5.3 The method of Lagrange multipliers

Consider two functions

$$f\colon \mathbb{R}^n \to \mathbb{R}, \qquad \mathbf{h}\colon \mathbb{R}^n \to \mathbb{R}^m,$$

and the associated optimisation problem to

$$\begin{aligned} \text{minimise} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{h}(\mathbf{x}) = \mathbf{0}. \end{aligned}$$

Lagrange's theorem gives necessary (first order) conditions for nonlinear problems with equality constraints. Second order conditions a la the SONC and the SOSC are also given. Inequality constraints are covered in the next section.

Joseph-Louis Lagrange

We lose no generality by assuming the right-hand side of the constraint equation is $\mathbf{0}$, as we can subtract any non-zero constant terms from the constraint. For example, the constraint $(x_1 - 0.3)^2 - 2x_2 = 3$ is equivalent to $(x_1 - 0.3)^2 - 2x_2 - 3 = 0$.

We will assume that the solution $\mathbf{x}^*$ is a regular point of $\mathbf{h}$.

Let $\mathcal{H}$ denote the $\mathbf{0}$-level set of $\mathbf{h}$,

$$\mathcal{H} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{h}(\mathbf{x}) = \mathbf{0}\}.$$

If $\mathbf{x}^*$ is not a regular point of f, then this implies $\nabla f(\mathbf{x}^*) = 0$. Otherwise, if $\mathbf{x}^*$ is a regular point of f, then with

$$\mathcal{F} = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) = f(\mathbf{x}^*)\},$$

the normal space $N\mathcal{F}(\mathbf{x}^*)$ is a 1-dimensional subspace of $\mathbb{R}^n$, and the normal space $N\mathcal{H}(\mathbf{x}^*)$ is an m-dimensional subspace of $\mathbb{R}^n$. The fundamental theorem of linear algebra then implies that the tangent space $T\mathcal{F}(\mathbf{x}^*)$ is $(n-1)$-dimensional and that $T\mathcal{H}(\mathbf{x}^*)$ is $(n-m)$-dimensional.

In the example given in Section 5.1, the key idea was to make the tangent spaces coincide. In general, we want them parallel, requiring $T\mathcal{H}(\mathbf{x}^*) \subset T\mathcal{F}(\mathbf{x}^*)$ instead. Equivalently, we require $N\mathcal{F}(\mathbf{x}^*) \subset N\mathcal{H}(\mathbf{x}^*)$. This will hold true if and only if

$$\nabla f(\mathbf{x}^*) \in N\mathcal{H}(\mathbf{x}^*).$$

As $N\mathcal{H}(\mathbf{x}^*) = \operatorname{Im}(Dh^T(\mathbf{x}^*))$, this amounts to finding a vector $\boldsymbol{\lambda}$ such that

$$\nabla f(\mathbf{x}^*) = Dh^T(\mathbf{x}^*)\boldsymbol{\lambda},$$

and by taking the transpose, we get

$$Df(\mathbf{x}^*) = \boldsymbol{\lambda}^T Dh(\mathbf{x}^*).$$

By observing that the sign of $\boldsymbol{\lambda}$ is irrelevant, this results in the next theorem, which provides a first order necessary condition. A more rigorous proof is given at the end of this section.

Theorem 13 (Lagrange's multiplier theorem)
Let $\mathbf{h}\colon \mathbb{R}^n \to \mathbb{R}^m$ with $m < n$, and let $\mathbf{x}^*$ be a regular point of $\mathbf{h}$. If $\mathbf{x}^*$ is a local extremiser of $f\colon \mathbb{R}^n \to \mathbb{R}$ subject to $\mathbf{h}(\mathbf{x}) = \mathbf{0}$, then there exists $\boldsymbol{\lambda} \in \mathbb{R}^m$ such that

$$Df(\mathbf{x}^*) + \boldsymbol{\lambda}^T D\mathbf{h}(\mathbf{x}^*) = \mathbf{0}^T. \tag{5.1}$$

We isolate the case $n = 2$, $m = 1$ for illustrative purposes.

Corollary 14 (Lagrange's multiplier theorem with $n = 2$ and $m = 1$)
Let $\mathbf{h}\colon \mathbb{R}^2 \to \mathbb{R}$ and let $\mathbf{x}^* \in \mathbb{R}^2$ with $\nabla h(\mathbf{x}^*) \neq \mathbf{0}$. If $\mathbf{x}^*$ is a local extremiser of $f\colon \mathbb{R}^2 \to \mathbb{R}$ subject to $h(\mathbf{x}) = 0$, then there exists a real number λ such that

$$\nabla f(\mathbf{x}^*) + \lambda \nabla h(\mathbf{x}^*) = \mathbf{0}.$$

Example: Lagrange's multiplier theorem

Consider the problem of Section 5.1,

$$\begin{aligned} \text{minimise} \quad & x_1^2 + x_2^2 \\ \text{subject to} \quad & (x_1 - 0.3)^2 - 2x_2 = 3. \end{aligned} \tag{5.2}$$

To express the problem in the right form, we take

$$\begin{aligned} f(\mathbf{x}) &= x_1^2 + x_2^2, \\ h(\mathbf{x}) &= (x_1 - 0.3)^2 - 2x_2 - 3. \end{aligned}$$

Then,

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}, \qquad \nabla h(\mathbf{x}) = \begin{pmatrix} 2(x_1 - 0.3) \\ -2 \end{pmatrix}.$$

The second component of $\nabla h(\mathbf{x})$ is -2, so $\nabla h(\mathbf{x}) \neq \mathbf{0}$, for all $\mathbf{x}$. Thus, if $\mathbf{x}^*$ is a local extremiser of (5.2), then there exists $\lambda \in \mathbb{R}$ such that

$$\begin{aligned} \nabla f(\mathbf{x}) + \lambda \nabla h(\mathbf{x}^*) = \mathbf{0} \iff & \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} + \lambda \begin{pmatrix} 2(x_1 - 0.3) \\ -2 \end{pmatrix} = \mathbf{0} \\ \iff & \begin{pmatrix} 2x_1 + 2\lambda(x_1 - 0.3) \\ 2x_2 - 2\lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \end{aligned}$$

This results in a pair of equations

$$2x_1 + 2\lambda(x_1 - 0.3) = 0, \tag{1}$$
$$2x_2 - 2\lambda = 0. \tag{2}$$

Equation (2) implies $\lambda = x_2$, and substituting this into equation (1) yields

$$2x_1 + 2x_2(x_1 - 0.3) = 0,$$

and the constraint $h(\mathbf{x}) = 0$ implies

$$x_2 = \frac{(x_1 - 0.3)^2 - 3}{2}.$$

Substituting the latter into the former then gives

$$\begin{aligned} & 2x_1 + 2 \cdot \frac{(x_1-0.3)^2-3}{2} \cdot (x_1 - 0.3) = 0 \\ \iff\ & 2x_1 + (x_1-0.3)^3 - 3(x_1-0.3) = 0. \end{aligned}$$

which is exactly the same equation we arrived at in Section 5.1. Using MATLAB to solve the equation numerically yields the optimiser. However, it remains to be seen how to verify whether it is a minimiser or a maximiser.

The function $\mathcal{L}\colon \mathbb{R}^n \to \mathbb{R}$ (or $\mathcal{L}\colon \mathbb{R}^{n+m} \to \mathbb{R}$, depending on whether you count the $\boldsymbol{\lambda} \in \mathbb{R}^m$ as a parameter or as a variable), defined by

$$\mathcal{L}(\mathbf{x};\boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T\mathbf{h}(\mathbf{x}) = f(\mathbf{x}) + \lambda_1 h_1(\mathbf{x}) + \ldots + \lambda_m h_m(\mathbf{x}),$$

is called the **Lagrangian** function of f and $\mathbf{h}$. Note that the condition of Theorem 13 expresses the vanishing of the derivative of the Lagrangian at $\mathbf{x}^*$. The derivative of the Lagrangian with respect to $\boldsymbol{\lambda}$ yields the constraint $\mathbf{h}(\mathbf{x}) = \mathbf{0}$.

Example: the Lagrangian function

For the problem

$$\begin{aligned} \text{minimise} \quad & x_1^2 + x_2^2 \\ \text{subject to} \quad & (x_1-0.3)^2 - 2x_2 = 3, \end{aligned}$$

the Lagrangian is

$$\mathcal{L}(\mathbf{x};\lambda) = x_1^2 + x_2^2 + \lambda((x_1-0.3)^2 - 2x_2 - 3),$$

so that

$$D\mathcal{L}(\mathbf{x};\lambda) = \begin{pmatrix} 2x_1 + 2\lambda(x_1-0.3) & 2x_2 - 2\lambda & (x_1-0.3)^2 - 2x_2 - 3 \end{pmatrix}.$$

Note that the third column is the derivative with respect to λ, which corresponds to the constraint $(x_1-0.3)^2 - 2x_2 = 3$.

By Lagrange's multiplier theorem, we must have $D\mathcal{L}(\mathbf{x};\lambda) = \mathbf{0}^T$, which results in the three equations

$$2x_1 + 2\lambda(x_1-0.3) = 0 \tag{1}$$
$$2x_2 - 2\lambda = 0 \tag{2}$$
$$(x_1-0.3)^2 - 2x_2 - 3 = 0 \tag{3}$$

Following exactly the same steps as the previous example will then find the optimiser.

Second order necessary and sufficient conditions are given below without proof. Recall that for an $n \times n$ matrix $\mathbf{A}$ and a set $S \subseteq \mathbb{R}^n$, the matrix $\mathbf{A}$ is:

- positive definite over S if $\mathbf{v}^T\mathbf{A}\mathbf{v} > 0$, for all $\mathbf{v} \in S$ with $\mathbf{v} \neq \mathbf{0}$,
- negative definite over S if $\mathbf{v}^T\mathbf{A}\mathbf{v} < 0$, for all $\mathbf{v} \in S$ with $\mathbf{v} \neq \mathbf{0}$,
- positive semi-definite over S if $\mathbf{v}^T\mathbf{A}\mathbf{v} \geqslant 0$, for all $\mathbf{v} \in S$,

- negative semi-definite over S if $\mathbf{v}^T\mathbf{A}\mathbf{v} \leqslant 0$, for all $\mathbf{v} \in S$.

In the results below, the derivatives in the Hessian $D^2\mathcal{L}(\mathbf{x}^*;\boldsymbol{\lambda})$ are with respect to $\mathbf{x}$ only.

Theorem 15 (SONC with equality constraints)

Let $\mathbf{h}\colon \mathbb{R}^n \to \mathbb{R}^m$ with $m < n$, let $f\colon \mathbb{R}^n \to \mathbb{R}$, and let $\mathcal{L}$ be the Lagrangian of f and $\mathbf{h}$. If $\mathbf{x}^*$ is a regular point of $\mathbf{h}$ and $\mathbf{x}^*$ is a local minimiser of $f(\mathbf{x})$ subject to $\mathbf{h}(\mathbf{x}) = \mathbf{0}$, then there exists $\boldsymbol{\lambda} \in \mathbb{R}^m$ such that $D\mathcal{L}(\mathbf{x}^*;\boldsymbol{\lambda}) = \mathbf{0}^T$ and $D^2\mathcal{L}(\mathbf{x}^*;\boldsymbol{\lambda})$ is positive semi-definite on $T\mathcal{H}(\mathbf{x}^*)$.

By symmetry, if $\mathbf{x}^*$ is a local minimiser, then $D^2\mathcal{L}(\mathbf{x}^*;\boldsymbol{\lambda})$ is negative semi-definite instead.

Theorem 16 (SOSC with equality constraints)

Let $\mathbf{h}\colon \mathbb{R}^n \to \mathbb{R}^m$ with $m < n$, let $f\colon \mathbb{R}^n \to \mathbb{R}$, and let $\mathcal{L}$ be the Lagrangian of f and $\mathbf{h}$. If there exists $\boldsymbol{\lambda} \in \mathbb{R}^m$ such that $D\mathcal{L}(\mathbf{x}^*;\boldsymbol{\lambda}) = \mathbf{0}^T$ and $D^2\mathcal{L}(\mathbf{x}^*;\boldsymbol{\lambda})$ is positive definite on $T\mathcal{H}(\mathbf{x}^*)$, then $\mathbf{x}^*$ is a strict local minimiser of $f(\mathbf{x})$ subject to $\mathbf{h}(\mathbf{x}) = 0$.

By symmetry, replacing positive definite with negative definite will result in a strict local maximiser.

Example: classification using the SOSC with equality constraints

For the problem

$$\begin{aligned} \text{minimise} \quad & x_1^2 + x_2^2 \\ \text{subject to} \quad & (x_1 - 0.3)^2 - 2x_2 = 3, \end{aligned}$$

we found the optimiser $\mathbf{x}^* = (x_1^*, x_2^*) \approx (-0.921, -0.754)$.

The Lagrangian is

$$\mathcal{L}(\mathbf{x};\lambda) = x_1^2 + x_2^2 + \lambda((x_1 - 0.3)^2 - 2x_2 - 3).$$

With respect to $\mathbf{x}$ only, we have

$$D\mathcal{L}(\mathbf{x};\lambda) = \begin{pmatrix} 2x_1 + 2\lambda(x_1 - 0.3) & 2x_2 - 2\lambda \end{pmatrix},$$

and then

$$D^2\mathcal{L}(\mathbf{x};\lambda) = \begin{pmatrix} 2 + 2\lambda & 0 \\ 0 & 2 \end{pmatrix}.$$

For $\mathbf{v} = \begin{pmatrix} v_1 & v_2 \end{pmatrix} \in \mathbb{R}^2$, we have

$$\mathbf{v}^T \begin{pmatrix} 2 + 2\lambda & 0 \\ 0 & 2 \end{pmatrix} \mathbf{v} = \begin{pmatrix} v_1 & v_2 \end{pmatrix} \begin{pmatrix} 2 + 2\lambda & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = (2 + 2\lambda)v_1^2 + 2v_2^2.$$

Earlier we found that $\lambda = x_2^*$, so $x_2^* > -1$ implies $2+2\lambda > 0$. Thus, if $\mathbf{v} \neq \mathbf{0}$, then $(2 + 2\lambda)v_1^2 + 2v_2^2 > 0$. Hence $D^2\mathcal{L}(\mathbf{x},\lambda)$ is positive definite (so we do not need to check it on the tangent space) and $\mathbf{x}^*$ is a local minimiser.

Example: applying both first and second-order conditions

Here we will find and classify all extremisers of $f(\mathbf{x}) = 2x_1x_2+2x_1x_3+x_2x_3$ subject to $x_1+x_2+x_3 = 28$. Let $h(\mathbf{x}) = x_1 + x_2 + x_3 - 4$. The Lagrangian of f and h is

$$\mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda h(\mathbf{x}) = 2x_1x_2 + 2x_1x_3 + x_2x_3 + \lambda(x_1 + x_2 + x_3 - 28),$$

which has Jacobian

$$D\mathcal{L}(\mathbf{x}, \lambda) = \begin{pmatrix} 2x_2 + 2x_3 + \lambda & 2x_1 + x_3 + \lambda & 2x_1 + x_2 + \lambda & x_1 + x_2 + x_3 - 28 \end{pmatrix}.$$

Note that $\nabla h(\mathbf{x}) = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}^T \neq \mathbf{0}$, so all points are regular.

If $\mathbf{x}^* = (x_1, x_2, x_3)$ satisfies Lagrange's multiplier theorem, then we must have $D\mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}) = \mathbf{0}^T$; this results in the following system of equations:

$$\begin{aligned} 2x_2 + 2x_3 + \lambda &= 0 && (1)\\ 2x_1 + x_3 + \lambda &= 0 && (2)\\ 2x_1 + x_2 + \lambda &= 0 && (3)\\ x_1 + x_2 + x_3 - 28 &= 0. && (4) \end{aligned}$$

Equations (2) and (3) give

$$2x_1 + x_3 + \lambda = 2x_1 + x_2 + \lambda \implies x_2 = x_3,$$

and then equation (1) results in $4x_2 + \lambda = 0 \implies \lambda = -4x_2$. Substituting this into equation (3) yields

$$2x_1 + x_2 - 4x_2 = 0 \implies x_2 = \tfrac{2}{3}x_1,$$

and then equation (4) gives

$$x_1 + \tfrac{2}{3}x_1 + \tfrac{2}{3}x_1 - 28 = 0 \implies \tfrac{7}{3}x_1 = 28 \implies x_1 = 12.$$

So the only possible optimiser is $\mathbf{x}^* = (12, 8, 8)$.

To apply the SOSC, we need the Hessian of $\mathcal{L}$ with respect to $\mathbf{x}$ only,

$$D^2\mathcal{L}(\mathbf{x}, \lambda) = \begin{pmatrix} 0 & 2 & 2 \\ 2 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix}.$$

This matrix is itself indefinite (as the first leading principal minor is zero), but this need not be the case on the tangent space $T\mathcal{H}(\mathbf{x}^*)$, where $\mathcal{H} = \{\mathbf{x} \in \mathbb{R}^3 : h(\mathbf{x}) = \mathbf{0}\}$. We have

$$T\mathcal{H}(\mathbf{x}^*) = \operatorname{Ker}(Dh(\mathbf{x}^*)) = \operatorname{Ker}(\begin{pmatrix} 1 & 1 & 1 \end{pmatrix}^T) = \{\mathbf{x} \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 0\},$$

so for $\mathbf{v} = (v_1, v_2, v_3) \in T\mathcal{H}(\mathbf{x}^*)$ with $\mathbf{v} \neq \mathbf{0}$, using $v_1 + v_2 + v_3 = 0$, we have

$$\begin{aligned} \mathbf{v}^T D^2\mathcal{L}(\mathbf{x}, \lambda)\mathbf{v} &= \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} \begin{pmatrix} 0 & 2 & 2 \\ 2 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \\ &= v_1(2v_2 + 2v_3) + v_2(2v_1 + v_3) + v_3(2v_1 + v_2) \\ &= 2v_1(v_2 + v_3) + v_1v_2 + v_2(v_1 + v_3) + v_1v_3 + v_3(v_1 + v_2) \\ &= 2v_1(-v_1) + v_2(-v_2) + v_3(-v_3) + v_1(v_2 + v_3) \\ &= -2v_1^2 - v_2^2 - v_3^2 - v_1^2 < 0. \end{aligned}$$

So $D^2\mathcal{L}(\mathbf{x}, \lambda)$ is negative definite on $T\mathcal{H}(\mathbf{x}^*)$, which implies $\mathbf{x}^*$ is a local maximiser.

Example: an example with multiple constraints

Consider the following problem:

$$\begin{aligned} \text{minimise} \quad & x_1 + x_2 + x_3 \\ \text{subject to} \quad & x_1^2 + x_2^2 + x_3^2 = 4, \\ & x_1 - x_2 - x_3 = 0. \end{aligned}$$

We take

$$f(\mathbf{x}) = 3x_1 + 2x_2 + x_3, \quad \mathbf{h}(\mathbf{x}) = \begin{pmatrix} x_1^2 + x_2^2 + x_3^2 - 4 \\ x_1 - x_2 - x_3 \end{pmatrix}.$$

Let $\boldsymbol{\lambda} = \begin{pmatrix} \lambda_1 & \lambda_2 \end{pmatrix}^T$. The Lagrangian is

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) &= f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}) \\ &= x_1 + x_2 + x_3 + \begin{pmatrix} \lambda_1 & \lambda_2 \end{pmatrix} \begin{pmatrix} x_1^2 + x_2^2 + x_3^2 - 4 \\ x_1 - x_2 - x_3 \end{pmatrix} \\ &= x_1 + x_2 + x_3 + \lambda_1(x_1^2 + x_2^2 + x_3^2 - 4) + \lambda_2(x_1 - x_2 - x_3). \end{aligned}$$

If $\mathbf{x}^* = (x_1, x_2, x_3)$ satisfies Lagrange's multiplier theorem, then we must have $D\mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}) = \mathbf{0}^T$; this results in the following system of equations:

$$\begin{aligned} 1 + 2\lambda_1 x_1 + \lambda_2 &= 0, && (1) \\ 1 + 2\lambda_1 x_2 - \lambda_2 &= 0, && (2) \\ 1 + 2\lambda_1 x_3 - \lambda_2 &= 0, && (3) \\ x_1^2 + x_2^2 + x_3^2 - 4 &= 0, && (4) \\ x_1 - x_2 - x_3 &= 0. && (5) \end{aligned}$$

From equation (1):

$$1 + 2\lambda_1 x_1 + \lambda_2 = 0 \implies -\lambda_2 = 1 + 2\lambda_1 x_1.$$

Substituting this into equations (2) and (3) gives

$$\begin{aligned} 1 + 2\lambda_1 x_2 - \lambda_2 = 0 &\implies 1 + 2\lambda_1 x_2 + 1 + 2\lambda_1 x_1 = 0 \\ &\implies 2 + 2\lambda_1(x_1 + x_2) = 0, && (6) \\ 1 + 2\lambda_1 x_3 - \lambda_2 = 0 &\implies 1 + 2\lambda_1 x_3 + 1 + 2\lambda_1 x_1 = 0 \\ &\implies 2 + 2\lambda_1(x_1 + x_3) = 0. && (7) \end{aligned}$$

From (6) and (7) we have $x_1 + x_2 = x_1 + x_3 \implies x_2 = x_3$. Then (5) yields $x_1 - 2x_2 = 0 \implies x_1 = 2x_2$. Substituting these into (4) gives

$$4x_2^2 + x_2^2 + x_2^2 - 4 = 0 \iff 6x_2^2 = 4 \iff x_2 = \pm\sqrt{\tfrac{2}{3}},$$

resulting in two candidates for minimisers:

$$\mathbf{p}_1 = \left(2\sqrt{\tfrac{2}{3}}, \sqrt{\tfrac{2}{3}}, \sqrt{\tfrac{2}{3}}\right) \text{ and } \mathbf{p}_2 = \left(-2\sqrt{\tfrac{2}{3}}, -\sqrt{\tfrac{2}{3}}, -\sqrt{\tfrac{2}{3}}\right).$$

With respect to $\mathbf{x}$ only, the Hessian of $\mathcal{L}$ is

$$D^2\mathcal{L}(\mathbf{x};\boldsymbol{\lambda}) = \begin{pmatrix} 2\lambda_1 & 0 & 0 \\ 0 & 2\lambda_1 & 0 \\ 0 & 0 & 2\lambda_1 \end{pmatrix}.$$

This is positive definite when $\lambda_1 > 0$ and negative definite when $\lambda_1 < 0$. To find λ_1, note that equation (6) implies

$$2 + 6\lambda_1 x_2 = 0 \implies \lambda_1 = -\frac{1}{3x_2}.$$

So the sign of λ_1 is opposite that of x_2.

Hence, $D^2\mathcal{L}(\mathbf{x};\boldsymbol{\lambda})$ is positive definite when x_2 is negative, and it is negative definite when x_2 is positive. We conclude that only $\left(2\sqrt{\frac{2}{3}}, \sqrt{\frac{2}{3}}, \sqrt{\frac{2}{3}}\right)$ is a local minimiser.

This proof of Lagrange's multiplier theorem is not examinable, but is included for your own reference.

Proof of Lagrange's multiplier theorem. After taking the transpose we obtain

$$\nabla f(\mathbf{x}^*) = -D\mathbf{h}^T(\mathbf{x}^*)\lambda.$$

If we can show that $\nabla f(\mathbf{x}^*) \in \operatorname{Im}(D\mathbf{h}^T(\mathbf{x}^*))$ then $\boldsymbol{\lambda}$ exists. As $\operatorname{Im}(D\mathbf{h}^T(\mathbf{x}^*)) = \operatorname{Ker}(D\mathbf{h}(\mathbf{x}^*))^\perp$ it suffices to show that $\nabla f(\mathbf{x}^*)$ is orthogonal to every vector in $\operatorname{Ker}(D\mathbf{h}(\mathbf{x}^*))$. Let $\mathbf{v} \in \operatorname{Ker}(D\mathbf{h}(\mathbf{x}^*))$ be such a vector. By the implicit function theorem, there exist a differentiable curve $\mathbf{x} : \mathbb{R} \to \mathcal{H}$, such that $\mathbf{h}(\mathbf{x}(t)) = \mathbf{c}$ and a point $t^* \in \mathbb{R}$ such that

$$\mathbf{x}^* = \mathbf{x}(t^*), \qquad D\mathbf{x}(t^*) = \mathbf{v}.$$

The point t^* should be a local extremiser of the composite function $\phi(t) = f(\mathbf{x}(t))$. By the FONC, we need $\frac{\mathrm{d}\phi}{\mathrm{d}t}(t^*) = 0$. Applying the chain rule gives

$$\frac{\mathrm{d}\phi}{\mathrm{d}t}(t^*) = D(f \circ \mathbf{x})(t^*) = Df(\mathbf{x}(t^*))D\mathbf{x}(t^*) = \nabla f^T(\mathbf{x}^*)\mathbf{v} = 0,$$

which shows $\nabla f^T(\mathbf{x}^*)$ is orthogonal to $\mathbf{v}$. □