5 Extremal Values **2021-22**

Extremal Values & Lagrange Multipliers

Stress we are only looking at **scalar-valued** functions.

Definition 1 *Suppose that $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ is a function. Given a point $\mathbf{a} \in U$,*

- *if $f(\mathbf{x}) \leq f(\mathbf{a})$ for all $\mathbf{x} \in U$ we say that f has a* ***global*** *or* ***absolute maximum*** *at $\mathbf{a}$;*
- *if $f(\mathbf{x}) \geq f(\mathbf{a})$ for all $\mathbf{x} \in U$ we say that f has a* ***global*** *or* ***absolute minimum*** *at $\mathbf{a}$;*
- *if there exists an open set $V \subseteq \mathbb{R}^n$ with $\mathbf{a} \in V$ such that $f(\mathbf{x}) \leq f(\mathbf{a})$ for all $\mathbf{x} \in V \cap U$ then we say that f has a* ***local maximum*** *at $\mathbf{a}$;* ***strict*** *if $f(\mathbf{x}) < f(\mathbf{a})$ for all $\mathbf{x} \in V \cap U \setminus \{\mathbf{a}\}$;*
- *if there exists an open set $V \subseteq \mathbb{R}^n$ with $\mathbf{a} \in V$ such that $f(\mathbf{x}) \geq f(\mathbf{a})$ for all $\mathbf{x} \in V \cap U$ then we say that f has a* ***local minimum*** *at $\mathbf{a}$;* ***strict*** *if $f(\mathbf{x}) > f(\mathbf{a})$ for all $\mathbf{x} \in V \cap U \setminus \{\mathbf{a}\}$;*
- *An* ***extremum*** *is a point which is either a maximum or minimum, so we can refer to a (strict) absolute extremum or (strict) local extremum.*

Recall the **Chain Rule** in the special situation $\mathbb{R} \xrightarrow{\mathbf{g}} \mathbb{R}^n \xrightarrow{f} \mathbb{R}$ which gives, under appropriate conditions on the functions,

$$(f \circ \mathbf{g})'(t) = df_{\mathbf{g}(t)}(\mathbf{g}'(t)) = \nabla f(\mathbf{g}(t)) \bullet \mathbf{g}'(t) = Jf(\mathbf{g}(t))\, \mathbf{g}'(t)\,.$$

The following is the generalisation of the result that the extrema for real-valued functions of one variable occur at the turning points.

Proposition 2 *Suppose that $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ is a Fréchet differentiable function on the open set U which has a local extremum at $\mathbf{a} \in U$. Then the derivative $df_{\mathbf{a}} = 0$ (i.e. it is the zero linear map from $\mathbb{R}^n$ to $\mathbb{R}$) or, equivalently, $\nabla f(\mathbf{a}) = \mathbf{0}$ or $Jf(\mathbf{a}) = 0$.*

Proof Let $\mathbf{v} \in \mathbb{R}^n$ a non-zero vector be given. Since $\mathbf{a} \in U$ and U is an open set there exists $\eta > 0$ such that for $|t| < \eta$ we have $\mathbf{a} + t\mathbf{v} \in U$. For such t define $\phi(t) = f(\mathbf{a} + t\mathbf{v})$.

Since f has a local extremum at $\mathbf{a}$, then f has a local extremum at $\mathbf{a}$ as you approach $\mathbf{a}$ in the direction of $\mathbf{v}$, that is $\phi(t)$ has a local extremum at $t = 0$. So, by second year analysis,

$$0 = \phi'(0) = \lim_{t \to 0} \frac{\phi(t) - \phi(0)}{t} = \lim_{t \to 0} \frac{f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a})}{t} = d_{\mathbf{v}} f(\mathbf{a}),$$

where, unlike earlier in the course, we are not restricting to unit $\mathbf{v}$ in $d_{\mathbf{v}} f$. Yet f is Fréchet differentiable at $\mathbf{a}$ so we have $df_{\mathbf{a}}(\mathbf{v}) = d_{\mathbf{v}} f(\mathbf{a})$. Thus $df_{\mathbf{a}}(\mathbf{v}) = 0$ for all non-zero $\mathbf{v}$. Hence $df_{\mathbf{a}} = 0$ ■

Definition 3 *Given a Fréchet differentiable function $f : U \subset \mathbb{R}^n \to \mathbb{R}$ then $\mathbf{a} \in U$ is a **critical point** if $df_{\mathbf{a}} = 0$ or, equivalently, $\nabla f(\mathbf{a}) = \mathbf{0}$ or $Jf(\mathbf{a}) = \mathbf{0}$.*

So, if extremums exist, they are a subset of the critical points, thus to find the extremums you first find the critical points.

Be aware that Proposition 2 says that IF an extremum exists then it is a critical point. It does **not** say that extremums exist. For a proof of existence you will need to apply other results. We have seen such a result in Second Year Analysis: a continuous function on a closed and bounded subset of $\mathbb{R}$ is bounded and attains its bounds.

A subset A of $\mathbb{R}^n$ is **closed** if its complement is open. A set B is **bounded** if there exists $M \in \mathbb{R}$ such that $|\mathbf{x}| \leq M$ for all $\mathbf{x} \in B$. Sets which are both closed and bounded are said to be **compact**. The fundamental existence theorem for maxima and minima is

Theorem 4 *A continuous real-valued function defined on a compact subset of $\mathbb{R}^n$ attains it's global maximum and attains it's global minimum.*

Proof Not given. ■

Example 5 *Find the critical points of* $f(\mathbf{x}) = x^2y^2+z^2+2x-4y+z$, $\mathbf{x} \in \mathbb{R}^3$.

Solution (*Problem Class*) The gradient vector is

$$\nabla f(\mathbf{x}) = \left(2xy^2 + 2,\, 2yx^2 - 4,\, 2z + 1\right)^T .$$

Then $\nabla f(\mathbf{x}) = \mathbf{0}$ iff

$$\begin{aligned} 2xy^2 + 2 &= 0, \\ 2yx^2 - 4 &= 0, \\ 2z + 1 &= 0. \end{aligned}$$

From the last equation $z = -1/2$. From the first two we see that x and y are neither 0. Multiply first equation by 2 and add the second to see that $0 = 4xy^2 + 2yx^2 = 2xy\,(2y + x)$. Since neither x nor y is 0 we must have $x = -2y$. In the second equation this gives $2y^3 - 1 = 0$. Therefore the only critical value is

$$\mathbf{a} = \left(-2^{2/3},\ 2^{-1/3},\ -2^{-1}\right)^T .$$

■

Example 6 *Find the critical points of* $f(\mathbf{x}) = x^2 - y^2 + z^2$. *Are they extremal?*

Solution $\nabla f(\mathbf{x}) = \mathbf{0}$ iff $2x = 0$, $-2y = 0$ and $2z = 0$ so the only critical point is $\mathbf{x} = (0,0,0)^T$.

The value $f(\mathbf{0}) = 0$ is not a local maximum because, in any open set $\mathbf{0} \in U \subseteq \mathbb{R}^3$ we can find $\mathbf{t} = (t,0,0)^T \in U$ with $t \neq 0$, in which case $f(\mathbf{t}) = t^2 > 0 = f(\mathbf{0})$.

Also $f(\mathbf{0}) = 0$ is not a local minimum because, in any open set $\mathbf{0} \in U \subseteq \mathbb{R}^3$ we can find $\mathbf{s} = (0,s,0)^T \in U$ with $s \neq 0$, in which case $f(\mathbf{s}) = -s^2 < 0 = f(\mathbf{0})$. ■

There can be no extremal values since there are no other critical points to examine. In particular, there is no global maximum or global minimum. We could have seen this directly; by choosing $\mathbf{x} = (x,0,0)^T$ we can make $f(\mathbf{x}) = x^2$ as large and positive as we like while, with $\mathbf{x} = (0,y,0)^T$, we can make $f(\mathbf{x}) = -y^2$ as large and negative as we like.

It is important to stress that if $\mathbf{a}$ is a local extremum then $\mathbf{a}$ is a critical point, while $\mathbf{a}$ critical does not imply $\mathbf{a}$ is a local extremum.

$$\boxed{\begin{array}{l} \mathbf{a} \text{ a local extremum } \implies \mathbf{a} \text{ critical,} \\ \mathbf{a} \text{ critical } \quad \not\Longrightarrow \mathbf{a} \text{ a local extremum} \end{array}}$$

Nonetheless one can find local extremum by searching amongst the critical points.

Problem Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ are Fréchet differentiable. How to find $\max\{f(\mathbf{x}) : \mathbf{g}(\mathbf{x}) = 0\}$ and $\min\{f(\mathbf{x}) : \mathbf{g}(\mathbf{x}) = 0\}$ or what we might group together as $\operatorname{ext}\{f(\mathbf{x}) : \mathbf{g}(\mathbf{x}) = 0\}$?

From Chapter 4 we know that $\{\mathbf{x} : \mathbf{g}(\mathbf{x}) = 0, J\mathbf{g}(\mathbf{x}) \text{ is of full-rank}\}$ is an example of a surface. So we could ask the more general question of how to find $\operatorname{ext}\{f(\mathbf{x}) : \mathbf{x} \in S\}$ for a surface $S \subseteq \mathbb{R}^n$. We call these the **extreme values of f subject to the constraint that $\mathbf{x}$ lies on the surface S.** The solution is rather nice.

Assume $\mathbf{a} \in S$. However, unless S is already given as a graph, we know by either the Implicit Function Theorem or Inverse Function Theorem, that there exists an open set $W \subseteq \mathbb{R}^n$ containing $\mathbf{a}$, and, for some $r \geq 1$, a subset $V \subseteq \mathbb{R}^r$ and a $\mathcal{C}^1$-function $\phi : V \to \mathbb{R}^{n-r}$ such that

$$S \cap W = \{\mathbf{F}(\mathbf{u}) : \mathbf{u} \in V\},$$

where the function $\mathbf{F} : V \to \mathbb{R}^n$ given by

$$\mathbf{F}(\mathbf{u}) = \begin{pmatrix} \mathbf{u} \\ \phi(\mathbf{u}) \end{pmatrix}.$$

for $\mathbf{u} \in V$. Since $\mathbf{a} \in S \cap W$ there exists $\mathbf{b} \in V$ such that $\mathbf{a} = \mathbf{F}(\mathbf{b})$.

Next,

$$\{f(\mathbf{x}) : \mathbf{x} \in S \cap W\} = \{f(\mathbf{F}(\mathbf{u})) : \mathbf{u} \in V\} = \{G(\mathbf{u}) : \mathbf{u} \in V\}, \qquad (1)$$

where $G : V \to \mathbb{R}$ is the composite function $G = f \circ \mathbf{F}$.

Thus

$$\operatorname{ext}\{f(\mathbf{x}) : \mathbf{x} \in S \cap W\} = \operatorname{ext}\{G(\mathbf{u}) : \mathbf{u} \in V\}.$$

We find the extremal points by searching through the critical points of G, i.e. finding $\mathbf{b} \in V : JG(\mathbf{b}) = \mathbf{0}$.

The Fundamental Observation What has happened in (1) is that in the left hand set the $\mathbf{x} \in \mathbb{R}^n$ are restricted to lie in a subset $S \subseteq \mathbb{R}^n$, with dimension *less* than n. In the right hand side set of (1) the $\mathbf{u} \in \mathbb{R}^r$ are restricted to a subset $V \subseteq \mathbb{R}^r$ of the *same* dimension as $\mathbb{R}^r$.

Definition 7 *We say that* $\mathbf{a}$ *is a* ***critical point for*** f ***restricted to*** S *if* $dG_{\mathbf{b}} = 0$, *or* $JG(\mathbf{b}) = 0$, *for the* G *and* $\mathbf{b}$ *arising as above.*

The 'nice' result referred to above is

Theorem 8 *Suppose that* $S \subseteq \mathbb{R}^n$ *is a surface. Let the scalar-valued* $f : \mathbb{R}^n \to \mathbb{R}$ *be a Fréchet differentiable function. If* $\mathbf{a} \in S$ *is a critical point of* f *restricted to* S *then*

$$\nabla f(\mathbf{a}) \in T_{\mathbf{a}}(S)^{\perp}.$$

That is, at a critical point $\mathbf{a}$ of f restricted to S the gradient vector of f is perpendicular to the Tangent Space to S at $\mathbf{a}$.

Proof Assume that $\mathbf{a} \in S$ is a critical point of f restricted to S. As above, there exists $W \subseteq \mathbb{R}^n$ with $\mathbf{a} \in W$ along with a C^1-function $\mathbf{F} : V \subseteq \mathbb{R}^r \to \mathbb{R}^n$ such that

$$S \cap W = \{\mathbf{F}(\mathbf{u}) : \mathbf{u} \in V\}.$$

Then

$$\{f(\mathbf{x}) : \mathbf{x} \in S \cap W\} = \{G(\mathbf{u}) : \mathbf{u} \in V\}.$$

where $G = f \circ \mathbf{F}$.

The definition of $\mathbf{a}$ is a critical point of f restricted to S is that $dG_{\mathbf{b}} = 0$ where $\mathbf{b} \in V$ with $\mathbf{F}(\mathbf{b}) = \mathbf{a}$. Apply the Chain Rule in the Jacobian matrix form. Since $dG_{\mathbf{b}} = 0$ the Jacobian matrix satisfies $JG(\mathbf{b}) = 0$ which implies

$$\begin{aligned}
\mathbf{0} &= JG(\mathbf{b}) \\
&= Jf(\mathbf{F}(\mathbf{b}))\, J\mathbf{F}(\mathbf{b}) \quad \text{by the Chain Rule} \\
&= Jf(\mathbf{a})\, J\mathbf{F}(\mathbf{b}) \quad \text{since } \mathbf{a} = \mathbf{F}(\mathbf{b}).
\end{aligned}$$

For a scalar-valued function $Jf(\mathbf{a}) = \nabla f(\mathbf{a})^T$, so

$$\nabla f(\mathbf{a})^T \, J\mathbf{F}(\mathbf{b}) = 0. \tag{2}$$

In particular, $\nabla f(\mathbf{a})$ is orthogonal to each column vector in $J\mathbf{F}(\mathbf{b})$. Yet, by a result in the chapter on surface the column vectors of $J\mathbf{F}(\mathbf{b})$ form a basis to $T_{\mathbf{a}}(S)$. Hence (2) implies $\nabla f(\mathbf{a})$ is orthogonal to all vectors in a basis of $T_{\mathbf{a}}(S)$ and thus $\nabla f(\mathbf{a}) \in T_{\mathbf{a}}(S)^{\perp}$. ■

Aside The result $\nabla f(\mathbf{a}) \in T_{\mathbf{a}}(S)^{\perp}$ means that $\nabla f(\mathbf{a})$ is normal to the Tangent Space and thus the surface at $\mathbf{a}$. If $\nabla f(\mathbf{a})$ were **not** normal to the surface then we could find a vector in the Tangent plane, $\mathbf{v}$ say, with $\nabla f(\mathbf{a}) \bullet \mathbf{v} > 0$ (actually $\neq 0$, but if < 0 replace $\mathbf{v}$ by $-\mathbf{v}$). Yet $\mathbf{v}$ in the tangent plane means there exists a curve in the surface $\boldsymbol{\gamma} : (-\eta, \eta) \subseteq \mathbb{R} \to S$ such that $\boldsymbol{\gamma}(0) = \mathbf{a}$ and $\boldsymbol{\gamma}'(0) = \mathbf{v}$. Thus $\nabla f(\mathbf{a}) \cdot \boldsymbol{\gamma}'(0) > 0$, and so (by the Chain Rule in reverse)

$$\left. \frac{d}{dt} f(\boldsymbol{\gamma}(t)) \right|_{t=0} > 0.$$

If we have continuity of derivative then $df(\boldsymbol{\gamma}(t))/dt > 0$ for all t sufficiently close to 0. Thus the value of f increases as you travel away from $\mathbf{a}$ on $\boldsymbol{\gamma}$ with increasing $t > 0$ and decreases as you travel away with increasing magnitude $t < 0$. This would contradict f having an extremum at $\mathbf{a}$.

End of Aside

Lagrange Multipliers In the majority of examples the surface is given as a level set, $S = \mathbf{g}^{-1}(\mathbf{0})$ for some $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$. As we saw in Chapter 3, at $\mathbf{a} \in S$ the space orthogonal to the Tangent Space of a level set, i.e. $T_{\mathbf{a}}(S)^{\perp}$, is spanned by the rows of $J\mathbf{g}(\mathbf{a})$, which are the gradient vectors $\nabla g^i(\mathbf{a})$, $1 \leq i \leq m$. Thus Theorem 8 says that for such S if $\mathbf{a} \in S$ is a critical point of f restricted to S then $\nabla f(\mathbf{a}) \in \text{Span}\{\nabla g^i(\mathbf{a}), 1 \leq i \leq m.\}$. Equivalently, there exist constants, the **Lagrange multipliers**, $\lambda_i, 1 \leq i \leq m$ such that

$$\nabla f(\mathbf{a}) = \lambda_1 \nabla g^1(\mathbf{a}) + \lambda_2 \nabla g^2(\mathbf{a}) + ... + \lambda_m \nabla g^m(\mathbf{a}).$$

The λ_i are useful but ultimately irrelevant constants.

Example 9 *Find the maximum and minimum values of* $2x + 3y - z$ *subject to* $x^2 + y^2/2 = 1$ *and* $x + z = 1$.

Solution Define $\mathbf{g} : \mathbb{R}^3 \to \mathbb{R}^2$ by

$$\mathbf{g}(\mathbf{x}) = \begin{pmatrix} x^2 + y^2/2 - 1 \\ x + z - 1 \end{pmatrix}.$$

for $\mathbf{x} \in \mathbb{R}^3$.

To apply Lagrange's method we require $\{\mathbf{x} : \mathbf{g}(\mathbf{x}) = \mathbf{0}\}$ to be a surface, i.e. $J\mathbf{g}(\mathbf{x})$ of full-rank. The Jacobian matrix of the level set is

$$J\mathbf{g}(\mathbf{x}) = \begin{pmatrix} 2x & y & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

The only time this is not of full-rank is when the top row is $\mathbf{0}$, i.e. $x = y = 0$. But if $x = y = 0$ then $\mathbf{g}(\mathbf{x}) \neq 0$. Hence $\{\mathbf{x} : \mathbf{g}(\mathbf{x}) = \mathbf{0}\}$ is a surface and we can apply the method of Lagrange multipliers and try to solve

$$\nabla f(\mathbf{x}) = \lambda \nabla g^1(\mathbf{x}) + \mu \nabla g^2(\mathbf{x})$$

for $\mathbf{x} : \mathbf{g}(\mathbf{x}) = 0$ and $\lambda, \mu \in \mathbb{R}$.

The first condition is

$$\begin{pmatrix} 2 \\ 3 \\ -1 \end{pmatrix} = \lambda \begin{pmatrix} 2x \\ y \\ 0 \end{pmatrix} + \mu \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \text{i.e.} \quad \begin{aligned} 2 &= 2\lambda x + \mu, \\ 3 &= \lambda y, \\ -1 &= \mu. \end{aligned}$$

These reduce to two: $\lambda x = 3/2$ and $\lambda y = 3$. These can be substituted into $g^1(\mathbf{x}) = 0$ written as $2 = 2x^2 + y^2$, to get

$$2\lambda^2 = 2(\lambda x)^2 + (\lambda y)^2 = 2\left(\frac{3}{2}\right)^2 + (3)^2 = 3^2\frac{3}{2}.$$

Thus $\lambda = \pm 3\sqrt{3}/2$. Hence $x = \pm 1/\sqrt{3}$, $y = \pm 2/\sqrt{3}$ and, from $g^2(\mathbf{x}) = 0$, $z = 1 \mp 1/\sqrt{3}$.

We have, therefore, two critical points of f restricted to the surface are

$$\mathbf{a}_1 = \left(\frac{1}{\sqrt{3}}, \frac{2}{\sqrt{3}}, 1 - \frac{1}{\sqrt{3}}\right)^T,$$

$$\mathbf{a}_2 = \left(-\frac{1}{\sqrt{3}}, -\frac{2}{\sqrt{3}}, 1 + \frac{1}{\sqrt{3}}\right)^T.$$

We need show these correspond to extremal values of f. First note that the set of $\mathbf{x} : \mathbf{g}(\mathbf{x}) = \mathbf{0}$ is **bounded**. From $x^2 + y^2/2 = 1$ we have $|x| \leq 1$, $|y| \leq \sqrt{2}$. Then from $x + z = 1$ we have $|z| = |1 - x| \leq 1 + |x| \leq 2$.

Next note that the set of $\mathbf{x} : \mathbf{g}\left(\mathbf{x}\right) = 0$ is **closed** since its compliment is open; if $\mathbf{g}(\mathbf{x}_0) \neq \mathbf{0}$ then $\mathbf{g}(\mathbf{x}) \neq \mathbf{0}$ for all $\mathbf{x}$ sufficiently close to $\mathbf{x}_0$, i.e. for $\mathbf{x} \in B_\delta\left(\mathbf{x}_0\right)$ for sufficiently small $\delta > 0$.

Hence, since $f(\mathbf{x}) = 2x + 3y - z$ is continuous it will have maximum and minimum values on this compact subset of $\mathbb{R}^3$. These will each correspond to a critical point and since we have only found two critical points they must correspond to these extrema.

All that remains are the calculations $f(\mathbf{a}_1) = 3\sqrt{3} - 1$, the maximum, and $f(\mathbf{a}_1) = -3\sqrt{3} - 1$, the minimum value. ■

Remark The method can be used to find extreme values on more general closed sets. For example, suppose we wish to find the maximum and minimum values of a Fréchet differentiable function $f : \mathbb{R}^3 \to \mathbb{R}$ on the **closed** *ball* $\{\mathbf{x} \in \mathbb{R}^3 : |\mathbf{x}| \leq 1\}$. We might use Lagrange multipliers to

- identify the maximum and minimum values **on** the sphere $\{\mathbf{x} \in \mathbb{R}^3 : |\mathbf{x}| = 1\}$ and then
- search amongst the critical values **in** the open ball $\{\mathbf{x} \in \mathbb{R}^3 : |\mathbf{x}| < 1\}$, using $df_{\mathbf{x}} = 0$, to find the maximum and minimum values in the open ball.

Example 10 *Find the maximum and minimum values of* $f(\mathbf{x}) = 4x^2 + 10y^2$ *on the closed disc* $\{\mathbf{x} \in \mathbb{R}^2 : x^2 + y^2 \leq 4\}$.

Solution (*Problem Class*) We first look for critical points **in the interior of the set**, i.e. $\mathbf{x} : x^2 + y^2 < 4$ and $\nabla f(\mathbf{x}) = \mathbf{0}$. But $\nabla f(\mathbf{x}) = (8x, 20y^2)^T$ equals $\mathbf{0}$ only when $\mathbf{x} = \mathbf{0}$. So $\mathbf{0}$ is the only critical point in the interior.

We now look for critical points **restricted to the boundary**. Let $g(\mathbf{x}) = x^2 + y^2 - 4$, we want the critical points of f restricted to $g(\mathbf{x}) = 0$. To apply Lagrange's method the Jacobian matrix of g has to be of full-rank. In this case, $Jg(\mathbf{x}) = (2x, 2y)$. The only way this cannot be of full-rank is if it is $\mathbf{0}$, i.e. $x = y = 0$. But $g(\mathbf{0}) = -4 \neq 0$, and so $Jg(\mathbf{x})$ is of full-rank for all $\mathbf{x} : g(\mathbf{x}) = 4$.

We need to solve $\nabla f(\mathbf{x}) = \lambda \nabla g(\mathbf{x})$ with $g(\mathbf{x}) = 0$. The condition on the gradient vectors is

$$\begin{pmatrix} 8x \\ 20y \end{pmatrix} = \lambda \begin{pmatrix} 2x \\ 2y \end{pmatrix},$$

so $4x = \lambda x$ and $10y = \lambda y$. That is, $(4 - \lambda)x = 0$ and $(10 - \lambda)y = 0$.

Then $(4 - \lambda)x = 0$ implies either $4 - \lambda = 0$ or $x = 0$. If $4 - \lambda = 0$ then $10 - \lambda \neq 0$ so the second equation implies $y = 0$. Then $g(\mathbf{x}) = 0$ implies $x = \pm 2$.

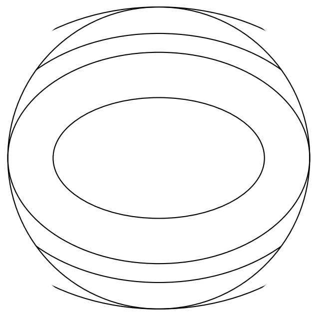
If $x = 0$ then $g(\mathbf{x}) = 0$ implies $y = \pm 2$ (and $(10 - \lambda)y = 0$ implies $\lambda = 10$, though this is not required).

Hence the five critical points are $(0,0)^T$, $(\pm 2, 0)^T$ and $(0, \pm 2)^T$.

The function f is continuous and the disc closed and bounded so f will attain it's extremal values on the disc. We search through these points to find the extremal values:

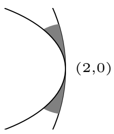
$$\begin{aligned} &f(\mathbf{0}) = 0, \quad \text{minimum value}, \\ &f\left((\pm 2, 0)^T\right) = 16, \\ &f\left((0, \pm 2)^T\right) = 40 \quad \text{maximum value}. \end{aligned}$$

Note this result should be no surprise. Consider the diagram of the circle $x^2 + y^2 = 4$ and the ellipse $4x^2 + 10y^2 = c$ drawn with a variety of values for c.

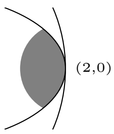


The inner ellipse has $c = 4$, say, when the ellipse and circle do not intersect. When $c = 16$ they intersect at $(\pm 2, 0)$. The forth ellipse has $c = 40$ when the points of intersection are $(0, \pm 2)$. So the minimum value of $4x^2 + 10y^2$ taken by points **on the circle** is 16, the maximum 40.

But further, we can see that 16, the value of $4x^2+10y^2$ at $(\pm 2, 0)$ is neither a local minimum or maximum **in the disc**. Looking closely at $(2, 0)$, if we move away from it within the gray region the values of $4x^2 + 10y^2$ increase.



Whereas if we move away from $(2, 0)$ within this gray region the values decrease.



Appendix for Section 5

Surfaces in $\mathbb{R}^3$

Consider the surface in $\mathbb{R}^3$ given as a graph of $f : \mathbb{R}^2 \to \mathbb{R}$, so $S = \{\mathbf{F}(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^2\}$ with

$$\mathbf{F}(\mathbf{x}) = \begin{pmatrix} \mathbf{x} \\ f(\mathbf{x}) \end{pmatrix}.$$

In an earlier Appendix we showed that $d_{\mathbf{v}} f(\mathbf{q})$ represents the rate of change of the z-coordinate (i.e. height) of the point $\mathbf{p} = \mathbf{F}(\mathbf{x})$ on the surface as $\mathbf{q}$ moves in the $\mathbf{v}$ direction.

If $\mathbf{a}$ is a critical point of f then $\nabla f(\mathbf{a}) = 0$ and so $d_{\mathbf{v}} f(\mathbf{q}) = \nabla f(\mathbf{a}) \bullet \mathbf{v} = 0$ for all $\mathbf{v}$. This means, essentially, that there is no change in height as you immediately move from the point $\mathbf{F}(\mathbf{a})$ on the surface, i.e. it is locally horizontal.

Another way to look at this is consider the Tangent plane which is the graph of

$$f(\mathbf{a}) + Jf(\mathbf{a})(\mathbf{x} - \mathbf{a}) = f(\mathbf{a}),$$

since $Jf(\mathbf{a}) = 0$. That is, the Tangent Plane is

$$\left\{ \begin{pmatrix} \mathbf{x} \\ f(\mathbf{a}) \end{pmatrix} : \mathbf{x} \in \mathbb{R}^2 \right\},$$

a horizontal plane.

Examples

Example *Find the extremal values of* $f(\mathbf{x}) = 7x^2 + 6xy + 7y^2$ *on the closed disc* $\{\mathbf{x} \in \mathbb{R}^2 : x^2 + y^2 \leq 8\}$.

Solution On the interior $x^2 + y^2 < 8$ the only critical point $\mathbf{x} : \nabla f(\mathbf{x}) = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$.

On the boundary $x^2 + y^2 = 8$ the critical points satisfy the system

$$14x + 6y = 2\lambda x \quad \text{and} \quad 6x + 14y = 2\lambda y. \tag{3}$$

Rewrite in matrix form as

$$\begin{pmatrix} 7-\lambda & 3 \\ 3 & 7-\lambda \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{0}.$$

If the determinant is non-zero then the only solution is $\mathbf{x} = \mathbf{0}$, which does not lie on $x^2 + y^2 = 8$.

If the determinant is zero then λ satisfies $(7-\lambda)^2 - 9 = 0$, This has two solutions, 4 and 10.

If $\lambda = 4$ then the system (3) reduces to the one equation $x + y = 0$. In $x^2 + y^2 = 8$ this gives the points $(2, -2)^T$ and $(-2, 2)^T$.

If $\lambda = 10$ then we get $x - y = 0$ which leads to $(2, 2)^T$ and $(-2, -2)^T$.

Thus we have five critical points $(\pm 2, \mp 2)^T$, $(\pm 2, \pm 2)^T$ and $\mathbf{0}$.

The function f is continuous and the disc closed and bounded so f will attain it's extremal values on the disc. We search through these points to find the extremal values:

$$\begin{aligned} & f(\mathbf{0}) = 0, \quad \text{minimum value,} \\ & f\left((\pm 2, \mp 2)^T\right) = 32, \\ & f\left((\pm 2, \pm 2)^T\right) = 80 \quad \text{maximum value.} \end{aligned}$$

■

A standard application of Lagrange's method is to give a proof of the arithmetic-geometric mean inequality.

Example 11 ***GM-AM inequality*** *For n positive real numbers*

$$x_1,\ x_2,\ \dots\ , x_n > 0$$

we have

$$(x_1 x_2 \dots x_n)^{1/n} \leq \frac{x_1 + x_2 + \dots + x_n}{n}.$$

Solution First assume $x_1, x_2, \dots, x_n > 0$ and $x_1 + x_2 + \dots + x_n = 1$. We will show that for such $\{x_i\}_{1 \leq i \leq n}$ we have

$$x_1 x_2 \dots x_n \leq \frac{1}{n^n}.$$

This entails finding the maximum of $f(\mathbf{x}) = x_1x_2...x_n$ subject to $g(\mathbf{x}) = x_1 + x_2 + ... + x_n = 1$ and $x_1, x_2, ..., x_n > 0$.

Note that if $\mathbf{x} : g(\mathbf{x}) = 1$ and $x_1, x_2, ..., x_n > 0$ then $0 < x_i \leq 1$ for all i so $|\mathbf{x}| \leq \sqrt{n}$, i.e. the set of such $\mathbf{x}$ is bounded. Also the set is closed, because its compliment is open: If $g(\mathbf{x}) \neq 1$ then $\exists\, \delta > 0 : \mathbf{x}' \in B_\delta(\mathbf{x})$ implies $g(\mathbf{x}') \neq 1$. Hence $S = \{\mathbf{x} : g(\mathbf{x}) = 1 \text{ and } x_1, x_2, ..., x_n \geq 0\}$ is a compact set. Since f is continuous it is bounded and attains its bounds.

If any $x_i = 0$ then $f(\mathbf{x}) = 0$, a minimum value for f. So assume all $x_i > 0$ and look for extrema in $S' = \{\mathbf{x} : g(\mathbf{x}) = 1 \text{ and } x_1, x_2, ..., x_n > 0\}$. We look for these within the critical points of f restricted to S', i.e. those $\mathbf{x}$ satisfying $\nabla f(\mathbf{x}) = \lambda \nabla g(\mathbf{x})$ for some λ, with $x_1, x_2, ..., x_n > 0$. Yet $\nabla f(\mathbf{x}) = \lambda \nabla g(\mathbf{x})$ becomes

$$\begin{aligned} \lambda(1, 1, ..., 1) &= (x_2...x_n,\ x_1x_3...x_n,\ x_1x_2x_4...x_n,\ ...,\ x_1x_2...x_{n-1}) \\ &= x_1x_2...x_n\left(\frac{1}{x_1}, \frac{1}{x_2}, ..., \frac{1}{x_n}\right). \end{aligned}$$

This implies $x_1 = x_2 = ... = x_n$. Within $g(\mathbf{x}) = 1$ this means $x_1 = x_2 = ... = x_n = 1/n$. Therefore at the critical point we find $f(\mathbf{x}) = 1/n^n$. Since there is only one critical point it must give the *maximum* value of $f(\mathbf{x})$. Hence $f(\mathbf{x}) \leq 1/n^n$ for all $\mathbf{x} \in S$.

Assume now that we only have the restriction $x_1, x_2, ..., x_n > 0$. Let $M = \sum_{i=1}^{n} x_i$ and define $y_i = x_i/M$ for all $1 \leq i \leq n$. Then $y_1, y_2, ..., y_n > 0$ and

$$y_1 + y_2 + ... + y_n = \frac{x_1 + x_2 + ... + x_n}{M} = \frac{M}{M} = 1.$$

So we can apply the result above to deduce that

$$y_1y_2...y_n \leq \frac{1}{n^n}, \text{ i.e. } \frac{x_1}{M}\frac{x_2}{M}...\frac{x_n}{M} \leq \frac{1}{n^n} \text{ so } x_1x_2...x_n \leq \frac{M^n}{n^n}.$$

Then

$$(x_1x_2...x_n)^{1/n} \leq \frac{M}{n} = \frac{x_1 + x_2 + ... + x_n}{n},$$

as required. ■

A subtle point. You may attempt this by looking for a lower bound for $f(\mathbf{x}) = x_1 + x_2 + ... + x_n$ subject to $x_1 x_2 ... x_n = 1$. Unfortunately the set of $\mathbf{x} : x_1 x_2 ... x_n = 1$ is not bounded and thus not compact. Then you could not apply Theorem 4 to deduce that there will be extrema.

Another standard application is to give a proof of a result used many times in this course, often in the form $|\mathbf{a} \bullet \mathbf{b}| \leq |\mathbf{a}| |\mathbf{b}|$ for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$.

Example 12 ***Cauchy-Schwartz*** *For arbitrary real numbers a_i, b_i, $1 \leq i \leq n$, we have*

$$\left(\sum_{i=1}^{n} a_i b_i \right)^2 \leq \sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2.$$

Solution Set up. We work inside $\mathbb{R}^n \times \mathbb{R}^n$, the space of ordered pairs of vectors. What does a general $\mathbf{x} \in \mathbb{R}^n \times \mathbb{R}^n$ look like? Simple, $\mathbf{x} = (\mathbf{a}, \mathbf{b})$ an ordered pair with $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$. But what are the coordinates of $\mathbf{x}$? Relate $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}^{2n}$ by the bijection

$$(\mathbf{a}, \mathbf{b}) \to \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}.$$

Reversing this map,

$$\begin{pmatrix} x^1 \\ \vdots \\ x^{2n} \end{pmatrix} \to \left(\begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix}, \begin{pmatrix} x^{n+1} \\ \vdots \\ x^{2n} \end{pmatrix} \right).$$

Thus, if $\mathbf{x} = (\mathbf{a}, \mathbf{b})$, then $x^i = a^i$ for $1 \leq i \leq n$ and $x^i = b^{i-n}$ for $n+1 \leq i \leq 2n$.

We first show that $|\mathbf{a} \bullet \mathbf{b}| \leq 1$ when $|\mathbf{a}| = 1$ and $|\mathbf{b}| = 1$.

Define $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, (\mathbf{a}, \mathbf{b}) \mapsto \mathbf{a} \bullet \mathbf{b}$. Let $S = \{(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^n \times \mathbb{R}^n : |\mathbf{a}| = |\mathbf{b}| = 1\}$. Then S is a closed and bounded set, i.e. compact. Also f is a continuous function, so f is bounded and attains its bounds. We look for these extrema within the critical points.

Define

$$\begin{aligned} g_1 &: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, (\mathbf{a}, \mathbf{b}) \mapsto \mathbf{a} \bullet \mathbf{a} = |\mathbf{a}|^2, \\ g_2 &: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, (\mathbf{a}, \mathbf{b}) \mapsto \mathbf{b} \bullet \mathbf{b} = |\mathbf{b}|^2. \end{aligned}$$

So

$$S = \{\, \mathbf{x} \in \mathbb{R}^n \times \mathbb{R}^n : g_1(\mathbf{x}) = g_2(\mathbf{x}) = 1 \} \, .$$

We can only apply Lagrange's method if the Jacobian matrix of this level set is of full-rank. Let $\mathbf{x} = (\mathbf{a}, \mathbf{b})$ for some $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$. From

$$\mathbf{a} \bullet \mathbf{a} = \sum_{i=1}^{n} \left(a^i\right)^2$$

we see that

$$\frac{d}{dx^i} g_1(\mathbf{x}) = \frac{d}{da^i} \mathbf{a} \bullet \mathbf{a} = 2a^i$$

for $1 \leq i \leq n$, while

$$\frac{d}{dx^i} g_1(\mathbf{x}) = \frac{d}{db^{i-n}} \mathbf{a} \bullet \mathbf{a} = 0$$

for $n+1 \leq i \leq 2n$. Thus $Jg_1(\mathbf{x}) = \left(2\mathbf{a}^T, \mathbf{0}^T\right)$, where $\mathbf{x} = \left(\mathbf{a}^T, \mathbf{b}^T\right)^T$.

Similarly for Jg_2 and then the $2 \times 2n$ Jacobian matrix is

$$\begin{pmatrix} Jg_1(\mathbf{x}) \\ Jg_2(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} 2\mathbf{a}^T & \mathbf{0}^T \\ \mathbf{0}^T & 2\mathbf{b}^T \end{pmatrix},$$

This is of full-rank whenever $\mathbf{a} \neq \mathbf{0}$ and $\mathbf{b} \neq \mathbf{0}$, which anyway are excluded by the conditions $|\mathbf{a}| = |\mathbf{b}| = 1$.

Thus we can apply Lagrange's method when the critical points of f restricted to S are $\mathbf{x} \in \mathbb{R}^n \times \mathbb{R}^n$ which satisfy $\nabla f(\mathbf{x}) = \lambda \nabla g_1(\mathbf{x}) + \mu \nabla g_2(\mathbf{x})$ for some real λ, μ. Here

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \partial f(\mathbf{x}) / \partial x^1 \\ \vdots \\ \vdots \\ \partial f(\mathbf{x}) / \partial x^{2n} \end{pmatrix} = \begin{pmatrix} \partial \mathbf{a} \bullet \mathbf{b} / \partial a^1 \\ \vdots \\ \partial \mathbf{a} \bullet \mathbf{b} / \partial a^n \\ \partial \mathbf{a} \bullet \mathbf{b} / \partial b^1 \\ \vdots \\ \partial \mathbf{a} \bullet \mathbf{b} / \partial b^n \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ \mathbf{a} \end{pmatrix}.$$

Thus

$$\begin{pmatrix} \mathbf{b} \\ \mathbf{a} \end{pmatrix} = \lambda \begin{pmatrix} 2\mathbf{a} \\ \mathbf{0} \end{pmatrix} + \mu \begin{pmatrix} \mathbf{0} \\ 2\mathbf{b} \end{pmatrix},$$

From the first line $\mathbf{b} = 2\lambda\mathbf{a}$. Since $|\mathbf{a}| = |\mathbf{b}| = 1$ we have $2\lambda = \pm 1$ and thus $\mathbf{b} = \pm\mathbf{a}$. So the critical points of f restricted to S are contained within the set of points $(\mathbf{a}, \mathbf{a})$ and $(\mathbf{a}, -\mathbf{a})$ with $|\mathbf{a}| = 1$. For these points $f((\mathbf{a}, \mathbf{a})) = |\mathbf{a}|^2 = 1$ and $f((\mathbf{a}, -\mathbf{a})) = -|\mathbf{a}|^2 = -1$. These, therefore, are the extreme values and in general

$$-1 \leq f((\mathbf{a}, \mathbf{b})) \leq 1.$$

This was proved subject to $|\mathbf{a}| = 1$ and $|\mathbf{b}| = 1$. If we only know that $\mathbf{a} \neq \mathbf{0}$ and $\mathbf{b} \neq \mathbf{0}$ we can apply the result just proved to $\mathbf{a}/|\mathbf{a}|$ and $\mathbf{b}/|\mathbf{b}|$ to get

$$-1 \leq f\left(\left(\frac{\mathbf{a}}{|\mathbf{a}|}, \frac{\mathbf{b}}{|\mathbf{b}|}\right)\right) \leq 1.$$

But

$$f\left(\left(\frac{\mathbf{a}}{|\mathbf{a}|}, \frac{\mathbf{b}}{|\mathbf{b}|}\right)\right) = \frac{\mathbf{a}}{|\mathbf{a}|} \bullet \frac{\mathbf{b}}{|\mathbf{b}|} = \frac{f((\mathbf{a}, \mathbf{b}))}{|\mathbf{a}|\,|\mathbf{b}|}.$$

Hence

$$-|\mathbf{a}|\,|\mathbf{b}| \leq f((\mathbf{a}, \mathbf{b})) \leq |\mathbf{a}|\,|\mathbf{b}|$$

That is $|f((\mathbf{a}, \mathbf{b}))| \leq |\mathbf{a}|\,|\mathbf{b}|$ as required. ■

Other examples are

Example 13 *Let $e_1, e_2, ..., e_n$ be positive numbers with $\sum_{i=1}^n e_i = 1$. Maximise the function $f(\mathbf{x}) = x_1^{e_1} x_2^{e_2} ... x_n^{e_n}$ subject to $\sum_{i=1}^n e_i x_i = 1$ and $x_i > 0$ for all $1 \leq i \leq n$.*

Deduce the extended GM-AM inequality:

$$x_1^{e_1} x_2^{e_2} ... x_n^{e_n} \leq e_1 x_1 + e_2 x_2 + ... + e_n x_n$$

for all non-negative x_i.

Solution The i-th component of $\nabla f(\mathbf{x}) = \lambda \nabla g(\mathbf{x})$ is

$$e_i \frac{x_1^{e_1} x_2^{e_2} ... x_n^{e_n}}{x_i} = \lambda e_i,$$

Thus $\lambda x_i = y$ say, is the same value for all i. From $\sum_{i=1}^n e_i x_i = 1$ we then get

$$\lambda = \sum_{i=1}^n e_i \lambda x_i = y \sum_{i=1}^n e_i = y,$$

since $\sum_{i=1}^n e_i = 1$. Thus $y = \lambda$ which in $\lambda x_i = y$ gives $x_i = 1$ for all i. Hence $f(\mathbf{x})$ is maximised at $\mathbf{x} = \mathbf{1}$ and $f(\mathbf{x}) \leq f(\mathbf{1}) = 1$ for all $\mathbf{x} : \sum_{i=1}^n e_i x_i = 1$ and $x_i > 0$ for all $1 \leq i \leq n$.

Given an $\mathbf{x}$ with $x_i > 0$ for all i, which does not satisfy $\sum_{i=1}^n e_i x_i = 1$ define $\widehat{\mathbf{x}} = \mathbf{x} / \sum_{i=1}^n e_i x_i$. The components of this now satisfy $\sum_{i=1}^n e_i \widehat{x}_i = 1$ and so, by above, $f(\widehat{\mathbf{x}}) \leq 1$. Yet

$$f(\widehat{\mathbf{x}}) = f\left(\frac{\mathbf{x}}{\sum_{i=1}^n e_i x_i}\right) = \frac{1}{\left(\sum_{i=1}^n e_i x_i\right)^{e_1+e_2+\ldots+e_n}} f(\mathbf{x}) = \frac{1}{\sum_{i=1}^n e_i x_i} f(\mathbf{x}),$$

since $\sum_{i=1}^n e_i = 1$. Then $f(\widehat{\mathbf{x}}) \leq 1$ rearranges to $f(\mathbf{x}) \leq \sum_{i=1}^n e_i x_i$. ■

Example 14 *Show that the maximum of $x_1 x_2 \ldots x_n$ on*

$$\sum_{m=1}^n \frac{x_m^2}{m^2} = 1 \tag{4}$$

is

$$\frac{n!}{n^{n/2}}.$$

Geometrically the surface (4) is of an ellipsoid in $\mathbb{R}^n$. Then $x_1 x_2 \ldots x_n$ is the volume of a box with one corner at the origin and the diagonally opposite corner on the surface of the ellipsoid.

Hint Write $X = x_1 x_2 \ldots x_n$ then the system of equations arising from Lagrange's method is

$$\frac{X}{x^i} = 2\lambda \frac{x_i}{i^2},$$

for $1 \leq i \leq n$ along with (4).

Example 15 *Maximise $x^2 y^2$ subject to the constraint*

$$\frac{x^{2p}}{p} + \frac{y^{2q}}{q} = r^2,$$

where p and q are real numbers greater than 1 which satisfy

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Show that the maximum is achieved when $x^{2p} = y^{2q} = r^2$. Now conclude that if $x, y > 0$ then

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q},$$

and equality is attained for some x and y.

Solution The x component of $\nabla f(\mathbf{x}) = \lambda \nabla g(\mathbf{x})$ is $2xy^2 = 2x^{2p-1}$, the y component $2x^2y = 2y^{2q-1}$. Thus $2y^{2q} = 2x^2y^2 = 2x^{2p}$, i.e. $y^{2q} = x^{2p}$. In

$$\frac{x^{2p}}{p} + \frac{y^{2q}}{q} = r^2 \text{ with } \frac{1}{p} + \frac{1}{q} = 1$$

gives $y^{2q} = x^{2p} = r^2$. The maximum is thus

$$x^2y^2 = r^{2/p}r^{2/q} = r^2$$

since $1/p + 1/q = 1$.

If $x, y > 0$ are given apply the above to s, t satisfying $s^2 = x$ and $t^2 = y$. Define r by

$$r^2 = \frac{s^{2p}}{p} + \frac{t^{2q}}{q}.$$

Then as proved above, $s^2t^2 \leq r^2$. Substituting back for x and y gives

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}.$$

$\blacksquare$

Example 16 *Let p and q be positive real numbers which satisfy*

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Maximise the function $g : \mathbb{R}^n \times \mathbb{R}^n, (\mathbf{a}, \mathbf{b}) \mapsto \mathbf{a} \bullet \mathbf{b}$, subject to $|\mathbf{a}|_p = 1$ and $|\mathbf{b}|_q = 1$ where, for $\mathbf{x} \in \mathbb{R}^n$ we define

$$|\mathbf{x}|_p = \left(\sum_{i=1}^{n} x_i^p \right)^{1/p}.$$

*Derive **Holder's Inequality**: for non-negative $a_1, ..., a_n, b_1, ..., b_n$ we have*

$$\sum_{i=1}^{n} a_i b_i \leq \left(\sum_{i=1}^{n} a_i^p \right)^{1/p} \left(\sum_{i=1}^{n} b_i^q \right)^{1/q}.$$

Returning to an example in the lectures.

Example 17 *The critical point of* $f(\mathbf{x}) = x^2y^2 + z^2 + 2x - 4y + z$ *is not an extremal value.*

Solution The critical point is $\mathbf{a} = \left(-2^{2/3},\ 2^{-1/3},\ -2^{-1}\right)^T$. Let $\mathbf{u} = (1,1,0)^T$ and $\mathbf{v} = (0,1,0)^T$. Then it can be shown that

$$f(\mathbf{a}{+}t\mathbf{u}) = f(\mathbf{a}) + \frac{1}{4}\left(2t - \left(\sqrt[3]{2}\right)^2\right)\left(2t + 3\left(\sqrt[3]{2}\right)^2\right)t^2 < f(\mathbf{a}),$$

provided $-3 \times 2^{-1/3} < t < 2^{-1/3.}$ and $t \neq 0$. This shows $f(\mathbf{a})$ is not a local minimum. Also

$$f(\mathbf{a}{+}t\mathbf{v}) = f(\mathbf{a}) + 2\sqrt[3]{2}t^2 > f(\mathbf{a})$$

for all $t \neq 0$. This shows that $f(\mathbf{a})$ is not a local maximum. ■