6. Static optimisation

Takuma Habu*
takumahabu@cornell.edu

23rd October 2024

1 Static optimisation problems

Our goal is to be able to solve the following type of problem, which we will call the *primal problem*.

$$\begin{aligned} \sup_{\mathbf{x}\in\mathbb{R}^d} \; & f(\mathbf{x}) \\ \text{s.t. } & h_k(\mathbf{x}) = 0 \; \forall k \in \{1,\ldots,K\}, \\ & g_j(\mathbf{x}) \geq 0 \; \forall j \in \{1,\ldots,J\}, \end{aligned}$$

where $f, h_k, g_j : \mathbb{R}^d \to \mathbb{R}$ for all $k \in \{1,\ldots,K\}$ and $j \in \{1,\ldots J\}$. By "solve", we want to obtain the set of maximisers, i.e., $\mathbf{x}$ that satisfy the constraints and maximises the objective function. This, in turn, allows us to compute the maximised objective function. We will also think about comparative statistics; i.e., how the maximisers and thus the objective function change as we "vary" the optimisation problem.

Since $\sup -f = -\inf f$, once we know how to solve maximisation problems, we also know how to solve minimisation problems. In economics, we tend to focus on maximisation problems, whereas in mathematics/computer science, the focus is on minimisation problems.

Define $h(\mathbf{x}) \coloneqq (h_k(\mathbf{x}))_{k=1}^K$ and $g(\mathbf{x}) \coloneqq (g_j(\mathbf{x}))_{J=1}^J$ (think of them as column vectors). Also define $\Gamma \subseteq \mathbb{R}^d$ as the set of all $\mathbf{x} \in \mathbb{R}^d$ that satisfies the constraints in the primal problem; i.e.,

$$\Gamma \coloneqq \left\{\mathbf{x} \in \mathbb{R}^d : h_k(\mathbf{x}) = 0 \; \forall k \in \{1,\ldots,K\}, \; g_j(\mathbf{x}) \geq 0 \; \forall j \in \{1,\ldots,J\}\right\}.$$

This allows us to write the primal problem succinctly as

$$\sup_{\mathbf{x}\in\Gamma} f(\mathbf{x}).$$

Given any $\mathbf{x} \in \mathbb{R}^d$, we say that a constraint is *binding* if it holds with equality, and *slack* if the constraint is satisfied but is not binding.

*Thanks to Giorgio Martini, Nadia Kotova and Suraj Malladi for sharing their lecture notes, on which these notes are heavily based.

Exercise 1. Suppose $\Gamma_1 \subseteq \Gamma_2 \subseteq \mathbb{R}^d$ and $f : \mathbb{R}^d \to \mathbb{R}$. Prove that

$$\sup_{\mathbf{x}\in\Gamma_1} f(\mathbf{x}) \leq \sup_{\mathbf{x}\in\Gamma_2} f(\mathbf{x}).$$

Exercise 2. Suppose $\Gamma \subseteq \mathbb{R}^d$ and $f : \mathbb{R}^d \to \mathbb{R}$. If $g : \mathbb{R} \to \mathbb{R}$ is a strictly increasing function, then

$$\left\{\mathbf{x} \in X : f(\mathbf{x}) = \sup_{\mathbf{z}\in\Gamma} f(\mathbf{z})\right\} = \left\{\mathbf{x} \in X : g(f(\mathbf{x})) = \sup_{\mathbf{z}\in\Gamma} g(f(\mathbf{z}))\right\}.$$

How does the result change if g was a weakly increasing function?

Remark 1. Note that $g(f(\mathbf{x})) = kf(\mathbf{x}) + c$ for any $k > 0$ and $c \in \mathbb{R}$ is a strictly increasing function so that above also tells us that multiplying the objective by a strictly positive constant and adding constant to the objective function leaves the set of maximisers unchanged.

2 Unconstrained optimisation

Let us first consider the problem of maximising a function without any constraints.

Proposition 1. *Suppose $f : X \subseteq \mathbb{R}^d \to \mathbb{R}$ has a local maximum or a local minimum at $\mathbf{x}^* \in \operatorname{int}(X)$ and that f is differentiable at $\mathbf{x}^*$. Then, $\mathbf{x}^*$ satisfies the first-order condition; i.e.,*

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

Proof. Suppose X is open. We already proved this for the case when $d = 1$ (Proposition 6 in 5. Differentiation). To extend the result to the case when $d > 1$, suppose that f has a local maximum (resp. minimum) at $\mathbf{x}^* \in \operatorname{int}(X)$. Fix any $\mathbf{v} \in \mathbb{R}^d\backslash\{\mathbf{0}\}$, $S_{\mathbf{v}} \coloneqq \{t \in \mathbb{R} : \mathbf{x}^* + t\mathbf{v} \in X\}$ and define $g : S_{\mathbf{v}} \to \mathbb{R}$ where $g(t) \coloneqq f(\mathbf{x}^* + t\mathbf{v})$. Observe that $g_{\mathbf{v}}$ must have a local maximum (resp. minimum) at 0. Since $g : \mathbb{R} \to \mathbb{R}$, we must have $g'(0) = \nabla f(\mathbf{x}^*)\mathbf{v} = 0$. Since $\mathbf{v} \in \mathbb{R}^d\backslash\{\mathbf{0}\}$ was chosen arbitrarily, this implies that $\nabla f(\mathbf{x}^*) = 0$. ■

Remark 2. The intuition for this comes from our discussion about gradient vectors. Recall that $\nabla f(\mathbf{x})$ is the direction that leads to the largest increase in the value of f. Together with the fact that $f(\mathbf{x}) + \nabla f(\mathbf{x})\mathbf{h}$ approximates $f(\mathbf{x} + \mathbf{h})$, it follows that if $\nabla f(\mathbf{x}) \neq \mathbf{0}$, we can move in the direction of $\nabla f(\mathbf{x})$, i.e., $\mathbf{h} = c\nabla f(\mathbf{x})$ for some $c > 0$, to increase the value of f. Thus, for a point to be maximum or a minimum, it must be that there is no direction in which we can move from the point to increase the value of f; i.e., $\nabla f(\mathbf{x}) = \mathbf{0}$.

Any $\mathbf{x} \in \operatorname{int}(X)$ that satisfies the first-order condition is called a *critical point of f*. Although first-order condition is necessary for a point to be a local maximum or a local minimum (assuming differentiability), it is not sufficient (e.g., $f(x) = x^3$). This leads us to the idea of second-order conditions that helps us distinguish between local maxima and minima.

Proposition 2. *Suppose f is $\mathbf{C}^2$ on $X \subseteq \mathbb{R}^d$. If f has a local maximum (resp. local minimum) at $\mathbf{x} \in \operatorname{int}(X)$, then $D^2 f(\mathbf{x})$ is negative semidefinite (resp. positive semidefinite).*

Proof. Suppose X is open and that f has a local maximum at $\mathbf{x} \in X$. Fix $\mathbf{v} \in \mathbb{R}^d\backslash\{\mathbf{0}\}$ and let $S_{\mathbf{v}} \coloneqq \{t \in \mathbb{R} : \mathbf{x} + t\mathbf{v} \in X\}$ and define $g : S_{\mathbf{v}} \to \mathbb{R}$ where $g(t) \coloneqq f(\mathbf{x} + t\mathbf{v})$. Define $h : S_{\mathbf{v}} \to X$ as

$h(t) \coloneqq \mathbf{x} + t\mathbf{v}$. Then, $g_{\mathbf{v}}(t) = f(h(t))$ and $h'(t) = \mathbf{v}$ so that

$$g'_{\mathbf{v}}(t) = \nabla f(h(t))\,\mathbf{v} = \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}(h(t))\,v_i.$$

(why can we replace total derivatives with partial derivatives?) Therefore,

$$g''_{\mathbf{v}}(t) = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(h(t))\,v_i v_j = \mathbf{v}^\top D^2 f(h(t))\,\mathbf{v}.$$

If f has a local maximum at $\mathbf{x}$, then g has a local maximum at 0. Hence, $g'(0) = 0$. Moreover, we cannot have $g''(0) > 0$: if it were, then x is also a local minimum (PS7) meaning that g' must be constant at x which contradicts that $g''(0) > 0$. Hence, it follows that $g''(0) \leq 0$. This, in turn, implies that $\mathbf{v}^\top D^2 f(\mathbf{x})\mathbf{v} \leq 0$. Since this holds for all $\mathbf{v} \in \mathbb{R}^d$, it follows that $D^2 f(\mathbf{x})$ is negative semidefinite. ■

Proposition 3. *Suppose f is $\mathbf{C}^2$ on $X \subseteq \mathbb{R}^d$. If $\nabla f(\mathbf{x}) = \mathbf{0}$ and $D^2 f(\mathbf{x})$ is negative definite (resp. positive definite) at some $\mathbf{x} \in X$, then $\mathbf{x}$ is a strict local maximum (resp. minimum).*

Remark 3. Focusing on maxima, so far, what we have shown the following. Suppose f is $\mathbf{C}^2$ on X and $\mathbf{x} \in \text{int}(X)$. We have shown the following necessary conditions for $\mathbf{x}$ to be a local maximum: the first-order necessary condition: $\nabla f(\mathbf{x}) = \mathbf{0}$; and the second-order necessary condition: $D^2 f(\mathbf{x})$ is negative semidefinite. We also have the following sufficient condition: If $\nabla f(\mathbf{x}) = \mathbf{0}$ and $D^2 f(\mathbf{x})$ is negative definite, then $\mathbf{x}$ is a strict local maximum. Notice, in particular, that if we have $\mathbf{x} \in \text{int}(X)$ such that $\nabla f(\mathbf{x}) = \mathbf{0}$ but $D^2 f(\mathbf{x})$ is negative semidefinite (but not negative definite), we cannot conclude that $\mathbf{x}$ is a local maximum; but we cannot rule out the possibility that $\mathbf{x}$ is a local maximum! Situation will be improved if we could either (i) strengthen the necessary second-order condition to be about negative definiteness of $D^2 f(\mathbf{x})$; or (ii) strengthen the sufficient second-order condition to require $D^2 f(\mathbf{x})$ to be negative semidefinite (while perhaps giving up on the strictness of local maximum). The following examples demonstrates that we can't.

(i) Consider $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) \coloneqq -x^4$. Since $f(\cdot) \leq 0$ and $f(0) = 0$, it follows that 0 is a (global) maximum of f. However, $f''(0) = 0$ so that, viewed as a 1×1 matrix, $f''(0)$ is negative semidefinite but not negative definite.

(ii) Consider $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) \coloneqq x^3$. Then, $f'(0) = f''(0) = 0$. Thus, $f''(0)$, viewed as a 1×1 matrix. is negative semidefinite but not negative definite. Observe that 0 would satisfy the condition for the sufficiency of second-order condition if we had relaxed the condition to allow for negative definiteness. However, 0 is neither a local maximum nor a local minimum of f.

The point of this (long) remark is that second-order condition isn't all that useful. All hope is not lost, however!

Proposition 4. *Suppose f is differentiable on X, where $\text{int}(X)$ is convex, and that f is concave. Fix $\mathbf{x}^* \in \text{int}(X)$. The following are equivalent:*

(i) $\nabla f(\mathbf{x}^*) = \mathbf{0}$.

(ii) f *has a local maximum at* $\mathbf{x}^*$*.*

(iii) f *has a global maximum at* $\mathbf{x}^*$*.*

Exercise 3 (PS9)**.** Prove Proposition 4. **Hint:** Use Proposition 14 from "5. Differentiation.'

Remark 4. Why does this result work when second-order conditions did not...? Remember that the second-order condition was about the property of the Hessian at a particular point. However, concavity of f is about the Hessian at all points in the domain of f.

Remark 5. First-order approach can only help to identify maximum/minimum in the interior of the domain of f. Thus, if the domain of f is closed, even if we know that an interior point maximises f, we must still check that that f is not maximised at some boundary point.

Remark 6. Recall that if f is strictly concave, then f is strictly quasiconcave (PS5). Thus, if we add to Proposition 4 that f is strictly concave, then we know that $\mathbf{x}^* \in \text{int}(X)$ such that $\nabla f(\mathbf{x}^*) = \mathbf{0}$ (if it exists) must be unique. Moreover, if we add that X is compact, then we know that that a maximum exists by Weierstrass theorem. Hence, we can conclude that f attains a global maximum at $\mathbf{x}^*$ such that $\nabla f(\mathbf{x}^*) = \mathbf{0}$or at some $\mathbf{x}^* \in X \backslash \text{int}(X)$.

3 Constrained optimisation

3.1 Necessity: Equality constraints

Theorem 1 (Theorem of Lagrange)**.** *Let* $f : \mathbb{R}^d \to \mathbb{R}$ *and* $h : \mathbb{R}^d \to \mathbb{R}^K$*, where* h_k *is* $\mathbf{C}^1$ *for each* $k \in \{1, \ldots, K\}$*. Suppose* $\mathbf{x}^*$ *is a local maximum or minimum of* f *on the constraint set*

$$\Gamma := \left\{\mathbf{x} \in \mathbb{R}^d : h\left(\mathbf{x}\right) = \mathbf{0}\right\}.$$

Suppose that[1]

$$\text{rank}\left(Dh\left(\mathbf{x}^*\right)\right) = K. \tag{1}$$

Then, there exists Lagrange multipliers $\boldsymbol{\mu}^* = (\mu_k^*)_{k=1}^K \in \mathbb{R}^K$ *such that*

$$\nabla f\left(\mathbf{x}^*\right) + \sum_{k=1}^K \mu_k^* \nabla h_k\left(\mathbf{x}^*\right) = \mathbf{0}_{1\times d}. \tag{2}$$

Proof. Note first that Dh is a $K \times d$ matrix and that $\text{rank}(Dh) \leq \min\{K, d\}$. Hence, if $K > d$, then (1) cannot be satisfied. Thus, we may assume that $K \leq d$ and further assume that the $K \times K$ submatrix of $Dh(\mathbf{x}^*)$ that has full rank consists of the first K rows and K columns of $Dh(\mathbf{x}^*)$. For each $\mathbf{x} \in \Gamma$, we write $\mathbf{x} = (\mathbf{w}, \mathbf{z}) \in \mathbb{R}^K \times \mathbb{R}^{d-K}$. Let $\nabla_{\mathbf{w}} f$ (a $1 \times K$ matrix) and $\nabla_{\mathbf{z}} f$ (a $1 \times (d-K)$ matrix) denote the derivative of f with respect to $\mathbf{w}$ variables along and $\mathbf{z}$ variables alone, respectively. $D_{\mathbf{w}} h$ (a $K \times K$ matrix) and $D_{\mathbf{z}} h$ (a $K \times (d-K)$ matrix) are defined analogously. We will treat $\boldsymbol{\mu}^* \in \mathbb{R}^K$ as a $1 \times K$ matrix. Let $\mathbf{x}^* = (\mathbf{w}^*, \mathbf{z}^*) \in \Gamma$ denote a local maximum (or

[1] Recall that any matrix $X \in \mathbb{R}^{m \times n}$ can be viewed as a collection of (row or column) vectors. Thus, we can consider the linear space that the vectors span. The linear space spanned by the columns of X is the *column space* of X. The *column rank* of X is the rank of the column space of X (recall that rank of a linear space equals the cardinality of (any) basis of that space). The linear space spanned by the rows of X, i.e., $\text{span}(\{\mathbf{x}_1, \ldots, \mathbf{x}_m\})$, is the *row space* of X. The *row rank* of X is the rank of the row space of X. Finally, recall that column and row ranks are equal.

a local minimum) of f in the constraint set that satisfies (1). We want to show that there exists $\boldsymbol{\mu}^* \in \mathbb{R}^K$ such that

$$\nabla_{\mathbf{w}} f\left(\mathbf{w}^*, \mathbf{z}^*\right) + \boldsymbol{\mu} D_{\mathbf{w}} h\left(\mathbf{w}^*, \mathbf{z}^*\right) = \mathbf{0}_{1\times K}, \tag{3}$$

$$\nabla_{\mathbf{z}} f\left(\mathbf{w}^*, \mathbf{z}^*\right) + \boldsymbol{\mu} D_{\mathbf{z}} h\left(\mathbf{w}^*, \mathbf{z}^*\right) = \mathbf{0}_{1\times (d-K)}. \tag{4}$$

The condition (1) allows us to appeal to the Implicit Function Theorem to deliver an open set $V \subseteq \mathbb{R}^{d-K}$ that contains $\mathbf{z}^*$ and a $\mathbf{C}^1$ function $b : V \to \mathbb{R}^K$ such that $b(\mathbf{z}^*) = \mathbf{w}^*$ and $h(b(\cdot), \cdot) = \mathbf{0}_{K\times 1}$ on V. Treating the two sides of the equality as functions of $\mathbf{z}$, we obtain

$$D_{\mathbf{w}} h\left(b\left(\mathbf{z}\right), \mathbf{z}\right) Db\left(\mathbf{z}\right) + D_{\mathbf{z}} h\left(b\left(\mathbf{z}\right), \mathbf{z}\right) = \mathbf{0}_{K\times(d-K)} \; \forall \mathbf{z} \in V.$$

Since $Dh_{\mathbf{w}}(\mathbf{w}^*, \mathbf{z}^*)$ has full rank, it is invertible so that

$$Db\left(\mathbf{z}^*\right) = -\left[D_{\mathbf{w}} h\left(\mathbf{w}^*, \mathbf{z}^*\right)\right]^{-1} D_{\mathbf{z}} h\left(\mathbf{w}^*, \mathbf{z}^*\right).$$

Define $\boldsymbol{\mu}^* \in \mathbb{R}^K$ by

$$\boldsymbol{\mu}^* \coloneqq -\nabla_{\mathbf{w}} f\left(\mathbf{w}^*, \mathbf{z}^*\right)\left[D_{\mathbf{w}} h\left(\mathbf{w}^*, \mathbf{z}^*\right)\right]^{-1}.$$

Then,

$$\sum_{k=1}^{K} \mu_k^* \nabla h_k\left(\mathbf{x}^*\right) = \boldsymbol{\mu}^* D_{\mathbf{w}} h\left(\mathbf{w}^*, \mathbf{z}^*\right) = -\nabla_{\mathbf{w}} f\left(\mathbf{w}^*, \mathbf{z}^*\right),$$

which gives (3). It remains to show (4). Define $F : V \to \mathbb{R}$ as $F(\mathbf{z}) \coloneqq f(b(\mathbf{z}), \mathbf{z})$ for all $\mathbf{z} \in V$. Since f has a local maximum at $(\mathbf{w}^*, \mathbf{z}^*) = (b(\mathbf{z}^*), \mathbf{z}^*)$, F also has a local maximum at $\mathbf{z}^*$. Since V is open, $\mathbf{z}^*$ is an unconstrained local maximum of F and the first-order conditions for an unconstrained maximum implies that $\nabla F(\mathbf{z}^*) = 0$; i.e.,

$$\begin{aligned}
0 = \nabla F\left(\mathbf{z}^*\right) &= \nabla_{\mathbf{w}} f\left(b\left(\mathbf{z}^*\right), \mathbf{z}^*\right) Db\left(\mathbf{z}^*\right) + \nabla_{\mathbf{z}} f\left(b\left(\mathbf{z}^*\right), \mathbf{z}^*\right) \\
&= \nabla_{\mathbf{w}} f\left(\mathbf{w}^*, \mathbf{z}^*\right)\Big(-\left[D_{\mathbf{w}} h\left(\mathbf{w}^*, \mathbf{z}^*\right)\right]^{-1} D_{\mathbf{z}} h\left(\mathbf{w}^*, \mathbf{z}^*\right)\Big) + \nabla_{\mathbf{z}} f\left(\mathbf{w}^*, \mathbf{z}^*\right) \\
&= \nabla_{\mathbf{z}} f\left(\mathbf{w}^*, \mathbf{z}^*\right) + \boldsymbol{\mu}^* D_{\mathbf{z}} h\left(\mathbf{w}^*, \mathbf{z}^*\right);
\end{aligned}$$

i.e., we have shown (4). ■

Remark 7. The condition (1) is called the *constraint qualification under equality constraints* and plays a central role in the proof to deliver the existence of the Lagrange multipliers $\boldsymbol{\mu}^* \in \mathbb{R}^K$.

Remark 8. Just as in the case of unconstrained optimisation, there exist second-order conditions that allows one to distinguish between local maximum and local minimum in equality constrained optimisation problem that are similar to Propositions (2) and (3).[2] Importantly, they have similar limitations as in the unconstrained case.

[2] See, for example, Sundaram Theorem 5.4.

3.2 Necessity: Inequality constraints

Let $f : \mathbb{R} \to \mathbb{R}$ be concave and $\mathbf{C}^1$ and consider the following problem:

$$\max_{x \in \mathbb{R}} f(x) \text{ s.t. } x \geq 0.$$

Ignoring the constraint, the solution $\overline{x}$ satisfies the first-order condition

$$f'(\overline{x}) = 0.$$

There are three cases to consider: (i) $\overline{x} < 0$; (ii) $\overline{x} = 0$; and (iii) $\overline{x} > 0$. Let x^* denote the solution to the constrained problem. In each of the three cases (try drawing!), we have

(i) $x^* = 0$ and $f'(x^*) < 0$;

(ii) $x^* = 0$ and $f'(x^*) = 0$;

(iii) $x^* > 0$ and $f'(x^*) = 0$.

Observe that the product of the two, i.e., $x^* f'(x^*)$, is zero in all three cases. However, $x^* f'(x^*) = 0$ is not a sufficient condition as you can see from case (iii). There, we see that the point $x = 0$ at which $f'(x) > 0$ also satisfies the condition, yet x is not an optimum. To rule this case out, we must add that $f'(x^*) \leq 0$. Together with the constraint itself, observe that we have identified three conditions:

$$\begin{aligned} x^* f'(x^*) &= 0, \\ f'(x^*) &\leq 0, \\ x^* &\geq 0. \end{aligned}$$

We will see that all these conditions are also important when we generalise the problem to multivariate-many-constraints case.

Theorem 2 (KKT Theorem)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ and $g_j : \mathbb{R}^d \to \mathbb{R}$ be $\mathbf{C}^1$ for each $j \in \{1, \ldots, J\}$. Suppose $\mathbf{x}^*$ is a local maximum of f on the constraint set*

$$\Gamma := \left\{ \mathbf{x} \in \mathbb{R}^d : g_j(\mathbf{x}) \geq 0 \ \forall j \in \{1, \ldots, J\} \right\}.$$

Let $E \subseteq \{1, \ldots, J\}$ denote the set of binding constraints at x^* and let $g_E := (g_j)_{j \in E}$. Suppose that[3]

$$\operatorname{rank}(Dg_E(\mathbf{x}^*)) = |E|. \tag{5}$$

[3] Recall that any matrix $X \in \mathbb{R}^{m \times n}$ can be viewed as a collection of (row or column) vectors. Thus, we can consider the linear space that the vectors span. The linear space spanned by the columns of X is the *column space* of X. The *column rank* of X is the rank of the column space of X (recall that rank of a linear space equals the cardinality of (any) basis of that space). The linear space spanned by the rows of X, i.e., $\text{span}(\{\mathbf{x}_1, \ldots, \mathbf{x}_m\})$, is the *row space* of X. The *row rank* of X is the rank of the row space of X. Finally, recall that column and row ranks are equal.

Then, there exists $\boldsymbol{\lambda}^* = (\lambda_j^*)_{j=1}^J \in \mathbb{R}^J$ *such that*

$$\lambda_j^* \geq 0 \ \forall j \in \{1, \ldots, J\}, \tag{6}$$

$$\lambda_j^* g_j(\mathbf{x}^*) = 0 \ \forall j \in \{1, \ldots, J\}, \tag{7}$$

$$\nabla f(x^*) + \sum_{j=1}^J \lambda_j^* \nabla g_j(\mathbf{x}^*) = \mathbf{0}_{1\times d}. \tag{8}$$

If x^* *is a local minimum of* f *on* Γ*, then (8) becomes* $\nabla f(x^*) - \sum_{j\in J} \lambda_j^* \nabla g_j(\mathbf{x}^*) = \mathbf{0}^\top$.

Proof. Let $\mathbf{x}^*$ be a local maximum of f on the set Γ such that (5) holds. Let E denote the set of binding constraints at $\mathbf{x}^*$. We want to show that there exists $\boldsymbol{\lambda}^* \in \mathbb{R}^J$ such that: (i) $\lambda_j \geq 0$ and $\lambda_j h_j(\mathbf{x}^*) = 0$ for all $j \in \{1, \ldots, J\}$; and (ii) $\nabla f(x^*) + \sum_{j=1}^J \lambda_j \nabla g_j(\mathbf{x}^*) = \mathbf{0}$.

With the exception of the nonnegative of the vector $\boldsymbol{\lambda}^*$ we can proceed as in the proof of Theorem of Lagrange. Without loss of generality, suppose that the first $J^* \coloneqq |E|$ constraints are binding and that the last $J - J^*$ constraints are slack. For each $j \in \{1, \ldots, J\}$, define

$$V_j \coloneqq \left\{\mathbf{x} \in \mathbb{R}^d : g_i(\mathbf{x}) > 0\right\},$$

and define $V \coloneqq \bigcap_{j=J^*+1}^J V_i$. Because g_i is continuous, V_j is open for each $j \in \{1, \ldots, J\}$ (why?) and so V is also open (why?). Let $\Gamma^* \subseteq \Gamma$ be the equality-constrained set given by

$$\Gamma^* \coloneqq V \cap \left\{\mathbf{x} \in \mathbb{R}^d : g_j(\mathbf{x}) = 0 \ \forall j \in \{1, \ldots, J^*\}\right\}.$$

By construction, $\mathbf{x}^* \in \Gamma^*$. Since $\mathbf{x}^*$ is a local maximum of f on Γ, it is also a local maximum of f on Γ^*. Together with (5), by the Theorem of Lagrange, there exists a vector $\boldsymbol{\mu}^* \in \mathbb{R}^{J^*}$ such that

$$\nabla f(x^*) + \sum_{j=1}^{J^*} \mu_j^* \nabla g_j(\mathbf{x}^*) = \mathbf{0}_{1\times d}.$$

Define $\boldsymbol{\lambda}^* = (\lambda_j^*)_{j=1}^J \in \mathbb{R}^J$ as

$$\lambda_j^* \coloneqq \begin{cases} \mu_j^* & \text{if } j \in \{1, \ldots, J^*\} \\ 0 & \text{if } j \in \{J^*+1, \ldots, J\} \end{cases}.$$

We will show that $\boldsymbol{\lambda}^*$ satisfies the required properties. First, observe that

$$\nabla f(x^*) + \sum_{j=1}^J \lambda_j^* \nabla g_j(\mathbf{x}^*) = \nabla f(x^*) + \sum_{j=1}^{J^*} \mu_j^* \nabla g_j(\mathbf{x}^*) = \mathbf{0}_{1\times d};$$

i.e., (8) is satisfied. Since $g_j(\mathbf{x}^*) = 0$ for all $j \in \{1, \ldots, J^*\}$, $\lambda_j g_j(\mathbf{x}^*) = 0$ for all $j \in \{1, \ldots, J^*\}$. For $j \in \{J^*+1, \ldots, J\}$, we have $\lambda_j^* = 0$ so that $\lambda_j g_j(\mathbf{x}^*) = 0$. Hence, we've shown (7). It remains to show (6). By construction of $\boldsymbol{\lambda}^*$, it suffices to show that $\lambda_j^* \geq 0$ for all $j \in \{1, \ldots, J^*\}$. Let us first show that $\lambda_1^* \geq 0$. To this end, define $\mathbf{x} \in \mathbb{R}^d$ and $\gamma \in \mathbb{R}$, and functions $G = (G_1, \ldots, G_{J^*})$:

$\mathbb{R}^{d+1} \to \mathbb{R}^{J^*}$ by

$$G_1(\mathbf{x},\gamma) \coloneqq g_1(\mathbf{x}) - \gamma,$$
$$G_j(\mathbf{x},\gamma) \coloneqq g_j(\mathbf{x}) \ \forall j \in \{2,\ldots,J^*\}.$$

Let $DG_{\mathbf{x}}$ denote the ($J^* \times d$ matrix) derivative of G with respect to the $\mathbf{x}$ variables alone and let ∇G_γ denote the ($k \times 1$ matrix) derivative of G with respect to γ. Note that $D_{\mathbf{x}}G(\cdot,\gamma) = Dg_E(\cdot)$ and $\nabla_\gamma G(\mathbf{x},\cdot) = (-1,0,\ldots,0)$ for all $\mathbf{x} \in \mathbb{R}^d$. By definition of G, $G(\mathbf{x}^*,0) = \mathbf{0}_{J^*\times 1}$ and rank of

$$\operatorname{rank}(D_{\mathbf{x}}G(\mathbf{x}^*,\gamma)) = \operatorname{rank}(Dg_E(\mathbf{x}^*)) = |E| = J^*.$$

By the Implicit Function Theorem, there exits an open ball around $0 \in \mathbb{R}$ denoted B and a function $\xi : B \to \mathbb{R}^d$ that is $\mathbf{C}^1$ such that $\xi(0) = \mathbf{x}^*$ and

$$G(\xi(\gamma),\gamma) = \mathbf{0} \ \forall \gamma \in B.$$

Treating both sides of the equation as functions of γ and differentiating both sides and evaluation at $\xi(0) = \mathbf{x}^*$ gives

$$D_{\mathbf{x}}G(\mathbf{x}^*,0)\,D\xi(0) + \nabla_\gamma G(\mathbf{x}^*,0) = \mathbf{0} \Leftrightarrow Dg_E(\mathbf{x}^*)\,D\xi(0) + (-1,0,\ldots,0) = \mathbf{0},$$

where $D\xi$ is a $d \times 1$ matrix. That is,

$$\nabla g_j(\mathbf{x}^*)\,D\xi(0) = \begin{cases} 1 & \text{if } j = 1 \\ 0 & \text{if } j \in \{2,\ldots,J^*\} \end{cases}.$$

Using (8) and the fact that $\lambda_j = 0$ for all $j \in \{J^*+1,\ldots,J\}$,

$$\nabla f(\mathbf{x}^*)\,D\xi(0) = -\left(\sum_{j=1}^{J} \lambda_j \nabla g_j(\mathbf{x}^*)\right) D\xi(0) = -\lambda_1.$$

To complete the proof we will show that $\nabla f(\mathbf{x}^*)D\xi(0) \le 0$ which implies $-\lambda_1 \le 0 \Leftrightarrow \lambda_1 \ge 0$. Toward this goal, we first show that there is a $\gamma^* > 0$ such that for all $\gamma \in [0,\gamma^*)$, $\xi(\gamma) \in \Gamma$; i.e., for each $j \in \{1,\ldots,J\}$, $g_j(\xi(\gamma)) \ge 0$ for all $\gamma \in [0,\gamma^*)$. If $\gamma > 0$, since $G_j(\xi(\gamma),\gamma) = 0$ for all $j \in \{1,\ldots,J^*\}$, we have

$$g_j(\xi(\gamma)) = \begin{cases} \gamma > 0 & \text{if } j = 1 \\ 0 & \text{if } j \in \{2,\ldots,J^*\} \end{cases}.$$

For $j \in \{J^*+1,\ldots,J\}$, we have $g_j(\xi(0)) = g_j(\mathbf{x}^*) > 0$. Since both g_j and ξ are continuous, we can choose γ sufficiently small, say $\gamma \in (0,\gamma^*)$, such that

$$g_j(\xi(\gamma)) > 0 \ \forall j = \{J^*+1,\ldots,J\}.$$

We have thus shown that there exists $\gamma^* > 0$ such that $\xi(\gamma) \in \Gamma$ for all $\gamma \in [0,\gamma^*)$.

Because $\xi(0) = \mathbf{x}^*$ is a local maximum of f on Γ and $\xi(\gamma) \in \Gamma$ for all $\gamma \in [0,\gamma^*)$, it follows that for γ sufficiently close to 0, we must have

$$f(\mathbf{x}^*) \geq f(\xi(\gamma))$$

Rearranging gives and dividing both sides by $\gamma > 0$ gives

$$0 \geq \frac{f(\xi(\gamma)) - f(\mathbf{x}^*)}{\gamma} = \frac{f(\xi(\gamma)) - f(\xi(0))}{\gamma}$$

Taking limits as $\gamma \searrow 0$, above gives us that

$$0 \geq \nabla f(\xi(0)) D\xi(0) = \nabla f(\mathbf{x}^*) D\xi(0).$$

Hence, we have now shown that $\lambda_1 \geq 0$. Analogous argument shows that $\lambda_j^* \geq 0$ for all $j \in \{2, \ldots, J^*\}$ and the proof is complete. ■

Remark 9. Suppose $d = 1$ and the only constraint is the nonnegativity constraint; i.e., $J = 1$ and $g_1(x) = x$. Then, KKT first-order conditions become

$$\lambda_1^* \geq 0,\ \lambda_1^* x^* = 0,\ f'(x^*) = -\lambda_1^*$$

and so we can write it as

$$f'(x^*) \leq 0,\ x^* f'(x^*) = 0,$$

and, of course, we must have $g_1(x^*) = x^* \geq 0$. These are the exact conditions we had before!

The properties (6), (7) and (8) are together referred to as the *KKT first-order conditions*. Let us go through the theorem carefully.

Complementary slackness (7) is referred to as the *complementary slackness conditions*. Since $g_j(x) \geq 0$ from the constraint and $\lambda_j^* \geq 0$, (7) tells us that: (i) if $g_j(x^*) > 0$, then $\lambda_j^* = 0$; and (ii) if $\lambda_j^* > 0$, then $g_j(x^*) = 0$. That is, if one inequality is not strict (i.e., "slack"), then the other cannot be.

Nonnegativity constraints Suppose that the ℓth constraint is a nonnegativity constraint on some x_i, $i \in \{1, \ldots, d\}$; i.e., $g_\ell(\mathbf{x}) = x_{i_\ell} \geq 0$. Since $\nabla g_\ell(\mathbf{x}^*) = \mathbf{e}_{i_\ell}$, (8) is given by

$$\nabla f(x^*) + \sum_{j \in J \setminus \{\ell\}} \lambda_j^* \nabla g_j(\mathbf{x}^*) - \lambda_\ell^* \mathbf{e}_{i_\ell} = \mathbf{0}.$$

Note that (6) ensures that $\lambda_\ell^* \geq 0$ and, together with (7), we know that if $x_i^* > 0$, then $\lambda_\ell^* = 0$. Thus, we can rewrite above as: for all $i \in \{1, \ldots, d\} \setminus \{i_\ell\}$

$$\frac{\partial f}{\partial x_i}(\mathbf{x}^*) + \sum_{j=1}^{J} \lambda_j^* \frac{\partial g_j}{\partial x_i}(\mathbf{x}^*) = \frac{\partial f}{\partial x_i}(\mathbf{x}^*) + \sum_{j \neq \ell} \lambda_j^* \frac{\partial g_j}{\partial x_i}(\mathbf{x}^*) = 0$$

and

$$\frac{\partial f}{\partial x_i}(\mathbf{x}^*) + \sum_{j \neq \ell} \lambda_j^* \frac{\partial g_j}{\partial x_i}(\mathbf{x}^*) \leq 0 \text{ with equality if } x_{i_\ell}^* > 0.$$

Constraint qualification The condition (5) is called the *constraint qualification.* To understand this, notice first that we can rewrite (8) as

$$\nabla f\left(\mathbf{x}^{*}\right)=\sum_{j \in J}\left(-\lambda_{j}^{*}\right) \nabla g_{j}\left(\mathbf{x}^{*}\right);$$

i.e., $\nabla f(\mathbf{x}^*)$ is a linear combination of the gradients $\nabla g_1(\mathbf{x}^*)$, …, $\nabla g_J(\mathbf{x}^*)$. Moreover, by complimentary slackness, since $\lambda_j = 0$ for all non-binding constraints, $\nabla f(\mathbf{x}^*)$ is, in fact, a linear combination of $(\nabla g_j)_{j\in E}$. Letting $E = \{1, \ldots, |E|\}$, the constraint qualification is the requirement that $|E| \times d$ matrix

$$\nabla g_{E}\left(\mathbf{x}^{*}\right)=\left[\begin{array}{ccc}\frac{\partial g_{1}}{\partial x_{1}}\left(\mathbf{x}^{*}\right) & \cdots & \frac{\partial g_{1}}{\partial x_{n}}\left(\mathbf{x}^{*}\right) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_{|E|}}{\partial x_{1}}\left(\mathbf{x}^{*}\right) & \cdots & \frac{\partial g_{|E|}}{\partial x_{n}}\left(\mathbf{x}^{*}\right)\end{array}\right]_{|E| \times d}$$

has rank $|E|$. Since $\operatorname{rank}(\nabla g_E(\mathbf{x}^*)) \leq \min\{|E|, d\}$, this is possible only if $d \geq |E|$. One implication of the rank qualification is therefore that the $\{\nabla g_j\}_{j\in E}$ are linearly independent.

Example 1. Let $f, g : \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x, y) \coloneqq -(x^2 + y^2)$ and $g(x, y) \coloneqq (x-1)^3 - y^2$, respectively. Consider the problem of maximising f on $\Gamma \coloneqq \{(x, y) \in \mathbb{R}^2 : g(x, y) \geq 0\}$. Let us argue that solution to this constrained problem is $(x^*, y^*) = (1, 0)$. first, the function f reaches a maximum when $x^2 + y^2$ reaches a minimum. Since the constraint requires $(x-1)^3 \geq y^2$, and $y^2 \geq 0$ for all $y \in \mathbb{R}$, the smallest absolute value of x in the constraint set is $x = 1$. And the smallest absolute value of y in the constraint set is $y = 0$. Thus, $(x^*, y^*) = (1, 0)$. Observe that the constraint is binding at the optimum. Note that

$$D g\left(x^{*}, y^{*}\right)=\left(3\left(x^{*}-1\right)^{2}, 2 y^{*}\right)=(0,0)$$

so that it has rank less than the number of binding constraints (i.e., 1). Thus, the constraint qualification fails in this case. Moreover, we have

$$D f\left(x^{*}, y^{*}\right)=\left(-2 x^{*},-2 y^{*}\right)=(-2,0)$$

and so there cannot exist $\lambda \geq 0$ such that $Df(x^*, y^*) + \lambda Dg(x^*, y^*) = (0, 0)$; i.e., conclusion of Theorem 2 also fails.

Remark 10. Since any equality constraints can be written as a two inequality constraints (i.e., $h_k(\mathbf{x}) = 0$ if and only if $h_k(\mathbf{x}) \geq 0$ and $-h_k(\mathbf{x}) \geq 0$), Theorem 3 can be applied to optimisation problems with both equality and inequality constraints. However, recall that for constraint qualification to be satisfied, we must have $d \geq |E|$, where $|E|$ is the number of binding constraints. Thus, treating equality constraints as inequality constraints may lead to violation of constraint qualification. Luckily, we can treat equality constraints separately from inequality constraints.

Theorem 3. *Let $f : \mathbb{R}^d \to \mathbb{R}$ and $g_j : \mathbb{R}^d \to \mathbb{R}$ be $\mathbf{C}^1$ for each $j \in \{1, \ldots, J\}$. Suppose x^* is a local maximum of f on the constraint set*

$$\Gamma \coloneqq\left\{\mathbf{x} \in \mathbb{R}^{d}: h_{k}(\mathbf{x})=0\ \forall k \in\{1, \ldots, K\},\ g_{j}(\mathbf{x}) \geq 0\ \forall j \in\{1, \ldots, J\}\right\}$$

Let $E \subseteq \{1, \ldots, J\}$ denote the set of binding constraints at x^* and let $g_E \coloneqq (g_j)_{j \in E}$. Suppose that[4]

$$\operatorname{rank}\left(D\begin{bmatrix} h(\mathbf{x}) \\ g_E(\mathbf{x}^*) \end{bmatrix}\right) = K + |E|. \tag{9}$$

Then, there exists $\boldsymbol{\mu}^* \in \mathbb{R}^K$ and $\boldsymbol{\lambda}^* = (\lambda_j^*)_{j=1}^J \in \mathbb{R}^J$ such that

$$\lambda_j^* \geq 0 \ \forall j \in \{1, \ldots, J\}, \tag{10}$$

$$\lambda_j^* g_j(\mathbf{x}^*) = 0 \ \forall j \in \{1, \ldots, J\}, \tag{11}$$

$$\nabla f(x^*) + \sum_{k=1}^{K} \mu_k^* \Delta h_k(\mathbf{x}^*) + \sum_{j=1}^{J} \lambda_j^* \nabla g_j(\mathbf{x}^*) = \mathbf{0}^\top. \tag{12}$$

3.3 Sufficiency: Inequality constraints

As in the first-order condition for unconstrained problems, the KKT first-order conditions are only necessary for a local maximum or a local minimum and not sufficient.

Example 2. Suppose $f(x) = x^3$ with the constraint that $g(x) \coloneqq x \geq 0$. As noted above $x^* = 0$ is not a local maximum or a minimum. At $x = 0$, the constraint is binding so that $E = \{1\}$. Then, $g'(x^*) = 1 = |E|$ so that constraint qualification is satisfied. Moreover, $\lambda^* = 0$ satisfies the KKT first-order conditions since

$$f'(x^*) + \lambda^* \nabla g(\mathbf{x}^*) = 3(0)^2 + 0 \cdot 1 = 0.$$

Theorem 4 (Sufficiency of KKT with concavity)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ and $g_j : \mathbb{R}^d \to \mathbb{R}$ for each $j \in \{1, \ldots, J\}$ are all be $\mathbf{C}^1$ and concave. Suppose there exists $\boldsymbol{\lambda}^* = (\lambda_j^*)_{j=1}^J \in \mathbb{R}^J$ that satisfies the KKT first-order conditions; i.e., (6), (7) and (8). Then, x^* is a global maximum of f on on the constraint set $\Gamma \coloneqq \{\mathbf{x} \in \mathbb{R}^d : g_j \geq 0 \ \forall j \in \{1, \ldots, J\}\}$.*

Proof. Let us prove an interim result first. Say that a point $\mathbf{y} \in \mathbb{R}^d$ *points into set* $X \subseteq \mathbb{R}^d$ *at* $\mathbf{x} \in X$ if there is $\omega > 0$ such that $(\mathbf{x} + \eta \mathbf{y}) \in X$ for all $\eta \in (0, \omega)$.

Lemma 1. *Suppose $X \subseteq \mathbb{R}^d$ is a convex set and $f : X \to \mathbb{R}$ is a concave function. Then, $\mathbf{x}^*$ maximises f on X if and only if $D_{\mathbf{y}} f(\mathbf{x}^*) \leq 0$ for all $\mathbf{y}$ pointing into X at $\mathbf{x}^*$.*

Proof. Suppose $\mathbf{x}^*$ maximises a concave $f : X \to \mathbb{R}$ on a convex $X \subseteq \mathbb{R}^d$. Let $\mathbf{y} \in \mathbb{R}^d$ point into X at $\mathbf{x}^*$. Since $\mathbf{x}^*$ is the maximiser, we have $f(\mathbf{x}^*) \geq f(\mathbf{x}^* + \eta \mathbf{y})$ for all $\eta > 0$ such that $\mathbf{x}^* + \eta \mathbf{y} \in X$ (and such η exists for sufficiently small η). Rearranging and taking limits yields

$$0 \geq \lim_{\eta \searrow 0} \frac{f(\mathbf{x}^* + \eta \mathbf{y}) - f(\mathbf{x}^*)}{\eta} = D_{\mathbf{y}} f(\mathbf{x}^*)$$

as desired.

[4] Recall that any matrix $X \in \mathbb{R}^{m \times n}$ can be viewed as a collection of (row or column) vectors. Thus, we can consider the linear space that the vectors span. The linear space spanned by the columns of X is the *column space* of X. The *column rank* of X is the rank of the column space of X (recall that rank of a linear space equals the cardinality of (any) basis of that space). The linear space spanned by the rows of X, i.e., $\operatorname{span}(\{\mathbf{x}_1, \ldots, \mathbf{x}_m\})$, is the *row space* of X. The *row rank* of X is the rank of the row space of X. Finally, recall that column and row ranks are equal.

Conversely, suppose $D_{\mathbf{y}}f(\mathbf{x}^*) \leq 0$ for all $\mathbf{y}$ pointing into X at $\mathbf{x}^*$. If $\mathbf{x}^*$ does not maximise f on X, there exists $\mathbf{z} \in X$ with $f(\mathbf{z}) > f(\mathbf{x}^*)$. Let $\mathbf{y} \coloneqq \mathbf{z} - \mathbf{x}^*$. Then, since $\mathbf{x}^* + 1 \cdot \mathbf{y} \in X$ and X is convex, $\mathbf{y}$ points into X at $\mathbf{x}^*$. But for $\eta \in (0,1)$, because f is concave,

$$\begin{aligned} f\left(\mathbf{x}^* + \eta\left(\mathbf{z} - \mathbf{x}^*\right)\right) &= f\left(\left(1-\eta\right)\mathbf{x}^* + \eta\mathbf{z}\right) \\ &\geq \left(1-\eta\right) f\left(\mathbf{x}^*\right) + \eta f\left(\mathbf{z}\right) \\ &= f\left(\mathbf{x}^*\right) + \eta\left[f\left(\mathbf{z}\right) - f\left(\mathbf{x}^*\right)\right] \end{aligned}$$

so that

$$\frac{f\left(\mathbf{x}^* + \eta\left(\mathbf{z} - \mathbf{x}^*\right)\right) - f\left(\mathbf{x}^*\right)}{\eta} \geq f\left(\mathbf{z}\right) - f\left(\mathbf{x}^*\right) > 0.$$

Observe that left-hand side converges to $D_{\mathbf{y}}f(\mathbf{x}^*)$ as $\eta \searrow 0$ so that $D_{\mathbf{y}}f(\mathbf{x}^*) > 0$; a contradiction. ■

We prove the sufficiency part first by using the fact that if a function is differentiable at a point, then directional derivative exists in all directions and is given by the dot product of the partial derivatives and the direction. So suppose there exists $\lambda^* \in \mathbb{R}^J_+$ that satisfies the KKT first-order conditions. Let

$$V_j \coloneqq \left\{\mathbf{x} \in \mathbb{R}^d : g_j\left(\mathbf{x}\right) \geq 0\right\}.$$

Suppose $\mathbf{x}_1, \mathbf{x}_2 \in X_j$, Pick any $\lambda \in (0,1)$ and let $\mathbf{z} \coloneqq \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$. Because g_j is concave, X_j is convex for all $j \in \{1, \ldots, J\}$ (why?). Hence, $\bigcap_{j=1}^J V_j = \Gamma$ is also convex (why?). Since f is concave, all that remains to show is that $\nabla f(\mathbf{x}^*)\mathbf{y} \leq 0$ for all $\mathbf{y}$ pointing into X at $\mathbf{x}^*$ and we can then appeal to the lemma above. Suppose some $\mathbf{y}$ points into Γ at $\mathbf{x}^*$. We will show that, for each $j \in \{1, \ldots, J\}$, we have $\lambda_j^* \nabla g_j(\mathbf{x}^*)\mathbf{y} \geq 0$, which by (8), would imply $\nabla f(\mathbf{x}^*)\mathbf{y} \leq 0$. Observe that, by definition of $\mathbf{y}$, there exists $\epsilon > 0$ such that $\mathbf{x}^* + t\mathbf{y} \in X$ for all $t \in (0, \epsilon)$. By our choice of Γ, it follows that $g_j(\mathbf{x}^* + t\mathbf{y}) \geq 0$ for all $j \in \{1, \ldots, J\}$ for all $t \in (0, \epsilon)$. Fix any $j \in \{1, \ldots, J\}$. There are two possibilities: either $g_j(\mathbf{x}^*) > 0$ or $g_j(\mathbf{x}^*) = 0$. In the first case, $\lambda_j^* = 0$ by (7) and so clearly $\lambda_j^* \nabla g_j(\mathbf{x}^*)\mathbf{y} \geq 0$. In the second case, we have

$$\frac{g_j\left(\mathbf{x}^* + t\mathbf{y}\right) - g_j\left(\mathbf{x}^*\right)}{t} \geq 0 \ \forall t \in \left(0, \epsilon\right).$$

Taking limits as $\epsilon \searrow 0$, we obtain $D_{\mathbf{y}}g(\mathbf{x}^*) = \nabla g_j(\mathbf{x}^*)\mathbf{y} \geq 0$. Since $\lambda_j^* \geq 0$ by (6), we have $\lambda_j^* \nabla g_j(\mathbf{x}^*)\mathbf{y} \geq 0$. ■

Remark 11. One can show that if f and g_j's are all $\mathbf{C}^1$ and concave and the following condition, called *Slater's condition*, holds

$$\exists \mathbf{x} \in \mathbb{R}^d, \ g_j\left(\mathbf{x}\right) > 0 \ \forall j \in \left\{1, \ldots, J\right\}. \tag{13}$$

Then, KKT first-order conditions are both necessary and sufficient. That is, the Slater's condition can be used as an alternative to the constraint qualification condition to ensure that KKT first-order conditions are necessary (when f and g_j's are all concave).

Theorem 5 (Sufficiency of KKT with quasiconcavity)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ and $g_j : \mathbb{R}^d \to \mathbb{R}$ for each $j \in \{1, \ldots, J$ be $\mathbf{C}^1$ and quasiconcave. Suppose there exists $\mathbf{x}^* \in \mathbb{R}^d$ and $\boldsymbol{\lambda} \in \mathbb{R}^J$ that satisfy all* *the constraints (i.e.,* $\mathbf{x}^* \in \Gamma \coloneqq \{\mathbf{x} \in \mathbb{R}^d : g_j \geq 0\ \forall j \in \{1, \ldots, J\}\}$*) and satisfy the KKT first-order conditions (i.e., (6), (7) and (8)). Then,* x^* *maximises* f *on* Γ *provided that at least one of the following condition holds:*

$$\nabla f(\mathbf{x}^*) \neq \mathbf{0} \text{ or } f \text{ is concave.} \tag{14}$$

Proof. First observe that the set $X_j \coloneqq \{\mathbf{x} \in \mathbb{R}^d : g_j(\mathbf{x}) \geq 0\}$ is convex because g_j is quasiconcave (why?). Thus, $X \coloneqq \bigcap_{j=1}^{J} X_j$ is also convex.

Lemma 2. *Under the hypothesis of the theorem,* $\nabla f(\mathbf{x}^*)(\mathbf{y} - \mathbf{x}^*) \leq 0$ *for all* $\mathbf{y} \in X$*.*

Proof. By hypothesis, we have

$$\nabla f(\mathbf{x}^*)(\mathbf{y} - \mathbf{x}^*) = -\sum_{j=1}^{J} \lambda_j \nabla g_j(\mathbf{x}^*)(\mathbf{y} - \mathbf{x}^*).$$

It suffices to show that $\lambda_j \nabla g_j(\mathbf{x}^*)(\mathbf{y} - \mathbf{x}^*) \geq 0$ for each $j \in \{1, \ldots, J\}$ and any $\mathbf{y} \in X$. Fix $j \in \{1, \ldots, J\}$ and $\mathbf{y} \in X$. We have that either $g_j(\mathbf{x}^*) > 0$ or $g_j(\mathbf{x}^*) = 0$. In the first case, $\lambda_j^* = 0$ by (7) and so clearly $\lambda_j^* \nabla g_j(\mathbf{x}^*)(\mathbf{y} - \mathbf{x}^*) = 0$. In the second case, because X is convex, $\mathbf{x}^* + t(\mathbf{y} - \mathbf{x}^*) = (1-t)\mathbf{x}^* + t\mathbf{y} \in X$ for all $t \in (0,1)$. Because g_j is quasiconcave,

$$g_j((1-t)\mathbf{x}^* + t\mathbf{y}) \geq \min\{g_j(\mathbf{x}^*), g_j(\mathbf{y})\} \geq 0$$

so that (recall $g_j(\mathbf{x}^*) = 0$),

$$0 \leq \frac{g_j((1-t)\mathbf{x}^* + t\mathbf{y}) - g_j(\mathbf{x}^*)}{t}.$$

Taking limits as $t \searrow 0$ establishes that $D_{\mathbf{y}-\mathbf{x}^*} g(\mathbf{x}^*) = \nabla g(\mathbf{x}^*)(\mathbf{y} - \mathbf{x}^*) \geq 0$. Since $\lambda_j^* \geq 0$ from (6), it follows that $\lambda_j^* \nabla g_j(\mathbf{x}^*)(\mathbf{y} - \mathbf{x}^*) \geq 0$. ■

Suppose first that $\nabla f(\mathbf{x}^*) \neq \mathbf{0}$. Then, there exists $\mathbf{w} \in \mathbb{R}^d$ such that $\nabla f(\mathbf{x}^*)\mathbf{w} < 0$. Let $\mathbf{z} \coloneqq \mathbf{x}^* + \mathbf{w}$ so that $\nabla f(\mathbf{x}^*)(\mathbf{z} - \mathbf{x}^*) < 0$. Pick any $\mathbf{y} \in X$. For $t \in (0,1)$, let

$$\mathbf{y}(t) \coloneqq (1-t)\mathbf{y} + t\mathbf{z},\ \mathbf{x}(t) \coloneqq (1-t)\mathbf{x}^* + t\mathbf{z}.$$

Fixing any $t \in (0,1)$, we have

$$\nabla f(\mathbf{x}^*)(\mathbf{x}(t) - \mathbf{x}^*) = t\nabla f(\mathbf{x}^*)(\mathbf{z} - \mathbf{x}^*) < 0$$

and, by lemma above, we also have

$$\nabla f(\mathbf{x}^*)(\mathbf{y}(t) - \mathbf{y}^*) = (1-t)\nabla f(\mathbf{x}^*)(\mathbf{y} - \mathbf{x}^*) \leq 0.$$

Summing the inequalities yield

$$\nabla f(\mathbf{x}^*)(\mathbf{y}(t) - \mathbf{x}^*) < 0.$$

Toward a contradiction, suppose that $f(\mathbf{y}(t)) \geq f(\mathbf{x}^*)$. Since f is quasiconcave, for any $\alpha \in (0,1)$,

$$f(\mathbf{x}^* + \alpha(\mathbf{y}(t) - \mathbf{x}^*)) = f((1-\alpha)\mathbf{x}^* + \alpha\mathbf{y}(t)) \geq \min\{f(\mathbf{x}^*), \mathbf{y}(t)\} = f(\mathbf{x}^*).$$

Hence,

$$\frac{f\left(\mathbf{x}^{*}+\alpha\left(\mathbf{y}\left(t\right)-\mathbf{x}^{*}\right)\right)-f\left(\mathbf{x}^{*}\right)}{\alpha}\geq0\;\forall\alpha\in\left(0,1\right).$$

Taking limits as $\alpha\searrow0$ (recall f is continuous), the left-hand side converges to $D_{\mathbf{y}(t)-\mathbf{x}^*}f(\mathbf{x}^*)=\nabla f(\mathbf{x}^*)(y(t)-\mathbf{x}^*)\geq0$, which is a contradiction. Hence, we must have $f(\mathbf{y}(t))<f(\mathbf{x}^*)$. Since this holds for all $t\in(0,1)$, taking limits as $t\to1$, we have $f(\mathbf{y})\leq f(\mathbf{x}^*)$ which establishes the optimality of $\mathbf{x}^*$.

Suppose now that f is concave. By repeating the arguments in the proof for the sufficiency of KKT with concavity, we can show that $\nabla f(\mathbf{x}^*)\mathbf{y}\leq0$ for all $\mathbf{y}$ pointing into X at $\mathbf{x}^*$. Since f is concave and X is convex, Lemma (1) then establishes that $\mathbf{x}^*$ is optimal. ■

Remark 12. Observe that Theorem 5 does not give necessary conditions for $\mathbf{x}^*$ to be an optimum—indeed the conditions in the theorem are not necessary unless, for example, the constraint qualification is met at $\mathbf{x}^*$.

Exercise 4. Let $f,g:\mathbb{R}\to\mathbb{R}$ be given by

$$f\left(x\right)\coloneqq\begin{cases}x^{3} & \text{if }x<0\\ 0 & \text{if }0\leq x\leq1\\ \left(x-1\right)^{2} & \text{if }x>1\end{cases},\;g\left(x\right)\coloneqq x.$$

Verify that f and g are both $\mathbf{C}^1$ and quasiconcave and that f is not concave. Show that, for any $x^*\in[0,1]$, we can find $\lambda^*\geq0$ such that (x^*,λ^*) satisfies the KKT first-order conditions. Finally, argue that no $x^*\in[0,1]$ can be a solution to the constrained optimisation problem of maximising f on $\Gamma\coloneqq\{x\in\mathbb{R}:g(x)\geq0\}$. What can you conclude about Theorem 5 from this?

Remark 13. Recall that the set of maximisers do not change when we transform the objective function using a strictly increasing function. Hence, even if f is not concave, if we can apply a strictly increasing transformation of f that is concave, then we can apply the theorem above to obtain the maximisers.