4. Optimization on Manifolds

In many situations one wants to find the (local or global) maxima or minima of a function $f : \mathbb{R}^n \to \mathbb{R}$ subject to constraints $g_1(x) = 0, \ldots, g_m(x) = 0$. Suppose that f takes such a local maximum or minimum at a point $p \in \mathbb{R}^n$ (which, of course, satisfies the constraints). This means that there is a neighborhood $U \subseteq \mathbb{R}^n$ of p such that $f(p) \geq f(x)$ (in the case of a maximum) or $f(p) \leq f(x)$ (in the case of a minimum) for all x in

$$(\star) \qquad \{x \in U \mid g_1(x) = 0, \ldots, g_m(x) = 0\}.$$

Now if $g_1'(p), \ldots, g_m'(p)$ are linearly independent then, by continuity, we can choose U so small that $g_1'(x), \ldots, g_m'(x)$ are linearly independent at each point $x \in U$, which implies that the set $(\star)$ is a manifold. Hence, after replacing U by an even smaller neighborhood if necessary, we can find a local parametrization $\varphi : \Omega \to U$ where $\Omega \subseteq \mathbb{R}^{n-m}$ is an open set. If $p = \varphi(\xi_0)$, then f takes a local extremum at p under the constraints $g_i = 0$ if and only if the function $f \circ \varphi : \Omega \to \mathbb{R}$ takes a local extremum at ξ_0. Thus, by using a parametrization, we can reduce a constrained optimization problem to an unconstrained optimization problem. However, sometimes a parametrization is hard to find or cumbersome to use, and we can ask whether or not it is possible to detect extrema under constraints directly from the constraints, without invoking a parametrization. This is indeed the case. A necessary condition is provided by the following result.

(4.1) Lagrange's Theorem. *Let $U \subseteq \mathbb{R}^n$ be an open set and let $f, g_1, \ldots, g_m : U \to \mathbb{R}$ be C^1-functions. Assume that f takes a maximum or minimum under the constraints $g_1(x) = \cdots = g_m(x) = 0$ at some point $p \in U$. If $(\nabla g_1)(p), \ldots, (\nabla g_m)(p)$ are linearly independent, then there are numbers $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ (called* **Lagrange multipliers***) such that*

$$(\nabla f)(p) \;=\; \lambda_1 (\nabla g_1)(p) + \cdots + \lambda_m (\nabla g_m)(p).$$

Proof. Interpret $g_1, \ldots, g_m$ as the coordinate functions of a mapping $g : U \to \mathbb{R}^m$. By assumption, the linearization $g'(p)$ has the rank m; by continuity, the rank of $g'(x)$ is m for all x sufficiently close to p. Hence (after replacing U by a smaller neighborhood of p if necessary) the set $M := \{x \in U \mid g(x) = 0\}$ is a manifold of dimension $d := n - m$. Let $\varphi : \Omega \to \mathbb{R}^n$ be a parametrization of M around p (with an open subset $\Omega \subseteq \mathbb{R}^d$), say $p = \varphi(\xi_0)$. Then for each C^1-curve $\xi : (-\varepsilon, \varepsilon) \to \Omega$ with $\xi(0) = \xi_0$, the curve $\alpha : t \mapsto \varphi\big(\xi(t)\big)$ is a curve in M with $\alpha(0) = p$. By assumption, the function $t \mapsto f\big(\alpha(t)\big)$ takes a maximum or minimum at $t = 0$. Consequently,

$$0 \;=\; \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} f\big(\alpha(t)\big) \;=\; f'\big(\alpha(0)\big)\alpha'(0) \;=\; f'(p)\varphi'(\xi_0)\xi'(0).$$

Since $\xi'(0)$ can be chosen to be an arbitrary vector $w \in \mathbb{R}^d$ (for example, we can choose $\xi(t) := \xi_0 + tw$), this means that $0 = f'(p)\varphi'(\xi_0)w = \langle (\nabla f)(p), \varphi'(\xi_0)w \rangle$ for all $w \in \mathbb{R}^d$ so that

$$\begin{aligned}(\nabla f)(p) \;&\in\; \big(\operatorname{im} \varphi'(\xi_0)\big)^{\perp} \;=\; (T_pM)^{\perp} \;=\; \big(\ker g'(p)\big)^{\perp} \\ &=\; \operatorname{im} g'(p)^T \;=\; \mathbb{R}(\nabla g_1)(p) + \cdots + \mathbb{R}(\nabla g_m)(p),\end{aligned}$$

which is the claim. (The last equation holds because the image of $g'(p)^T$ is the vector subspace of $\mathbb{R}^n$ spanned by the columns of $g'(p)^T$, which are exactly the gradients of the functions g_i at p.) ■

(4.2) Example. We want to find the local minima and maxima of the function $f(x,y) := x^4 - 2x^2 + y^2$ on the circle around $(0,0)$ with radius $r > 0$. This circle is parametrized by $\varphi(t) := (r\cos t, r\sin t)$ where $0 \le t \le 2\pi$; hence we are looking for the minima and maxima of the function $g : [0, 2\pi] \to \mathbb{R}$ defined by $g(t) := f\big(\varphi(t)\big) = f(r\cos t, r\sin t)$, i.e.,

$$g(t) \;=\; r^4\cos^4 t - 2r^2\cos^2 t + r^2\sin^2 t \;=\; \left(r^2\cos^2 t - \frac{3}{2}\right)^2 + r^2 - \frac{9}{4}.$$

It is clear that g can take a local extremum only for those t for which $\cos^2(t)$ equals 1, 0 oder $3/(2r^2)$ (where the last case is possible only if $r^2 \ge 3/2$). Since $(x,y) = (r\cos t, r\sin t)$, these three possible cases correspond to the points

$$(\star) \qquad P_{1,2} = (\pm r,\, 0), \quad P_{3,4} = (0,\, \pm r), \quad P_{5,6,7,8} = (\pm 3/2,\, \pm\sqrt{r^2 - (3/2)});$$

these are the only candidates for points at which a local minimum or maximum can occur. The same conclusion can be reached using Lagrange's Theorem. Letting $g(x,y) := x^2 + y^2 - r^2$, this theorem tells us to look for solutions of the system $(\nabla f)(x,y) = \lambda\,(\nabla g)(x,y)$, which reads

$$\begin{bmatrix} 4x^3 - 4x \\ 2y \end{bmatrix} = 2\lambda \begin{bmatrix} x \\ y \end{bmatrix}.$$

Eliminating λ, we find that $(4x^3 - 4x)y = 2xy$, i.e., $4x^3y = 6xy$ and hence $x = 0$, $y = 0$ or $x^2 = 3/2$. Using the condition $x^2 + y^2 = r^2$, these three possibilities give rise to the possible solution candidates $(\star)$.

It remains to check which of these candidates represent local minima, local maxima, or saddle points. Here we can argue as follows: Since the circle $\{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = r^2\}$ is a compact set and since f is continuous, we know *a priori* that f attains a global minimum amd a glocal maximum. We now simply calculate the values

$$f(\pm r,\, 0) = r^4 - 2r^2, \quad f(0,\, \pm r) = r^2, \quad f\big(\pm 3/2,\, \pm\sqrt{r^2 - (3/2)}\big) = r^2 - (9/4),$$

and distinguish the various possible cases.

- **First case:** $r \le \sqrt{3/2}$. In this case we have $r^4 - 2r^2 < r^2$ so that f takes its (global) minimum $r^4 - 2r^2$ at P_1 and P_2 and its global maximum r^2 at P_3 and P_4.
- **Second case:** $\sqrt{3/2} < r < \sqrt{3}$. In this case we have $r^2 - (9/4) < r^4 - 2r^2 < r^2$ so that f takes its global minimum $r^2 - (9/4)$ at each of the four points P_5, P_6, P_7, P_8 and its global maximum r^2 at P_3 and P_4. Since each of the points P_1 and P_2 is the only critical point between two points at which a global minimum is taken, we conclude that f must take a local maximum at P_1 and at P_2.

- **Third case:** $r \geq \sqrt{3}$. In this case we have $r^2 - (9/4) < r^2 \leq r^4 - 2r^2$ so that f takes its global minimum $r^2 - (9/4)$ at each of the four points P_5, P_6, P_7, P_8 and its global maximum $r^4 - 2r^2$ at P_1 and P_2. Since each of the points P_3 and P_4 is the only critical point between two points at which a global minimum is taken, we conclude that f must take a local maximum at P_1 and P_2.

Note that this discussion invoked arguments extraneous to the local differential theory which was used to locate the critical points. Namely, we used the fact that a continuous function on a compact set takes a global minimum and a global maximum, and we used the fact that a critical point of a function on a one-dimensional manifold which is located between two adjacent local minima of this function is by necessity a local maximum. (The notion of a point "lying between" two other points is, of course, specific to the one-dimensional situation.) The results obtained are visualized in the following diagrams.

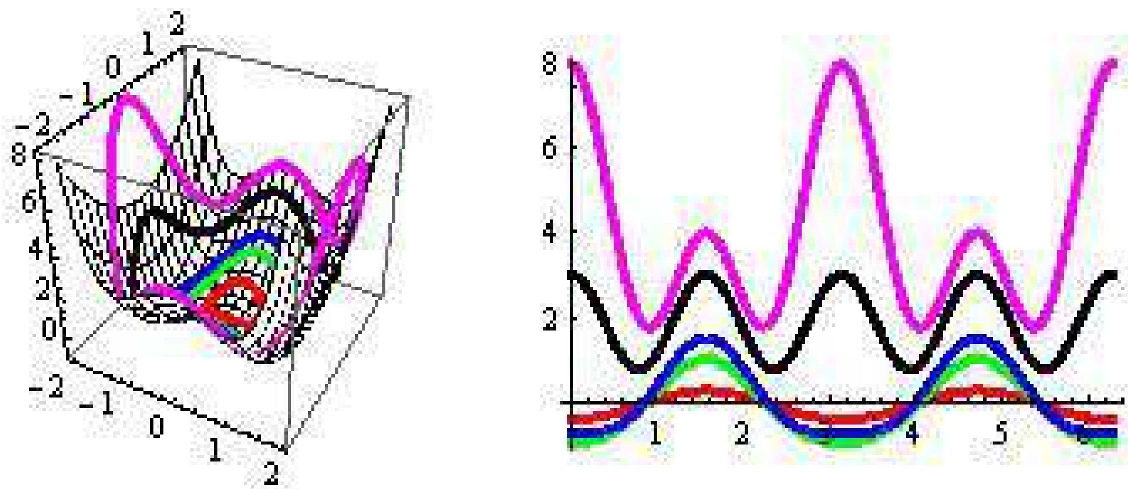

Fig. 4.1: Restriction of the function $f(x, y) = x^4 - 2x^2 + y^2$ to the circles $x^2 + y^2 = r^2$ where $r \in \{1/2,\ 1,\ \sqrt{3/2},\ \sqrt{3},\ 2\}$.

An interpretation of Lagrange's Theorem can be easily given for the case of a function $f : \mathbb{R}^2 \to \mathbb{R}$ to be optimized subject to a constraint $g(x, y) = 0$. We interpret the surface $z = f(x, y)$ as a mountain range and the set $M = \{(x, y) \in \mathbb{R}^2 \mid g(x, y) = 0\}$ as the course of a street through this mountain range along which we want to determine the locally highest and lowest points. Now at a point $p \in \mathbb{R}^2$ at which such a local minimum or maximum occurs, the curve $g(x, y) = 0$ cannot cross the level set $\{(x, y) \in \mathbb{R}^2 \mid f(x, y) = f(p)\}$, because this would mean that the values of f along the curve $g(x, y) = 0$ increase or decrease as we move along this curve starting from p. Thus at a point p at which f takes a local extremum subject to $g = 0$, the curves $f(x, y) = f(p)$ and $g(x, y) = p$ must be tangent to each other, which implies that the gradients $(\nabla f)(p)$ and $(\nabla g)(p)$ are linearly dependent, which, if $(\nabla g)(p) \neq 0$, implies that $(\nabla f)(p) = \lambda\, (\nabla g)(p)$ for some $\lambda \in \mathbb{R}$.

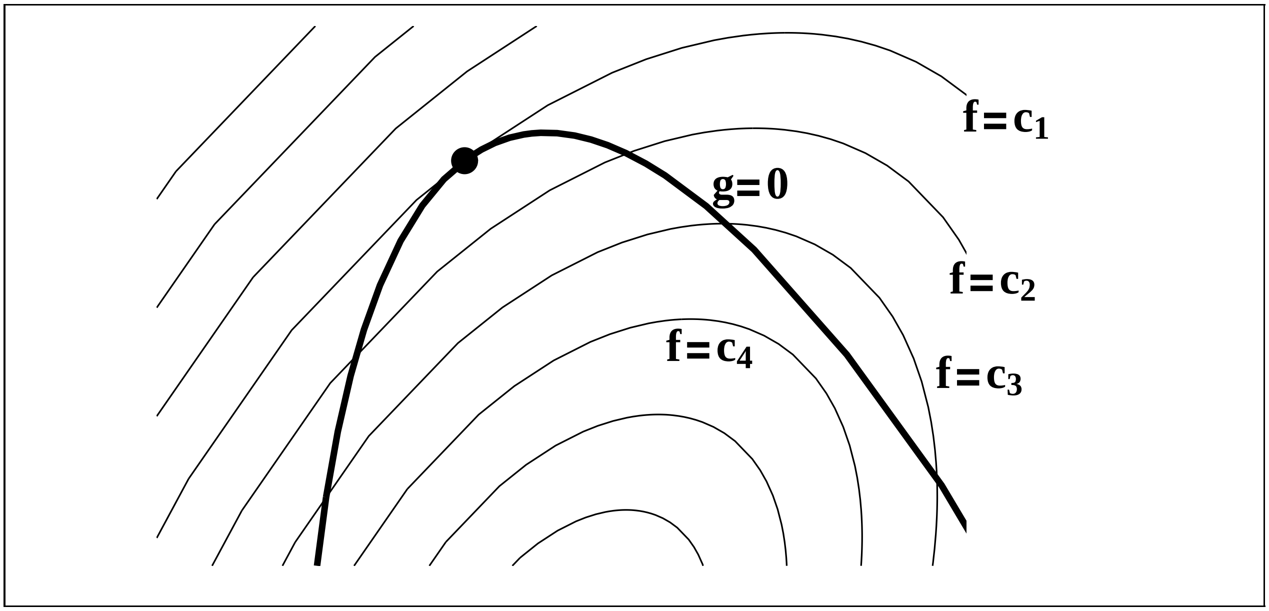

Fig. 4.2: Interpretation of Langrange's Theorem for a function $f : \mathbb{R}^2 \to \mathbb{R}$ under a constraint $g(x, y) = 0$.

In the case at hand it is also clear that Lagrange's Theorem provides only a necessary, not a sufficient for local optimality, because the curves $f(x, y) = p$ and $g(x, y) = 0$ can be tangent to each other at p even if the values of f increase or decrease monotonically as we pass through p while moving along the curve $g(x, y) = 0$.

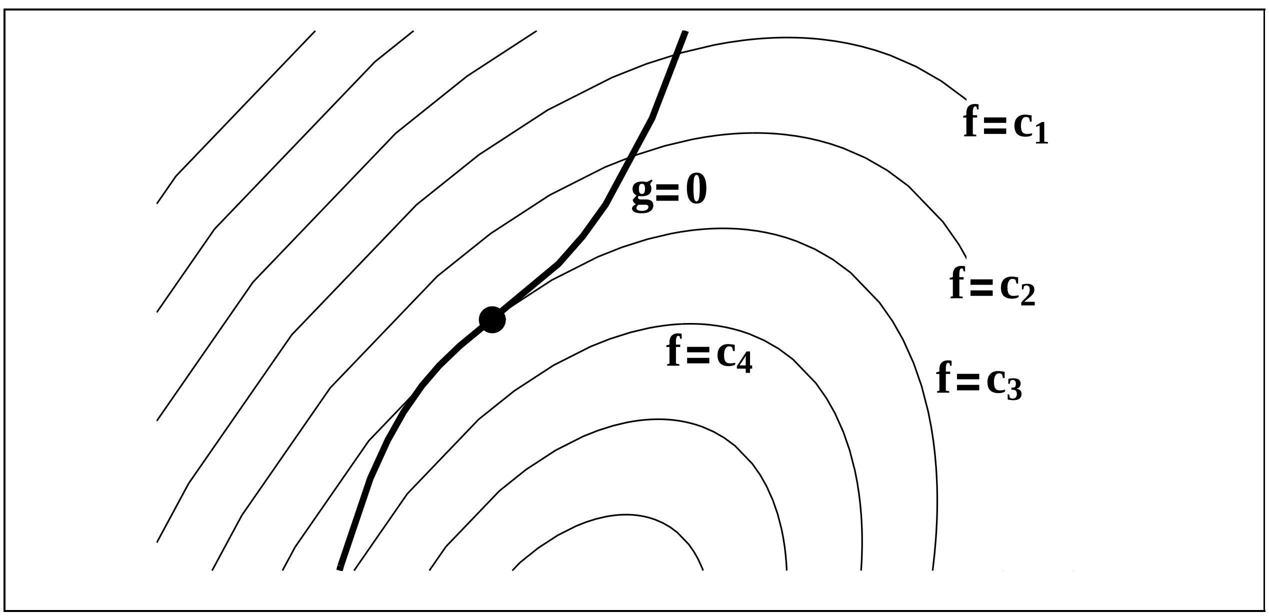

Fig. 4.3: Example which shows that, in general, the existence of Lagrange multipliers is necessary, but not sufficient for the existence of a local optimum.

If $(\nabla g)(p) = 0$, then the statement of Lagrange's Theorem is not generally true: The function f may take a local extremum at p subject to a constraint $g(x, y) = 0$ even though there is no Lagrange multiplier λ such that $(\nabla f)(p) = \lambda\,(\nabla g)(p)$. For a simple example, take $f(x, y) = x$, $g(x, y) = x^3 - y^2$ und $p = (0, 0)$. (See the exercises.)

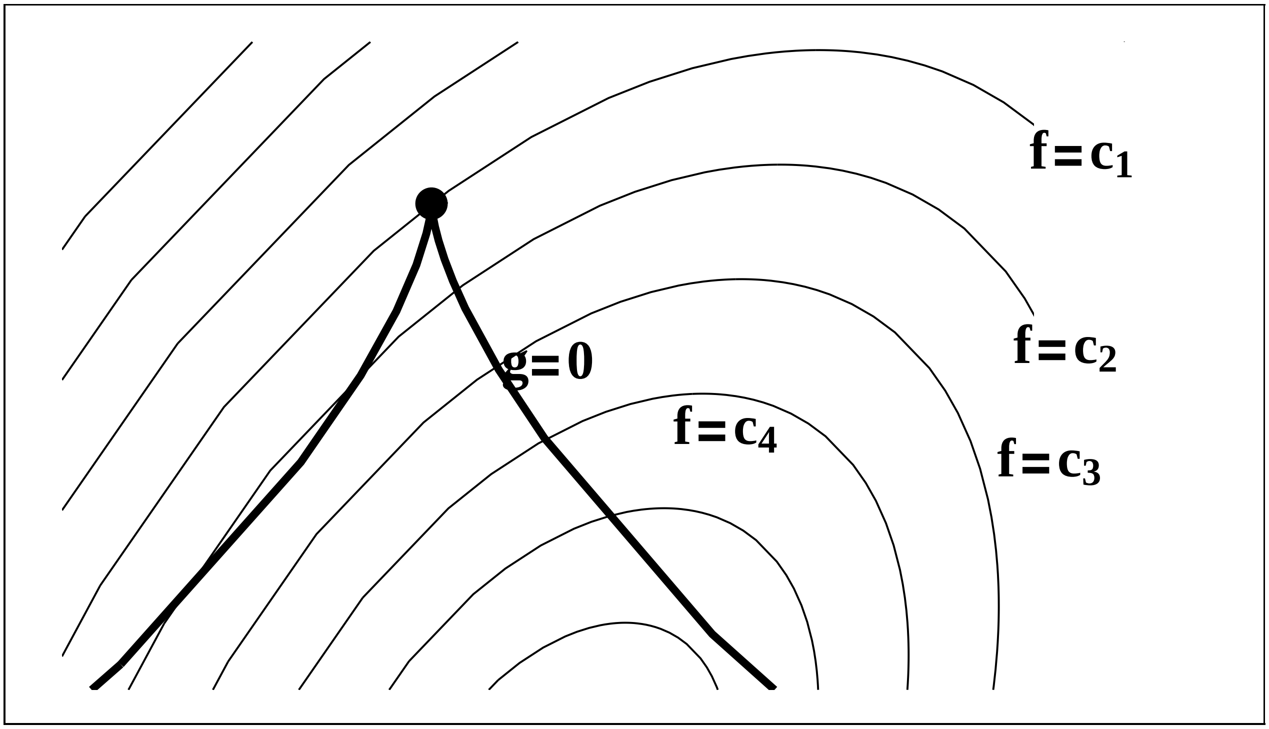

Fig. 4.4: Example which shows that the linear independence of the gradients $(\nabla g_i)(p)$ is an indispensible hypothesis in Lagrange's Theorem.

Note that Lagrange's Theorem can be rephrased as follows: Given the functions $f, g_1, \ldots, g_m$, we can form the **Lagrange function** L defined by

$$L(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m) \ := \ f(x) - \lambda_1 g_1(x) - \cdots - \lambda_m g_m(x)$$

where $x = (x_1, \ldots, x_n)$. If f takes a local extremum at p subject to the constraints $g_i = 0$, then there is a vector $\lambda \in \mathbb{R}^m$ such that (p, λ) is a critical point of L. (In fact, the conditions $(\partial L/\partial x_i)(p) = 0$ where $1 \leq i \leq n$ express the condition $(\nabla f)(p) = \sum_{i=1}^m \lambda_i (\nabla g_i)(p)$ derived in Lagrange's Theorem, and the conditions $0 = (\partial L/\partial \lambda_j)(p) = -g_j(p)$ where $1 \leq j \leq m$ express the fact that p satisfies the constraints $g_j(p) = 0$.) Once the Lagrange multipliers λ_i have been determined, we can consider $L = f - \sum_{i=1}^m \lambda_i g_i$ as a function of the x-variables alone and conclude that a critical point p of f subject to the constraints $g_i = 0$ must be a critical point of the unconstrained function L. This allows for an interesting interpretation: Assume that $x = (x_1, \ldots, x_n)$ is a vector of economic variables which determine the profit $f(x) = f(x_1, \ldots, x_n)$ of a company, and assume that this company wants to maximize its profit. The conditions $g_i(x) = 0$ may describe constraints which are imposed on the company by the political system (such as taxation, environmental or social standards, and so on). Lagrange's Theorem indicates† that if we assign to the cost factors $g_i(x)$ appropriate weights λ_i, then maximizing the profit f subject to the constraints $g_i = 0$

† We write "indicates" rather than "implies" because Lagrange's Theorem only guarantees that p is a critical point of L, not necessarily a maximum, even if p is truly a maximum of f under the given constraints.

is tantamount to maximizing the modified profit $L = f - \sum_i \lambda_i g_i$ in the absence of any constraints. Hence policymakers may try to ensure the standards $g_i = 0$ not by legally enforcing them, but by imposing financial penalties or incentives, which lead the company to automatically satisfy the constraints as a by-product of its profit maximization. Theories along these lines date back to the Scottish economist Adam Smith (1723-1790), who coined the expression of the "invisible hand" of market forces which lead to unintended social benefits resulting from individual actions.

Lagrange's Theorem only yields a necessary crition for p to be a critical point. We now want to derive necessary and sufficient criteria for the restriction $f = F|_M$ to take a local minimum or maximum at a point $p \in M$.

(4.3) Theorem. *Let $M \subseteq \mathbb{R}^n$ be a d-dimensional manifold of class C^2 and let $f : \mathbb{R}^n \to \mathbb{R}$ be a function of class C^2; we want to find the local extrema of the restriction $f|_M$ of f to M. Assume that M is given by a system of $m = n - d$ equations $g_1(x) = \cdots = g_m(x) = 0$ in a neighborhood of $p \in M$ on which $g_1'(x), \ldots, g_m'(x)$ are linearly independent. Assume Lagrange multipliers $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ are given such that $L(x) := f(x) - \lambda_1 g_1(x) - \cdots - \lambda_m g_m(x)$ satisfies $L'(p) = 0$. Let $(v_1, \ldots, v_d)$ be a basis of $T_pM = \big(\mathbb{R}(\nabla g_1)(p) + \cdots + \mathbb{R}(\nabla g_m)(p)\big)^\perp$, and let $A \in \mathbb{R}^{n\times d}$ be the matrix with columns $v_1, \ldots, v_d$. Denoting by $L''(p) \in \mathbb{R}^{n\times n}$ the Hessian of L at p, we form the matrix*

$$B \;:=\; A^T L''(p) A \in \mathbb{R}^{d\times d}\,.$$

Then the following statements hold true.
(a) *If $f|_M$ has a local minimum at p then B is positive semidefinite.*
(b) *If B is positive definite then $f|_M$ has a local minimum at p.*
(c) *If $f|_M$ has a local maximum at p then B is negative semidefinite.*
(d) *If B is negative definite then $f|_M$ has a local maximum at p.*
(e) *If B is indefinite then $f|_M$ has no local extremum at p.*

Proof. Let φ be a parametrization of M in a neighborhood of $\varphi(u^\star) = p$, say

$$\varphi(u_1, \ldots, u_d) \;=\; \begin{bmatrix} x_1(u_1, \ldots, u_d) \\ \vdots \\ x_n(u_1, \ldots, u_d) \end{bmatrix}$$

where $d = \dim(M)$. Then $f|_M$ has a local minimum at p if and only if $f \circ \varphi$ has a local minimum at $u^\star$. Given Lagrange multipliers $\lambda_1, \ldots, \lambda_m$ such that $L := f - \lambda_1 g_1 - \cdots - \lambda_m g_m$ satisfies $L'(p) = 0$, the functions f and L coincide on M; hence $f \circ \varphi$ takes a local minimum at $u^\star$ if and only if $L \circ \varphi$ does. A necessary or, respectively, sufficient criterion for this to happen is that $(L \circ \varphi)''(u^\star)$ is positive semidefinite respectively positive definite. Given any vectors $\xi, \eta \in \mathbb{R}^d$, we want to compute

$$(\star) \qquad \langle \xi, (L \circ \varphi)''(u^\star)\, \eta\rangle \;=\; \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=0} L\big(\varphi(u^\star + s\xi + t\eta)\big).$$

First of all, using the chain rule, we see that $(\mathrm{d}/\mathrm{d}s)L\big(\varphi(u^\star+s\xi+t\eta)\big)$ equals

$$\sum_{i=1}^n\sum_{k=1}^d \frac{\partial L}{\partial x_i}\big(\varphi(u^\star+s\xi+t\eta)\big)\frac{\partial\varphi_i}{\partial u_k}(u^\star+s\xi+t\eta)\,\xi_k\,.$$

Letting $s=0$ und plugging this into $(\star)$ yields

$$\begin{aligned}\langle\xi,(L\circ\varphi)''(u^\star)\,\eta\rangle &= \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} L\big(\varphi(u^\star+s\xi+t\eta)\big)\\ &= \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}\sum_{i=1}^n\sum_{k=1}^d \frac{\partial L}{\partial x_i}\big(\varphi(u^\star+t\eta)\big)\frac{\partial\varphi_i}{\partial u_k}(u^\star+t\eta)\,\xi_k\,.\end{aligned}$$

Using the chain and the product rule, we find that

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial x_i}\big(\varphi(u^\star+t\eta)\big)\frac{\partial\varphi_i}{\partial u_k}(u^\star+t\eta)\xi_k\right) =\\ &\sum_{j=1}^n\sum_{\ell=1}^d \frac{\partial^2 L}{\partial x_i\partial x_j}\big(\varphi(u^\star+t\eta)\big)\frac{\partial\varphi_j}{\partial u_\ell}(u^\star+t\eta)\frac{\partial\varphi_i}{\partial u_k}(u^\star+t\eta)\xi_k\eta_\ell\\ &\qquad+\sum_{\ell=1}^d \frac{\partial L}{\partial x_i}\big(\varphi(u^\star+t\eta)\big)\frac{\partial^2\varphi}{\partial u_k\partial u_\ell}(u^\star+t\eta)\xi_k\eta_\ell\,.\end{aligned}$$

Letting $t=0$ and plugging this into the above equations, we find that

$$\begin{aligned}\langle\xi,(L\circ\varphi)''(u^\star)\,\eta\rangle &= \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} L\big(\varphi(u^\star+s\xi+t\eta)\big)\\ &= \sum_{i,j=1}^n\sum_{k,\ell=1}^d \frac{\partial^2 L}{\partial x_i\partial x_j}\big(\varphi(u^\star)\big)\frac{\partial\varphi_j}{\partial u_\ell}(u^\star)\frac{\partial\varphi_i}{\partial u_k}(u^\star)\xi_k\eta_\ell+\sum_{i=1}^n\sum_{k,\ell=1}^d \frac{\partial L}{\partial x_i}\big(\varphi(u^\star)\big)\frac{\partial^2\varphi}{\partial u_k\partial u_\ell}(u^\star)\xi_k\eta_\ell\\ &= \langle\varphi'(u^\star)\xi, L''(p)\varphi'(u^\star)\eta\rangle + L'(p)\begin{bmatrix}\langle\xi,\varphi_1''(u^\star)\eta\rangle\\ \vdots\\ \langle\xi,\varphi_n''(u^\star)\eta\rangle\end{bmatrix} = \langle\varphi'(u^\star)\xi, L''(p)\varphi'(u^\star)\eta\rangle\,.\end{aligned}$$

(The last equation holds because $L'(p)=0$, and this is exactly the reason why we executed the calculation for L rather than for f! Note that f does *not* satisfy the condition $f'(p)=0$, because $f'(p)$ vanishes on T_pM, but not necessarily on all of $\mathbb{R}^n$.) Hence we obtain the following chain of equivalent conditions:

- $(L\circ\varphi)''(u^\star)$ is positive definite (resp. semidefinite);
- $\langle\varphi'(u^\star)\xi, L''(p)\varphi'(u^\star)\xi\rangle>0$ (resp. ≥ 0) for all $\xi\in\mathbb{R}^d\setminus\{0\}$;
- $\langle v, L''(p)v\rangle>0$ (resp. ≥ 0) for all $v\in(T_pM)\setminus\{0\}$;
- $\langle\sum_i\lambda_i v_i, L''(p)\sum_i\lambda_i v_i\rangle>0$ (resp. ≥ 0) for all $\lambda\in\mathbb{R}^d\setminus\{0\}$;
- $\lambda^T A^T L''(p)A\lambda>0$ (resp. ≥ 0) for all $\lambda\in\mathbb{R}^d\setminus\{0\}$;

- $\langle \lambda, B\lambda \rangle > 0$ (resp. ≥ 0) for all $\lambda \in \mathbb{R}^d \setminus \{0\}$;
- B is positive definite (resp. semidefinite).

This yields the claim for local minima. The claim for local maxima is established in a completely analogous manner (or by applying the claim for local minima to $-f$ instead of f). Finally, if B is indefinite, then B is neither positive semidefinite nor negative semidefinite, so that f can take neither a local minimum nor a local maximum at p, due to parts (a) and (c). ■

(4.4) Example. Let us return to the example given in (4.2), in which we sought the local extrema of $f(x,y) := x^4 - 2x^2 + y^2$ on $M := \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = r^2\}$ for a given number $r > 0$. We try to determine the nature of the critical points found in (4.2) by using Theorem (4.3) above.

• Plugging the point $p = (\pm r, 0)$ into the condition $(\nabla f)(p) = \lambda\,(\nabla g)(p)$ yields $\lambda = 2r^2 - 2$; hence we consider the Lagrange function

$$L(x,y) \;=\; x^4 - 2x^2 + y^2 - (2r^2-2)(x^2+y^2-r^2).$$

It is readily checked that $L''(p) = \mathrm{diag}(8r^2, 6-4r^2) \in \mathbb{R}^{2\times 2}$ and that the tangent space T_pM is spanned by $(0,1)^T$; hence the matrix to be considered is the (1×1)-matrix

$$B \;=\; \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 8r^2 & 0 \\ 0 & 6-4r^2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \;=\; \begin{bmatrix} 6-4r^2 \end{bmatrix}.$$

If $r^2 < 3/2$ this matrix is positive definite, so that f has a local minimum at p. If $r^2 > 3/2$ this matrix is negative definite, so that f has a local maximum at p.

• Plugging the point $p = (0, \pm r)$ into the condition $(\nabla f)(p) = \lambda\,(\nabla g)(p)$ yields $\lambda = 1$; hence we consider the Lagrange function

$$L(x,y) \;=\; x^4 - 2x^2 + y^2 - 1\cdot(x^2+y^2-r^2) \;=\; x^4 - 3x^2 + r^2.$$

It is readily checked that $L''(p) = \mathrm{diag}(-6, 0) \in \mathbb{R}^{2\times 2}$ and that the tangent space T_pM is spanned by $(1,0)^T$; hence the matrix to be considered is the (1×1)-matrix

$$B \;=\; \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} -6 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \;=\; \begin{bmatrix} -6 \end{bmatrix}.$$

This matrix is negative definite, so that f has a local maximum at p.

• Plugging the point $p = (\pm\sqrt{3/2}, \pm\sqrt{r^2-(3/2)})$ (which is a critical point if $r^2 > 3/2$) into the condition $(\nabla f)(p) = \lambda\,(\nabla g)(p)$ yields $\lambda = 1$; hence we consider the Lagrange function

$$L(x,y) \;=\; x^4 - 2x^2 + y^2 - 1\cdot(x^2+y^2-r^2) \;=\; x^4 - 3x^2 + r^2.$$

It is readily checked that $L''(p) = \mathrm{diag}(12, 0) \in \mathbb{R}^{2\times 2}$ and that the tangent space T_pM is spanned by $(\mp\sqrt{r^2-3/2}, \pm\sqrt{3/2})^T$; hence the matrix to be considered is the (1×1)-matrix

$$B \;=\; \begin{bmatrix} \mp\sqrt{r^2-(3/2)} & \pm\sqrt{3/2} \end{bmatrix} \begin{bmatrix} 12 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mp\sqrt{r^2-(3/2)} \\ \pm\sqrt{3/2} \end{bmatrix} \;=\; \begin{bmatrix} 12r^2 - 18 \end{bmatrix}.$$

Since $r^2 > 3/2$, this matrix is positive definite, so that f has a local minimum at p.