derivatives at 0 are differentially equivalent there, but two monomials \( c_1 x^{m_1} \) and \( c_2 x^{m_2} \) are asymptotically equal at 0 if and only if they are equal.

For the function \( \mu_F(x) \) in equation (2.1), asymptotic equality implies differential equivalence. The proof of this implication in lemma 2.6 depends on a modified implicit function theorem in lemma 2.5, and on the Lipschitz continuity of \( \mu_F(x) \) at \( x_0 \).

**Hypothesis 2.1.** Hypothesis 1–4 are used throughout this paper, while 5 or 6 are used occasionally.

1. Norms are given for \( \mathbb{R}^m, \mathbb{R}^n \) and \( \mathbb{R}^p \).
2. \( \mathcal{D} \subseteq \mathbb{R}^m \times \mathbb{R}^n \) is a neighborhood of \( (y_0, x_0) \).
3. \( F : \mathcal{D} \to \mathbb{R}^p \) is continuously Fréchet differentiable.
4. \( F(y_0, x_0) = 0 \).
5. \( D_1F(y_0, x_0) : \mathbb{R}^m \to \mathbb{R}^p \) is onto.
6. \( D_2F(y_0, x_0) : \mathbb{R}^n \to \mathbb{R}^p \) is one-to-one.

**Lemma 2.5 (Modified implicit function theorem).** Under hypotheses 2.1 (1–5), there is a neighborhood \( N \) of \( x_0 \) and a Fréchet differentiable function \( \phi : N \to \mathbb{R}^m \) with \( \phi(x_0) = y_0 \) and \( F(\phi(x), x) = 0 \) for all \( x \in N \).

**Proof:** The proof applies the usual theorem, which requires that \( D_1F(y_0, x_0) \) be one-to-one. In the present case the mapping is onto, so there are \( p \) vectors in \( \mathbb{R}^m \) that map to linearly independent vectors in \( \mathbb{R}^p \), and there are \( m - p \) additional vectors that complete a basis for \( \mathbb{R}^m \). Let \( y = y^{(p)} + y^{(m-p)} \) be the decomposition of \( y \in \mathbb{R}^m \) into the subspaces spanned by the respective sets of basis vectors. \( D_1F(y_0, x_0) \) restricted to \( (\mathbb{R}^m)^{(p)} \) is one-to-one. This fact and hypotheses 2.1 (1–4) suffice to invoke the implicit function theorem for the function defined by \( F(y + y_0^{(m-p)}, x) \) on the domain \( \mathcal{D} \cap [(\mathbb{R}^m)^{(p)} \times \mathbb{R}^n] \). There is a neighborhood \( N \) of \( x_0 \) in \( \mathbb{R}^n \) on which there is a continuously differentiable function \( \phi : N \to (\mathbb{R}^m)^{(p)} \) such that \( \phi(x_0) = y_0^{(p)} \) and \( F(\phi(x) + y_0^{(m-p)}, x) = 0 \). The implicit function in the lemma is given by \( \phi(x) + y_0^{(m-p)} \).

**Lemma 2.6 (Existence of \( \mu_F(x) \) and properties).** Under hypotheses 2.1 (1–5), there is a constant \( L > 0 \) and a neighborhood \( N_{x_0}^{(2,6)} \) of \( x_0 \) where the function \( \mu_F(x) \) of equation (2.1) exists, and \( \mu_F(x) \leq L\|x - x_0\| \). Further, for any function \( f \),

\[
f \equiv^\epsilon x_0 \mu_F \implies f \equiv^\epsilon x_0 \mu_F.
\]

**Proof:** Hypotheses 2.1 (1–5) suffice to invoke the version of the implicit function theorem in lemma 2.5: \( x_0 \) has a neighborhood \( N \) on which there is a continuously differentiable function \( \phi : N \to \mathbb{R}^m \) such that \( (\phi(x), x) \) is always a root of \( F \). Thus the minimization problems for \( \mu_F(x) \) have feasible points for all \( x \in N \). The feasible sets are closed because \( F \) is continuous, so the minimal distance to \( y_0 \) is attained because the spaces have finite
dimension. This means $\mu_F$ is well defined on $N$. Since $\phi$ is continuously differentiable, it is Lipschitz continuous on compact sets. Choose a compact neighborhood $N_{x_0}^{(2.6)} \subseteq N$ with Lipschitz constant $L$. Thus $\mu_F(x) \leq \|\phi(x) - y_0\| = \|\phi(x) - \phi(x_0)\| \leq L\|x - x_0\|$ for every $x$ in the neighborhood.

Given $\epsilon > 0$, let $N(\epsilon)$ be the neighborhood in definition 2.3 for $f \equiv \frac{\partial}{\partial x_0} \mu_F$. If $x \in N(\epsilon) \cap N_{x_0}^{(2.6)}$, then $(1 - \epsilon)\mu_F(x) \leq f(x) \leq (1 + \epsilon)\mu_F(x)$ by the equivalence, so $|f(x) - \mu_F(x)| \leq \epsilon \mu_F(x) \leq \epsilon L\|x - x_0\|$ and thus the limit in equation (2.2) vanishes.  

2.3. Linearized Problems with Equivalent Minimal Perturbations. It is instructive to compare the present situation with the implicit function theorem. Under hypotheses 2.1 (1–4) and if $D_1F(y_0, x_0) : \mathbb{R}^m \rightarrow \mathbb{R}^p$ is invertible, then some roots of $F(y, x) = 0$ are given by a smooth parameterization $(\phi(x), x)$. These roots can be located to first order in $x - x_0$ by considering the linearization,

\[(2.4) \quad F(\phi(x), x) = 0 \quad \Rightarrow \quad [D_1F(y_0, x_0) D\phi(x) + D_2F(y_0, x_0)] (x - x_0) = 0.\]

The parameterized roots are approximated by,

\[\phi(x) - y_0 \approx -[D_1F(y_0, x_0)]^{-1} [D_2F(y_0, x_0)] (x - x_0).\]

In contrast, if $D_1F(y_0, x_0) : \mathbb{R}^m \rightarrow \mathbb{R}^p$ is not invertible, the smallest change $y - y_0$ as a function of $x$ can still be approximated from the linearizations $F^{(i)}$ of $F$ in Table 2.1.

| $i$ | $F^{(i)}(y, x)$ |
|-----|----------------|
| 0   | $F(y, x)$      |
| 1   | $F^{(1)}(y, x) = D_1F(y_0, x_0) \Delta y + F(y_0, x)$ |
| 2   | $F^{(2)}(y, x) = D_1F(y_0, x_0) \Delta y + F(y_0, x)$ |
| 3   | $F^{(3)}(y, x) = D_1F(y_0, x_0) \Delta y + D_2F(y_0, x_0) \Delta x$ |

The different linearizations have different uses. For example, $F^{(1)}$ and (2.1) do not require $x_0$. The several approximations are treated in a progression of equivalences for $F$ and $F^{(1)}$, then $F^{(1)}$ and $F^{(2)}$, and so on. The last $F^{(3)}$ is the full linearization (2.4) of the implicit function theorem. The proof of asymptotic equality for $F^{(i)}$ and $F^{(i+1)}$ is carried out with the dual mathematical programs. All the optimization problems are listed in Table 2.2, and the network of equivalences to be established is shown in Figure 2.1.

If $F$ satisfies hypotheses 2.1 (1–5), then all the functions of Table 2.1 satisfy the same hypotheses, so they also satisfy the conclusions of lemma 2.6.

**Corollary 2.7 (Existence of $\mu_i(x)$ and properties).** Under hypotheses 2.1 (1–5) for $F$, for each function $F^{(i)}$ of Table 2.1 there is a constant $L > 0$ and a neighborhood $N_{x_0}^{(2.7, i)}$ of $x_0$ where the following distance function is well defined

\[(2.5) \quad \mu_i(x) = \min_{y : F^{(i)}(y, x) = 0} \|y - y_0\|,\]

and $\mu_i(x) \leq L\|x - x_0\|$. Further, for any function $f$,

\[f \equiv \frac{\partial}{\partial x_0} \mu_i \quad \Rightarrow \quad f \equiv \frac{\partial}{\partial x_0} \mu_i.\]

**Proof.** The linearizations satisfy the same hypotheses as $F$, so lemma 2.6 applies.
Table 2.2
Optimization problems parameterized by $x$ and their duals. The values of problems ($P$), ($P_1$), ($P_2$) are asymptotically equal at $x_0$ under hypotheses 2.1 (1–5). The value of ($P_3$) is differentially equivalent to the other under these hypotheses, and is asymptotically equal under hypotheses 2.1 (1–6). See Table 7.1 for the problems in matrix notation. In these formulas, $\Delta x = x - x_0$ and $\Delta y = y - y_0$.

| name | value | constraint function | minimization form | dual, maximization form |
|------|------|-------------------|-----------------|------------------------|
| ($P$) | $\mu_F(x)$ | $F(y, x)$ | $\min_{y : F(y, x) = 0} \|\Delta y\|$ | $\max_f(F(y_0, x))$ |
| ($P_1$) | $\mu_1(x)$ | $F^{(1)}(y, x)$ | $\min_{y : D_1 F(y_0, x)\Delta y + F(y_0, x) = 0} \|\Delta y\|$ | $\max_f(F(y_0, x))$ |
| ($P_2$) | $\mu_2(x)$ | $F^{(2)}(y, x)$ | $\min_{y : D_1 F(y_0, x_0)\Delta y + F(y_0, x) = 0} \|\Delta y\|$ | $\max_f(F(y_0, x))$ |
| ($P_3$) | $\mu_3(x)$ | $F^{(3)}(y, x)$ | $\min_{y : DF(y_0, x_0)(\Delta y, \Delta x) = 0} \|\Delta y\|$ | $\max_f(D_2 F(y_0, x_0)\Delta x)$ |

3. First Equivalence, $(P)_{\min} \equiv (P_1)_{\min}$. The preparations to establish the first equivalence are the most elaborate in this paper. Several aspects of the difference between $F$ and the tangent function for $y$, $F^{(1)}$, are uniform in $x - x_0$: mean values, Fréchet differentials, and level sets. The first equivalence thus requires giving a uniform parameterization to many basic concepts in real analysis, which are indicated in Figure 3.1. The mean value theorem and Fréchet quotient are discussed in section 3.1, the matrix lower bound is in section 3.2, level sets are in section 3.3, and finally the proof of the first equivalence is in section 3.4.
3.1. Uniformly Parameterized Mean Value Theorem. It is well known that if \( f \) is continuously differentiable, then for every \( y_3 \) and every \( \epsilon > 0 \) there is a neighborhood \( N_{y_3}(\epsilon) \) of \( y_3 \) where

\[
    (3.1) \quad y_1, y_2 \in N_{y_3}(\epsilon) \Rightarrow \|f(y_1) - f(y_2) - Df(y_3)(y_1 - y_2)\| \leq \epsilon \|y_1 - y_2\|.
\]

This serves as a mean value theorem in multiple dimensions. Luenberger [4, p. 212] remarks that it has been discussed many times. Bartle [1, p. 377] calls (3.1) the "key lemma" for theorems like the implicit function theorem. Ortega and Rheinboldt [5, p. 72] show that (3.1) is equivalent to the continuity of the derivative. Here, this surrogate mean value theorem is generalized to parameterized functions.

**Lemma 3.1** (Uniformly parameterized mean value theorem). Under hypotheses 2.1 (1–5), for every \( \epsilon > 0 \) there is a neighborhood \( N_{y_0}^{(3,1)}(\epsilon) \times N_{x_0}^{(3,1)}(\epsilon) \subseteq D \) such that for all \( y_1, y_2, y_3 \in N_{y_0}^{(3,1)}(\epsilon) \) and \( x \in N_{x_0}^{(3,1)}(\epsilon) \),

\[
    (3.2) \quad \|F(y_1, x) - F(y_2, x) - D_1F(y_3, x)(y_1 - y_2)\| \leq \epsilon \|y_1 - y_2\|.
\]

**Proof.** The topology of the product space \( \mathbb{R}^m \times \mathbb{R}^n \) can be generated from the products of the open sets, so it is possible to choose a compact, convex neighborhood \( Y \) around \( y_0 \), and a compact neighborhood \( X \) around \( x_0 \), so that \( Y \times X \subseteq D \). All norms for a finite dimensional space generate the same topology, so without loss of generality let the norm for \( \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^n \) be \( \max \{\|y_1\|, \|y_2\|, \|x\|\} \). Since \( D_1F(y, x) \) is continuous, hence \( g(y_1, y_2, x) = D_1F(y_1, x) - D_1F(y_2, x) \) is uniformly continuous on the compact set \( K = Y \times X \). The uniform continuity means, for every \( \epsilon > 0 \) there is a \( \delta(\epsilon) > 0 \) so that if \( (y_1, y_2, x), (y'_1, y'_2, x') \in K \) with \( \max \{\|y_1 - y'_1\|, \|y_2 - y'_2\|, \|x - x'\|\} \leq \delta(\epsilon) \), then \( \|g(y_1, y_2, x) - g(y'_1, y'_2, x')\| \leq \epsilon \).

Choose the neighborhoods in the statement of the lemma to be \( N_{y_0}^{(3,1)}(\epsilon) = B_{y_0}(\delta(\epsilon)) \cap Y \) and \( N_{x_0}^{(3,1)}(\epsilon) = B_{x_0}(\delta(\epsilon)) \cap X \). Note, these sets are convex. If \( y_1, y_2, y_3 \) and \( x \) are from the respective sets, then

\[
    \|D_1F(ty_1 + (1 - t)y_2, x) - D_1F(y_3, x)\| = \|g(ty_1 + (1 - t)y_2, y_3, x)\|
\]

Choose the neighborhoods in the statement of the lemma to be \( N_{y_0}^{(3,1)}(\epsilon) = B_{y_0}(\delta(\epsilon)) \cap Y \) and \( N_{x_0}^{(3,1)}(\epsilon) = B_{x_0}(\delta(\epsilon)) \cap X \). Note, these sets are convex. If \( y_1, y_2, y_3 \) and \( x \) are from the respective sets, then

\[
    \|D_1F(ty_1 + (1 - t)y_2, x) - D_1F(y_3, x)\| = \|g(ty_1 + (1 - t)y_2, y_3, x)\|
\]

\[
    \leq \epsilon.
\]
It is well known from [1, p. 376, lemma 41.3] or from [5, p. 70, lemma 3.2.5] that if \( D \subseteq \mathbb{R}^m \) is a convex, open set, and if \( f : D \to \mathbb{R}^p \) is continuously differentiable, then for any \( y_1, y_2, y_3 \in D \),
\[
\|f(y_1) - f(y_2) - Df(y_3)(y_1 - y_2)\| \leq \sup_{0 \leq t \leq 1} \|Df(ty_1 + (1-t)y_2) - Df(y_3)\| \|y_1 - y_2\| .
\]

Applying this inequality to the parameterized function \( F(y, x) \) for the previously chosen \( y_1, y_2, y_3, x \) gives
\[
\|F(y_1, x) - F(y_2, x) - D_1F(y_3, x)(y_1 - y_2)\| \\
\leq \sup_{0 \leq t \leq 1} \|D_1F(ty_1 + (1-t)y_2, x) - D_1F(y_3, x)\| \|y_1 - y_2\| \\
\leq \epsilon \|y_1 - y_2\| .
\]

Lemma 3.1 gives conditions under which the Fréchet differential for \( y \) is uniformly approximating with respect to the parameter \( x \).

**Corollary 3.2 (Uniformly approximating differential).** The neighborhoods of lemma 3.1 also satisfy, for all \( y \in N_{x_0}^{(3.1)}(\epsilon) \) and \( x \in N_{y_0}^{(3.1)}(\epsilon) \),
\[(3.3)
\|F(y, x) - F^{(1)}(y, x)\| \leq \epsilon \|y - y_0\| ,
\]
where \( F^{(1)}(y, x) \) is the parameterized tangent function of Table 2.1.

**Proof.** Choose \( y_1 = y, y_2 = y_0 \) and \( y_3 = y_0 \) so that the formula in equation (3.2) becomes
\[
F(y_1, x) - F(y_2, x) - D_1F(y_3, x)(y_1 - y_2) = F(y, x) - F(y_0, x) - D_1F(y_0, x)(y - y_0) \\
= F(y, x) - F^{(1)}(y, x) .
\]

**3.2. Matrix Lower Bound.** The matrix lower bound, \( \|A\|_\epsilon \), is analogous to the matrix norm but with reversed inequalities. The following are from [3, p. 205, def. 2.1 and lem. 2.2; p. 212, cor. 4.3].

**Definition 3.3 (Matrix lower bound).** Let \( A \) be a nonzero matrix. The matrix lower bound, \( \|A\|_\epsilon \), is the largest of the numbers, \( m \), such that for every \( y \) in the column space of \( A \), there is some \( x \) with \( Ax = y \) and \( m \|x\| \leq \|y\| .
\]

**Lemma 3.4.** The matrix lower bound exists and is positive for every nonzero matrix.

**Lemma 3.5.** The matrix lower bound is continuous on the open set of full rank matrices.

The present use of the lower bound is in the following lemma.

**Lemma 3.6 (Uniform lower bounds for partial derivatives).** Under hypotheses 2.1 (1–5), there is a neighborhood \( N_{x_0}^{(3.6)} \) of \( x_0 \) where \( D_1F(y_0, x) : \mathbb{R}^m \to \mathbb{R}^p \) is onto for every \( x \in N_{x_0}^{(3.6)} \). There is also a number \( m^{(3.6)} > 0 \) such that every \( x \in N_{x_0}^{(3.6)} \) and \( u \in \mathbb{R}^p \) have some \( w \in \mathbb{R}^m \) (which depends on \( x \) and \( u \)) so that \( D_1F(y_0, x)w = u \) and \( m^{(3.6)} \|w\| \leq \|u\| .
\]

**Proof.** Choose some bases for \( \mathbb{R}^m \) and \( \mathbb{R}^p \) so that these spaces are represented by real column vectors. The linear transformations \( D_1F(y_0, x) \) are then represented by \( p \times m \) matrices, \( A(x) \). By Hypothesis 2.1 (3) \( F \) is continuously differentiable and (5) \( D_1F(y_0, x_0) \) is onto, which means \( A(x) \) is a continuous function of \( x \) and the column space of \( A(x_0) \) is all of \( \mathbb{R}^p \), or equivalently \( A(x_0) \) has full row rank. For a matrix \( M \) to have full row rank means \( \det(MM') \) does not vanish. The determinant is a continuous function of the matrix,
so $A(x_0)$ has a neighborhood of matrices $N_{A(x_0)}$ all of which have full row rank. From the continuity of $A(x)$, there is a neighborhood $N_{x_0}$ for which all matrices lie in $N_{A(x_0)}$. Hence for all $x \in N_{x_0}$ the mappings $D_1F(y_0, x)$ are onto, or equivalently the column space of each matrix $A(x)$ is all of $\mathbb{R}^p$.

Choose a compact neighborhood $N_{x_0}^{(3.6)} \subseteq N_{x_0}$. Since $\|A(x)\|_\epsilon$ is continuous and positive on $N_{x_0}$ by lemmas 3.4 and 3.5, $\|A(x)\|_\epsilon$ is uniformly bounded below on $N_{x_0}^{(3.6)}$ by some $m^{(3.6)} > 0$. If $x \in N_{x_0}^{(3.6)}$ and $u \in \mathbb{R}^p$, then since the column space of $A(x)$ is all of $\mathbb{R}^p$, by definition 3.3 there is $w \in \mathbb{R}^m$ so $D_1F(y_0, x)w = A(x)w = u$ and $\|A(x)\|_\epsilon \|w\| \leq \|u\|$. Further, $m^{(3.6)} \leq \|A(x)\|_\epsilon$ by the choice of $N_{x_0}^{(3.6)}$. □

3.3. Uniformly Colocated Level Sets. Suppose $D$ is an open set in $\mathbb{R}^m$, on which $f : D \to \mathbb{R}^p$ is continuously differentiable. By analogy with real-valued functions, the set $f^{-1}(a)$ may be called a level set of $f$. It is possible to make a geometric comparison between the level sets of $f$ and those of its tangent function at $y_0$. For functions such as $F$ that vary smoothly with a parameter, the distance between the corresponding level sets is uniformly bounded with respect to changes in the parameter. The proof is a modification of a construction apparently due to L. M. Graves [2], see also [1, p. 378, theorem 41.6].

**Lemma 3.7** (Uniformly colocated level sets). Under hypotheses 2.1 (1–5), for every $\epsilon > 0$ there is a radius $r(\epsilon) > 0$ and a neighborhood $N_{x_0}^{(3.7)}(\epsilon)$ of $x_0$ so $\text{cl}(B_{y_0}(r(\epsilon))) \times N_{x_0}^{(3.7)}(\epsilon) \subseteq D$. For each pair $(y, x) \in B_{y_0}(r(\epsilon)/(1 + \epsilon)) \times N_{x_0}^{(3.7)}(\epsilon)$:

| (a) there exists | (b) with | (c) and with |
|------------------|----------|-------------|
| $(1)$ $y_1 \in B_{y_0}(r(\epsilon))$ | $F^{(1)}(y_1, x) = F(y, x)$ | $\|y_1 - y\| \leq \epsilon \|y - y_0\|$ |
| $(2)$ $y_F \in \text{cl}(B_{y_0}(r(\epsilon)))$ | $F(y_F, x) = F^{(1)}(y, x)$ | $\|y_F - y\| \leq \epsilon \|y - y_0\|$ |

**Proof.** Lemma 3.7 has the first of the two most complicated proofs in this paper. Let $\delta = \epsilon/(1 + \epsilon) < 1$. Let $m^{(3.6)}$ be the lower bound for the neighborhood $N_{x_0}^{(3.6)}$ in lemma 3.6. Choose a radius $r(\epsilon) > 0$ so that

$$
(3.4) \quad \text{cl}(B_{y_0}(r(\epsilon))) \subseteq N_{y_0}^{(3.1)}(\delta m^{(3.6)}).
$$

The neighborhoods from which the lemma is allowed to choose $y$ and $x$ are

$$
(3.5) \quad y \in B_{y_0}(r(\epsilon)/(1 + \epsilon)) \subseteq \text{cl}(B_{y_0}(r(\epsilon))) \subseteq N_{y_0}^{(3.1)}(\delta m^{(3.6)}),
$$

$$
(3.6) \quad x \in N_{x_0}^{(3.7)}(\epsilon) := N_{x_0}^{(3.6)} \cap N_{x_0}^{(3.1)}(\delta m^{(3.6)}) \subseteq N_{x_0}^{(3.1)}(\delta m^{(3.6)}).
$$

Note the product $B_{y_0}(r(\epsilon)) \times N_{x_0}^{(3.7)}(\epsilon)$ is a subset of $N_{y_0}^{(3.1)}(\delta m^{(3.6)}) \times N_{x_0}^{(3.1)}(\delta m^{(3.6)})$ in $D$ by lemma 3.1.

(Part 1.) Because $D_1F(y_0, x) : \mathbb{R}^m \to \mathbb{R}^p$ is onto, the range of the transformation contains the vector $F(y, x) - F^{(1)}(y, x)$, and because $x \in N_{x_0}^{(3.6)}$ by (3.6), lemma 3.6 finds a $\hat{y}$ with

$$
(3.7) \quad D_1F(y_0, x) \hat{y} = F(y, x) - F^{(1)}(y, x),
$$

$$
(3.8) \quad m^{(3.6)} \|\hat{y}\| \leq \|F(y, x) - F^{(1)}(y, x)\|.
$$

Let $y_1 = \hat{y} + y$ so $\hat{y} = y_1 - y$. The equality (3.7) and some algebra imply

$$
F^{(1)}(y_1, x) = D_1F(y_0, x)(y_1 - y_0) + F(y_0, x) \quad \text{by definition of } F^{(1)} \text{ in Table 2.1}
$$

$$
= [D_1F(y_0, x)(y_1 - y)] + [D_1F(y_0, x)(y - y_0) + F(y_0, x)] \quad \text{inserting } \pm y
$$

$$
= [F(y, x) - F^{(1)}(y, x)] + F^{(1)}(y, x) \quad \text{by (3.7) and by definition of } F^{(1)}
$$

$$
= F(y, x).
$$

7
which is part (1b). Further,
\[ \|y_1 - y\| \leq \frac{\|F(y, x) - F^{(1)}(y, x)\|}{m^{(3.6)}} \]  
from (3.8)
\[ \leq \frac{(\delta m^{(3.6)}) \|y - y_0\|}{m^{(3.6)}} \]  
by (3.3) and \( y \in N_{y_0}^{(3.1)}(\delta m^{(3.6)}) \) in (3.5)
\[ = \delta \|y - y_0\| \]
\[ < \epsilon \|y - y_0\| \]  
by the choice \( \delta = \epsilon/(1 + \epsilon) \),
which is part (1c). Finally,
\[ \|y_1 - y_0\| \leq \|y_1 - y\| + \|y - y_0\| \]  
from (1c)
\[ \leq (1 + \epsilon) \|y - y_0\| \]
\[ = r(\epsilon) \]  
from the choice \( y \in B_{y_0}(r(\epsilon)/(1 + \epsilon)) \) in (3.5).

Therefore \( y_1 \in B_{y_0}(r(\epsilon)) \), which is part (1a).

(Part 2.) Let \( y_1 = y \) from (3.5). This \( y_1 \) and \( y_0 \) begin a sequence \( \{y_n\} \) to be built subject to the conditions:
\[ (1_n) \|y_{n+1} - y_n\| \leq \delta^n \|y - y_0\| , \]
\[ (2_n) \|F(y_{n+1}, x) - F^{(1)}(y, x)\| \leq (\delta m^{(3.6)}) \|y_{n+1} - y_0\| . \]
Condition (1) is just \( \|y - y_0\| \leq \|y - y_0\| \). Condition (2) is (3.3) in corollary 3.2 which is applicable by the choices of \( y_1 = y \) and \( x \) in equations (3.5) and (3.6).

Suppose \( y_0, y_1, \ldots, y_k \) have been constructed to satisfy (1_n) and (2_n) for \( 0 \leq n \leq k - 1 \). The selection of \( y_{k+1} \) proceeds as for \( y_1 \) in the first half of the proof. Again because \( D_1F(y_0, x) : \mathbb{R}^m \rightarrow \mathbb{R}^n \) is onto, the transformation maps to \( -[F(y_k, x) - F^{(1)}(y, x)] \), and because \( x \in N_{y_0}^{(3.6)} \) by (3.6), it is possible to invoke lemma 3.6 to find a \( \hat{y} \) with
\[ D_1F(y_0, x) \hat{y} = -[F(y_k, x) - F^{(1)}(y, x)] , \]
\[ m^{(3.6)} \|\hat{y}\| \leq \|F(y_k, x) - F^{(1)}(y, x)\| . \]
Let \( y_{k+1} = \hat{y} + y_k \) so \( \hat{y} = y_{k+1} - y_k \). For this choice of \( y_{k+1} \),
\[ \|y_{k+1} - y_k\| \leq \frac{\|F(y_k, x) - F^{(1)}(y, x)\|}{m^{(3.6)}} \]  
from (3.10)
\[ \leq \frac{(\delta m^{(3.6)}) \|y_k - y_{k-1}\|}{m^{(3.6)}} \]  
from (2_{k-1})
\[ = \delta \|y_k - y_{k-1}\| \]
\[ < \delta^k \|y - y_0\| \]  
from (1_k),
which is (1_k). Summing (1_n) for \( 0 \leq n \leq k \) gives
\[ \|y_{k+1} - y_0\| \leq \sum_{n=0}^{k} \|y_{n+1} - y_n\| \leq \frac{1 - \delta^{k+1}}{1 - \delta} \|y - y_0\| < (1 + \epsilon) \|y - y_0\| , \]
which easily follows from the choice \( \delta = \epsilon/(1 + \epsilon) \). This inequality combines with \( y \in B_{y_0}(r(\epsilon)/(1 + \epsilon)) \) to place \( y_{k+1} \in B_{y_0}(r(\epsilon)) \subseteq N^{(3.1)}(\delta n^{(3.6)}) \) from equation (3.4), and then \((y_{k+1}, x) \in D\). Thus, the evaluation of \( F(y_{k+1}, x) \) is well defined. Further,

\[
\|F(y_{k+1}, x) - F^{(1)}(y, x)\| = \|F(y_{k+1}, x) - F(y_{k+1}, x) - \{ - [F(y_k, x) - F^{(1)}(y, x)] \} \| \quad \text{inserting } \pm F(y_k, x)
\]

\[
= \|F(y_{k+1}, x) - F(y_k, x) - D_1 F(y_0, x)(y_{k+1} - y_k)\| \quad \text{from (3.9)}
\]

\[
\leq (\delta n^{(3.6)}) \|y_{k+1} - y_k\| \quad \text{from (3.2)},
\]

which is (2c).

In this way a sequence \( \{y_n\} \subseteq B_{y_0}(r(\epsilon)) \) is constructed that satisfies conditions (1n) and (2n) for all \( n \). The sequence is a Cauchy sequence by (1n), so it has a limit \( y_F \in \text{cl}(B_{y_0}(r(\epsilon))) \), which is part (2a). Passing to the limit in \( (2n) \) shows \( F(y_F, x) = F^{(1)}(y, x) \), which is part (2b). Summing (1n), now for \( 1 \leq n \leq k \), gives

\[
\|y_{k+1} - y_F\| = \|y_{k+1} - y_{n}\| \leq \sum_{n=1}^{k} \|y_{n+1} - y_{n}\| \leq \delta \frac{1 - \delta^k}{1 - \delta} \|y - y_0\|,
\]

which in the limit becomes (2c), \( \|y_F - y\| \leq \delta \frac{1 - \delta}{1 - \delta} \|y - y_0\| = \epsilon \|y - y_0\| \).

### 3.4. Proof of the First Equivalence.

**Theorem 3.8** \( (P)_{\min} \equiv (P_1)_{\min} \). Under hypotheses 2.1 (1–5), there is a neighborhood of \( x_0 \) where both optimization problems \( (P)_{\min} \) and \( (P_1)_{\min} \) of Table 2.2 are well defined. Their values are asymptotically equal at \( x_0 \) in the sense of definition 2.3.

**Proof.** By lemma 2.6, \( x_0 \) has a neighborhood \( N^{(2.6)}_{x_0} \) where problem \( (P)_{\min} \) is well defined for every \( x \in N^{(2.6)}_{x_0} \), and the optimal value, \( \mu_F(x) \), is Lipschitz continuous at \( x_0 \) with constant \( L \).

By corollary 2.7 similarly, \( x_0 \) has a neighborhood \( N^{(2.7.1)}_{x_0} \) where problem \( (P_1)_{\min} \) is well defined for every \( x \in N^{(2.7.1)}_{x_0} \), and the optimal value, \( \mu_1(x) \), is Lipschitz continuous at \( x_0 \) with constant \( L_1 \).

Let \( B_{y_0}(r(\epsilon)/(1 + \epsilon)) \times N^{(3.7)}_{x_0}(\epsilon) \) be the neighborhood of \((y_0, x_0)\) in lemma 3.7, and let

\[
N(\epsilon) = N^{(2.6)}_{x_0} \cap N^{(2.7.1)}_{x_0} \cap N^{(3.7)}_{x_0}(\epsilon) \cap B_{x_0} \left( \min \{ L^{-1}, L_1^{-1}, \frac{r(\epsilon)}{1 + \epsilon} \} \right).
\]

Note the ball in this formula is around \( x_0 \) rather than \( y_0 \).

Suppose \( x \in N(\epsilon) \). Let \( \mu_F(x) \) be attained at \( y \). By lemma 2.6 and \( x \in B_{x_0}(L^{-1}r(\epsilon)/(1 + \epsilon)) \), therefore

\[
\|y - y_0\| = \mu_F(x) \leq L \|x - x_0\| < r(\epsilon)/(1 + \epsilon),
\]

which places \((y, x) \in B_{y_0}(r(\epsilon)/(1 + \epsilon)) \times N^{(3.7)}_{x_0}(\epsilon) \). Part 1 of lemma 3.7 now asserts there is a \( y_1 \in B_{y_0}(r(\epsilon)) \) with

\[
\|y_1 - y\| \leq \epsilon \|y - y_0\| \quad \text{and} \quad F^{(1)}(y_1, x) = F(y, x) = 0.
\]

Thus

\[
\mu_1(x) \leq \|y_1 - y_0\| \leq \|y_1 - y\| + \|y - y_0\| \leq (1 + \epsilon) \|y - y_0\| = (1 + \epsilon) \mu_F(x)
\]

which is the upper side of (2.3) in definition 2.3. The inequality with \( \mu_F \) and \( \mu_1 \) exchanged is established by the same argument using \( L_1 \) instead of \( L \), corollary 2.7 instead of lemma 2.6, and lemma 3.7 part 2 instead of part 1. The two upper-side inequalities imply (2.3). \( \square \)
4. Equalities for the Dual Problems. The duality theory for best linear approximation guarantees that the three pairs of dual problems in Table 2.2 have equal values. Equalities like these are well known and can be established in many ways. These are derived from the following duality theorem that Luenberger [4, p. 119, thm. 1] proves directly from the Hahn-Banach theorem.

Theorem 4.1 (Best linear approximation). If \( S \) is a subspace and \( y_0 \) is an element of a real, normed linear space, then

\[
\inf_{y \in S} \|y - y_0\| = \max_{f \in S^*, \|f\| \leq 1} f(y_0).
\]

Corollary 4.2 (Best affine approximation). If \( A \) is an affine subspace and \( y_0 \) is an element of a real, normed linear space, then

\[
\inf_{y \in A} \|y - y_0\| = \max_{f \in (A - a)^+, \|f\| \leq 1} f(y_0 - a)
\]
in which \( a \) is any element of \( A \).

Proof. Replace \( y, y_0, S \) in theorem 4.1 by \( y - a, y_0 - a, A - a \). □

Corollary 4.3. Let \( T : \mathbb{R}^m \to \mathbb{R}^p \) be a linear transformation. For every \( y_0 \in \mathbb{R}^m \), each optimization problem below is well defined if and only if \( h \in T(\mathbb{R}^m) \), in which case the optimal values are equal.

\[
\min_{y \in \mathbb{R}^m : Ty = h} \|y - y_0\| = \max_{g \in (\mathbb{R}^n)^* : \|T^* g\| \leq 1} g(Ty_0 - h)
\]

Proof. The minimization is well-posed whenever \( h \) is in the image of \( T \). The same can be proved for the maximization. If \( h \in T(\mathbb{R}^m) \), then \( h = Tu \) for some \( u \), so the objective function,

\[
g(Ty_0 - h) = g(T(y_0 - u)) = (T^* g)(y_0 - u) \leq \|T^* g\| \|y_0 - u\| \leq \|y_0 - u\|,
\]
is bounded above for every \( g \in (\mathbb{R}^n)^* \). The maximum is attained because the feasible set is closed in a finite dimensional space.

Conversely, suppose the maximization is well posed. If \( g \in T(\mathbb{R}^n)^\perp = \ker(T^*) \), then \( g \) and all its multiples are feasible. Hence \( g(h) = 0 \), lest by scaling \( g \) it would be possible to make \( g(Ty_0 - h) = g(h) \) arbitrarily large. Thus \( h \in T(\mathbb{R}^m)^\perp = T(\mathbb{R}^m) \).

All that remains is to establish the equality using corollary 4.2. Choose \( A = \{ y \in \mathbb{R}^m : Ty = h \} \) and \( a \in A \). Now \( A - a = \ker(T) \), so

\[
(A - a)^\perp = [\ker(T)]^\perp = [\perp(T^* (\mathbb{R}^n)^*)]^\perp = T^* (\mathbb{R}^n)^*.
\]

This means \( f \in (A - a)^\perp \) if and only if \( f = T^* g \) for some \( g \in (\mathbb{R}^n)^* \). Thus the maximization in corollary 4.2 is over all such \( g \) with \( \|T^* g\| = \|f\| \leq 1 \). Finally, the objective function is

\[
f(y_0 - a) = (T^* g)(y_0 - a) = g(Ty_0 - a) = g(Ty_0 - Ta) = g(Ty_0 - h).
\]

□

Theorem 4.4 \(((P_i)_{\text{min}} \equiv (P_i)_{\text{max}}, i = 1, 2, 3)\). Under hypotheses 2.1 \((1-5)\), there is a neighborhood of \( x_0 \) where problems \((P_i)_{\text{min}} \) and \((P_i)_{\text{max}} \) of Table 2.2 are well defined and their values are equal, and similarly for the \((P_2)\) and \((P_3)\) pairs of dual problems.

Proof. By lemma 3.6, \( D_1 F(y_0, x) \) is onto for every \( x \in N_{x_0}^{(3,6)} \). Therefore by corollary 4.3 the following problems are well defined and their values are equal for every \( h \in \mathbb{R}^p \).

\[
\min_{y : D_1 F(y_0, x)y - h = 0} \|y - y_0\| = \max_{f : \|D_1 F(y_0, x)^* f\| \leq 1} f(D_1 F(y_0, x)y_0 - h)
\]
Choosing $h = D_1 F(y_0, x)y_0 - F(y_0, x)$ gives the conclusion of the theorem for the $(P_1)$ dual problems.

In particular $D_1 F(y_0, x_0)$ is onto, so also by corollary 4.3 the following problems are well defined and their optimal values are equal for every $h \in \mathbb{R}^p$.

$$\min_{y : D_1 F(y_0, x_0) y - h = 0} \| y - y_0 \| = \max_{f : \| D_1 F(y_0, x_0)^* f \| \leq 1} f(D_1 F(y_0, x_0)y_0 - h)$$

The choice $h = D_1 F(y_0, x_0)y_0 - F(y_0, x)$ gives the conclusion for the $(P_2)$ dual problems; similarly $h = D_1 F(y_0, x_0)y_0 - D_2 F(y_0, x_0)(x - x_0)$ for the $(P_3)$ problems.

5. Second Equivalence, $(P_1)_{\max} \equiv (P_2)_{\max}$. The second equivalence to be proved, in the notation of Table 2.2, says that the feasible set $\{ f : \| D_1 F(y_0, x)^* f \| \leq 1 \}$ can be replaced by one that is independent of $x$. The proof is self-contained and is the second of the two most complicated proofs in this paper.

**Theorem 5.1** $(P_1)_{\max} \equiv (P_2)_{\max}$. Under hypotheses 2.1 (1–5), there is a neighborhood of $x_0$ where both optimization problems $(P_1)_{\max}$ and $(P_2)_{\max}$ of Table 2.2 are well defined. Their values are asymptotically equal at $x_0$ in the sense of definition 2.3.

**Proof.** The hypotheses suffice to invoke theorem 4.4 which says $(P_1)_{\max}$ and $(P_2)_{\max}$ are well defined on some neighborhood $N(1)$ of $x_0$. The feasible sets are given by

$$\mathcal{C}(x) = \{ f \in (\mathbb{R}^m)^* : \| D_1 F(y_0, x)^* f \| \leq 1 \}$$

$$\mu_1(x)_{\max} = \max_{f \in \mathcal{C}(x)} f (F(y_0, x)) \quad \mu_2(x)_{\max} = \max_{f \in \mathcal{C}(x)} f (F(y_0, x))$$

The proof has three steps that culminate in equations (5.2), (5.3) and (5.4), respectively.

(Step 1) If $f_1 \in \text{bd}(\mathcal{C}(x_0)) = \{ f : \| D_1 F(y_0, x_0)^* f \| = 1 \}$, then

$$\| D_1 F(y_0, x)^* f_1 \| - 1 = \| D_1 F(y_0, x)^* f_1 \| - \| D_1 F(y_0, x_0)^* f_1 \| \leq \| D_1 F(y_0, x)^* f_1 - D_1 F(y_0, x_0)^* f_1 \|$$

$$\leq \| D_1 F(y_0, x)^* - D_1 F(y_0, x_0)^* \| \| f_1 \|$$

$$= \| D_1 F(y_0, x) - D_1 F(y_0, x_0) \| \| f_1 \|$$

$$\leq \| D_1 F(y_0, x) - D_1 F(y_0, x_0) \| \max_{f \in \text{bd}(\mathcal{C}(x_0))} \| f \| .$$

The linear transformation $D_1 F(y_0, x_0)$ is onto, so its adjoint $D_1 F(y_0, x_0)^*$ is one-to-one. Hence $\| D_1 F(y_0, x)^* f \|$ defines a norm on the dual space whose closed unit ball is $\mathcal{C}(x_0)$. Thus, in the last bound of equation (5.1), the maximum is finite because $\mathcal{C}(x_0)$ is compact. There also, the difference term converges to 0 as $x \to x_0$ because $F$ is continuously differentiable. Altogether, $\| D_1 F(y_0, x)^* f_1 \|$ converges to 1 uniformly on $\text{bd}(\mathcal{C}(x_0))$ as $x \to x_0$. This means, for every $\epsilon > 0$, there is a neighborhood $N(2)(\epsilon)$ of $x_0$, such that

$$x \in N(2)(\epsilon) \text{ and } f_1 \in \text{bd}(\mathcal{C}(x_0)) \Rightarrow 1 - \epsilon \leq \| D_1 F(y_0, x)^* f_1 \| \leq 1 + \epsilon$$

(Step 2) Choose $x \in N(1) \cap N(2)(\epsilon)$, and then choose any nonzero $f \in \mathcal{C}(x)$, and finally let $f_1 = f / \| D_1 F(y_0, x_0)^* f \| \in \text{bd}(\mathcal{C}(x_0))$. Assume without loss of generality that $\epsilon < 1$. It is now possible to calculate

$$\| (1 - \epsilon) D_1 F(y_0, x_0)^* f \| = (1 - \epsilon) \| D_1 F(y_0, x_0)^* f \|$$

$$\leq \| D_1 F(y_0, x)^* f_1 \| \| D_1 F(y_0, x_0)^* f \|$$

$$= \| D_1 F(y_0, x)^* f \|$$

$$\leq 1 \quad \text{because } f \in \mathcal{C}(x).$$
This proves \((1 - \epsilon)f \in \mathcal{C}(x_0)\). Similarly, choose any nonzero \(f_2 \in \mathcal{C}(x_0)\) and let \(f_1 = f_0/\|D_1F(y_0, x_0)^*f_0\| \in \text{bd}(\mathcal{C}(x_0))\). It now follows that
\[
\|(1 + \epsilon)^{-1}D_1F(y_0, x)^*f_0\| = (1 + \epsilon)^{-1}\|D_1F(y_0, x)^*f_0\|
\]
\[
= (1 + \epsilon)^{-1}\|D_1F(y_0, x)^*f_1\|\|D_1F(y_0, x_0)^*f_0\|
\]
\[
\leq \|D_1F(y_0, x_0)^*f_0\| \quad \text{from equation (5.2)}
\]
\[
\leq 1 \quad \text{because } f_0 \in \mathcal{C}(x_0).
\]

This proves \((1 + \epsilon)^{-1}f_0 \in \mathcal{C}(x)\). These two calculations establish the next implication.

\[(5.3) \quad x \in N^{(1)} \cap \mathcal{N}^{(2)}(\epsilon) \Rightarrow (1 - \epsilon)\mathcal{C}(x) \subseteq \mathcal{C}(s_0) \subseteq (1 + \epsilon)\mathcal{C}(x)\]

(Step 3.) Choose \(x \in N^{(1)} \cap \mathcal{N}^{(2)}(\epsilon)\), and then choose \(f_1 \in \mathcal{C}(x)\) that attains \(\mu_1(x)_{\max}\). Equation (5.3) asserts \((1 - \epsilon)f_1 \in \mathcal{C}(x_0)\), so
\[
\mu_2(x)_{\max} = \max_{f \in \mathcal{C}(x_0)} f(F(y_0, x)) \geq (1 - \epsilon)f_1(F(y_0, x)) = (1 - \epsilon)\mu_1(x)_{\max}.
\]

Similarly, choose \(f_2 \in \mathcal{C}(x_0)\) that attains \(\mu_2(x)_{\max}\). Now equation (5.3) asserts \((1 + \epsilon)^{-1}f_2 \in \mathcal{C}(x)\), so
\[
\mu_1(x)_{\max} = \max_{f \in \mathcal{C}(x)} f(F(y_0, x)) \geq (1 + \epsilon)^{-1}f_2(F(y_0, x)) = (1 + \epsilon)^{-1}\mu_2(x)_{\max}.
\]

Together these two inequalities provide the final implication,

\[(5.4) \quad x \in N^{(1)} \cap \mathcal{N}^{(2)}(\epsilon) \Rightarrow (1 - \epsilon)\mu_1(x)_{\max} \leq \mu_2(x)_{\max} \leq (1 + \epsilon)\mu_1(x)_{\max},\]

which is (2.3) in definition 2.3.

6. Third Equivalence, \((P_2)_{\max} \equiv (P_3)_{\max}\). The proof of the last equivalence involves a class of norms that has been used already in the proof of theorem 5.1. If a linear mapping \(T : \mathbb{R}^m \rightarrow \mathbb{R}^p\) is onto, then its adjoint \(T^*\) is one-to-one, so \(\|T^*f\|\) defines a norm on the dual space, \((\mathbb{R}^p)^*\). The dual of this norm, viewed as a norm on \(\mathbb{R}^p\), is given by the following construction. All the maximization problems in Table 2.2 are norms of this kind.

**Lemma 6.1.** If a linear transformation \(T : \mathbb{R}^m \rightarrow \mathbb{R}^p\) is onto, then
\[
\|v\|_T := \max_{f : \|f\| \leq 1} f(v)
\]
is a norm on \(\mathbb{R}^p\). (The proof is clear.)

**Theorem 6.2** ((\(P_2)_{\max} \equiv (P_3)_{\max}\)). Under hypotheses 2.1 (1–5), there is a neighborhood of \(x_0\) where both of optimization problems \((P_2)_{\max}\) and \((P_3)_{\max}\) of Table 2.2 are well defined. Their values are differentially equivalent at \(x_0\) in the sense of definition 2.1.

Under hypotheses 2.1 (1–6), the values of the problems \((P_2)_{\max}\) and \((P_3)_{\max}\) are asymptotically equal at \(x_0\) in the sense of definition 2.3.

**Proof.** (Part 1.) Let \(\| \cdot \|_T\) be the norm given in lemma 6.1 for the linear transformation \(T = D_1F(y_0, x_0)\). Let \(T(y, x) = D_2F(y, x_0)(x - x_0) + F(y, x_0)\) be the linear function parameterized by \(y\) whose graph is tangent to the graph of \(F(y, x)\) at \(x = x_0\). (Note this is not the \(T^{(1)}\) of Table 2.1.) In this notation, \(\mu_2(x) _ {\max} = \|F(y_0, x)\|_T\) and \(\mu_3(x) _ {\max} = \|T(y_0, x)\|_T\). Thus by the triangle inequality,
\[
|\mu_2(x)_{\max} - \mu_3(x)_{\max}| = \|F(y_0, x)\|_T - \|T(y_0, x)\|_T \leq \|F(y_0, x) - T(y_0, x)\|_T.
\]
The approximations can be found very simply using the matrix QR factorization. \[ \| \cdot \|_T \] is a norm for \( x \) as noted \( A \). That is, the values \( \mu_2(x) \) approximate \( \mu_F(x) \) increasingly well as \( x \) nears \( x_0 \). For 2-norms, the approximations can be found very simply using the matrix QR factorization. 

\[ \lim_{x \to x_0} \frac{\mu_2(x) - \mu_3(x)}{\|x - x_0\|} \leq \lim_{x \to x_0} \frac{\|F(y_0, x) - T(y_0, x)\|_T}{\|x - x_0\|} = 0, \]

which is (2.2) in definition 2.1.

(Part 2.) The Fréchet differentiability of \( F \) with \( F(y_0, x, 0) = 0 \) imply

\[ F(y_0, x) = D_2F(y_0, x_0)(\Delta x) + R(\Delta x) \]

where \( \Delta x = x - x_0 \) and the remainder \( R(\Delta x) \) is \( o(\|\Delta x\|) \). Again by the triangle inequality,

\[ \|F(y_0, x)\|_T = \|D_2F(y_0, x_0)(\Delta x) + R(\Delta x)\|_T \leq \|D_2F(y_0, x_0)(\Delta x)\|_T + \|R(\Delta x)\|_T. \]

where as noted \( \mu_2(x) \) and \( \mu_3(x) \) are equivalent in finite dimensional spaces. Therefore

\[ \frac{\left| \mu_2(x)_{\max} - \mu_3(x)_{\max} \right|}{\|x - x_0\|} \leq \frac{\|R(\Delta x)\|_T}{\|D_2F(y_0, x_0)(\Delta x)\|_T}. \]

Again by the equivalence of all norms for a finite dimensional space, the upper bound vanishes in the limit \( x \to x_0 \) because \( R(\Delta x) \) is \( o(\|\Delta x\|) \). The vanishing limit implies (2.3) in definition 2.3.

7. Summary in Matrix Notation and for 2-Norms. Suppose bases have been chosen for \( \mathbb{R}^m, \mathbb{R}^n, \mathbb{R}^p \) and a norm has been chosen to measure perturbations in \( \mathbb{R}^m \). These choices express the optimization problems in matrix notation:

1. \( J(x) \) is the \( p \times n \) Jacobian matrix for \( D_2F(y_0, x) \). The entries are the partial derivatives of \( F(y, x) \) with respect to \( x \) evaluated at \( (y_0, x) \).
2. \( K(x) \) is the \( p \times m \) Jacobian matrix for \( D_1F(y_0, x) \). Entries are partial derivatives of \( F(y, x) \) with respect to \( y \) evaluated at \( (y_0, x) \).
3. The residual vector of the equations is \( r(x) = F(y_0, x) \in \mathbb{R}^p \).
4. \( \| \cdot \| \) is the chosen norm for \( \mathbb{R}^m \), and \( \| \cdot \|_\ast \) is the dual norm.

The matrix versions of the problems are in Table 7.1. If \( K(x_0) \) has full row rank, then this paper has shown:

1. The minimizations and maximizations of Table 7.1 are duals (theorem 4.4).
2. The optimal values \( \mu_F(x) \), \( \mu_1(x) \), \( \mu_2(x) \) are asymptotically equal at \( x_0 \) (theorems 3.8 and 5.1).
3. \( \mu_3(x) \) is differentially equivalent to the other values (theorem 6.2).

That is, the values \( \mu_3(x) \) approximate \( \mu_F(x) \) increasingly well as \( x \) nears \( x_0 \). For 2-norms, the approximations can be found very simply using the matrix QR factorization.

**Lemma 7.1.** Let \( A \in \mathbb{R}^{m \times p} \) and \( s, u \in \mathbb{R}^p \). If \( A \) has full column rank, then for the \( A = QR \) factorization,

\[ \max_{\|Au\|_2 \leq 1} u^T s = \|R^{-T} s\|_2. \]

**Proof.** Because \( u^T s = u^T R^T R^{-T} s \) and \( \|Ru\|_2 = \|Au\|_2 = 1 \) therefore \( u^T s \leq \|R^{-T} s\|_2 \) with equality when \( u = R^{-1} R^{-T} s/\|R^{-T} s\|_2 \).
### Table 7.1

Optimization problems of Table 2.2 in matrix notation. In these formulas, $\Delta x = x - x_0$.

| name | value | minimization form | dual maximization |
|------|-------|-------------------|-------------------|
| $(P)$ | $\mu_p(x)$ | $\min_{F(y, x) = 0} \| y - y_0 \|$ |  |
| $(P_1)$ | $\mu_1(x)$ | $\min_{\textbf{K}(x)} \| \Delta y \|$ | $\max_{\| \textbf{K}(x) u \| \leq 1} u^T r(x)$ |
| $(P_2)$ | $\mu_2(x)$ | $\min_{\textbf{K}(x_0)} \| \Delta y \|$ | $\max_{\| \textbf{K}(x_0) u \| \leq 1} u^T r(x)$ |
| $(P_3)$ | $\mu_3(x)$ | $\min_{\textbf{K}(x_0)} \| \Delta y \|$ | $\max_{\| \textbf{K}(x_0) u \| \leq 1} u^T J(x_0) \Delta x$ |

### Appendix A. Nomenclature and Notation.

This appendix lists some standard notation that is used without comment throughout the paper.

1. For $f : D \subseteq \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^p$, the Fréchet derivative of $f$ evaluated at $(y, x)$ is $Df(y, x) \in \text{hom}(\mathbb{R}^{m+n}, \mathbb{R}^p)$. The partial Fréchet derivative of $f$ with respect to the first space $\mathbb{R}^m$ and evaluated at $(y, x)$ is $D_1f(y, x) \in \text{hom}(\mathbb{R}^m, \mathbb{R}^p)$, and similarly for the second space $\mathbb{R}^n$ and $D_2f$.

2. The dual space of a normed linear space $\mathbb{R}^m$ is the space of functionals $(\mathbb{R}^m)^* = \text{hom}(\mathbb{R}^m, \mathbb{R})$ with the induced norm. The annihilator of a set $S \subseteq \mathbb{R}^m$ is the subspace $S^\perp \subseteq (\mathbb{R}^m)^*$. The subspace annihilated by a set $S \subseteq (\mathbb{R}^m)^*$ is $^*S \subseteq \mathbb{R}^m$. The transpose of $T \in \text{hom}(\mathbb{R}^m, \mathbb{R}^p)$ is $T^* \in \text{hom}((\mathbb{R}^p)^*, (\mathbb{R}^m)^*)$.

3. The interior, boundary, and closure of a set $S$ are indicated by $\text{int}(S)$, $\text{bd}(S)$, and $\text{cl}(S)$.

4. The open ball with center $c$ and radius $r$ is $B_r(c)$.

5. Six lemmas assert the existence of neighborhoods that are indicated by placing the lemma number in a superscript, the point around which the neighborhood lies in a subscript, and any parameterization of the neighborhood in parentheses:

$$N_{x_0}^{(2,6)}(\epsilon) \quad N_{x_0}^{(2,7,j)}(\epsilon) \quad N_{x_0}^{(3,1)}(\epsilon) \quad N_{y_0}^{(3,1)}(\epsilon) \quad N_{x_0}^{(3,6)}(\epsilon) \quad N_{x_0}^{(3,7)}(\epsilon).$$

### REFERENCES

[1] R. G. Bartle, *The Elements of Real Analysis*, John Wiley & Sons, New York, 2nd ed., 1976.

[2] L. M. Graves, Some mapping theorems, Duke Math. J., 17 (1950), pp. 111–114.

[3] J. F. Grcar, A matrix lower bound, Linear Algebra Appl., 433 (2010), pp. 203–220.

[4] D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley & Sons, New York, 1969.

[5] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.