Abstract. Variational methods find solutions of equations by considering a solution as a critical point of an appropriately chosen function. Local minima and maxima are well-known types of critical points. We explore methods for finding critical points that are neither local maxima or minima, but instead are mountain passes or saddle points. Criteria for the existence of minima or maxima are well-known, but those for mountain passes or saddle points are less well-known. We give an accessible treatment of some criteria for the existence of such points (including the Mountain Pass Lemma), as well as describe a method that could be used to find such points.

Key words. critical point theory, minimax methods, mountain pass, saddle point

AMS subject classifications. 58E05, 35A15, 47J50

1. Introduction. A basic problem of applied mathematics is to provide solutions (or approximations to solutions) of various equations. One very useful approach to this goal is the so-called variational method, in which the solutions of an equation are sought by finding the critical points of an appropriately chosen function.

Example: Solutions of the equation $2x e^{x^2} - 4x + \sin x = 3$ correspond to critical points of the function $f(x) = e^{x^2} - 2x^2 - \cos x - 3x$, since the equation $f'(x) = 0$ is equivalent to the given equation. Thus, solutions of the equation can be found by minimizing the function $f$.

Example: Suppose $A$ is an $n \times n$ symmetric matrix, and suppose $b \in \mathbb{R}^n$ is fixed. A very important type of equation to solve is $Ax = b$. If we define $F_b : \mathbb{R}^n \to \mathbb{R}$ by $F_b(x) = \frac{1}{2}Ax \cdot x - x \cdot b$ (where $\cdot$ represents the dot product on $\mathbb{R}^n$), then it can be shown that $\nabla F_b(x) = Ax - b$. In particular, $\nabla F_b(x) = 0$ if and only if $Ax = b$. Thus, solutions of $Ax = b$ correspond to critical points of the function $F_b$.

In many useful instances, the problem of finding solutions of some type of equation is equivalent to the problem of finding critical points of an appropriately chosen function. From elementary calculus, a familiar type of critical point is a local minimum (or maximum). These can be found by picking an initial point, and then moving to another point at which the function is lower. However, not every function has a minimum or maximum.

Example: Suppose $A$ is a $2 \times 2$ symmetric matrix, with eigenvalues $\lambda_1$ and $\lambda_2$ such that $\lambda_1 < 0 < \lambda_2$. As mentioned above, solutions of $Ax = b$ can be found as critical points of $F_b(x) = \frac{1}{2}Ax \cdot x - x \cdot b$. In this case, if $x_1$ is an eigenvector for $\lambda_1$, then for any $t \in \mathbb{R}$, we will have

$$F_b(tx_1) = \frac{1}{2}tA x_1 \cdot x_1 - t x_1 \cdot b = \frac{\lambda_1 t^2}{2} \|x_1\|^2 - t(x_1, b) \to -\infty$$

as $t \to \infty$ since $\lambda_1 < 0$. If $x_2$ is an eigenvector for $\lambda_2$, a similar calculation shows that $F_b(tx_2) \to +\infty$ as $t \to \infty$, since $\lambda_2 > 0$. Thus, in this case, $F_b$ will have neither a minimum or a maximum on $\mathbb{R}^n$. See Figure 1.1 for a graph of $F_0$.

Variational methods can be used to find critical points that are neither minimums nor maximums. Another useful feature of solving equations by looking for critical
points is that it is sometimes possible to use the existence of one critical point together with other information about the function to prove that there is at least one other critical point. For example, suppose $F_1 : \mathbb{R} \to \mathbb{R}$ is a smooth real valued function, and we know that $F$ has a local minimum at 0 and there is an $x_1$ with $F_1(x_1) < F_1(0)$. Then, by Rolle’s Theorem and the intermediate value theorem, there is a critical point $x_2$ between $x_1$ and 0 (see Figure 1.2a). Alternatively, if $F_2 : \mathbb{R} \to \mathbb{R}$ is smooth and has two local minima at $x_1$ and $x_2$, then $F_2$ must have another critical point $x_3$ between $x_1$ and $x_2$ (see Figure 1.2b). However, it is important to note that these last two examples fail once we move from $\mathbb{R}$ to $\mathbb{R}^2$. In fact, there are polynomial counterexamples!

**Example:** If $G(x, y) = x^2(1 + y)^3 + 7y^2$, then $G$ has a single critical point (at $(0,0)$), which is a local minimum, but not a global minimum. ([7]; see the contours in Figure 1.3a.)

**Example:** If $F(x, y) = (x^2y - x - 1)^2 + (x^2 - 1)^2$, then $F$ has exactly two critical points, both of which are local minima. ([11]; see the contours in Figure 1.3b.)

**2. The Palais-Smale Condition and Palais-Smale sequences.** As the last two examples in the previous section suggest, some type of extra assumption in higher dimensions will be needed. From a naive point of view, if $F(x, y)$ has two local minima,
it seems reasonable that $F$ should have another critical point. After all, if we think of the two local minima as being the low points in two adjacent valleys, there should be a mountain pass linking the two valleys, which would correspond to a third critical point. Using an idea from [2], if we think of pouring water into a valley containing one of the critical points, the water level should rise until it fills that valley and then overflow into the adjacent valley. The point at which the water begins flowing into the adjacent valley would be the third critical point.

However, as we know from the last examples in the previous section, our intuition is incorrect. We consider now $F(x,y) = (x^2y - x - 1)^2 + (x^2 - 1)^2$. A routine calculation shows that $F$ has two critical points at $(1,2)$ and $(-1,0)$, both of which are local minima. A look at the contours of $F$ shows that there is a ridge running parallel to the positive $y$-axis which separates the valleys that $(1,2)$ and $(-1,0)$ lie in. What happens as we pour water into the valley containing the local minimum at $(1,2)$? As the water level rises, the shoreline advances up the valley, in the direction of the positive $y$-axis. Now, notice first that a shoreline is part of a contour of $F$. 

Fig. 1.3: Examples of “bad” behavior

Fig. 2.1: Shorelines at different levels ($\Delta h = .3$) - note how they spread out along the positive $y$-axis!
(because the water level along a shore is constant), and second that along the positive $y$-axis, the portion of the shorelines at water level $h$ and $h + \Delta h$ get farther and farther apart. (See Figure 2.1.) Next, recall that $\nabla F(x, y)$ tells us two geometric things the graph of $F$ at the point $(x, y)$:

(i) The direction of $\nabla F(x, y)$ tells us the direction of maximum increase of $F$ (so which way is straight uphill),

(ii) The length gives us the size of that increase, i.e. the size of $\nabla F(x, y)$ tells us about the steepness of the landscape at $(x, y)$.

If the contours of $F$ are far apart, then the length of $\nabla F(x, y)$ must be small. Since the shorelines along the positive $y$-axis get farther apart as the water level rises, there must be a sequence $(x_n, y_n)$ on those shorelines for which $|\nabla F(x_n, y_n)| \to 0$. Notice also that along this sequence $F(x_n, y_n)$ increases (since the water level rises) and yet $F(x_n, y_n)$ is bounded from above (since the water level remains below the ridge separating the valleys). These types of sequences play such an important role in variational methods that they have their own name:

**Definition 2.1.** Suppose that $F : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function. A sequence of points $x_n$ such that

(i) $F(x_n)$ is bounded and

(ii) $|\nabla F(x_n)| \to 0$

is called a Palais-Smale sequence (or simply a (PS)-sequence) for $F$.

**Example:** For $F(x, y) = (x^2 - y - 1)^2 + (x^2 - 1)^2$, then $(x_n, y_n) = (1/n^2, 2n)$ is a PS-sequence for $F$. Notice that this sequence does not converge! That $F$ has a non-convergent PS-sequence is a sign of its “bad” behavior. Note also that $(x_1, y_1) = (1, 2)$, and that $(x_n, y_n)$ is in the valley that contains $(1, 2)$.

**Definition 2.2 (PS condition).** If $F : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, we say that $F$ satisfies the Palais-Smale condition (or simply $F$ satisfies (PS)) if every Palais-Smale sequence of $F$ has a convergent subsequence.

Note that neither of the last examples in the previous section satisfies (PS)! Geometrically, the Palais-Smale condition can be thought of as a “steepness” condition on the landscape given by the graph of $F$: away from critical points, the landscape has at least some minimal steepness. From this point of view, the contours for the last examples in the previous section make it clear that those functions don’t satisfy (PS): in certain directions, their landscapes get flatter and flatter and yet don’t have critical point. Analytically, the Palais-Smale condition is a compactness requirement on the function $F$. Much more information about the history and genesis of the (PS) condition may be found in [17].

**3. Mountain Passes.** Throughout this section, we will assume that $F : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable. Suppose now that $F$ satisfies the following:

(MP1) $F(0) = 0$.

(MP2) There is an $r > 0$ and an $\alpha > 0$ such that $F(x) \geq \alpha$ for all $x$ with $|x| = r$.

(MP3) There is an $x_1$ such that $|x_1| > r$ and $F(x_1) \leq 0$.

Geometrically, we think of 0 as lying in a valley, surrounded by a range of mountains whose minimum height at distance $r$ from 0 is at least $\alpha$, see Figure 3.1. Ideally, there should be a mountain pass over the mountains. Unfortunately, as $F$ in the last section shows, this may not be true. But we can show the following:

**Theorem 3.1.** Suppose $F$ satisfies (MP1-3). Then, there is a (PS)-sequence $x_n$ such that $F(x_n) \to c$, where $c \geq \alpha$. 
The idea of the proof is: pick a path $\gamma$ that connects $0$ and $x_1$, and think of it as a long rubber string. Next, move all the points on this string “downhill” (decreasing the value of $F$). As we do this, the string will move and stretch over the landscape. Notice however that since all points on the string are moved downhill, points that start in the valley will remain in the valley and points that are close to $x_1$ must remain outside the valley. Thus, every time we move the string, it must still cross the mountain range, and so if $x_n$ is the highest point along the deformed string, $F(x_n) \geq \alpha$. To see why $|\nabla F(x_n)| \to 0$, recall that the length of the gradient describes the steepness of the landscape and if the landscape is very steep, then a small change in $x$ can make a large change in the value of $F(x)$. Therefore, if $|\nabla F(x_n)|$ did not go to zero, then every time the string was deformed, the maximum of $F$ along the string would decrease too much and eventually we would have $F(x_n) < \alpha$.

4. How to move points “downhill”. A key ingredient in the proof of Theorem 3.1 is the idea of “pushing” a point “downhill” on the graph of $F$. Before we give the proof, we need to be more explicit as to what this function is. Given a point $x$, we will want to push $x$ in such a fashion that the direction of movement is in the same direction as the negative gradient of $F$ at $x$. That is, if $\varphi_t(x)$ represents the function (in $t$) that moves a point $x$, we ideally would use:

$$\frac{d}{dt} \varphi_t(x) = -\nabla F(\varphi_t(x)),$$

$$\varphi_0(x) = x.$$

Since $F$ is twice continuously differentiable, the standard existence and uniqueness theorems for initial value problems apply. However, there is a difficulty: these solutions may not exist for all $t > 0$ (for example, if $F(x, y) = \frac{x^2}{3}$, solutions may “blow-up” in finite time). To simplify things, we would like solutions to exist for all time. This can be guaranteed by making the right side of the differential equation bounded. With
this in mind, let \( w(x) := \frac{\|\nabla F(x)\|}{1 + \|\nabla F(x)\|^2} \) for \( x \in \mathbb{R}^n \), and consider

\[
\frac{d}{dt} \varphi_t(x) = -w(\varphi_t(x)) \nabla F(\varphi_t(x))
\]

(4.1)

\[ \varphi_0(x) = x. \]

Note that the right side of the first equation in (4.1) satisfies

\[
\|w(\varphi_t(x))\nabla F(\varphi_t(x))\| = \frac{\|\nabla F(\varphi_t(x))\|}{1 + \|\nabla F(\varphi_t(x))\|^2} \|\nabla F(\varphi_t(x))\|
\]

\[
= \frac{\|\nabla F(\varphi_t(x))\|^2}{1 + \|\nabla F(\varphi_t(x))\|^2} \leq 1,
\]

and so solutions of (4.1) will exist for all time. (The inequality above also implies that the maximum speed at \( \varphi_t(x) \) moves \( x \) is 1.) Note that since \( w(x) \geq 0 \) for all \( x \in \mathbb{R}^n \), if \( \varphi_t(x) \) solves this initial value problem, then the first equation of 4.1 says that the tangent vector to the curve \( \varphi_t(x) \) has the same direction as the negative gradient (so we push in the direction of the negative gradient), and the second equation of 4.1 says that the initial position is \( x \). Thus, for a given initial position \( x \), \( \varphi_t(x) \) describes how \( x \) moves in time if we always push straight downhill. Note also that \( \varphi_t(x) = x \) if \( x \) is a critical point of \( f \). There are two important features of \( \varphi_t(x) \):

(i) \( \varphi_t(x) \) is a continuous function of \( t \) and \( x \), and

(ii) \( \varphi_{t_2}(\varphi_{t_1}(x)) = \varphi_{t_1+t_2}(x) \).

Property (i) means that if \( x \) and \( y \) are sufficiently close together then after pushing them for the same amount of time, they must remain close. (A word of caution: the larger \( t \) is, the closer together the points must be at the beginning! (i) does not say that two points that begin close together must remain close for all \( t > 0 \).) Property (ii) says that if we begin at \( x \), push \( x \) for some time \( t_1 \) to the point \( \varphi_{t_1}(x) \), and then use \( \varphi_{t_1}(x) \) as a new initial value to push downhill for some amount of time \( t_2 \) and end up at \( \varphi_{t_2}(\varphi_{t_1}(x)) \), we have the same effect as if we had just let \( x \) be pushed down for \( t_1 + t_2 \) to get to \( \varphi_{t_1+t_2}(x) \). Property (ii) is often referred to as the semi-group property.

5. Proving Theorem 3.1. We are now ready to prove Theorem 3.1:

Proof. Notice that for any fixed \( x \in \mathbb{R}^n \), we have the following:

\[
\frac{d}{dt} F(\varphi_t(x)) = \langle \nabla F(\varphi_t(x)), \frac{d}{dt} \varphi_t(x) \rangle
\]

(5.1)

\[
= \langle \nabla F(\varphi_t(x)), -w(\varphi_t(x)) \nabla F(\varphi_t(x)) \rangle
\]

\[
= -w(\varphi_t(x)) \|\nabla F(\varphi_t(x))\|^2 \leq 0.
\]

(5.1) says that \( F \) always decreases along \( \varphi_t(x) \), which means that \( \varphi_t(x) \) moves \( x \) downhill.

Next, we claim that \( \|\varphi_t(0)\| < r \) and \( \|\varphi_t(x_1)\| > r \) for all \( t \geq 0 \). Geometrically (see Figure 3.1), this says that \( \varphi_t(0) \) and \( \varphi_t(x_1) \) remain inside and outside respectively of the mountain range, which is geometrically clear since \( \varphi_t \) moves points downhill and both \( 0 \) and \( x_1 \) are below the minimum height of the mountain range. Analytically, if there was a \( t > 0 \) such that \( \|\varphi_t(0)\| = r \), then assumptions (MP1) and (MP2) and (5.1) would imply

\[ \alpha \leq F(\varphi_t(0)) \leq F(0) < \alpha, \]
which is a contradiction. A similar argument implies that \( \| \varphi_t(x_1) \| > r \) for all \( t > 0 \).

Suppose now that \( \gamma : [0, 1] \to \mathbb{R}^n \) is continuous, \( \gamma(0) = 0 \) and \( \gamma(1) = x_1 \). (Thus, the image of the interval \([0, 1]\) is a path that connects \( 0 \) and \( x_1 \).) For any \( i \in \mathbb{N} \), consider \( \gamma_i : [0, 1] \to \mathbb{R}^n \) given by \( \gamma_i := \varphi_i \circ \gamma \). (\( \gamma_i \) is \( \gamma \) deformed by \( \varphi_i \) for \( i \) units of time, see Figure 5.1.) Because \( \gamma \) and \( \varphi_i \) are continuous, so too is \( \gamma_i \), and so there is an

\[
\begin{align*}
\text{Fig. 5.1: } & \gamma \text{ (magenta), } \gamma_1 \text{ (cyan) and } \gamma_2 \text{ (green). Note that where the contours are close together (steep terrain), the path is deformed more than where the contours are far apart (flat terrain).} \\
s_i \in [0, 1] \text{ such that } & F(\gamma_i(s_i)) = \max_{s \in [0, 1]} F(\gamma_i(s)). \text{ Note that } \gamma_i(s_i) \text{ is a high point along the path } \gamma_i. \text{ Because } \gamma_i(0) = \varphi_i(\gamma(0)) = \varphi_i(0) \text{ and } \gamma_i(1) = \varphi_i(\gamma(1)) = \varphi_i(x_1), \text{ we know that } \| \gamma_i(0) \| < r < \| \gamma_i(1) \|. \text{ Thus, by the intermediate value theorem, there is an } s \text{ such that } \| \gamma_i(s) \| = r. \text{ Therefore, by assumption (ii), we know that} \\
\alpha \leq F(\gamma_i(s)) & \leq \max_{s \in [0, 1]} F(\gamma_i(s)) \leq F(\gamma_i(s_i)). \\
\end{align*}
\]

Notice that we have a sequence \( s_i \) in \([0, 1]\). Because \([0, 1]\) is compact, there is a subsequence \( s_i \) which converges to some \( s^* \in [0, 1] \). Let \( x^* := \gamma(s^*) \).

We now claim that \( F(\varphi_t(x^*)) \geq \alpha \) for all \( t \geq 0 \). If not, there must be a \( \tau > 0 \) such that \( F(\varphi_{\tau}(x^*)) < \alpha \). Since \( x^* = \lim_{s \to \infty} \gamma(s_i) \) and \( F \) is continuous, we know that \( F(\varphi_{\tau}(\gamma(s_i))) < \alpha \)

for all large \( j \). By picking \( J \) sufficiently large, we may assume that \( i_J > \tau \). We have

\[
\begin{align*}
\alpha & \leq F(\varphi_{i_J}(\gamma(s_i))) \quad \text{(by (5.2))} \\
& \leq F(\varphi_{\tau}(\gamma(s_i))) \quad \text{(by (5.1), since } \tau < i_J) \\
& < \alpha,
\end{align*}
\]

which is impossible. Thus, \( F(\varphi_t(x^*)) \geq \alpha \) for all \( t \geq 0 \).
Since \( F(\varphi_t(x^*)) \) is decreasing as a function of \( t \) and is bounded from below, we know that \( \lim_{t \to \infty} F(\varphi_t(x^*)) \) exists. Since
\[
\int_0^\infty \frac{d}{dt} F(\varphi_t(x^*)) \, dt = \left( \lim_{t \to \infty} F(\varphi_t(x^*)) \right) - F(x^*),
\]
we know that the integral
\[
\int_0^\infty -\frac{d}{dt} F(\varphi_t(x^*)) \, dt
\]
is finite. By (5.1), we then have
\[
\int_0^\infty w(\varphi_t(x^*)) \| \nabla F(\varphi_t(x^*)) \|^2 \, dt < \infty,
\]
and so there is a sequence \( t_n \) such that \( t_n \to \infty \) and \( w(\varphi_{t_n}(x^*)) \| \nabla F(\varphi_{t_n}(x^*)) \|^2 \to 0 \).

By definition of \( w(x) \), we have
\[
\| \nabla F(\varphi_{t_n}(x^*)) \|^3 \over 1 + \| \nabla F(\varphi_{t_n}(x^*)) \|^2 \to 0,
\]
as \( n \to \infty \), and so \( \| \nabla F(\varphi_{t_n}(x^*)) \| \to 0 \). Since \( F(\varphi_t(x^*)) \geq \alpha \) and \( F(\varphi_t(x^*)) \) is decreasing, \( x_n := \varphi_{t_n}(x^*) \) is a (PS)-sequence with the correct behavior if we define \( c := \lim_{t \to \infty} F(\varphi_t(x^*)) \).

We have the following automatic corollary:

**Corollary 5.1.** Suppose that \( F \) satisfies the conditions of Theorem 3.1 and \( F \) satisfies the (PS)-condition. Then \( F \) has a critical point \( x_2 \) such that \( F(x_2) \geq \alpha \).

### 6. A closer analysis of the proof of Theorem 3.1.

In the proof of Theorem 3.1, we used the “high points” \( \gamma_i(s_i) \) along the deformed paths \( \gamma_i \) to find an appropriate point \( x^* \) on the initial curve \( \gamma \). Using these deformed paths, we can actually get a better bound on the critical value \( F(x_2) \) than that provided in Corollary 3.1. For this, recall the geometric picture of a valley surrounded by a range of mountains. If \( F \) satisfies the (PS)-condition, we know by Corollary 5.1 that there is a mountain pass leading out of the valley. In some sense, a mountain pass between \( 0 \) and \( x_1 \) should occur at the lowest high point among all possible paths connecting \( 0 \) and \( x_1 \). That is:

\[
\hat{c} := \inf_{\gamma \in \Gamma} \max_{s \in [0,1]} F(\gamma(s))
\]

should be the elevation of the lowest mountain pass leading out of the valley containing \( 0 \) if \( \Gamma \) is the set of paths which begin in the valley containing \( 0 \) and end outside the valley at a level lower than \( F(x_1) \), i.e.

\[
\Gamma := \{ \gamma \in C([0,1], \mathbb{R}^n) : \| \gamma(0) \| < r < \| \gamma(1) \|, F(\gamma(0)) \leq F(0), \text{and } F(\gamma(1)) \leq F(x_1) \}.
\]

We can show the following:

**Lemma 6.1.** Suppose that \( F \) satisfies the conditions of Theorem 3.1. Then

1. \( \hat{c} \geq \alpha \), where \( \hat{c} \) is from 6.1.
2. If \( x_n \) is the (PS)-sequence from Theorem 3.1 and \( c := \lim_{n \to \infty} F(x_n) \), then \( c \geq \hat{c} \), where \( \hat{c} \) is from (6.1).
\textbf{Proof.} To prove (1), we show that if $\gamma \in \Gamma$, then there is an $\hat{s}$ such that $\|\gamma(\hat{s})\| = r$. Assuming that, we will have $F(\gamma(\hat{s})) \geq \alpha$ by (MP2). Since this is true for all $\gamma \in \Gamma$, $\hat{c} \geq \alpha$ follows from (6.1). To see that there is an appropriate $\hat{s}$, note that since $\|\gamma(0)\| < r < \|\gamma(1)\|$ and $\|\gamma(s)\|$ is a continuous function of $s$, the intermediate value theorem implies the existence of an appropriate $\hat{s}$.

For (2), we need only show that the high points $\gamma_i(s_i)$ on the curves $\gamma_i$ satisfy $F(\gamma_i(s_i)) \geq \hat{c}$. But this follows immediately from the fact that $\gamma_i \in \Gamma$ and the definition of $\hat{c}$. \qed

An immediate corollary:

\textbf{Corollary 6.2.} Suppose $F$ satisfies the conditions of Theorem 3.1 and the (PS) condition. Then there is a critical point $x_2$ such that $F(x_2) \geq \hat{c}$.

In fact, we have the following:

\textbf{Lemma 6.3.} If $F$ satisfies the conditions of Theorem 3.1 and the (PS) condition, there is a critical point $z$ with $F(z) = \hat{c}$.

Lemma 6.3 is often called the Mountain Pass Lemma, and is due to Ambrosetti and Rabinowitz ([1]). It is one of the most important tools in finding critical points of functions that may not be bounded from below, and has a huge number of applications. For example, the bibliography of [13] has over a thousand references!

\textbf{Proof.} By the definition of $\hat{c}$, for each $j \in \mathbb{N}$, there is a curve $\gamma_j$ with $\hat{c} \leq \max_{s \in [0,1]} F(\gamma_j) \leq \hat{c} + \frac{1}{j}$. For each of these curves $\gamma_j$, we may repeat the procedure in the proof of Theorem 3.1, use the assumption that $F$ satisfies the (PS) condition and Lemma 6.1 to show that there exists a critical point $x_j$ with $\hat{c} \leq F(x_j) \leq \hat{c} + \frac{1}{j}$. Notice that since $\nabla F(x_j) = 0$, the previous inequality implies that $x_j$ is a (PS)-sequence for $F$. Since $F$ satisfies the (PS) condition, there is a subsequence $x_{j_k}$ that converges to some $z$. Since $F$ is twice continuously differentiable, $\nabla F(z) = 0$. Since $\hat{c} \leq F(x_{j_k}) \leq \hat{c} + \frac{1}{j_k}$, we must also have $F(z) = \hat{c}$, as desired. \qed

In the Mountain Pass Lemma, we begin with a known critical point surrounded by mountains (the local minimum at $0$) and find a second one by picking a path connecting the known critical point to the region outside the mountains and pushing the points on the path downhill. Unfortunately, there are situations when the function doesn’t have this sort of geometry. For example, if $A$ is an $n \times n$ symmetric matrix with positive and negative eigenvalues, then the function $F_b(x) = \frac{1}{2}Ax \cdot x - x \cdot b$ doesn’t satisfy the conditions to apply the Mountain Pass Lemma.

\textbf{7. Saddle Points.} When $A$ is an $n \times n$ symmetric matrix with positive and negative eigenvalues and $F_b(x) = \frac{1}{2}Ax \cdot x - x \cdot b$, there are complementary subspaces $A_+$ and $A_-$ of $\mathbb{R}^n$ on which $F$ goes to $+\infty$ or $-\infty$, respectively. $A_+$ is the subspace spanned by the eigenvectors with positive eigenvalues and $A_-$ is the subspace spanned by the eigenvectors with negative eigenvalues. More generally, throughout this section we assume that $F$ is twice continuously differentiable and that $F$ satisfies the following:

\begin{itemize}
  \item[(SP1)] There are two subspaces $A_+$ and $A_-$ of $\mathbb{R}^n$ such that $\mathbb{R}^n = A_+ \oplus A_-$. \\
  \item[(SP2)] There exists an $\alpha$ such that $F(x) \geq \alpha$ for all $x \in A_+$. \\
  \item[(SP3)] There exist $r > 0$ and $\beta < \alpha$ so that $F(x) \leq \beta$ for all $x \in A_-$ with $\|x\| = r$.
\end{itemize}

In Figure 7.1, we suppose that $F : \mathbb{R}^3 \to \mathbb{R}$, $A_+$ is one dimensional and $A_-$ is two dimensional, which is why $A_- \cap B_r(0)$ is a circle. The plusses along $A_+$ mean that along $A_+$, $F$ is positive (so $\alpha > 0$), while the minus signs around $A_- \cap B_r(0)$ mean that there $F$ is negative ($\beta \leq 0$). Recall that $\mathbb{R}^n = A_+ \oplus A_-$ means that every $x \in \mathbb{R}^n$ may be written as $x = y + z$ for unique $y \in A_+$ and $z \in A_-$. 

Fig. 7.1: Saddle Point Geometry for $F: \mathbb{R}^3 \to \mathbb{R}$.

**Example:** Suppose $F(x, y) = x^2 - y^2$. We have

$$\mathbb{R}^2 = \{(x, 0) : x \in \mathbb{R}\} \oplus \{(0, y) : y \in \mathbb{R}\} = A_+ \oplus A_-.$$  

Note that if $|x| = r$, then $F(x, 0) = r^2 > 0$, while $F(0, y) = -y^2 \leq 0$. Thus, we may take $\alpha = r^2$ and $\beta = 0$. See Figure 7.2.

Fig. 7.2: Saddle Point Geometry for $F(x, y) = x^2 - y^2$.

**Example:** More generally, suppose that $A$ is an $n \times n$ matrix in block form:

$$A = \begin{bmatrix} P & B^T \\ B & N \end{bmatrix},$$
where $P$ is a $j \times j$ symmetric positive definite matrix, $N$ is a $(n-j) \times (n-j)$ symmetric negative definite matrix and $B$ is a $(n-j) \times j$ matrix. Notice that this means $A$ is a symmetric matrix, and so solutions of $Ax = b$ may be found by looking for critical points of $F_b(x) = \frac{1}{2} \langle x, Ax \rangle - \langle x, b \rangle$. Writing $x \in \mathbb{R}^n$ as

$$x = \begin{bmatrix} y \\ z \end{bmatrix}$$

where $y$ consists of the first $j$ components of $x$ and $z$ consists of the remaining $n-j$ components, we may decompose $\mathbb{R}^n$ as

$$\mathbb{R}^n = \mathbb{R}^j \times \{0\} \oplus \{0\} \times \mathbb{R}^{n-j} = A_+ \oplus A_-.$$ 

If $x \in A_+$ (i.e. $x = \begin{bmatrix} y \\ 0 \end{bmatrix}$), then $F_b(x) = \frac{1}{2} \langle y, Py \rangle - \langle y, P\Pi b \rangle$, where $\Pi$ is the projection onto the first $j$ components. Since $P$ is symmetric and positive definite, for any $x \in A_+$, we will have

$$F_b(x) \geq \lambda_1 \|y\|^2 - \|b\| \cdot \|y\| = \lambda_1 \|x\|^2 - \|b\| \cdot \|x\|,$$

where $\lambda_1$ is the smallest eigenvalue of $P$. Now, the smallest that $\lambda_1 \|x\|^2 - \|b\| \cdot \|x\|$ can be is $-\frac{\|b\|^2}{4\lambda_1}$ which means

$$F_b(x) \geq -\frac{\|b\|^2}{4\lambda_1} \text{ for all } x \in A_+,$$

and so we may take $\alpha = -\frac{\|b\|^2}{4\lambda_1}$. Similarly, if $x \in A_-$ (i.e. $x = \begin{bmatrix} 0 \\ z \end{bmatrix}$), we will have

$$F_b(x) \leq \hat{\lambda}_1 \|x\|^2 + \|b\| \cdot \|x\|,$$

where $\hat{\lambda}_1$ is the largest eigenvalue of $N$. Since $\hat{\lambda}_1$ is negative, there is an $\tilde{r}$ such that $\lambda_1 \tilde{r}^2 + \|b\| \tilde{r} < \alpha$. Taking $r = \tilde{r}$ and $\beta = \lambda_1 \tilde{r}^2 + \|b\| \tilde{r}$, we have

$$F_b(x) \leq \beta < \alpha \text{ for all } x \in A_- \text{ with } \|x\| = \tilde{r}.$$ 

Therefore, $F_b(x) = \frac{1}{2} \langle Ax, x \rangle - \langle x, b \rangle$, $F$ satisfies (SP1-3). Note that the first example takes $P = [1]$, $N = [-1]$ and $B = [0]$.

We can show the following

**Theorem 7.1.** Suppose $F$ satisfies (SP1-3). Then, there is a (PS)-sequence $x_n$ such that $F(x_n) \to c$, where $c \geq \alpha$.

The idea of the proof is similar to what was done in the previous section for a mountain pass. There, we began with a path $\gamma$ and deformed it downhill by using $\varphi_\gamma$ as defined in (4.1). In the saddle point setting, if $A_-$ is $j$ dimensional, we begin with a $j$ dimensional sub-surface in $\mathbb{R}^n$ whose boundary is the ball of radius $r$ in the subspace $A_-$, and deform it. By keeping track of the largest value of $F$ (the “high” points) on these deformations, we can find an appropriate initial point $x^\ast$ for which $F(\varphi_\gamma(x^\ast))$ is bounded from below.

An important ingredient in the mountain pass was an intersection property: every deformed path intersected the mountain range, which enabled us to show that our (PS) sequence satisfied $F(x_n) \to c \geq \alpha$. We will need a similar ingredient in this setting, namely that every deformed sub-surface intersects $A_+$. (See Figure 7.3.) Intuitively,
Fig. 7.3: Because $F$ on $\partial B_r(0) \cap A_-$ is smaller than $F$ on $A_+$, the intersection of the surface and $A_+$ cannot be removed by deforming the surface if the deformation must decrease $F$.

the idea is clear: notice that on the boundary of the initial surface, $F(x) \leq \beta < \alpha$, while on the line (representing the subspace $A_+$), $F(x) \geq \alpha$. If we deform the surface using (4.1), the points from the boundary cannot cross the subspace $A_+$ since $\varphi_t$ moves points in such a fashion as to make $F$ smaller, and $F$ on $A_+$ is larger than $F$ on the boundary. Thus, the deformed surface must also always intersect $A_+$.

8. Proof of Theorem 7.1. The proof of Theorem 7.1 is much more technical than the corresponding proof of Theorem 3.1, since in the Mountain Pass setting we can use the intermediate value theorem to show the deformed paths intersect the mountain range. Here, since we are higher dimensions, we will need more sophisticated machinery.

Proof. Let $\gamma : \overline{B_r(0)} \cap A_- \to \mathbb{R}^n$ satisfy $\gamma(x) = x$ for all $x \in \partial B_r(0) \cap A_-$. (We could for example take $\gamma$ to be the identity.) For each $i \in \mathbb{N}$, let $\gamma_i : \overline{B_r(0)} \cap A_- \to \mathbb{R}^n$ be defined by $\gamma_i(x) := \varphi_i(\gamma(x))$. (Recall that $\varphi_i(x)$ is the solution of (4.1), and so $\varphi_i(x)$ pushes $x$ “downhill”.) Let us assume for the moment that

(8.1) for each $i \in \mathbb{N}$, there is an $\hat{x}_i \in \overline{B_r(0)} \cap A_-$ such that $\varphi_i(\gamma(\hat{x}_i)) \in A_+$.

Since $\overline{B_r(0)} \cap A_-$ is compact, there is an $x_i \in \overline{B_r(0)} \cap A_-$ such that

\[
F(x_i) = \max_{x \in \overline{B_r(0)} \cap A_-} F(x) \geq \alpha,
\]

where for the inequality we have used (8.1) and assumption (SP2). Since $\overline{B_r(0)} \cap A_-$ is compact, there is a subsequence $x_{i_j}$ such that $x_{i_j} \to x^* \in \overline{B_r(0)} \cap A_-$. We claim now that $F(\varphi_{\tau}(\gamma(x^*)))$ is bounded from below. If not, then there is a $\tau$ such that

\[
F(\varphi_{\tau}(\gamma(x^*))) < \beta.
\]

Since $x_{i_j} \to x^*$ and $\varphi_{\tau}$ and $\gamma$ are continuous, we then know that for all suitably large $j$ that

\[
F(\varphi_{\tau}(\gamma(x_{i_j}))) < \beta.
\]
Taking \( J \) sufficiently large that \( i_J > \tau \), (8.2) would imply that
\[
\alpha \leq F(\varphi_{i_J}(\gamma(x_{i_J}))) \leq F(\varphi_*(\gamma(x_{i_J}))) < \beta,
\]
which contradicts assumption (SP3) that \( \beta < \alpha \). Since \( F(\varphi_*(\gamma(x^*))) \) is bounded from below, we may make the same argument as in the proof of Theorem 3.1 to conclude the existence of an appropriate (PS)-sequence \( x_\alpha \).

It remains only to prove (8.1). We will need the notion of the degree of a continuous mapping, which makes the remainder of this proof more technical. The intuitive idea of why (8.1) holds is explained in Figure 7.3, and beginning readers should feel no compulsion to read the following details.

Suppose \( U \) is a bounded open subset of \( \mathbb{R}^l \), \( f : \overline{U} \to \mathbb{R}^l \) is twice continuously differentiable, \( c \notin f(\partial U) \) and \( f'(x) \) is invertible for all \( x \in f^{-1}(c) \). For any such \( f, U, c \), we define
\[
d(f, U, c) := \sum_{x \in f^{-1}(c)} \text{sgn}(\det f'(x)).
\]
Notice that \( d(f, U, c) \) has a couple of “obvious” properties:

(d1) \( d(id, U, c) = 1 \) if \( c \in U \), and \( d(id, U, c) = 0 \) if \( c \notin U \), where \( id \) is the identity mapping.

(d2) If \( d(f, U, c) \neq 0 \), then there is at least one \( x \in U \) with \( f(x) = c \).

(d3) \( f \) is jointly continuous in \( s \) and \( x \) and there is no \( (s, x) \in [0, 1] \times \partial U \) for which \( h(s, x) = c \), then \( d(h(s, \cdot), U, c) \) is independent of \( s \).

(d4) If \( f(x) = g(x) \) for all \( x \in \partial U \), then \( d(f, U, c) = d(g, U, c) \). (In fact, (d4) is a consequence of (d3).)

This definition of degree can be extended to continuous functions \( f : \overline{U} \to \mathbb{R}^n \) and any \( c \notin f(\partial U) \); see [10], [19], or [15]. Even though degree theory may be an unfamiliar topic for students, it is very useful beyond its application here. Additional applications can be found in the above references.

Let \( P : \mathbb{R}^n \to A_- \) be the projection onto \( A_- \). By assumption (SP1), \( Pu = 0 \) if and only if \( u \in A_+ \). Using (d2), (8.1) will then follow by showing \( d(P\varphi_*(\gamma(\cdot)), B_r(0) \cap A_-) \neq 0 \).

Notice that since \( \gamma(x) = x \) for all \( x \in \partial B_r(0) \cap A_- \), \( P\gamma(x) = Px = x \). Thus, \( id(x) = P\gamma(x) \) for all \( x \in B_r(0) \cap A_- \). By (d1) and (d4), we then have
\[
d(P\gamma(\cdot), B_r(0) \cap A_-) = d(id, B_r(0) \cap A_-, 0) = 1.
\]

Next, we will use (d3), so we need to find an appropriate \( h \). For \( x \in \overline{B_r(0)} \cap A_- \) and \( s \in [0, 1] \), let
\[
h : [0, 1] \times \overline{B_r(0)} \cap A_- \to A_- \text{ be given by } h(s, x) := P\varphi_{is}(\gamma(x)).
\]
Note that \( h(0, x) = P\gamma(x), h(1, x) = P\varphi_{i}(\gamma(x)) \), and \( h(t, x) \) is continuous in \( t \) and \( x \).

If we can show that \( h(s, x) \neq 0 \) for all \( (s, x) \in [0, 1] \times \partial B_r(0) \cap A_- \), then
\[
d(P\varphi_*(\gamma(\cdot)), B_r(0) \cap A_-) = d(h(1, \cdot), B_r(0) \cap A_-, 0) \quad (\text{by (8.4)})
\]
\[
= d(h(0, \cdot), B_r(0) \cap A_-, 0) \quad (\text{by (d3)})
\]
\[
= d(P\gamma(\cdot), B_r(0) \cap A_-, 0) \quad (\text{by (8.4)})
\]
\[
= 1 \neq 0, \quad (\text{by (8.3)})
\]
as desired.

It remains to show that \( h(s,x) \neq 0 \) for all \((s,x) \in [0,1] \times \partial B_r(0) \cap A_- \). Suppose that there is in fact a \( \tau \in [0,1] \) and \( x_0 \in \partial B_r(0) \cap A_- \) such that \( h(\tau,x_0) = 0 \). By (8.4), \( P\varphi_{\tau}(\gamma(x_0)) = 0 \), and so (by definition of \( P \)) \( \varphi_{\tau}(\gamma(x_0)) \in A_+ \). \( \text{(SP2)} \) then implies that \( F(\varphi_{\tau}(h(x_0))) \geq \alpha \). By (5.1), we then have

\[
\alpha \leq F(\varphi_{\tau}(\gamma(x_0))) \leq F(\gamma(x_0)) \leq \beta,
\]
since \( x_0 \in \partial B_r(0) \cap A_- \) and by (SP3). This contradicts \( \alpha > \beta \), and so there can be no such \( \tau \) and \( x_0 \).

We have the following immediate corollary:

**Corollary 8.1.** Suppose \( F \) satisfies the assumptions of Theorem 7.1 and the \( (PS) \) condition. Then \( F \) has a critical point \( \hat{x} \) such that \( F(\hat{x}) \geq \alpha \).

As it turns out, there is a minimax value similar to (6.1) in this setting:

\[
\hat{c} = \inf_{\gamma \in \Gamma} \max_{x \in \partial B_r(0) \cap A_-} F(\gamma(x))
\]

where

\[
\Gamma = \left\{ \gamma \in C(\partial B_r(0) \cap A_-; \mathbb{R}^n) : \gamma(x) = x \text{ for all } x \in \partial B_r(0) \cap A_- \right\}
\]

It can be shown that statements analogous to Lemma 6.1, Corollary 6.2 and Lemma 6.3 are true if \( F \) satisfies (SP1-2) and (PS). Proofs may be found in [5].

**9. From \( \mathbb{R}^n \) to Hilbert space.** So far, we’ve worked primarily in the setting of \( \mathbb{R}^n \), but as it turns out, with a slight modification of assumption (SP1), every statement we’ve given holds true in Hilbert space as well! Throughout this section, we suppose \( E \) is a Hilbert space. If \( F : E \to \mathbb{R} \) is differentiable, what is \( \nabla F(x) \)?

Recall that in \( \mathbb{R}^n \), \( \nabla F(x) \) has the following relationship to \( F'(x) \):

\[
\langle \nabla F(x), h \rangle = F'(x)h \text{ for all } h \in \mathbb{R}^n,
\]

where \( F'(x) \) is a linear operator on \( \mathbb{R}^n \) and \( \langle \cdot, \cdot \rangle \) is an inner product on \( \mathbb{R}^n \). Thus, on \( E, \nabla F(x) \) should have the same property. If \( F \) is (Frechet-)differentiable at \( x \), \( F'(x) \) defines a linear functional on \( E \) by \( h \mapsto F'(x)h \), and so the Riesz Representation Theorem guarantees the existence of an appropriate \( \nabla F(x) \). Thus, if \( F \) is twice continuously differentiable, we may define an appropriate \( \varphi_t \) using (4.1). In fact, we can get away with slightly weaker conditions on \( F \): we need only assume that \( x \mapsto F'(x) \) is locally Lipschitz continuous to define \( \varphi_t \) by (4.1).

Critical points in infinite dimensions often correspond to solutions of differential equations:

**Example:** Consider the boundary value problem

\[
\begin{align*}
 u''(x) &= \cos(u(x)) \text{ for all } x \in (0,1) \\
 u(0) &= 0 \text{ and } u(1) = 0
\end{align*}
\]

If we multiply the differential equation by any smooth function \( h(x) \) satisfying \( h(0) = 0 = h(1) \) and integrate over the interval \([0,1] \), we have

\[
\int_0^1 u''(x)h(x) \, dx = \int_0^1 \cos(u(x))h(x) \, dx.
\]
Integrating the left hand side by parts (note that the boundary terms vanish, since $h(0) = 0$ and $h(1) = 0$), we then have

$$- \int_0^1 u'(x)h'(x) \, dx = \int_0^1 \cos(u(x))h(x) \, dx,$$

or equivalently

$$\int_0^1 u'(x)h'(x) + \cos(u(x))h(x) \, dx = 0.$$

Since this must be true for any smooth $h$ that vanishes at 0 and 1, any solution of 9.2 has the property that

$$\int_0^1 u'(x)h'(x) + \cos(u(x))h(x) \, dx = 0$$

for all smooth $h$ with $h(0) = 0 = h(1)$, which is the weak form of the Euler-Lagrange equations for the functional

$$F(w) := \int_0^1 \frac{1}{2} |w'(x)|^2 + \sin(w(x)) \, dx.$$

In fact, under appropriate assumptions, $F$ is differentiable and

$$F'(w)h = \int_0^1 w'\, h' + \cos(w)h \, dx.$$

Thus, (9.3) means $F'(u) = 0$, which in turn means that $\nabla F(u) = 0$. Thus, a solution of (9.2) is a critical point of $F$, and under appropriate assumptions on the admissible functions $w$, any critical point of $F$ will be a solution of (9.2).

**Example**: Suppose $\Omega$ is an open subset of $\mathbb{R}^n$ and $\partial \Omega$ is a smooth $n-1$ dimensional surface. Under appropriate assumptions on $f : \mathbb{R} \to \mathbb{R}$, if we follow the procedure above, it can be shown that solutions of

$$\Delta u = f'(u) \text{ in } \Omega$$
$$u = 0 \text{ on } \partial \Omega$$

correspond to critical points of the functional

$$F(w) = \int_{\Omega} \frac{1}{2} |\nabla w|^2 - f(w) \, dx.$$

The biggest change in transitioning to infinite dimensions involves the intersection properties that were used to show that the (PS) sequences in Theorems 3.1 and 7.1 satisfied $F(x_n) \geq \alpha$. No change is necessary for the results of Section 3, since the intersection property there uses the intermediate value theorem. In Section 4, we needed to use the degree for functions $f : \partial B_r(0) \cap A_- \to A_-$, and $A_-$ was a finite dimensional space (being a subspace of $\mathbb{R}^n$). However, degree doesn’t generalize to infinite dimensions in a straightforward fashion. (This is related to the rather remarkable fact that the closed unit “sphere” $\{x \in E : \|x\| = 1\}$ in an infinite dimensional Hilbert space is deformable to the the unit ball $\{x \in E : \|x\| \leq 1\}$, which is false in finite dimensions!) A simple option is:

$$(\text{SP1})^* : E = A_+ \oplus A_-, \text{ where } A_- \text{ is finite dimensional.}$$
With assumption (SP1)*, all the conclusions of Section 4 hold even for an infinite dimensional Hilbert space $E$. It should be noted that there are alternatives to (SP1)*. For example, we can restrict the types of functionals $F$ that we consider, so that $\varphi_t$ is of a type that allows us to calculate a degree.

An interesting question is how to implement algorithms for finding mountain passes or saddle points. The treatment here suggests beginning with an appropriate $\gamma$, deforming it by the negative gradient and keeping track of the high points along each deformation. In principle, these high points should converge to a critical point. In the finite dimensional setting, calculating $\nabla F$ is straightforward. In the infinite dimensional setting, calculating $\nabla F$ is less obvious.

**Example:** Suppose 

$$F(w) := \int_0^1 \frac{1}{2}|w'(x)|^2 + \sin(w(x)) \, dx.$$ 

As we’ve shown, critical points of $F$ correspond to solutions of (9.2). An important question is what functions $w$ will be considered. We will take the Sobolev space $E := W^{1,2}_0(0,1)$. (Two excellent references for students unfamiliar with Sobolev spaces and their uses in differential equations are [6] and [12].) In this case, it can be shown that 

$$F'(v)h = \int_0^1 w'(x)h'(x) + \cos(w(x))h(x) \, dx. \tag{9.4}$$

Our question is: for a fixed $w$, what is $\nabla F(w)$? Note that (9.4) defines a linear operator on $E$, and so by the Riesz Representation Theorem, there is a $v \in E$ such that 

$$F'(w)h = \langle v, h \rangle \text{ for all } h \in E,$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product on $E$. Thus, $v = \nabla F(w)$. In our case, 

$$\langle v, h \rangle = \int_0^1 v'(x)h'(x) + v(x)h(x) \, dx.$$ 

Therefore, $v$ must satisfy 

$$\int_0^1 w'(x)h'(x) + \cos(w(x))h(x) \, dx = \int_0^1 v'(x)h'(x) + v(x)h(x) \, dx$$

for all $h \in E$, or equivalently 

$$\int_0^1 (v'(x) - w'(x))h'(x) + (v(x) - \cos(w(x)))h(x) \, dx = 0$$

for all $h \in E$. Suppressing the explicit depending on $x$ and assuming that $w$ and $v$ are sufficiently smooth that we may integrate by parts in the first terms, we then have 

$$\int_0^1 -(v'' - w'')h + (v - \cos(w))h \, dx = 0$$

for all $h \in E$. Thus, $v$ must satisfy 

$$-v''(x) + v(x) = -w''(x) + \cos(w(x)) \text{ for all } x \in (0,1)$$

$$v(0) = 0 = v(1)$$
Notice that this is a linear equation for \( v \). Thus, finding \( \nabla F(w) \) involves solving a linear differential equation. Note also that \( \nabla F(w) \) depends on the inner product. For this particular problem and choice of \( E \), we could have used \( \langle u, v \rangle := \int_0^1 u'(x)v'(x) \, dx \), in which case \( v = \nabla F(w) \) would satisfy

\[
-\nu''(x) = -w''(x) + \cos(w(x)) \quad \text{for all } x \in (0, 1)
\]

\[
v(0) = 0 = v(1).
\]

For more information on how the inner product affects the gradient, the reader is encouraged to consult the excellent paper [14].

10. Decreasing regularity and moving to Banach spaces. Throughout, we have assumed that the function \( F \) is twice continuously differentiable, which was primarily done to explain why there is a unique solution of (4.1). A more general existence and uniqueness theorem implies that (4.1) has a unique solution if we assume only that \( F' \) is locally Lipschitz. Remarkably, it turns out that Theorems 3.1 and 7.1 hold under the assumption that \( F' \) is merely continuous! The necessary tool is the pseudo-gradient. At a point \( x \), we want a vector \( v \) whose length is “comparable” to the size of \( F'(x) \) and so that when we move from \( x \) in the direction of \( v v \), \( F \) increases. More precisely:

**Definition 10.1.** If \( F : \mathbb{R}^n \to \mathbb{R} \) is differentiable at \( x \), we say that \( v \in \mathbb{R}^n \) is a pseudo-gradient for \( F \) at \( x \) if

1. \( \| v \| \leq 2\| F'(x) \| \) and
2. \( F'(x)v \geq \| F'(x) \|^2 \) (recall that \( F'(x) \) is a linear map from \( \mathbb{R}^n \) into \( \mathbb{R} \)).

Since \( F'(x)v \leq \| F'(x) \| \| v \| \), conditions (1) and (2) imply that when \( F'(x) \neq 0 \), we have \( \| F'(x) \| \leq \| v \| \leq 2\| F'(x) \| \), and so the length of \( v \) is comparable to the size of \( F'(x) \). In addition, if \( g(t) := F(x + tv) \), then the chain rule implies that

\[
g'(0) = F'(x)v \geq \| F'(x) \|^2 \quad \text{(by (2))},
\]

and so when \( F'(x) \neq 0 \), \( F \) will increase if we move in the direction of \( v \). Thus, a pseudo-gradient carries the same information as a gradient vector. There are two advantages to using pseudo-gradient vectors:

1. It can be shown that if \( F' \) is merely continuous on \( \mathbb{R}^n \), there exists a locally Lipschitz continuous pseudo-gradient field for \( F \) on \( \{ x \in \mathbb{R}^n : F'(x) \neq 0 \} \).
   At the cost of some extra technicalities, a pseudo-gradient field can be used in place of a gradient field to prove Theorems syz and syz. For details, see [18] or [22].

2. A pseudo-gradient doesn’t need an inner-product, only the norm. Therefore, pseudo-gradients can be used in Banach spaces. This is particularly useful when working on differential equations in the Sobolev spaces \( W^{1,p}(\Omega) \) for \( p \neq 2 \). For details, see [18] or [22].

11. Further Reading. The mountain pass and saddle point theorems are extraordinarily useful results for solving nonlinear equations. They have a huge number of applications in differential equations, of which we have just barely scratched the surface here. At a level accessible to advanced undergraduates and/or beginning graduates without much exposure to functional analysis is [19]. At a more advanced level, there are several books: [3], [9], [18] or [22]. Even more advanced are [21] or [20]. In addition, [13] has an extensive bibliography. It should also be pointed out that it is possible to relax several of the assumptions in Section 3 or 4 and get the same type of results. For example, it is possible to relax the amount of smoothness that \( F \) possesses, or relax the requirements on the geometry. Many of these issues are taken up in the references. Moreover, there are alternatives for “pushing” functions. For
example, in [8] or [4], a flow arising from a semi-linear heat equation is used (although this can also be thought of as changing the inner product, see [14]). In addition, instead of trying to prove that a function $F$ satisfies (PS), a more common approach is to get a (PS) sequence, and then use information about that particular sequence to extract a convergent subsequence (the approach followed in many of the references, in particular [22]). From that point of view, the approach here is useful, since different initial $\gamma$ may lead to different solutions.

REFERENCES