Differentiability Formalism and Sufficient Conditions for Weak or Locally Weak Minima

If you only care for the sufficient conditions, but not for the differentiability formalism, proceed through the 1st bullet, then skip to eqn (3)

Assume $L \in C^2([a,b] \times G \times \mathbb{R}^n \to \mathbb{R})$, where $G \subset \mathbb{R}^n$ is open, and consider the functional I defined by

$$I[y] := \int_a^b L(t, y(t), y'(t))\, dt\ , \quad I : C^1([a,b] \to G \subset \mathbb{R}^n) \to \mathbb{R} \tag{0}$$

is in C^2. More precisely we claim that the 1st and 2nd derivatives are

$$DI[y]\varphi = \int_a^b \Big\{ L_y(t, y(t), y'(t))\varphi(t) + L_{y'}(t, y(t), y'(t))\varphi'(t) \Big\}\, dt \tag{1}$$

and

$$\begin{aligned} D^2I[y](\varphi,\psi) = \int_a^b \Big\{ & \varphi(t)^T L_{yy}(t,y(t),y'(t))\psi(t) + \\ & + \varphi(t)^T L_{yy'}(t,y(t),y'(t))\psi'(t) + \psi(t)^T L_{yy'}(t,y(t),y'(t))\varphi'(t) + \\ & + \varphi'(t)^T L_{y'y'}(t,y(t),y'(t))\psi'(t) \Big\}\, dt \end{aligned} \tag{2}$$

respectively.

• First we must show that the linear mapping $\varphi \mapsto DI[y]\varphi$ defined by (1) is *continuous* from the space $X := C^1([a,b] \to G)$ to $\mathbb{R}$. This we need to show because the very definition of 'derivative', requires a *continuous* linear map as the analog of the Jacobi matrix. This step is therefore preliminary to *differentiability*. It has nothing to do with *continuous* differentiability, which is about continuity of the map $y \mapsto DI[y]$.

Note that the range $R := \{(t, y(t), y'(t)) : a \le t \le b\} \subset [a,b] \times G \times \mathbb{R}^n$ is compact for each fixed function y. For sufficiently small δ, a closed δ-neighbourhood $R_\delta := \{(t,u,v) : a \le t \le b\, ,\ |u - y'(t)| \le \delta\, ,\ |u - y(t)| \le \delta\}$ of this range will therefore still lie in $[a,b] \times G \times \mathbb{R}^n$, and R_δ is still a compact set. In particular, given any such function y, the set

$$\{(t, z(t), z'(t)) : a \le t \le b\, ,\ \|z - y\| \le \delta\}$$

lies in the compact set R_δ. Here $\|\cdot\|$ of course refers to the norm in $C^1([a,b] \to \mathbb{R}^n)$, namely $\|\varphi\| := \max\{|\varphi(t)|, |\varphi'(t)| : a \le t \le b\}$. The *continuous* functions L_y, $L_{y'}$, L_{yy}, $L_{yy'}$, $L_{y'y'}$ are *uniformly* continuous on the compact set R_δ.

The continuity of the linear map $DI[y]$ follows from

$$|DI[y]\varphi| \le |b-a| \Big(\max_{R_\delta} |L_y| + \max_{R_\delta} |L_{y'}| \Big) \|\varphi\|$$

(and the modulus of continuity is locally uniform).

The continuity (with locally uniform modulus of continuity) of the bilinear map $(\varphi,\psi) \mapsto D^2I[y](\varphi,\psi)$ that is defined by (2) follows analogously.

• At this stage, we have not shown yet that $DI[y]$ and $D^2I[y]$ indeed are the derivatives which by name they claim to be, even though it is clear from the directional derivative arguments that they are the only candidates for the job.

• We next show the continuity of the maps $y \mapsto DI[y]$ and $y \mapsto D^2I[y]$.

By uniform continuity of L_y and $L_{y'}$ on R_δ, we conclude that, given any ε, we can find $\eta < \delta$ such that $|L_y(t, y(t), y'(t)) - L_y(t, z(t), z'(t))| < \varepsilon$ (and a similar formula with $L_{y'}$) provided $\|y - z\| < \eta$. Then $|DI[y]\varphi - DI[z]\varphi| < 2\varepsilon|b-a|\|\varphi\|$, i.e. the norm of the linear maps satisfies $\|DI[y] - DI[z]\| < 2\varepsilon|b-a|$, provided $\|y - z\| < \eta < \delta$. An analogous argument can be made for D^2I.

• Now we show that $DI[y]$ is indeed the derivative of I at y, i.e., that

$$\Big| I[y+\varphi] - I[y] - DI[y]\varphi \Big| / \|\varphi\| \to 0 \quad \text{as} \quad \|\varphi\| \to 0$$

Indeed

$$\begin{aligned}
I[y+\varphi] &- I[y] - DI[y]\varphi = \\
&= \int_a^b \Big(L(t, y(t) + \varphi(t), y'(t) + \varphi'(t)) - L(t, y(t), y'(t)) \\
&\qquad - L_y(t, y(t), y'(t))\varphi(t) - L_{y'}(t, y(t), y'(t))\varphi'(t) \Big)\, dt \\
&= \int_a^b \int_0^1 \frac{d}{ds} \Big(L(t, y(t) + s\varphi(t), y'(t) + s\varphi'(t)) \\
&\qquad - sL_y(t, y(t), y'(t))\varphi(t) - sL_{y'}(t, y(t), y'(t))\varphi'(t) \Big)\, ds\, dt \\
&= \int_a^b \int_0^1 \Big\{ \Big(L_y(\ldots y + s\varphi \ldots) - L_y(\ldots y \ldots) \Big) \varphi(t) \\
&\qquad + \Big(L_{y'}(\ldots y + s\varphi \ldots) - L_{y'}(\ldots y \ldots) \Big) \varphi'(t) \Big\}\, ds\, dt
\end{aligned}$$

Now if $\|\varphi\|$ is sufficiently small, then all occurring arguments to L_y and $L_{y'}$ lie in the set R_δ, where the functions L_y and $L_{y'}$ are uniformly continuous. Then for every ε, there exists some η such that the differences in the big parentheses will be uniformly smaller than ε provided only $\|\varphi\| < \eta$. We conclude that $|I[y+\varphi] - I[y] - DI[y]\varphi| \le 2(b-a)\varepsilon\|\varphi\|$, which was to be shown.

By the same method it can be shown that $D^2I[y]$ is indeed the second derivative it claims to be. Namely, the estimate needs to prove that

$$\Big| DI[y+\psi]\varphi - DI[y]\varphi - D^2I[y](\psi, \varphi) \Big| / \|\varphi\|\, \|\psi\| \to 0$$

as $\|\psi\| \to 0$, and two invocations of the fundamental theorem of calculus and the uniform continuity of the 2nd derivatives of L on R_δ do the trick.

• So we have seen that $I \in C^2(X \to \mathbb{R})$, where X is itself a Banach space, namely $X = C^1([a,b] \to \mathbb{R}^n)$. This implies the same approximation of I by a 2nd degree Taylor 'polynomial' as in multivariable calculus, namely:

$$\Big| I[y+\varphi] - I[y] - DI[y]\varphi - \tfrac{1}{2} D^2I[y](\varphi, \varphi) \Big| / \|\varphi\|^2 \to 0 \quad \text{as} \quad \|\varphi\| \to 0\,. \tag{3}$$

Proof of (3):

$$I[y+\varphi] - I[y] = \int_0^1 \frac{d}{ds} I[y + s\varphi]\, ds = \int_0^1 DI[y + s\varphi]\varphi\, ds$$

Hence

$$
\begin{aligned}
I[y+\varphi]-I[y]-DI[y]\varphi &= \int_0^1 \Big(DI[y+s\varphi]-DI[y]\Big)\varphi\,ds \\
&= \int_0^1\int_0^s \frac{d}{d\sigma}DI[y+\sigma\varphi]\varphi\,d\sigma\,ds \\
&= \int_0^1\int_\sigma^1 D^2I[y+\sigma\varphi](\varphi,\varphi)\,ds\,d\sigma \\
&= \int_0^1 (1-\sigma)D^2I[y+\sigma\varphi](\varphi,\varphi)\,d\sigma
\end{aligned}
$$

and

$$
\begin{aligned}
&I[y+\varphi]-I[y]-DI[y]\varphi-\tfrac12 D^2I[y](\varphi,\varphi)= \\
&\qquad = \int_0^1 (1-\sigma)\Big(D^2I[y+\sigma\varphi](\varphi,\varphi)-D^2I[y](\varphi,\varphi)\Big)\,d\sigma
\end{aligned}
\tag{4}
$$

Using formula (2) for D^2I and the uniform continuity of L_{yy}, $L_{yy'}$, $L_{y'y'}$ on R_δ again, claim (3) is immediate.

We therefore have a rigorous analog of the MV-Calculus argument for relative minima. Having shown that, if $L\in C^2([a,b]\times G\times\mathbb{R}^n\to\mathbb{R})$, then $I\in C^2(X\to\mathbb{R})$ (where $X=C^1([a,b]\to\mathbb{R}^n)$), we obtain from (3) the 1st part of the following result immediately:

Theorem:
(a) If $L\in C^2([a,b]\times G\times\mathbb{R}^n\to\mathbb{R})$ and y_* satisfies the EL-eqn and if $D^2I[y_*](\varphi,\varphi)\ge c\|\varphi\|^2$ with $c>0$ for all $\varphi\in C_0^1([a,b]\to\mathbb{R}^n)$ in case of fixed boundary conditions, or else for all $\varphi\in C^1([a,b]\to\mathbb{R}^n)$ in case of free boundary, then y_* is a weak minimum of I given in (0).

(b) If however we only have the positive definiteness of $L_{y'y'}(t,y_*(t),y_*'(t))$ for all $t\in[a,b]$ (which implies the uniform positive definiteness in a neighbourhood R_δ), then y_* is locally *a weak minimum, i.e., sufficiently short segments of y_* are weak minima.*

As far as proving the 2nd part is concerned, we suppose that the support of φ has (small) length h, and recall that by continuity, there is an upper bound M for L_{yy} and $L_{yy'}$, and by uniform positive definiteness, there is a lower bound $m>0$ such that $\varphi'(t)^TL_{y'y'}(\ldots)\varphi'(t)\ge m|\varphi'(t)^2|$.

For such φ, we estimate taking a worst-case (negative) scenario for L_{yy} and $L_{yy'}$ relying only on $L_{y'y'}$ to give something positive in the end (more explanations after the formula):

$$
\begin{aligned}
D^2I[y_*](\varphi,\varphi) &\ge -M\int\varphi^2-2M\int|\varphi|\,|\varphi'|+m\int\varphi'^2 \\
&\ge -M\int\varphi^2-M\Big(\frac{2M}{m}\int\varphi^2+\frac{m}{2M}\int\varphi'^2\Big)+m\int\varphi'^2 \\
&\ge -\hat M\int\varphi^2+\frac{m}{2}\int\varphi'^2\ge\Big(\frac{m}{2}-\frac{\hat Mh^2}{4}\Big)\int\varphi'^2 \\
&\ge \Big(\frac{m}{2}-\frac{\hat Mh^2}{4}\Big)(1+h^2/4)^{-1}\|\varphi\|^2_{W^{1,2}}
\end{aligned}
\tag{5}
$$

In line 2, we have estimated $2|\varphi|\,\varphi'|$ under the integral using the famous and immensely useful inequality $2uv\le Au^2+v^2/A$. This inequality is true because $(u\sqrt{A}-v/\sqrt{A})^2\ge0$

In line 3 we have introduced the abbreviation $\hat M:=M+2M^2/m$ and then used the fact that for φ supported on an interval of length h, the estimate $\int\varphi(t)^2\,dt\le h^2/4\int\varphi'(t)^2\,dt$ holds (as mentioned in the notes on pg 18, and reproved the pedestrian way in a lemma below). This is why 'small support' always makes the tiniest platoon of $\int \varphi'^2$ win against the largest army of $\int \varphi^2$.

In line 4, we have introduced a NEW norm $\|\varphi\|^2_{W^{1,2}} := \int \varphi^2 + \int \varphi'^2$.

We need to (and can) take h so small that $\frac{m}{2} - \frac{\hat{M}h^2}{4} > 0$, and then (3), together with the EL eqn $DI[y_*]\varphi = 0$, *ought to imply* $I[y_* + \varphi] > I[y_*]$ (but doesn't, yet...).

A little nuisance arises here: Our positivity estimate for $D^2 I[y_*]$ hinges on the fact $\int L_{y'y'}\varphi'^2 \geq \int m\varphi'^2$, and we cannot do better than this, under the assumption $L_{y'y'} > 0$. All the other terms merely bite a bit off the constant m, so $D^2 I[y_*](\varphi,\varphi) \geq \tilde{m} \int \varphi'^2$. So we haven't squanderend anything essential in estimating the L_{yy} and $L_{yy'}$ terms and therefore cannot gain by improving there. We were thus forced to introduce the new norm $\|\varphi\|_{W^{1,2}}$, because an estimate in terms of $\|\varphi\|$ is just not available. But the error term in the Taylor polynomial estimate (3) is $\leq \varepsilon\|\varphi\|^2$, where $\|\varphi\|$ denotes our old C^1 norm $\max\{|\varphi|, |\varphi'|\}$. Now there is no way how we could prove $\tilde{m}\int \varphi'^2 - \varepsilon\|\varphi\|^2 > 0$, because this estimate is simply not true. Whatever small $\varepsilon/\tilde{m}$ is, we can still find φ such that $\tilde{m}\int \varphi'^2 - \varepsilon\|\varphi\|^2 < 0$. It seems we have labored in vain.

All the estimates up to (3) were ROUTINE estimates in the sense that you just write down what ought to be true in analogy to multivariable calculus, and it is indeed true by a very straightforward argument. (Fair enough if it doesn't seem routine or straightforward to you, if you are still struggling to absorb the formalism, but trust me it IS routine, once you master the formalism.) But then it turns out, as it often happens when lofty function spaces are used for down-and-dirty real problems, that the automatic machinery doesn't quite do the expected job, because the norm that the machinery likes is not the norm the question likes. So we have to return and do some adjustments manually. Luckily, in this problem this is easy: We need an improved variant of (3), namely:

$$\Big| I[y+\varphi] - I[y] - DI[y]\varphi - \tfrac{1}{2}D^2 I[y](\varphi,\varphi) \Big| \Big/ \|\varphi\|^2_{W^{1,2}} \to 0 \quad \text{as} \quad \|\varphi\| \to 0\,. \tag{6}$$

Note new norm $\nearrow$ $\qquad$ $\nwarrow$ still old C^1 norm!

This is much stronger than (3) because we divide by a potentially much smaller expression. It is also non-routine, because it blends different norms. Luckily the proof follows readily from (4): For all $\varepsilon > 0$ there exists $\eta < \delta$ such that

$$|I[y+\varphi] - I[y] - DI[y]\varphi - \frac{1}{2}D^2 I[y](\varphi,\varphi)| \leq \int_0^1 (1-\sigma)(\varepsilon\varphi^2 + 2\varepsilon|\varphi\varphi'| + \varepsilon\varphi'^2)d\sigma \leq \varepsilon\|\varphi\|^2_{W^{1,2}}$$

if $\|\varphi\| < \eta < \delta$. Now we can indeed use (5) with (6) instead of (3) and argue $I[y_* + \varphi] > I[y_*]$ for φ with small C^1 norm and small support. Hence y_* is locally a weak minimum, as claimed. ■

Let's review the small parameters we have used: Given y_* (solution of the EL eqn) we called R its 'range' in $[a,b] \times G \times \mathbb{R}^n$. First we chose δ so that still $R_\delta \subset [a,b] \times G \times \mathbb{R}^n$, and we have uniform continuity of L and all its derivatives up to order 2 on R_δ. We get a large constant M bounding the continuous functions L_{yy} and $L_{yy'}$ above, and a small constant m giving a lower bound for $L_{y'y'}$. From the onset, we commit at least to $\|\varphi\| < \delta$, but we will later require a smaller bound $\|\varphi\| < \eta$. From M and m we choose a small h, so that $m/2 - \hat{M}h^2/4 > 0$, and then we choose an ε so that $\tilde{m} := (m/2 - \hat{M}h^2/4)(1 + h^2/4)^{-1} > \varepsilon$. For this ε, we find $\eta < \delta$ so that $\|\varphi\| < \eta$ guarantees that the ratio in (6) is less than ε. This gives $I[y_* + \varphi] - I[y_*] \geq (\tilde{m} - \varepsilon)\|varphi\|^2_{W^{1,2}} > 0$.

Let's prove the lemma we used: Without loss of generality, we take the interval of length h to be $[0, h]$.

Lemma: If $\varphi(0) = 0 = \varphi(h)$, then $\int_0^h \varphi^2 \leq h^2/4 \int_0^h \varphi'^2$.

Proof: Since $\varphi(0) = 0$, we get by menas of the Cauchy Schwarz inequality (refer to it below)

$$|\varphi(t)| = |\int_0^t 1 \cdot \varphi'(s)\, ds| \leq (\int_0^t 1^2\, ds)^{1/2} (\int_0^t \varphi'^2(s)\, ds)^{1/2} \leq t^{1/2} (\int_0^h \varphi'^2(s)\, ds)^{1/2}$$

We use this for $0 \leq t \leq h/2$. For $h/2 \leq t \leq h$, we integrate from t to h and get

$$|\varphi(t)| \leq (h-t)^{1/2} (\int_0^h \varphi'^2(s)\, ds)^{1/2}$$

instead. We conclude

$$\varphi(t)^2 \leq \min\{t, h-t\} \times A \quad \text{where} \quad A := \int_0^h \varphi'^2(t)\, dt$$

Integrating this over $[0, h]$ we get

$$\int_0^h \varphi(t)^2\, dt \leq \frac{h^2}{4} \int_0^h \varphi'^2(t)\, dt$$

Note: A better constant is possible using a more sophisticated proof technology. Indeed, $\int_0^h \varphi(t)^2\, dt \leq \frac{h^2}{\pi^2} \int_0^h \varphi'^2(t)\, dt$. But the best possible constant was not our concern.

The Cauchy Schwarz inequality: For any inner product, the following inequality holds (and is called the Cauchy Schwarz inequality):

$$|\langle u, v\rangle| \leq \|u\|\, \|v\| \quad \text{with} \quad \|u\| := \sqrt{\langle u, u\rangle}$$

It has a very simple albeit tricky proof: Since $\langle tu+v, tu+v\rangle \geq 0$ for every number t, and since we can calculate $\langle tu+v, tu+v\rangle = t^2\langle u, u\rangle + 2t\langle u, v\rangle + \langle v, v\rangle$, we have a quadratic polynomial in t that is never negative. Therefore, if you try to find zeros t_1, t_2 with the quadratic formula, the term under the square root in the quadratic formula must be ≤ 0. This latter condition is the Cauchy Schwarz inequality. For continuous (or merely square integrable) functions defined on an interval (or any domain in $\mathbb{R}^n$), we have the inner product $\langle u, v\rangle := \int u(x)v(x)\, dx$, and then Cauchy Schwarz reads

$$\left| \int u(x)v(x)\, dx \right| \leq \left(\int u(x)^2\, dx \right)^{1/2} \left(\int v(x)^2\, dx \right)^{1/2}$$

Jochen Denzler, Feb 2006,
Lecture on Calculus of Variations