Constrained optimization: Lagrange multipliers.

A linear algebra lemma. (You can safely skip it)

Let $\varphi_1, \dots, \varphi_p, \psi : \mathbb{R}^n \to \mathbb{R}$ be linear

Then

$$\bigcap_{i=1}^{p} \ker(\varphi_i) \subset \ker(\psi) \iff \exists a_1, \dots, a_p \in \mathbb{R},\ \psi = \sum_{i=1}^{p} a_i \varphi_i$$

Δ $\Leftarrow$: Assume that $\psi = \sum_{i=1}^{p} a_i \varphi_i$ for some $a_i \in \mathbb{R}$

Let $x \in \bigcap_{i=1}^{p} \ker \varphi_i$ then

$$\psi(x) = \sum_{i=1}^{p} a_i \varphi_i(x) = \sum_{i=1}^{p} a_i 0 = 0$$

Hence $x \in \ker \psi$

We proved that $\bigcap_{i=1}^{p} \ker(\varphi_i) \subset \ker(\psi)$

$\Rightarrow$: We define $\Phi : \mathbb{R}^n \to \mathbb{R}^p$ by $\Phi(x) = (\varphi_1(x), \dots, \varphi_p(x))$

Notice that Φ is linear since the φ_i are

Claim 1: $\ker \Phi \subset \ker \psi$

Indeed, let $x \in \ker \Phi$, then $\vec{0} = \Phi(x) = (\varphi_1(x), \dots, \varphi_p(x))$
and $x \in \bigcap_{i=1}^{p} \ker \varphi_i \subset \ker \psi$

Hence $\ker \Phi \subset \ker \psi$ as claimed.

Claim 2: $\exists f : \mathbb{R}^p \to \mathbb{R}$ linear such that $\psi = f \circ \Phi$

Set $r = \operatorname{rank}(\Phi)$, then by the rank-nullity theorem, $\dim \ker \Phi = n - r$

Hence we may find a basis $(v_1, \dots, v_n)$ of $\mathbb{R}^n$ such that $(v_{r+1}, \dots, v_n)$ is a basis of $\ker \Phi$

Then $u_1 = \Phi(v_1), \dots, u_r = \Phi(v_r)$ are linearly dependent,

indeed $\sum_{i=1}^{r} a_i \Phi(v_i) = 0 \Rightarrow \Phi\left(\sum_{i=1}^{r} a_i v_i\right) = 0$

$\Rightarrow \sum_{i=1}^{r} a_i v_i \in \ker \Phi$

$\Rightarrow \forall i,\ a_i = 0$ since $\mathbb{R}^m = \langle v_1, \dots, v_r \rangle \oplus \ker \Phi$

So we can extend $(u_1, \dots, u_r)$ in a basis $(u_1, \dots, u_r, u_{r+1}, \dots, u_p)$ of $\mathbb{R}^p$.

Now we define $f: \mathbb{R}^p \to \mathbb{R}$ linear by:

$$f(u_1) = \psi(v_1), \dots, f(u_r) = \psi(v_r), \quad f(u_{r+1}) = \cdots = f(u_p) = 0$$

Let's check that $\psi = f \circ \Phi$

Let $x \in \mathbb{R}^m$, then $x = \sum_{i=1}^{m} x_i v_i$, and

$$\begin{aligned}
f \circ \Phi(x) &= f\left(\sum_{i=1}^{m} x_i \Phi(v_i)\right) \\
&= f\left(\sum_{i=1}^{r} x_i u_i\right) \text{ since } \begin{cases} \Phi(v_i) = u_i \text{ for } i = 1, \dots, r \\ \Phi(v_i) = 0 \text{ for } i = r+1, \dots, m \end{cases} \\
&= \sum_{i=1}^{r} x_i f(u_i) \\
&= \sum_{i=1}^{r} x_i \psi(v_i) \\
&= \sum_{i=1}^{m} x_i \psi(v_i) \text{ since for } i \geq r+1,\ v_i \in \ker \Phi \subset \ker \psi \text{ by claim 1} \\
&= \psi\left(\sum x_i v_i\right) \\
&= \psi(x)
\end{aligned}$$

And the claim is proved

Now, since $f: \mathbb{R}^p \to \mathbb{R}$ is linear, $f(y_1, \dots, y_p) = \sum_{i=1}^{p} y_i f(e_i)$

and $\psi(x) = f(\Phi(x)) = f(\varphi_1(x), \dots, \varphi_p(x)) = \sum_{i=1}^{p} f(e_i) \varphi_i(x) = \sum_{i=1}^{p} a_i \varphi_i(x)$

for $a_i = f(e_i)$

$\square$

(Extra curricular) (You can safely skip it)

<u>Comment</u>: If you are familiar with duality then the proof of "$\Leftarrow$" is very natural:

Δ $\varphi_1, \dots, \varphi_p$ are vectors of the n-dim space $\mathcal{L}(\mathbb{R}^n, \mathbb{R})$

so we may find a linearly independent subfamily $\varphi_1, \dots, \varphi_q$ in $\mathcal{L}(\mathbb{R}^n, \mathbb{R})$ such that $\mathrm{Vect}(\varphi_1, \dots, \varphi_q) = \mathrm{Vect}(\varphi_1, \dots, \varphi_p)$

Then we extend $(\varphi_1, \dots, \varphi_q)$ in a basis $(\varphi_1, \dots, \varphi_n)$ of $\mathcal{L}(\mathbb{R}^n, \mathbb{R})$

Hence $\psi = \sum_{i=1}^{n} a_i \varphi_i$

Let $(e_1, \dots, e_n)$ the basis of $\mathbb{R}^n$ dual to $(\varphi_1, \dots, \varphi_n)$, ie $\varphi_i(e_j) = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{otherwise} \end{cases}$

For $j \geq q+1$, $e_j \in \ker\left(\bigcap_{i=1}^{q} \varphi_i\right) = \ker\left(\bigcap_{i=1}^{p} \varphi_i\right) \subset \ker \psi$

Hence $0 = \psi(e_j) = \sum_{i=1}^{n} a_i \varphi_i(e_j) = a_j \Rightarrow \forall j \geq q+1,\ a_j = 0$

and $\psi = \sum_{j=1}^{q} a_j \varphi_j$ $\square$

Theorem: (Lagrange multipliers) ⚠ Main result of this chapter

$\mathcal{U} \subset \mathbb{R}^n$ open, $f, g_1, \dots, g_p : \mathcal{U} \longrightarrow \mathbb{R}$ of class C^1.

Define $X = \{x \in \mathcal{U} : g_1(x) = \dots = g_p(x) = 0\}$

If : $\begin{cases} f_{|X} : X \to \mathbb{R} \text{ has a local extremum at } a \in X \\ \text{and} \\ \nabla g_1(a), \dots, \nabla g_p(a) \text{ are linearly independant} \end{cases}$

then there exist $\lambda_1, \dots, \lambda_p \in \mathbb{R}$ s.t. $\nabla f(a) = \sum_{i=1}^{p} \lambda_i \nabla g_i(a)$

Comment:

⚠ $x \in X$ not $x \in \mathcal{U}$

- $f_{|X}$ has a local min at $a \in X$ means $\exists r > 0, \underbrace{\forall x \in X}, \|x - a\| < r \Rightarrow f(a) \leq f(x)$
- $f_{|X}$ has a local max at $a \in X$ means $\exists r > 0, \forall x \in X, \|x - a\| < r \Rightarrow f(a) \geq f(x)$

△ **Sketch of proof: the geometric idea** (You can safely skip it)

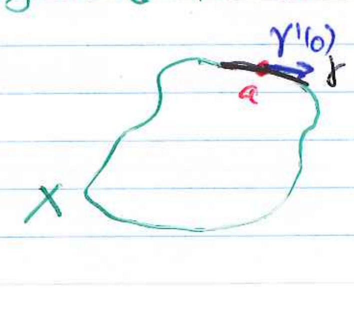
Fact: $\bigcap_{i=1}^{p} \ker(d_a g_i) = \{v \in \mathbb{R}^n : v = \gamma'(0) \text{ for a } C^1 \ \gamma : (-1,1) \to \mathbb{R}^n \text{ s.t. } \forall t \in (-1,1), \gamma(t) \in X \text{ and } \gamma(0) = a\}$

We admit this fact, but you can convince yourself that these two sets describe the tangent space of X at a

- v is tangent to $g_i = 0$ at a means $0 = \nabla g_i(a) \cdot v = d_a g_i(v)$, ie $v \in \ker d_a g_i$

  so v is tangent to X at a if v is tangent to all the $g_i = 0$, ie $v \in \bigcap \ker d_a g_i$

- v is tangent to X at a if $v = \gamma'(0)$ for γ has above



Let $\gamma : (-1,1) \to \mathbb{R}^n$ s.t. $\gamma(0) = a$, $\forall t \in (-1,1), \gamma(t) \in X$ and γ C^1

if $f_{|X}$ has an extremum at a, then $f \circ \gamma$ has an extremum at 0

So

$$0 = (f\circ\gamma)'(0) = d_0(f\circ\gamma)(1) = d_{\gamma(0)} f \circ d_0\gamma(1) = d_a f(\gamma'(0))$$

Hence $\gamma'(0) \in \ker d_a f$ for any γ as above

By the fact $\bigcap_{i=1}^{p} d_a g_i = \{\gamma'(0) : \gamma \text{ as above}\} \subset \ker d_a f$

Hence by the linear algebra lemma:

$$d_a f = \sum_{i=1}^{p} \lambda_i \, d_a g_i \quad \text{for some } \lambda_i \in \mathbb{R}$$

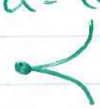
i.e. $$\nabla f(a) = \sum_{i=1}^{p} \lambda_i \nabla g_i(a)$$

$\square$

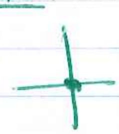
Remark: the assumption $(\nabla g_1(a), \dots, \nabla g_p(a))$ linearly independant ensures that $X = g_1^{-1}(0) \cap \dots \cap g_p^{-1}(0)$ is a "submanifold" at a so that the tangent space of X at a is well defined in the above proof:

Ex: $X = \{x^2 + y^2 = 1\}$
$a = (0,1)$

Non-ex: $X = \{x^3 - y^2 = 0\}$
$a = (0,0)$



Non-ex: $X = \{xy = 0\}$ $a = (0,0)$



Special case: $p = 1$

Theorem:

$\mathcal{U} \subset \mathbb{R}^m$ open, $f, g : \mathcal{U} \to \mathbb{R}$ C^1

Let $X = g^{-1}(0) := \{x \in \mathcal{U},\ g(x) = 0\}$

If $\begin{cases} f_{|X} \text{ has a local extremum at } a \in X \\ \nabla g(a) \neq \vec{0} \end{cases}$

then $\nabla f(a) = \lambda \nabla g(a)$ for some $\lambda \in \mathbb{R}$

If the constraint is given by an inequality:

$\mathcal{U} \subset \mathbb{R}^m$ open, $f, g : \mathcal{U} \to \mathbb{R}$ C^1

$X = g^{-1}((-\infty, 0)) := \{x \in \mathcal{U},\ g(x) \leq 0\}$

We look for local extrema of f on X

Notice that $X = \underbrace{\{x \in \mathcal{U} : g(x) = 0\}}_{X_1} \cup \underbrace{\{x \in \mathcal{U} : g(x) < 0\}}_{X_2}$

$= X_1 \cup X_2$.

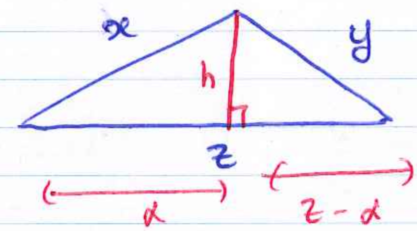
Step 1: Use Lagrange multipliers on X_1

Step 2: Notice that X_2 is open so you can use the results from the previous chapters here.

Homework: examples + questions from 2.8

A first application

What's the largest area that we can obtain with a triangle of perimeter p?



$$\begin{cases} x^2 = h^2 + \alpha^2 \\ y^2 = h^2 + (z-\alpha)^2 \end{cases} \Rightarrow \quad x^2 - y^2 = \alpha^2 - (z-\alpha)^2 = 2\alpha z - z^2$$

$$\Rightarrow \quad \alpha = \frac{x^2 - y^2 + z^2}{2z}$$

$$h^2 = x^2 - \alpha^2 = x^2 - \frac{(x^2 - y^2 + z^2)^2}{(2z)^2}$$

$$= \frac{(2xz)^2 - (x^2 - y^2 + z^2)}{(2z)^2}$$

$$= \frac{(2xz - x^2 + y^2 - z^2)(2xz + x^2 - y^2 + z^2)}{4z^2}$$

$$= \frac{(y^2 - (x-z)^2)((x+z)^2 - y^2)}{4z^2}$$

$$= \frac{(y - x + z)(y + x - z)(x + z - y)(x + z + y)}{4z^2}$$

$$= \frac{p(p-2x)(p-2y)(p-2z)}{4z^2}$$

$$\mathcal{A} = \frac{zh}{2} \Rightarrow \mathcal{A}^2 = \frac{z^2h^2}{4} = \frac{p(p-2x)(p-2y)(p-2z)}{16}$$

Since $t \mapsto \sqrt{t}$ is increasing on $[0,+\infty)$, it is enough to maximize $\mathcal{A}^2$ with the constraint $x+y+z=p$

(Get rid of the square root when you can...)

$$f(x,y,z) = p(p-2x)(p-2y)(p-2z)$$

$$g(x,y,z) = x+y+z-p$$

So we want to maximize f with the constraint $g=0$ for $x\geq 0, y\geq 0, z\geq 0$.

$$S = \{(x,y,z)\in\mathbb{R}^3 : x\geq 0,\ y\geq 0,\ z\geq 0,\ x+y+z=p\}$$

is compact hence f has a max on S.

If x, y or $z=0$ then $\mathcal{A}=0$, hence we study $f: \mathcal{U}\to\mathbb{R}$ on the open set $\{(x,y,z) : x>0, y>0, z>0\}$ with the constraint $g(x,y,z)=0$

By Lagrange multiplier theorem, at a local max a on $\mathcal{U}\cap g^{-1}(0)$ we have $a=(x_0,y_0,z_0)$

$\nabla f(a) = \lambda \nabla g(a)$ for some λ, assuming $\nabla g(a) \neq \vec{0}$ which is the case

$$\Leftrightarrow \begin{pmatrix} -2(p-2y_0)(p-2z_0) \\ -2(p-2x_0)(p-2z_0) \\ -2(p-2x_0)(p-2z_0) \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Hence we have to solve

$$\begin{cases} (p-2y_0)(p-2z_0) = \mu \\ (p-2x_0)(p-2z_0) = \mu \\ (p-2x_0)(p-2y_0) = \mu \\ p = x_0+y_0+z_0 \end{cases}$$

here $\mu = -\dfrac{\lambda}{2}$

eg: $(p-2y_0)(p-2z_0) = \mu = (p-2y_0)(p-2x_0)$

$\Rightarrow p - 2z_0 = p - 2x_0$

$\Rightarrow x_0 = z_0$

$$\Rightarrow \begin{cases} x_0 = y_0 = z_0 \\ p = x_0+y_0+z_0 \end{cases}$$

$\Rightarrow x_0 = y_0 = z_0 = p/3$

it is the only local extremum in $\{(x,y,z) : x>0, y>0, z>0, x+y+z=p\}$ and $f(p/3, p/3, p/3) > 0$ the min on S is 0 for $x=0$ or $y=0$ or $z=0$

So the only local max is at $(p/3, p/3, p/3)$ and it has to be a global max

$\rightsquigarrow$ We get the max area for an equilateral triangle and $A^2 = \dfrac{p\left(p-\frac{2}{3}p\right)^3}{16} = \dfrac{p^4}{27\times 16}$

ie $A = \dfrac{p^2}{12\sqrt{3}}$

Homework: questions from section 2.8.

A fancy proof of the AM-GM inequality

$\forall x_1, \dots, x_m \in \mathbb{R}_{\geq 0}$, $\sqrt[m]{x_1 \cdots x_m} \leq \frac{x_1 + \cdots + x_m}{m}$

Δ $\Gamma = \{(x_1, \dots, x_m) \in \mathbb{R}^m,\ x_i \geq 0,\ x_1 + \cdots + x_m = 1\}$ is closed and bounded

Hence it is compact and $f : \Gamma \longrightarrow \mathbb{R}$ defined by

$f(x_1, \dots, x_m) = x_1 \cdots x_m$ has a max on Γ since it is C^0 on a compact set.

If one of the $x_i = 0$ then $f(x_1, \dots, x_m) = 0$ so the max of f on Γ must be in $U \cap X$ where $U = \{(x_1, \dots, x_m) \in \mathbb{R}^m,\ x_i > 0\}$ is open

and $X = \{(x_1, \dots, x_m) \in \mathbb{R}^m : g(x_1, \dots, x_m) = 0\}$

where $g(x_1, \dots, x_m) = x_1 + \cdots + x_m - 1$

Notice that $\nabla g(x_1, \dots, x_m) = (1, \dots, 1) \neq \vec{0}$ hence, by Lagrange's multipliers theorem, if a is a max of f on X then $\exists \lambda \in \mathbb{R}$ s.t.

$\nabla f(a) = \lambda \nabla g(a)$ ie $(f(a)/a_1, \dots, f(a)/a_m) = \lambda(1, \dots, 1)$

Hence $\forall i, j$, $a_i = a_j$.

Moreover $g(a) = 0$ ie $a_1 + \cdots + a_m = 1 \Rightarrow \forall i, a_i = 1/m$.

Hence $f(1/m, \dots, 1/m) = \frac{1}{m^m}$ has to be the max of f on X

(it is > 0 and the only local extremum here, the min on Γ is 0 when some $x_i = 0$)

Now let $x_1, \dots, x_m \in \mathbb{R}_{>0}$ and set $x_i' = \frac{x_i}{\sum_{j=1}^m x_j}$ then $\sum_{i=1}^m x_i' = 1$

(if a $x_i = 0$ then the statement is obvious)

so that $\underbrace{f(x_1', \dots, x_m')}_{x_1' \cdots x_m' = \frac{x_1 \cdots x_m}{(\sum x_i)^m}} \leq \frac{1}{m^m} \Rightarrow x_1 \cdots x_m \leq \frac{(\sum_i x_i)^m}{m^m} \Rightarrow \sqrt[m]{x_1 \cdots x_m} \leq \frac{\sum x_i}{m}$

$\square$

Ex: Let $L = \{(x,y,z) \in \mathbb{R}^3 : x+y+z+\frac{7}{2}=0,\ x-y+2z=0\}$

Find $p \in L$ minimizing the distance to the origin

Δ Notice that for $g_1(x,y,z) = x+y+z+\frac{7}{2}$ and $g_2(x,y,z) = x-y+2z$

$\nabla g_1(x,y,z) = (1,1,1)$ and $\nabla g_2(x,y,z) = (1,-1,2)$ are linearly independent

hence L is a line

Let $f(x,y,z) = x^2+y^2+z^2$

By Lagrange multipliers theorem, if $p=(x_0,y_0,z_0)$ is a min of $f_{|L}$

then $\exists \lambda_1, \lambda_2 \in \mathbb{R}$ st. $\nabla f(P) = \lambda_1 \nabla g_1(P) + \lambda_2 \nabla g_2(P)$

$$\Rightarrow \begin{pmatrix} 2x_0 \\ 2y_0 \\ 2z_0 \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$$

then $x_0 = \dfrac{\lambda_1+\lambda_2}{2}$, $y_0 = \dfrac{\lambda_1-\lambda_2}{2}$, $z_0 = \dfrac{\lambda_1+2\lambda_2}{2}$

and $\begin{cases} g_1(x_0,y_0,z_0)=0 \\ g_2(x_0,y_0,z_0)=0 \end{cases} \Rightarrow \begin{cases} 3\lambda_1 + 2\lambda_2 = -7 \\ \lambda_1 + 3\lambda_2 = 0 \end{cases} \Rightarrow \begin{cases} \lambda_1 = -3 \\ \lambda_2 = +1 \end{cases}$

Therefore $(x_0,y_0,z_0) = (-1,-2,-\frac{1}{2})$ is the only possible $p \in L$ minimizing the distance to the origin

$\square$

Ex : Find the min and max of $f(x,y,z) = x + 2y + z$

on $X = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2=1,\ y-z=2\}$

Δ X is compact and f continuous, hence $f_{|X}$ has a min and a max

Define $g_1(x,y,z) = x^2+y^2-1$ $\quad g_2(x,y,z) = y-z-2$

then $Dg_1(x,y,z) = (2x, 2y, 0)$ $\quad \nabla g_2(x,y,z) = (0,1,-1)$

Hence $Dg_1(x,y,z)$ and $Dg_2(x,y,z)$ are linearly independent on X

Let $p=(x_0,y_0,z_0)$ be a local extremum of f on X then

by Lagrange multipliers theorem, $\exists \lambda_1, \lambda_2 \in \mathbb{R}$ s.t.

$$\nabla f(x_0,y_0,z_0) = \lambda_1 Dg_1(x_0,y_0,z_0) + \lambda_2 Dg_2(x_0,y_0,z_0)$$

$$\Rightarrow \quad (1,2,1) = \lambda_1(2x_0, 2y_0, 0) + \lambda_2(0,1,-1)$$

from the last component, we get that $\lambda_2 = -1$

Hence $\lambda_1(2x_0, 2y_0, 0) = (1,2,1) + (0,1,-1) = (1,3,0)$

and $x_0 = \frac{1}{2\lambda_1}$, $y_0 = \frac{3}{2\lambda_1}$, $z_0 = y_0 - 2 = \frac{3-4\lambda_1}{2\lambda_1}$ or $\frac{3}{2\lambda_1} - 2$

From $x_0^2 + y_0^2 = 1$ we get $\lambda_1 = \pm \frac{\sqrt{10}}{2}$

the local extrema has to be at $P_1 = (1/\sqrt{10}, \frac{3}{\sqrt{10}}, 3/\sqrt{10} - 2)$

and $P_2 = (-1/\sqrt{10}, -3/\sqrt{10}, -2-3/\sqrt{10})$

$f(P_1) = \sqrt{10} - 2 > -2 - \sqrt{10} = f(P_2)$

$\hookrightarrow$ max $\qquad \hookrightarrow$ min

$\square$

Theorem (Spectral theorem)

Let $A \in M_{n,n}(\mathbb{R})$ be a symmetric matrix (i.e. $A^t = A$)

Then there is an orthogonal basis of $\mathbb{R}^n$ made of eigenvectors of A

Δ Proof by induction on n: if $n=1$ OK.

Assume that the statement holds for $n-1$.

Let $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ be defined by $f(x) = x^t A x$

then f is differentiable and $\nabla f(x) = 2Ax$

Indeed: $f(x+h) = (x+h)^t A (x+h)$

$$= x^t A x + h^t A x + x^t A h + h^t A h$$

$$= f(x) + (Ax) \cdot h + (Ax)^t h + h^t A h$$

$$= f(x) + 2(Ax) \cdot h + h^t A h$$

and $|h^t A h| = |h \cdot (Ah)| \leq ||h|| \cdot ||Ah||$ by Cauchy-Schwarz

hence $\dfrac{|h^t A h|}{||h||} = ||Ah|| \xrightarrow[h \to 0]{} 0$ by continuity.

Define $g: \mathbb{R}^n \to \mathbb{R}$ by $g(x) = ||x||^2 = x^t x$

Then $X = \{x \in \mathbb{R}^n : g(x) = 1\}$ is compact and $f_{|X}$ has a max v

Recall that $\nabla g(x) = 2x \neq \vec{0}$ for $x \neq \vec{0}$ ~~$\in$~~ X, hence, by Lagrange multipliers theorem $\exists \lambda \in \mathbb{R}$, $\nabla f(v) = \lambda \nabla g(v)$

$$\Rightarrow 2Av = 2\lambda v$$

$$\Rightarrow Av = \lambda v$$

hence v is an eigenvector of A

Now, if $x \in \langle v \rangle^\perp$ then $(Ax)\cdot v = (Ax)^t v = x^t A^t v = x A v = \lambda x \cdot v = 0$

so $x \in \langle v \rangle^\perp \Rightarrow Ax \in \langle v \rangle^\perp$

Hence, in a basis w.r.t. $\mathbb{R}^m = \langle v \rangle \oplus \langle v \rangle^\perp$

$$A = \begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & B & \\ 0 & & & \end{pmatrix}$$

with $\underset{M_{m-1,m-1}(\mathbb{R})}{\overset{\cap}{B}}$ symmetric, so we may conclude by the induction hypothesis $\square$