Formalising Mathematics (2022-2023)

Coursework 3

(01495449)

Imperial College London

April 2, 2023

1 Introduction

I will be proving the following theorem from my 4th year Optimisation module.

Theorem (Necessary Second Order Optimality Conditions). Let $f : U \to \mathbb{R}$ be a function defined on an open set $U \subseteq \mathbb{R}^n$ Suppose that f is twice continuously differentiable over U and that $\mathbf{x}^*$ is a stationary point. Then

- if $\mathbf{x}^*$ is a local minimum point, then $\nabla^2 f(\mathbf{x}^*) \succcurlyeq 0$.

Proof. There exists a ball $B(\mathbf{x}^*, r) \subseteq U$ for which $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x} \in B(\mathbf{x}^*, r)$. Next, let $\mathbf{d} \in \mathbb{R}^n$ be a nonzero vector. For any $0 < \alpha < \frac{r}{\|\mathbf{d}\|}$, we have $\mathbf{x}_\alpha^* \equiv \mathbf{x}^* + \alpha\mathbf{d} \in B(\mathbf{x}^*, r)$ and for any such α, $f(\mathbf{x}_\alpha^*) \geq f(\mathbf{x}^*)$. On the other hand, there exists a vector $\mathbf{z}_\alpha \in [\mathbf{x}^*, \mathbf{x}_\alpha^*]$ such that

$$f(\mathbf{x}_\alpha^*) - f(\mathbf{x}^*) = \frac{\alpha^2}{2}\mathbf{d}^\top \nabla^2 f(\mathbf{z}_\alpha)\,\mathbf{d}$$

This implies that for any $\alpha \in \left(0, \frac{r}{\|\mathbf{d}\|}\right)$ the inequality $\mathbf{d}^\top \nabla^2 f(\mathbf{z}_\alpha)\,\mathbf{d} \geq 0$ holds. Since $\mathbf{z}_\alpha \to \mathbf{x}^*$ as $\alpha \to 0^+$, we obtain that $\mathbf{d}^\top \nabla^2 f(\mathbf{x}^*)\,\mathbf{d} \geq 0$, which leads to $\nabla^2 f(\mathbf{x}^*) \succcurlyeq 0$.

From the onset, I tried using as many of the built-in lean definitions, and theorems and to make the proof as general as possible such that the complete proof is concise and maybe even suitable to be added to the lean library someday (where it is currently missing).

Rather than using $f : U \to R$, where $U \subseteq \mathbb{R}^n$. I defined $f : U \to R$, where $U \subseteq E$. Here E is a general normed space over the reals and is an additive commutative group.

The first challenge of this project was in setting up the problem. In Lean, the Hessian of a function f can be represented without using matrices by working with linear maps.

The advantages of the linear maps over matrices in my work are 3 fold:
1) I can work without choosing a basis.
2) In lean linear algebra proofs become simpler when working with linear maps instead of matrices. This abstraction can lead to cleaner and more intuitive proofs. For example, proving the associativity of the composition of linear maps is straightforward compared to the corresponding proof for matrix multiplication.
3) By working with linear maps and avoiding matrices, you can work with more general normed vector spaces, not just $\mathbb{R}^n$.

With this in mind, I define the functions f, f' and f'' as follows:

- $f : E \to \mathbb{R}$.
- $f : E \to (E \overset{\mathcal{L}(\mathbb{R})}{\to} \mathbb{R})$.
- $f'' : E \to (E \overset{\mathcal{L}(\mathbb{R})}{\to} (E \overset{\mathcal{L}(\mathbb{R})}{\to} \mathbb{R}))$.

Note that $\mathcal{L}(\mathbb{R})$ represents a linear map in reals. Also, the proof technically is for $f : U \to R$, where $U \subseteq E$, because I only use points in U. Hence it's equivalent to replacing the E's in the definition of the functions with a U.

2 Setting up the problem

```
/-- The necessary second-order optimality condition for a local minimum in a multivariable function:
Given a function `f` with continuous second-order derivatives in an open set `u`, if `x_star` is a
local minimum of `f` in `u`, then the Hessian of `f` at `x_star` is positive semi-definite. -/
theorem necess_2nd_order_optimality_condit
    (u : set E) (h_open : is_open u)  -- u is an open subset of E
    (h_d_nonzero : ∃ d ∈ u, d ≠ (0 : E) ) -- d is in u and d is a not zero vector
    (x_star : E) (h_x_star : x_star ∈ u) -- x_star is in u
    (f : E → ℝ) (f' : E → E →L[ℝ] ℝ) (f'' : E → (E →L[ℝ] (E →L[ℝ] ℝ)))
    (h_deriv_exists : ∀ y ∈ u, has_fderiv_at f (f' y) y) -- f' is the first derivative of f
    (h_second_deriv_exists : ∀ y ∈ u, has_fderiv_at f' (f'' y) y) -- f'' is the second derivative of f
    (h_stationary : f' x_star = 0) -- x_star is a stationary point of f
    (h_twice_continuous_diffable : cont_diff_on ℝ 2 f u): --  f is twice continuously differentiable on u
    is_local_min_on f u x_star → ∀ y, 0 ≤ (f'' x_star) y y := -- if local min point then f'' at point is positive semi-definite
begin
```

I will proceed from here dividing the proof into sub-lemmas that I prove. This helps structure the proof and also speeds up the computation. I left some 'have' statements in the main body of the proof whose proofs I thought were too short to make into sub-lemmas.

3 First milestone

We first prove that:
There exists a ball $B(\mathbf{x}^*, r) \subseteq U$ for which $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x} \in B(\mathbf{x}^*, r)$.

For this, I start by proving that there is a metric ball with an associated r around x_star such that within it x_star is a local minimum. I have a hypothesis that says that x_star is a local minimum. Since I used builtin 'is_local_min_on f u x_star'. This definition gave me a subset V that is in the neighbourhood of x_star and a superset of U called hv_subset such that inside $v \cap hv\_subset$, x_star is the minimum point. I proceed with:

1) Showing *hv_subset* is in a neighbourhood of x_star in with lemma hv_subset_nbhd_x_star
2) U is in the neighbourhood of x_star with u_in_nbhd_x_star
3) Since V, U and hv_subset are in the neighbour of x_star then so is their intersection.
4) Then I showed that there exists an open set J in $v \cap hv\_subset \cap u$ such that j contains x_star.
5) Since j is open and contains x_star, there is a metric ball in it with some radius r that is centred at x_star using metric.is_open_iff.

4 Second milestone

Next, I need to prove:
For any $0 < \alpha < \frac{r}{\|\mathbf{d}\|}$, we have $\mathbf{x}_\alpha^* \equiv \mathbf{x}^* + \alpha\mathbf{d} \in B(\mathbf{x}^*, r)$ and for any such α, $f(\mathbf{x}_\alpha^*) \geq f(\mathbf{x}^*)$.
This was quite easy and follows from the first milestone. I fix an arbitrary α such that $0 < \alpha < \frac{r}{\|\mathbf{d}\|}$. My proof follows the following steps:
1) Since $\mathbf{x}_\alpha^*$ is in metric ball centred at x_star with radius r, then the point is in set J. Since the metric ball is in j.
2) $\mathbf{x}_\alpha^*$ is $v \cap hv\_subset \cap u$ since j is in $v \cap hv\_subset \cap u$.
3) Since x_star is a minimum point in $v \cap hv\_subset \cap u$, we have that $f(\mathbf{x}_\alpha^*) \geq f(\mathbf{x}^*)$.

5 Intermediary step

I now prove that For any $0 < \alpha < \frac{r}{\|\mathbf{d}\|}$, we have $[\mathbf{x}^*, \mathbf{x}_\alpha^*]$ is in U. This will come helpful later as we already have a hypothesis that f is twice continuously differentiable in U. I proceed in the following manner:

1) First I show that every point in $[\mathbf{x}^*, \mathbf{x}_\alpha^*]$ is in the metric ball centered at x_star with radius r.
2) Showed that the metric ball centred at x_star with radius r is a subset of U.
3) Then by transitivity every point in $[\mathbf{x}^*, \mathbf{x}_\alpha^*]$ is in U.

6 More detail on the next steps of the proof

Next, I want to show that: There exists a vector $\mathbf{z}_\alpha \in [\mathbf{x}^*, \mathbf{x}_\alpha^*]$ such that

$$f(\mathbf{x}_\alpha^*) - f(\mathbf{x}^*) = \frac{\alpha^2}{2}\mathbf{d}^\top \nabla^2 f(\mathbf{z}_\alpha)\,\mathbf{d}$$

There is a lot more going on here than the simple statement. Mathlib library has the Taylor theorem with the Lagrange form of the remainder for functions from $\mathbb{R}$ to $\mathbb{R}$. I utilise this to proceed with the proof.

Proof sketch:

1) Set the function $g : \mathbb{R} \to \mathbb{R}$ to be $g(\alpha) \equiv f(\mathbf{x}^* + \alpha\mathbf{d})$.
2) Then by Taylor theorem with Lagrange form of the remainder, $\exists$ c $\in [0, \alpha]$ such that $g(\alpha) = g(0) + \alpha g'(0) + \frac{\alpha^2}{2} g''(c)$
3) We have that $g(\alpha) \equiv f(\mathbf{x}^* + \alpha\mathbf{d})$, $g(0) \equiv f(\mathbf{x}^*)$, $g'(0) \equiv \nabla f(\mathbf{x}^*) \cdot d = 0$. Since $\mathbf{x}^*$ is a stationary point of f is one of the hypotheses I start the proof detailed in the Introduction section with. $g''(c) = d\frac{\nabla f(\mathbf{x}^* + cd) \cdot d}{d\alpha}$. Let $\mathbf{z}_\alpha = \mathbf{x}^* + cd$. Hence in matrix form $g''(c) = \mathbf{d}^\top \nabla^2 f(\mathbf{z}_\alpha)\,\mathbf{d}$.
4) By substitution in the equation in step 2. We get:

$$f\left(\mathbf{x}_{\alpha}^{*}\right)-f\left(\mathbf{x}^{*}\right)=\frac{\alpha^{2}}{2}\mathbf{d}^{\top}\nabla^{2}f\left(\mathbf{z}_{\alpha}\right)\mathbf{d} \tag{1}$$

7 Third milestone

To prove the step in the previous section. I had to set $g(x) \equiv f(\mathbf{x}^* + x\mathbf{d})$, where x is in [0, α] and then use the lean library taylor_theorem_with_lagrange_form_remainder on the function g.

To use this theorem I first had to show that:
1) g is once continuously differentiable on [0, α].
2) The first derivative of g on [0, α] is differentiable on (0, α)

The proof for 1) is straightforward. Since we have a hypothesis that f is twice continuously differentiable. I showed that $f(\mathbf{x}^* + x\mathbf{d})$ is twice continuously differentiable for x in [0, α]. This was implemented by decomposition of into two functions 'f' on U and '$\mathbf{x}^*$+$x\mathbf{d}$' on x in [0, α]. I proved that both are twice continuously differentiable and then used cont_diff_on.comp. cont_diff_on.comp essentially states: If two functions f and g are smooth up to a certain level on their sets s and t, then their composition ($g \circ f$) is also smooth up to that level on set s, provided s is within the preimage of t under f. Now that I have proven $f(\mathbf{x}^* + x\mathbf{d})$ is twice continuously differentiable on x in [0, α], I then use cont_diff_on.one_of_succ, to prove that it is once continuously differentiable. Proving our goal of 1).

Achieving the proof of 2 was more involved. I first showed that g is differentiable on [0, α]. In 1) I have already shown that g is continuously differentiable hence I can use cont_diff_on.differentiable_on, to show that g is differentiable on [0, α]. cont_diff_on.differentiable_on states that If a function is continuously differentiable up to a level n on a set, it is also differentiable on that set. Next, I aimed to show that the first derivative of g is $f'(\mathbf{x}^* + x\mathbf{d})d$, this clearly requires the use of the chain rule. I used 'fderiv_within.comp' which is a version of the chain rule in the context of differentiable functions within sets. To use I show the following:

1) Show differential belonging to the set [0, α] is unique. This is done by showing that the tangent cone to [0, α] for all points inside spans a dense subset of the entire space.
2) The function k(x)= x_star + x • d maps the closed interval [0, α] to the line segment between the points $\mathbf{x}^*$ and $(\mathbf{x}^* + \alpha\mathbf{d})$.
3) Show k(x)= x_star + x • d is differentiable on the closed interval [0, α].
4) f(x) is differentiable at all points within the line segment connecting $\mathbf{x}^*$ and $(\mathbf{x}^* + \alpha\mathbf{d})$.

In the proof for 1) I showed that [0, α] is a non-empty convex set, hence using unique_diff_on_convex, [0, α] is a unique differentiability set.

In the proof for 2) I re-wrote the goal using segment_eq_image' and then proved that the line segment between $\mathbf{x}^*$ and $(\mathbf{x}^* + \alpha\mathbf{d})$, can be represented as the set of all points obtained by adding a scaled difference of $\mathbf{x}^*$ and $(\mathbf{x}^* + \alpha\mathbf{d})$ to $\mathbf{x}^*$, where the scaling factor, x', varies between 0 and 1.

For the proof of 3), in differentiating k(x)= x_star + x • d, I can remove x_star since it is a constant and using differentiable_within_at_const_add_iff. This just states that if a function is differentiable, then the function with its constant part removed must also be differentiable. I then used differentiable_within_at.smul_const to close the goal, this showed that since the scalar identity function is differentiable, we have that the identity function multiplied by a constant vector is differentiable. Hence p(x)=x • d is differentiable.

Unfortunately, I ran out of time to complete the proof for 4). This proof would have been fairly straightforward as we know that f' is the derivative of f on the set U and since the line segment is in U (already proved). f must be differentiable on the line segment.

8 Closing comments

I enjoyed this proof very much. Especially since I have almost completely focused on applied modules since 3rd year, this module got me back into thinking about mathematics more rigorously, which I have found to be stimulating and rewarding. I am looking forward to sharing my insights at the lean conference next month.

9 Credits

I want to thank Eric Weiser for suggesting that I use bilinear maps rather than matrices. This took some time for me to understand how the mappings worked for f'' and he was very helpful there. Later, Prof. Buzzard sent me a message detailing this, which helped cement my understanding. Prof. Buzzard also let me know that there was a theorem in the lean library, which says that for a neighbourhood of a point in a metric space, there exists an epsilon.

I also want to thank Sebastien Gouezel for his comment about using normed space rather than $\mathbb{R}^n$, making for stronger proof. He also introduced me to cont_diff_on.comp.

I am impressed by the community around Lean. I think it's amazing how quickly and thoroughly people reply on lean and that it's quite a global community.