Fixed Point Theorems, supplementary notes
APPM 5440 Applied Analysis

Stephen Becker; 9/27/2019

A brief summary of the standard fixed point theorems, since the course text does not go into detail; we unify notation and naming. The names of theorems themselves are confusing since we have both the "Brouwer" and "Browder" as well as "Schauder" and "Schaefer" theorems. Parts of these notes may be extremely useful for certain types of research, but it is not a focus of the class or the qualifying exams, so this is a bit orthogonal to the main thrust of the class. We also touch on some applications to optimization.

1 Basic results from the class textbook and some basic definitions

From Hunter and Nachtergaele [8], modified slightly:

Definition 1. *Let (X, d) be a metric space. A mapping $T : X \to X$ is a **contraction** if there exists a constant c with $0 \le c \le 1$ such that*

$$(\forall x, y \in X)\ d(T(x), T(y)) \le c\, d(x, y),$$

*i.e., it is Lipschitz continuous with constant $c \le 1$. It is a **strict contraction** if $c < 1$. We will also refer to the case $c \le 1$ as being **non-expansive** to avoid confusion.*

Let us introduce some additional notation that is standard but not in the book:

Definition 2 (Fixed point set)**.** *The set $\mathrm{Fix}(T) \subset X$ is the set of fixed points of the operator $T : X \to X$, that is,*

$$\mathrm{Fix}(T) = \{x \mid x = Tx\}.$$

For example, if $T = I$ is the identity, then $\mathrm{Fix}(I) = X$. If $T = -I$, then $\mathrm{Fix}(-I) = \{0\}$. In the above definition, X is any general space (topological, metric, normed, etc.), but in the theorem below, it must be a complete metric space.

Theorem 3 (Contraction mapping theorem, due to Banach 1922; also known as Banach-Picard)**.** *Let $T : X \to X$ and let X be a complete metric space. If T is a strict contraction, then $\mathrm{Fix}(T)$ consists of exactly one element $x^\star$.*

The proof is constructive and creates the sequence $x_{n+1} = Tx_n$, which is known as the **Picard iteration**. The result is very easy, and there are easy implications, such as $d(x_n, x^\star) \le c^n d(x_0, x^\star)$, which means the sequence (x_n) converges at a *linear* rate to $x^\star$ (confusingly, in some fields, this is called *exponential* convergence). Note that the theorem applies whenever T is a strict contraction in *some* metric, so you are free to "search" for good metrics.

Definition 4. *Let (X, d) be a complete metric space. A function T is said to be a **weak contraction** if*

$$d(T(x), T(y)) < d(x, y) \quad \forall\, x \ne y,\ x, y \in X.$$

Note that a weak contraction is slightly stronger being non-expansive, but weaker than being a strict contraction. The following question about weak contractions was asked on the Aug 2016 prelim in applied analysis:

1. Prove the following variant of the contraction mapping theorem: if T is a *weak contraction* and the space X is *compact*, then T has a unique fixed point in X. Hint: consider the function $g(x) = d(x, T(x))$ over X. (answer: see the online solution hints, at Aug 2016 prelim solutions).

2. Let X be $\ell^2(\mathbb{N})$ and $T(x) = L(x) + b$ be an affine function defined by mapping $L : x \mapsto y$ where $x = (x_n)_{n\in\mathbb{N}}$ and $y = (y_n)_{n\in\mathbb{N}}$ with $y_n = (1 - \frac{1}{n})x_n$, and $b = (\frac{1}{n})_{n\in\mathbb{N}}$. Prove or disprove that T has a fixed point in X (answer: no, it need not). Does your answer change if X is the closed unit ball in $\ell^2(\mathbb{N})$? (answer: no, it doesn't help, it still need not have a fixed point).

Some definitions In the rest of the notes, we also use the language of Hilbert space to state results in more generality, though our focus in this section is not on infinite-dimensional spaces, and therefore there is no harm in thinking of Hilbert space as just $\mathbb{R}^n$. Formally, a Hilbert space is a Banach space with an inner product and the Banach space norm is the norm induced by the inner product; in particular, a Hilbert space is a normed vector space, and it is complete. For a Hilbert space $\mathcal{H}$ *over the real numbers*, we define the inner product $\langle \cdot, \cdot \rangle$ as a functional on $\mathcal{H} \times \mathcal{H} \to \mathbb{R}$ such that it is (1) symmetric, so $\langle x, y\rangle = \langle y, x\rangle$, (2) linear in each argument separately, so $\langle \alpha x + y, z\rangle = \alpha\langle x, z\rangle + \langle y, z\rangle$, and (3) positive-definite so that $\langle x, x\rangle \geq 0$ and $\langle x, x\rangle = 0$ only when $x = 0$. The induced norm is defined by $\|x\| = \sqrt{\langle x, x\rangle}$. N.B. if $\mathcal{H}$ is over $\mathbb{C}$, the inner product takes on complex values and the definitions change slightly.

2 Other classical fixed point theorems

Brief mentions of the Brouwer, Kakutani and Schauder fixed point theorems. The reference is "Methods of Mathematical Economics: Linear and Nonlinear Programming, Fixed-Point Theorems" by Joel Franklin [6] (1980, republished by SIAM, available for free online with CU access at `http://epubs.siam.org/ebooks/`), which is a readable introduction. The Kakutani theorem in full generality is covered by Royden's "Real Analysis" [10] and relies on the Alaoglu Theorem.

The **contraction mapping theorem** is the simplest fixed point theorem and is unusual because it is constructive. It is due to Banach, 1922, and implies the implicit function theorem.

The **Brouwer** fixed-point theorem is one of the most important results in modern mathematics, and is easy to state but hard to prove. It its most basic form, it is completely topological.

Theorem 5 (**Brouwer**, topological version)**.** *Let K be a compact convex set (inside a topological space X), and $f : K \to K$ a continuous function. Then* $\text{Fix}(f)$ *is nonempty.*

We'll pay more attention to this version, which is clearly a corollary of the topological version:

Theorem 6 (**Brouwer**, Euclidean space version)**.** *Let X be a Banach space and f be a continuous function defined on the unit ball B. If f maps B to itself (that is, $\|x\| \leq 1 \implies \|f(x)\| \leq 1$), then some point of the ball is mapped to itself, i.e., there is a fixed point of f.*

Franklin introduces a short proof based on Green's theorem (summarized by the "Stokes' Theorem proof" on the Brouwer page at Wikipedia).

A further corollary of the topological Brouwer theorem (and slightly more abstract than our Euclidean space version) is known as the Schauder fixed point theorem:

Theorem 7 (Brouwer, Banach space version, aka **Schauder** fixed point theorem)**.** *Let X be a Banach space and f be a continuous function defined on a convex subset K such that $f(K)$ is compact. If f maps K to itself, then some point of K is mapped to itself, i.e., there is a fixed point of f.*

Another extension of Brouwer's theorem is the Kakutani theorem, which concerns more general mappings that need not be functions. We consider set-valued mappings $F(x)$, e.g., $F : X \to \mathcal{P}(X)$. A very important example of such a mapping is the sub-differential of a convex function.

Theorem 8 (**Kakutani**)**.** *Let X be a closed, bounded, convex set inside a Euclidean space. For every $x \in X$, let $F(X)$ equal a non-empty convex subset $Y \subset X$. Assume the graph*

$$\{(x, y) \mid y \in F(x)\}$$

is closed. Then there is some $x^\star$ that solves $x^\star \in F(x^\star)$.

The graph being closed is similar to a function being lower semicontinuous.

The Kakutani theorem is used to prove the famous **Nash equilibrium** theorem for n-person games.

3 Fixed point theorems for optimization

The reference for this section is "Convex analysis and monotone operator theory in Hilbert Spaces" by Bauschke and Combettes, 2011 [1], mainly from chapters 4 and 5. The book is free on the CU campus via SpringerLink; see http://link.springer.com/book/10.1007%2F978-1-4419-9467-7. A much more concise yet still rigorous summary is in the book chapter "Proximal Splitting Methods in Signal Processing" by Combettes and Pesquet, 2011 [2], available at http://dx.doi.org/10.1007/978-1-4419-9569-8_10 (and preprints are on the arXiv).

Suppose T is **non-expansive but not strictly non-expansive**, that is, perhaps $\mathrm{Lip}(T) = c = 1$. Then the Banach-Picard iteration may fail. For example, take $T = -I$ and start at any $x_0 \neq 0$, then the iteration never converges (but there is a unique fixed point, which is 0). Franklin describes this case as unimportant since he uses a limit argument (of $c_n \to 1$ with $c_n < 1$) to show that a solution exists, but this is unsatisfactory because (1) this case is very different, since the fixed points may not be unique, as shown by $T = I$, and (2) it is not a constructive proof, so not helpful for numerical methods.

One way to handle this is the following variant on the Banach-Picard iteration, which in turn forms the basis of many optimization methods.

Theorem 9 (Krasnosel'skiĭ-Mann algorithm (aka **KM** algorithm))**.** *Let D be a nonempty closed convex subset of a Hilbert space $\mathcal{H}$, let $T : D \to D$ be a non-expansive operator such that $\mathrm{Fix}(T)$ is non-empty, and pick $\lambda \in (0, 1)$ and $x_0 \in D$. Then the sequence defined by*

$$x_{n+1} = x_n + \lambda(Tx_n - x_n)$$

is such that (a) $(Tx_n - x_n)$ converges to 0, and (b) (x_n) converges weakly to a point in $\mathrm{Fix}(T)$.

This can be used to prove convergence of the classical gradient descent method, as well as generalizations known as the "proximal gradient method" or "forward-backward" method (not to be confused with various other "forward-backward" methods in other fields).

Now, we briefly consider projections onto sets. For any set closed set C, define $\mathcal{P}_C(x)$ the projection of x onto C; that is, $p = \mathcal{P}_C(x)$ if $\|p - x\| \leq \|y - x\|$ for all $y \in C$. We require the set to be closed so that the point exists. The projection depends implicitly on the norm. If the projection is always unique, the set is called a Chebyshev set, and a basic result is that all closed (and non-empty) convex sets are Chebyshev sets. In a Hilbert space, using the induced norm, we have the further fundamental result:

Proposition 10. *Let $C \subset \mathcal{H}$ be a non-empty closed, convex set. Then $p = \mathcal{P}_C(x)$ if and only if $\langle x - p, y - p \rangle \leq 0$ for all $y \in C$.*

Using this, we can prove that $\mathcal{P}_C$ is a non-expansive (and in fact, firmly non-expansive) operator. However, if C contains more than one element, it is not strictly contractive, since then there are more than one points such that $x = \mathcal{P}_C(x)$. Combining the non-expansive properties with the KM algorithm, we also recover some classical projection algorithms:

Theorem 11 (Projection onto convex sets (POCS))**.** *Let $(C_i)_{i=1}^m$ be a finite collection of nonempty closed convex sets inside a Hilbert space $\mathcal{H}$, and let $\mathcal{P}_i$ be the projection onto C_i (which exists and is single-valued). Let $T = P_1 \circ P_2 \circ \ldots \circ P_m$ and assume $\mathrm{Fix}(T) \neq \emptyset$. Then the POCS algorithm $x_{n+1} = Tx_n$ will weakly converge to some point in C_1 (for full statement, see Corollary 5.23 in Bauschke and Combettes [1]).*

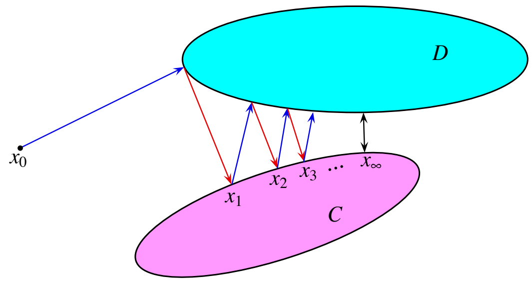

Figure 1: POCS algorithm for the case $C_1 = C$ and $C_2 = D$ with $C_1 \cap C_2 = \emptyset$, figure from Combettes and Pesquet 2011 [2]

The full POCS algorithm has mild assumptions. If we further assume that $\mathcal{C} \stackrel{\text{def}}{=} \cap C_i \neq \emptyset$, then (x_n) converges weakly to some point in $\mathcal{C}$. See Fig. 1 for an illustration of the algorithm in the case when $\cap C_i = \emptyset$. We also can strengthen the result to strong convergence if the sets are closed affine subspaces; this result is known as the **von Neumann** projection theorem (or the von Neumann-Halperin theorem).

We also note that there is a "mixture" of the Banach and Schauder theorems that is known as the **Browder** theorem (not to be confused with Brouwer), which we state:

Theorem 12 (**Browder-Göhde-Kirk**). *Let D be a nonempty bounded closed convex subset of $\mathcal{H}$ and let $T : D \to D$ be a nonexpansive operator. Then the set of fixed points of T is non-empty.*

Another other fixed point theorems, related to Ekeland's variational principle (which is used when level sets are not compact so that Bolzano-Weierstrass does not apply), is [3]

Theorem 13 (**Caristi** fixed-point theorem). *Let (X, d) be a complete metric space and $T : X \to X$, and $f : X \to \mathbb{R}$ be a non-negative lower semicontinuous function. If for all $x \in X$ we have $d(x, T(x)) \leq f(x) - f(T(x))$ then T has a fixed point.*

The choice of f is arbitrary, so one can search for a good f, just as one may search for a good norm, in order to prove existence of a fixed point of T.

3.1 Specific applications for convex optimization

This section uses more of Bauschke and Combettes' book [1], as well as Y. Nesterov's "Introductory Lectures on Convex Optimization." [9]

1. Let $X = \mathbb{R}^n$ and suppose $f : X \to \mathbb{R}$ is continuous and also has continuous derivatives ∇f. We denote this set of functions as $C^1(X)$. We let $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$ be the standard Euclidean inner product, and $\|x\| = \sqrt{\langle x, x \rangle}$ the standard Euclidean norm.

2. Suppose additionally ∇f is Lipschitz continuous with constant L; we done such class of functions as $C^1_L(x)$. Then for $f \in C^1_L(X)$ and all $x, y \in X$,
$$|f(x) - f(y) - \langle \nabla f(x), y - x \rangle| \leq \frac{L}{2} \|y - x\|^2.$$
The proof is elementary and uses the fundamental theorem of calculus.

3. We can characterize the *convex* functions in $C^1(X)$, denoted $\mathcal{F}^1(X)$, by those $f \in C^1(X)$ such that for all $x, y \in X$,
$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle. \tag{1}$$
This means that at every point x, the tangent line at $f(x)$ is a global lower bound for the function. By adding two copies of (3) but switching the roles of x and y, we have
$$\langle x - y, \nabla f(x) - \nabla f(y) \rangle \geq 0. \tag{2}$$

In general, a convex function need not be in C^1, so we have not characterized *all* convex functions. It turns out we can generalize the notion of a gradient to the notion of a subgradient, and we require exactly the same property as (3). We define a **subgradient** d to be any vector in X satisfying

$$f(y) \geq f(x) + \langle d, y - x \rangle \quad \forall x, y \in X. \tag{3}$$

The collection of all subgradients is called the **subdifferential**, denoted ∂f, and when f is convex, this has exactly one element if and only if f is differentiable. This is extremely useful, since we have **Fermat's rule** which says that the set of minimizers of f is the same as the set of zeros of ∂f; when f is differentiable and convex, this just means that $\nabla f(x) = 0$ whenever x is a minimizer. (Note: do not confuse the subdifferential of a function, ∂f, with the notion of the boundary of a set X, which is written ∂X)

4. Denote the set of C^1 convex functions with Lipschitz continuous gradients to be $\mathcal{F}^1_L(X) \stackrel{\text{def}}{=} \mathcal{F}^1(X) \cap C^1_L(X)$. Then $f \in \mathcal{F}^1_L(X)$ implies the following for all $x, y \in X$ (see Nesterov Thm. 2.1.5):
   a) $0 \leq f(y) - f(x) - \langle \nabla f(x), y - x \rangle \leq \frac{L}{2}\|x - y\|^2$
   b) $f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2 \leq f(y)$
   c) $\frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2 \leq \langle \nabla f(x) - \nabla f(y), x - y \rangle$
   d) $\langle \nabla f(x) - \nabla f(y), x - y \rangle \leq L\|x - y\|^2$

   The proof is elementary but some parts are a bit clever.

5. Consider the problem of minimizing f over X, assuming $f \in \mathcal{F}^1_L(X)$. As we'll prove in the homework, this is equivalent to finding a point x such that $0 = \nabla f(x)$, which can be thought of as a fixed point equation to find x such that $x = \underbrace{(I - T)}_{T'}(x)$, i.e., $x \in \text{Fix}(T')$.

6. The operator $T'$is generally not strictly non-expansive, even if we tried to scale T. The homework asks you to provide such an example. In a simpler case when we works with quadratic functions, so that T is a matrix, our condition is that the maximum eigenvalue in absolute value of $(I - T)$ is bounded by 1. If we do convex optimization, that means the smallest eigenvalue of T is at least 0. If it is 0, then the maximum eigenvalue (in absolute value) of $(I - T)$ is at least 1, so there is no chance of being strictly nonexpansive.

7. The notion of firmly non-expansive, introduced in the homework, is stronger than being non-expansive, and it rules out the "nasty"' non-expansive operators like $-I$. See Fig. 2. To summarize the homework: let $D \subset \mathcal{H}$ for a Hilbert space $\mathcal{H}$, and consider a function $T : D \to \mathcal{H}$. We say T is **firmly non expansive** if for all $x, y \in D$, $\|T(x) - T(y)\|^2 + \|(I - T)(x) - (I - T)(y)\|^2 \leq \|x - y\|^2$. Here, we use I to denote the identity mapping. Then the following are equivalent:
   a) T is firmly nonexpansive
   b) $I - T$ is firmly nonexpansive
   c) $2T - I$ is nonexpansive
   d) For all $x, y \in D$, $\|Tx - Ty\|^2 \leq \langle x - y, T(x) - T(y) \rangle$.

8. Working with minimizing quadratic functions, or equivalently, assuming that the derivatives are linear, and working in finite dimensional space, we encroach on the territory of linear system solving. There are many classical results here, such as Jacobi's method, Gauss-Seidel, SOR, etc. The relevant quantity is the **spectral radius** $\rho(A)$ of a matrix A, defined as the maximum eigenvalue of A in absolute value. That is, if A has eigenvalues (λ_i), then $\rho(A) = \max_i |\lambda_i|$. We collect a few facts below. For more reference on the spectral radius, see §5.6 in Horn and Johnson's classic "Matrix Analysis" book (1985) [7].
   a) The spectral radius is not a norm because (1) $\rho(A) = 0$ is possible for nonzero A, and (2) it does not satisfy the triangle inequality (unless both matrices are normal). To prove these results, consider the matrices $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

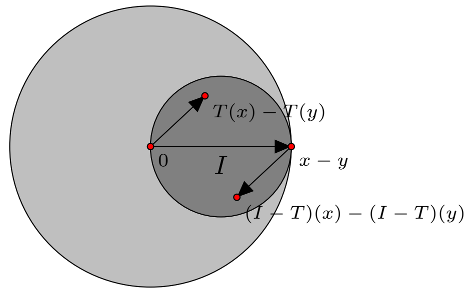

Figure 2: Figure and text from Eckstein and Bertsekas [4]. Illustration of the action of firmly non expansive operators in Hilbert space. If T is non expansive, then $T(x) - T(y)$ must lie in the larger sphere, which has radius $\|x - y\|$ and is centered at 0. If T is also *firmly* non expansive, then $T(x) - T(y)$ must lie in the smaller sphere, which has radius $\frac{1}{2}\|x - y\|$ and is centered at $\frac{1}{2}(x - y)$. This characterization follows directly from T being of the form $\frac{1}{2} + \frac{1}{2}C$ where C is nonexpansive (point (c)). Note that if $T(x) - T(y)$ lies inside the smaller sphere, then so does $(I - T)(x) - (I - T)(y)$, as we already saw (point (b)).
Note: see the 2019 paper Scaled Relative Graph: Nonexpansive operators via 2D Euclidean Geometry by Ernest Ryu, Robert Hannah, and Wotao Yin, for a nice theory that formalizes drawings like this, and from which you can use drawings to prove nonexpansiveness, this giving more intuition to manipulating inequalities.

b) In some special cases, the spectral radius is particularly simple: suppose each entry of the matrix A is non-negative, and if the columns or rows all have constant sum, then spectral radius is just the column or row sum.

c) For any sub-multiplicative norm (sometimes called a *matrix norm*; this means that $\|AB\| \le \|A\|\|B\|$) $\|\cdot\|$, then $\rho(A) \le \|A\|$. Furthermore, for any fixed A and fixed $\epsilon > 0$, there is some sub-multiplicative norm such that $\rho(A) \le \|A\| \le \rho(A) + \epsilon$. The connection with Picard iterations is the following: $\lim_{n\to\infty} A^n = 0$ if and only if $\rho(A) < 1$. The "if" part of the theorem is obvious, but the "only if" is surprisingly strong, and we do not have equivalent results for non-linear operators or for infinite dimensional spaces.

d) As a corollary, for any sub-multiplicative norm, we have $\rho(A) = \lim_{n\to\infty} \|A^n\|^{1/n}$.

e) Another consequence is the Neumann series: a matrix A is invertible if there is some sub-multiplicative norm such that $\|I - A\| < 1$, and if so, then $A^{-1} = \sum_{k=0}^{\infty}(I - A)^k$. Likewise, $\rho(A) < 1$ implies $(I - A)^{-1} = \sum_{k=0}^{\infty} A^k$. (side note: the Neumann series is named for Carl Neumann, not the more famous John von Neumann)

4 Fixed point theorems for Differential Equations

We follow L. C. Evans' classic "Partial Differential Equations." [5] The reference to section and theorem numbers comes from Evan's first edition, but we note that his second edition is now available. The appendices of his book are a very concise summary of analysis, and are a good review. His book is a rigorous approach to proving the existence and smoothness of solutions to PDE.

Banach's fixed point theorem, on Banach spaces, is Theorem 1 in §9.2.1; see Example 1 in that section for an application used to prove existence of a weak solution to a reaction-diffusion system of PDE. Brouwer's fixed point theorem is Theorem 3 in §8.1.4. It is extended to Schauder's fixed point theorem as Theorem 3 in §9.2.2 (with the variant that K is both convex and compact, hence $f(K)$ is necessarily compact). Below we present an alternative form of Schauder's theorem, known as **Schaefer's fixed point theorem**, which is more useful in applications to nonlinear PDE.

First, define a function $f : X \to Y$ to be **compact** if $f(B)$ is precompact in Y whenever B is a bounded subset of X (see definition 5.42 in Hunter and Nachtergaele). Equivalently, it means for every bounded sequence (x_n), then $f(x_n)$ contains a convergent subsequence (which converges in $\overline{f(B)}$, not necessarily $f(B)$).

Theorem 14 (**Schaefer**, Thm. 4 in §9.2.2 of Evans)**.** *Suppose $f : X \to X$ is continuous and compact, where X is a Banach space over the reals, and also assume the set*

$$B \stackrel{def}{=} \{x \in X \mid x = \lambda f(x), \text{ for some } \lambda \in [0,1]\}$$

is bounded. Then f has a fixed point.

Unlike Schauder's theorem, we don't need to explicitly identify an explicit convex, compact set.

Note that the theorem sounds a bit bizarre. The set B actually contains any possible fixed points, since if we take $\lambda = 1$ we include the points $x = f(x)$, so we are somewhat saying that if such a set is bounded, then the set exists. This is the method of *a priori* estimates, and for the subject of PDE, the general method is "if we can prove appropriate estimates for solutions of a nonlinear PDE, under the assumptions that such solutions exist, then in fact these solutions do exist." (Evans, §9.2.2).

Proof of Schaefer's theorem. We use Schauder's theorem. First, let M be a strict bound on all elements $x \in B$, which exists by assumption, so $\|x\| < M$ for all $x \in B$ (note that we are using whatever norm is associated with the generic Banach space X). Now define another function

$$\tilde{f}(x) \stackrel{\text{def}}{=} \begin{cases} f(x) & \|f(x)\| \le M \\ \frac{M}{\|f(x)\|} f(x) & \|f(x)\| \ge M \end{cases}.$$

Thus $\tilde{f}$ maps the ball $B_M(0)$ to itself. Define K to be the closed convex hull of $\tilde{f}(B_M(0))$. The function f is compact by assumption, and so is $\tilde{f}$, and we can consider $\tilde{f} : K \to K$.

Now we are in position to invoke Schauer's fixed point theorem, so there is some x such that $x = \tilde{f}(x)$. If this x is not also a fixed point of f, then it means that $\|f(x)\| > M$ otherwise f and $\tilde{f}$ coincide. Thus by definition of $\tilde{f}$ and the fact x is a fixed point, we have

$$x = \tilde{f}(x) = \frac{M}{\|f(x)\|} f(x) = \lambda f(x)$$

for $\lambda = \frac{M}{\|f(x)\|} < 1$, hence $x \in B$. On the other hand, $\|x\| = \|\tilde{f}(x)\| = \frac{M}{\|f(x)\|}\|f(x)\| = M$, so this contradicts our assumption that $\|x\| < M$ for all $x \in B$. $\square$

References

[1] Bauschke, H. H., and Combettes, P. L., *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* Springer, New York, 2011.

[2] Combettes, P. L. and Pesquet, J.-C., Proximal splitting methods in signal processing, in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering,* (H. H. Bauschke et al., eds), pp. 185–212. Springer, New York, 2011.

[3] Delfour, M. C., *Introduction to Optimization and Semidifferential Calculus*, SIAM, 2012.

[4] Eckstein, J. and Bertsekas, D. P., On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, *Math. Programming,* vol. 55, pp. 293–318, 1992.

[5] Evans, L. C., *Partial Differential Equations*, American Math Society, 1998.

[6] Franklin, J. N., *Methods of Mathematical Economics: Linear and Nonlinear Programming, Fixed-Point Theorems*, republished by SIAM, 2002.

[7] Horn, R. A. and Johnson, C. R., *Matrix Analysis*, Cambridge University Press, 1985.

[8] Hunter, J. and Nachtergaele, B., *Applied Analysis*, World Scientific, 2001.

[9] Nesterov, Y., *Introductory Lectures on Convex Optimization: A Basic Course*, Springer, 2004.

[10] Royden, H. L., *Real Analysis*, 3rd ed., Pearson, 1988.