Aalto University

Nonlinear optimisation, fundamentals (B)

CHEM-E7225 (was E7195), 2023

Francesco Corona (¬_¬)

Chemical and Metallurgical Engineering
School of Chemical Engineering

The Lagrangian function

Nonlinear optimisation

The Lagrangian function

Consider the nonlinear optimisation problem in the standard form

$$\begin{aligned} \min_{w \in \mathcal{R}^N} \quad & f(w) \\ \text{subject to} \quad & g(w) = 0 \\ & h(w) \geq 0 \end{aligned}$$

- Objective function

$$f : \mathcal{R}^N \to \mathcal{R}, \text{ with } f \in \mathcal{C}^2\left(\mathcal{R}^N\right)$$

- Equality constraint function

$$g : \mathcal{R}^N \to \mathcal{R}^{N_g}, \text{ with } g \in \mathcal{C}^2\left(\mathcal{R}^N\right)$$

- Inequality constraint function

$$h : \mathcal{R}^N \to \mathcal{R}^{N_h}, \text{ with } h \in \mathcal{C}^2\left(\mathcal{R}^N\right)$$

We denote a problem in this form as **primal optimisation problem**

The Lagrangian function (cont.)

$$\begin{aligned} \min_{w \in \mathcal{R}^N} \quad & f(w) \\ \text{subject to} \quad & g(w) = 0 \\ & h(w) \geq 0 \end{aligned}$$

The globally optimal (min) value of the objective function subjected to the constraints

$$p^* = \left(\min_{w \in \mathcal{R}^N} \quad f(w), \text{ s.t. } g(w) = 0, h(w) \geq 0 \right)$$

Remember that there can be a multiplicity of points $w^* \in \Omega$ such that $f(w^*) = p^*$

- ⇝ The globally optimal value p^* of the objective function is unique
- ⇝ The globally optimal value is called the **primal optimal value**

We are interested in a lower-bound (for minimisation problems) on the optimal value p^*

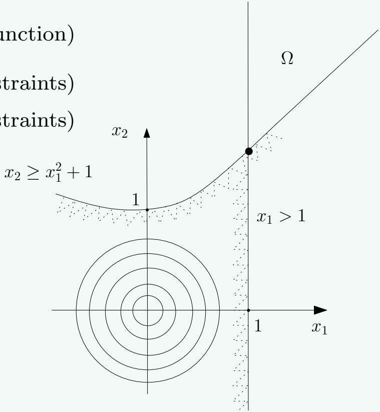
Overview (cont.)

Example

$$\min_{x \in \mathcal{R}^2} \quad x_1^2 + x_2^2 \qquad \text{(Objective function)}$$

$$\text{subject to} \quad x_1 - 1 = 0 \qquad \text{(Equality constraints)}$$

$$x_2 - 1 - x_1^2 \geq 0 \qquad \text{(Inequality constraints)}$$



- ⇝ $f : \mathcal{R}^2 \rightarrow \mathcal{R}$, with $f \in \mathcal{C}^2\left(\mathcal{R}^2\right)$
- ⇝ $g : \mathcal{R}^2 \rightarrow \mathcal{R}$, with $g \in \mathcal{C}^2\left(\mathcal{R}^2\right)$
- ⇝ $h : \mathcal{R}^2 \rightarrow \mathcal{R}$, with $h \in \mathcal{C}^2\left(\mathcal{R}^2\right)$

The feasible set of decision variables

$$\Omega = \{x \in \mathcal{R}^2 | h(x) \geq 0, g(x) = 0\}$$

The minimiser x^*, at point •

The Lagrangian function (cont.)

$$\begin{aligned} \min_{w \in \mathcal{R}^N} \quad & f(w) \\ \text{subject to} \quad & g(w) = 0 \\ & h(w) \geq 0 \end{aligned}$$

We can define an auxiliary function and we denote it as the **Lagrangian function**

$$\mathcal{L}(w, \lambda, \mu) = f(w) - \lambda^T g(w) - \mu^T h(w)$$

The Lagrangian function depends on w and two sets of auxiliary variables

⇝ The **Lagrangian multipliers**, or **dual variables**

- The inequality multipliers, $\mu \in \mathcal{R}^{N_h}$
- The equality multipliers, $\lambda \in \mathcal{R}^{N_g}$

$$\mathcal{L}(w, \lambda, \mu) = f(w) - \sum_{n_g=1}^{N_g} \lambda_{n_h} g_{n_g}(w) - \sum_{n_h=1}^{N_h} \mu_{n_g} h_{n_h}(w)$$

The Lagrangian function is a scalar function,

$$\mathcal{L} : \mathcal{R}^N \times \mathcal{R}_g^N \times \mathcal{R}_{\geq 0}^{N_h} \rightarrow \mathcal{R}$$

The Lagrangian function (cont.)

$$\begin{aligned} \min_{w \in \mathcal{R}^N} \quad & f(w) \\ \text{subject to} \quad & g(w) = 0 \\ & h(w) \geq 0 \end{aligned}$$

In expanded form, we have the Lagrangian function

$$\begin{aligned} \mathcal{L}(w, \lambda, \mu) &= f(w) - \lambda^T g(w) - \mu^T h(w) \\ &= f(w) - \begin{bmatrix} \lambda_1 & \cdots & \lambda_{N_g} \end{bmatrix} \begin{bmatrix} g_1(w) \\ \vdots \\ g_{N_g}(w) \end{bmatrix} - \begin{bmatrix} \mu_1 & \cdots & \mu_{N_h} \end{bmatrix} \begin{bmatrix} h_1(w) \\ \vdots \\ h_{N_h}(w) \end{bmatrix} \end{aligned}$$

The number of multipliers must match the number of constraints

⇝ (For the products $\lambda^T g(w)$ and $\mu^T h(w)$ to be defined)

While λ can take any value, we require the inequality multipliers to be positive ($\mu \geq 0$)

$$\mu \geq 0 = \begin{bmatrix} \mu_1 \geq 0 \\ \vdots \\ \mu_{N_h} \geq 0 \end{bmatrix}$$

Overview (cont.)

Example

$$\min_{x \in \mathcal{R}^2} \quad x_1^2 + x_2^2 \qquad \text{(Objective function)}$$

$$\text{subject to} \quad x_1 - 1 = 0 \qquad \text{(Equality constraints)}$$

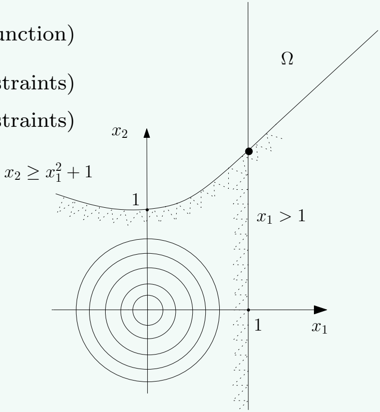
$$x_2 - 1 - x_1^2 \geq 0 \qquad \text{(Inequality constraints)}$$

The feasible set, the set of feasible decisions

$$\Omega = \{x \in \mathcal{R}^2 | h(x) \geq 0, g(x) = 0\}$$

For point $\widetilde{x} \in \Omega$, the Lagrangian function

$$\mathcal{L}(\widetilde{x}, \lambda, \mu) = f(\widetilde{x}) - \lambda^T g(\widetilde{x}) - \mu^T h(\widetilde{x})$$



The Lagrangian function (cont.)

$$\begin{aligned}
\min_{x\in\mathcal{R}^2} \quad & \underbrace{x_1^2 + x_2^2}_{f(x)} && \text{(Objective function)}\\
\text{subject to} \quad & \underbrace{x_1 - 1}_{g(x)} = 0 && \text{(Equality constraints)}\\
& \underbrace{x_2 - 1 - x_1^2}_{h(x)} \geq 0 && \text{(Inequality constraints)}
\end{aligned}$$

The Lagrangian function in expanded form, for any feasible pair $\widetilde{x} = (\widetilde{x}_1, \widetilde{x}_2) \in \Omega$

$$\begin{aligned}
\mathcal{L}(\widetilde{x}, \lambda, \mu) &= f(\widetilde{x}) - \lambda^T g(\widetilde{x}) - \mu^T h(\widetilde{x})\\
&= f(\widetilde{x}) - [\lambda_1]^T [g_1(\widetilde{x})] - [\mu_1]^T [h_1(\widetilde{x})]\\
&= (\widetilde{x}_1^2 + \widetilde{x}_2^2) - \lambda_1(\widetilde{x}_1 - 1) - \mu_1(\widetilde{x}_2 - 1 - \widetilde{x}_1^2)
\end{aligned}$$

□

The Lagrangian function (cont.)

Lower-bound property of the Lagrangian function

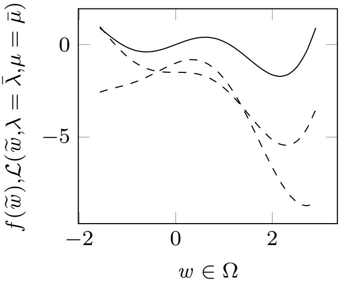
For any feasible point $\widetilde{w} \in \Omega$, for any λ and for any $\mu \geq 0$, we have the lower-bound

$$\mathcal{L}(\widetilde{w}, \lambda, \mu) = f(\widetilde{w}) \underbrace{-\lambda^T \underbrace{g(\widetilde{w})}_{=0}}_{=0} \underbrace{- \underbrace{\mu^T}_{\geq 0} \underbrace{h(\widetilde{w})}_{\geq 0}}_{\leq 0}$$

$$\leq f(\widetilde{w})$$

Because $w^* \in \Omega$, thus we also have

$$\mathcal{L}(w^*, \lambda, \mu) \leq f(w^*)$$



For w in the feasible set, the objective function is larger than the Lagrangian function

- (If $\widetilde{w}$ is a primal minimiser, then the lower-bound will be retained)

Example

$$\min_{x \in \mathcal{R}^2} \quad x_1^2 + x_2^2 \qquad \text{(Objective function)}$$

$$\text{subject to} \quad x_1 - 1 = 0 \qquad \text{(Equality constraints)}$$

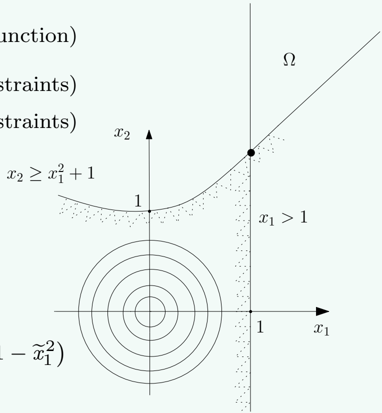
$$x_2 - 1 - x_1^2 \geq 0 \qquad \text{(Inequality constraints)}$$

The feasible set

$$\Omega = \{x \in \mathcal{R}^2 | h(x) \geq 0, g(x) = 0\}$$

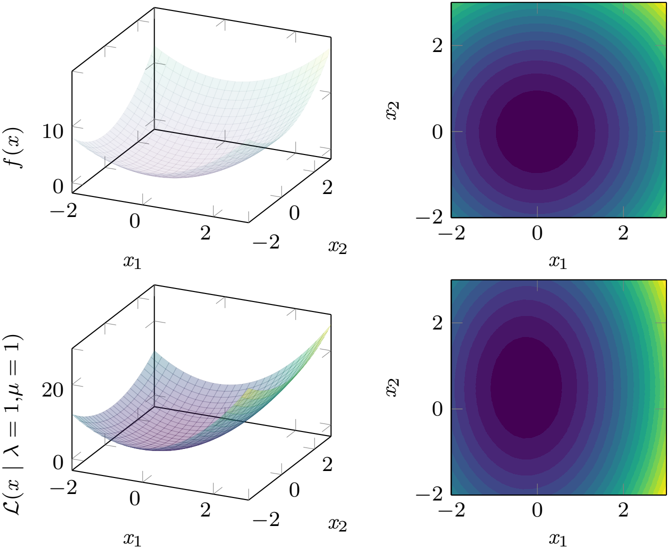
The Lagrangian function

$$\mathcal{L}(\widetilde{x}, \lambda, \mu) = \widetilde{x}_1^2 + \widetilde{x}_2^2 - \lambda_1 (\widetilde{x}_1 - 1) - \mu_1 \left(\widetilde{x}_2 - 1 - \widetilde{x}_1^2\right)$$

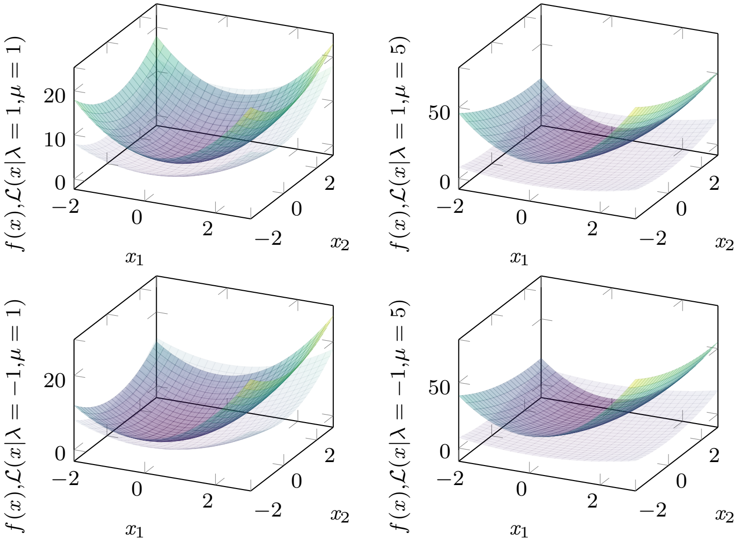


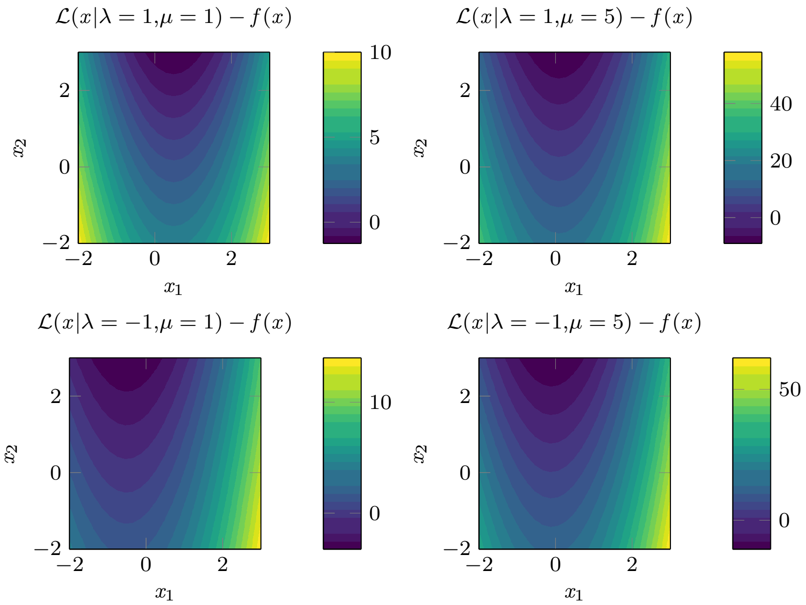
For any point $\widetilde{x} \in \Omega$ and for any λ and any $\mu \geq 0$, we have the lower-bound property

$$\mathcal{L}(\widetilde{x}, \lambda, \mu) \leq f(\widetilde{x})$$


$f(x)$
10
0
−2
0
2
x_1
−2
0
2
x_2
2
0
−2
x_2
−2
0
2
x_1
$\mathcal{L}(x \mid \lambda = 1, \mu = 1)$
20
0
−2
0
2
x_1
−2
0
2
x_2
2
0
−2
x_2
−2
0
2
x_1

For different pairs $(\lambda, \mu_{\geq 0})$ and for any $\tilde{x} \in \Omega$, we always have that $\mathcal{L}(\tilde{x}, \lambda, \mu) \leq f(\tilde{x})$




$\mathcal{L}(x|\lambda = 1, \mu = 1) - f(x)$
$\mathcal{L}(x|\lambda = 1, \mu = 5) - f(x)$
$\mathcal{L}(x|\lambda = -1, \mu = 1) - f(x)$
$\mathcal{L}(x|\lambda = -1, \mu = 5) - f(x)$
x_1
x_2
−2
0
2
10
5
0
40
20
50

The Lagrangian function | Duality

Consider some fixed pair of multipliers $\bar{\lambda}$ and $\bar{\mu} \geq 0$, we define the **Lagrange dual function**

$$q\left(\bar{\lambda}, \bar{\mu}\right) = \inf_{w \in \mathcal{R}^N} \quad \mathcal{L}\left(w | \lambda = \bar{\lambda}, \mu = \bar{\mu}\right)$$

Also the Lagrange dual function is a scalar function

$$q : \mathcal{R}^{N_g} \times \mathcal{R}^{N_h}_{\geq 0} \rightarrow \mathcal{R}$$

Let w^* be the unconstrained (in $\mathcal{R}^N$) minimiser of the Lagrangian function $\mathcal{L}\left(w | \bar{\lambda}, \bar{\mu}\right)$

$$w^* = w^*\left(\bar{\lambda}, \bar{\mu}\right)$$

Because we minimised out w, the infimum is $\mathcal{L}\left(w^*(\bar{\lambda}, \bar{\mu}) | \bar{\lambda}, \bar{\mu}\right) = q\left(\bar{\lambda}, \bar{\mu}\right)$

- At any feasible point $\widetilde{w} \in \Omega$ and fixed multipliers $(\bar{\lambda}, \bar{\mu})$, we have

$$\mathcal{L}\left(\widetilde{w} | \bar{\lambda}, \bar{\mu}\right) \geq \underbrace{\mathcal{L}\left(w^*(\bar{\lambda}, \bar{\mu}) | \bar{\lambda}, \bar{\mu}\right)}_{= q\left(\bar{\lambda}, \bar{\mu}\right)}$$

The Lagrangian function | Duality (cont.)

Lower-bound property of the Lagrange dual function

For any pair of multipliers λ and $\mu \geq 0$ and for any feasible point $\widetilde{w} \in \Omega$, we have that

$$\underbrace{\mathcal{L}\left(\widetilde{w}, \lambda, \mu\right) \leq f\left(\widetilde{w}\right)}_{\text{lower-bound property}}$$

For some pair $(\bar{\lambda}, \bar{\mu} \geq 0)$ and for any feasible point $\widetilde{w}$, we have

$$\underbrace{q\left(\bar{\lambda}, \bar{\mu}\right) = \mathcal{L}\left(w^*(\bar{\lambda}, \bar{\mu})|\bar{\lambda}, \bar{\mu}\right) \leq \mathcal{L}\left(\widetilde{w}, \bar{\lambda}, \bar{\mu}\right)}_{\text{infimum property}}$$

Combining these two inequalities, we have

$$q\left(\bar{\lambda}, \bar{\mu}\right) \leq \mathcal{L}\left(\widetilde{w}, \bar{\lambda}, \bar{\mu}\right) \leq f\left(\widetilde{w}\right)$$

Because $p^* = f(w^*)$ and $f(w^*) \leq f(\widetilde{w})$, we have $q\left(\bar{\lambda}, \bar{\mu}\right) \leq \mathcal{L}\left(\widetilde{w}, \bar{\lambda}, \bar{\mu}\right) \leq f(w^*) \leq f\left(\widetilde{w}\right)$

$$q\left(\bar{\lambda}, \bar{\mu}\right) \leq p^*$$

Lagrange dual functions $q\left(\lambda, \mu\right)$ provide a lower-bound to primal optimal values p^*

At the global minimiser $\widetilde{w} = w^*$, a feasible point, we have

$$\begin{aligned} q\left(\lambda, \mu\right) &\leq f\left(w^*\right) \\ &= p^* \end{aligned}$$

The Lagrangian function | Duality

The Lagrange dual function $q(\lambda, \mu)$ does not depend on primal decision variables w

- Sometimes it is possible to compute the Lagrange dual function explicitly

---

Concavity of the Lagrange dual function

The Lagrange dual function is always a concave function, also for non-convex problems

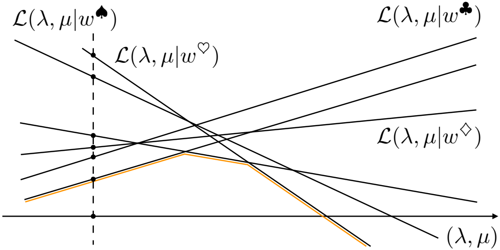
- Therefore, we have that $-q(\lambda, \mu)$ is a convex function

The Lagrangian function | Duality

For any fixed w, the Lagrangian function $\mathcal{L}(\lambda, \mu|w)$ is an affine function of λ and μ

$$\mathcal{L}\left(\lambda, \mu|w\right) = f\left(w\right) - \lambda^T g\left(w\right) - \mu^T h\left(w\right)$$

Visually, consider a set of points $\{w\}$ and associated Lagrangian functions $\{\mathcal{L}(\lambda, \mu|w)\}$



For fixed λ, μ, the dual function

$$q\left(\lambda, \mu\right) = \inf_{w \in \mathcal{R}^N} \quad \mathcal{L}\left(w|\lambda, \mu\right)$$

Or, equivalently

$$-q\left(\lambda, \mu\right) = \sup_{w \in \mathcal{R}^N} -\mathcal{L}\left(w|\lambda, \mu\right)$$

$-q\left(\lambda, \mu\right)$ is the supremum of affine, thus convex, functions in the dual variables (λ, μ)

- The supremum over a set of convex functions is a convex function
- (The epigraph is the intersection of convex sets)

The Lagrangian function | Duality (cont.)

The Lagrange dual function provides an understimate of the primal global minimiser

- The value of the dual function that is closest is achieved when q is maximised
- It is interesting to understand how close to p^* does $q(\lambda, \mu)$ actually get

---

Dual optimisation problem

The best lower-bound d^* is obtained by maximising the Lagrange dual function $q(\lambda, \mu)$

$$\begin{aligned} \max_{\substack{\lambda \in \mathcal{R}^{N_g} \\ \mu \in \mathcal{R}^{N_h}}} \quad & q(\lambda, \mu) \\ \text{subject to} \quad & \mu \geq 0 \end{aligned}$$

The dual optimisation problem is itself a constrained optimisation problem

- It is defined as a convex (concave) maximisation problem
- The decision variables are the dual variables λ and μ

The convexity of the dual optimisation problem is independent of the primal problem

The Lagrangian function | Duality (cont.)

The best lower-bound d^* is obtained by maximising the Lagrange dual function $q(\lambda, \mu)$

$$d^* = \left(\max_{\substack{\lambda \in \mathcal{R}^{N_g} \\ \mu \in \mathcal{R}^{N_h}}} \quad q(\lambda, \mu), \text{ s.t. } \mu \geq 0 \right)$$

For any general nonlinear programs, we have the **weak-duality** result

$$d^* \leq p^*$$

For any convex nonlinear programs[1], we have **strong-duality** result

$$d^* = p^*$$

[1] Slater's constraint qualification conditions must also be satisfied.

The Lagrangian function | Duality (cont.)

Example

Strictly convex quadratic program

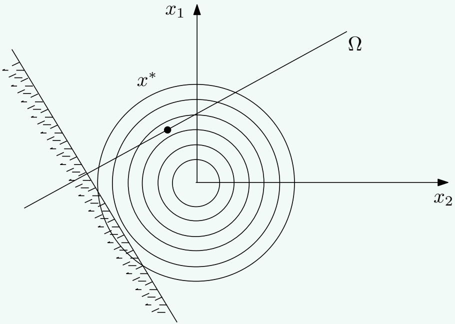
Consider a strictly convex quadratic program ($B \succ 0$) in primal form

The primal optimisation problem

$$\min_{x \in \mathcal{R}^N} \quad c^T x + \frac{1}{2} x^T B x$$
$$\text{subject to} \quad Ax - b = 0$$
$$Cx - d \geq 0$$

The primal global minimum

$$\rightsquigarrow p^*$$



We are interested in the Lagrange dual function $q(\lambda, \mu)$

The Lagrangian function | Duality (cont.)

$$\begin{aligned} \min_{x \in \mathcal{R}^N} \quad & \underbrace{c^T x + \frac{1}{2} x^T B x}_{f(w)} \\ \text{subject to} \quad & \underbrace{Ax - b}_{g(x)} = 0 \\ & \underbrace{Cx - d}_{h(x)} \geq 0 \end{aligned}$$

For the Lagrangian function, we have

$$\begin{aligned} \mathcal{L}(x, \lambda, \mu) &= \underbrace{c^T x + \frac{1}{2} x^T B x}_{f(x)} - \underbrace{\lambda^T (Ax - b)}_{\lambda^T g(x)} - \underbrace{\mu^T (Cx - d)}_{\mu^T h(x)} \\ &= c^T x + \frac{1}{2} x^T B x - \lambda^T A x + \lambda^T b - \mu^T C x + \mu^T d \\ &= \underbrace{\lambda^T b + \mu^T d}_{\text{constant in } x} + \underbrace{(c - A^T \lambda - C^T \mu)^T x}_{\text{linear in } x} + \underbrace{\frac{1}{2} x^T B x}_{\text{quadratic in } x} \end{aligned}$$

The Lagrangian function | Duality (cont.)

$$\mathcal{L}(x, \lambda, \mu) = \lambda^T b + \mu^T d + (c - A^T\lambda - C^T\mu)^T x + \frac{1}{2}x^T B x$$

The Lagrange dual function $q(\lambda, \mu)$ is defined as infimum of the Lagrangian function

- The minimisation is with respect to the primal variables x

We have,

$$\begin{aligned} q(\lambda, \mu) &= \inf_{x \in \mathcal{R}^N} \left(\lambda^T b + \mu^T d + (c - A^T\lambda - C^T\mu)^T x + \frac{1}{2}x^T B x \right) \\ &= \lambda^T b + \mu^T d + \underbrace{\inf_{x \in \mathcal{R}^N} \left((c - A^T\lambda - C^T\mu)^T x + \frac{1}{2}x^T B x \right)}_{\text{unconstrained quadratic program}} \\ &= \lambda^T b + \mu^T d - \frac{1}{2}\left(c - A^T\lambda - C^T\mu\right)^T B^{-1} \left(c - A^T\lambda - C^T\mu\right) \end{aligned}$$

We used the fact that for general unconstrained quadratic problems $f(x^*) = \frac{1}{2}c^T B^{-1} c$

The Lagrangian function | Duality (cont.)

$$q(\lambda, \mu) = \lambda^T b + \mu^T d - \frac{1}{2}\left(c - A^T\lambda - C^T\mu\right)^T B^{-1}\left(c - A^T\lambda - C^T\mu\right)$$

After rearranging, we formulate the dual optimisation problem

$$\max_{\substack{\lambda \in \mathcal{R}^{N_h} \\ \mu \in \mathcal{R}^{N_g}}} \quad -\frac{1}{2}c^T B^{-1} c + \begin{bmatrix} b + AB^{-1}c \\ d + CB^{-1}c \end{bmatrix}^T \begin{bmatrix} \lambda \\ \mu \end{bmatrix} - \frac{1}{2}\begin{bmatrix} \lambda \\ \mu \end{bmatrix}^T \begin{bmatrix} A \\ C \end{bmatrix} B^{-1} \begin{bmatrix} A \\ C \end{bmatrix}^T \begin{bmatrix} \lambda \\ \mu \end{bmatrix}$$

$$\text{subject to} \quad \mu \geq 0$$

The objective function is concave, the dual problem is a convex quadratic program

The term $(-1/2)c^T B^{-1} c$ is constant with respect to the dual variables

- It is retained to verify the strong duality result, $d^* = p^*$

□

The Lagrangian function | Duality (cont.)

Example

Linear program

The primal optimisation problem

$$\begin{aligned} \min_{w \in \mathcal{R}^N} \quad & c^T w \\ \text{subject to} \quad & Aw - b = 0 \\ & Cx - d \geq 0 \end{aligned}$$

The primal global minimum

$$\rightsquigarrow p^*$$

We are interested in the Lagrange dual function $q(\lambda, \mu)$

The Lagrangian function | Duality (cont.)

$$\begin{aligned} \min_{w \in \mathcal{R}^N} \quad & c^T w \\ \text{subject to} \quad & Aw - b = 0 \\ & Cx - d \geq 0 \end{aligned}$$

For the Lagrangian function, we can write

$$\begin{aligned} \mathcal{L}(w, \lambda, \mu) &= c^T w - \lambda^T (Aw - b) - \mu^T (Cw - d) \\ &= \underbrace{\lambda^T b + \mu^T d}_{\text{constant in } x} + \underbrace{\left(c - A^T \lambda - C^T \mu\right) w}_{\text{linear in } x} \end{aligned}$$

The Lagrange dual function, as infimum of the Lagrangian function

$$\begin{aligned} q(\lambda, \mu) &= \lambda^T b + \mu^T d + \underbrace{\inf_{w \in \mathcal{R}^N} \quad \left(c - A^T \lambda - C^T \mu\right) w}_{\text{unconstrained linear program}} \\ &= \lambda^T b + \mu^T d + \begin{cases} 0, & c - A^T \lambda - C^T \mu = 0 \\ -\infty, & \text{otherwise} \end{cases} \end{aligned}$$

The Lagrangian function | Duality (cont.)

$$q(\lambda, \mu) = \lambda^T b + \mu^T d + \begin{cases} 0, & c - A^T\lambda - C^T\mu = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

The Lagrange dual function $q(\lambda, \mu)$ equals $-\infty$ at all points $(\widetilde{\lambda}, \widetilde{\mu})$ that do not satisfy the linear equality $c - A^T\lambda - C^T\mu = 0$, these points can be treated as infeasible points

We use this observation to formulate the the dual optimisation problem,

$$\begin{aligned} \max_{\substack{\lambda \in \mathcal{R}^{N_h} \\ \mu \in \mathcal{R}^{N_g}}} \quad & \begin{bmatrix} b & d \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \\ \text{subject to} \quad & c - A^T\lambda - C^T\mu = 0 \\ & \mu \geq 0 \end{aligned}$$

□

Optimality conditions

Nonlinear optimisation

Optimality conditions | Unconstrained problems

Consider the unconstrained optimisation problem with $f : \mathcal{R}^N \rightarrow \mathcal{R}$ and $f \in \mathcal{C}^1(\mathcal{R}^N)$

$$\min_{w \in \mathcal{R}^N} \quad f(w)$$

We are imprecisely assuming that the domain of definition of function f is $\mathcal{R}^N$

- More precisely, the function is defined only on some set $\mathcal{D} \subseteq \mathcal{R}^N$

That is, we re-write the unconstrained optimisation problem

$$\min_{w \in \mathcal{D}} \quad f(w)$$

Optimality conditions | Unconstrained problems (cont.)

$$\min_{w \in \mathcal{D}} \quad f(w)$$

First-order necessary optimality conditions

If point $w^* \in \mathcal{D}$ is a local minimiser, then the first-order necessary condition holds

$$\nabla f(w^*) = 0$$

A point w^* such that $\nabla f(w^*) = 0$ is a **stationary point**

By contradiction, assume that the local minimiser w^* would be such that $\nabla f(w^*) \neq 0$

- Then, there is a direction $-\nabla f(w^*)$ that would be a descent direction

$$\nabla f(w^*)^T (-\nabla f(w^*)) = \underbrace{-\underbrace{\|\nabla f(w^*)\|_2^2}_{>0}}_{<0}$$

In the vicinity of w^*, for a point $\widetilde{w} = w^* + \lambda(w' - w^*)$ along such descent direction

$$f\left(w^* + \lambda(w' - w^*)\right) \approx f(w^*) + \lambda \underbrace{\nabla f(w^*)^T (w' - w^*)}_{<0}$$

$$< f(w^*) \qquad \text{(a contradiction for a local minimiser)}$$

Optimality conditions | Unconstrained problems (cont.)

$$\min_{w \in \mathcal{D}} \quad f(w)$$

Second-order necessary optimality conditions

If point $w^* \in \mathcal{D}$ is a local minimiser, then the second-order necessary condition holds

$$\nabla^2 f(w^*) \succeq 0$$

Assume the existence of direction $(w' - w^*)$ such that $(w' - w^*)^T \nabla^2 f(w^*)(w' - w^*) < 0$

- Along direction $(w' - w^*)$ the value of the objective function would diminish

In the vicinity of w^*, for a point $\widetilde{w} = w^* + \lambda(w' - w^*)$ along such descent direction

$$f\left(w^* + \lambda(w' - w^*)\right) \approx$$

$$f(w^*) + \lambda \underbrace{\nabla f(w^*)^T (w' - w^*)}_{=0} + \frac{1}{2}\lambda^2 \underbrace{(w' - w^*)^T \nabla^2 f(w^*)(w' - w^*)}_{<0}$$

$$< f(w^*) \qquad \text{(a contradiction for a local minimiser)}$$

Optimality conditions | Unconstrained problems (cont.)

Second-order sufficient optimality conditions

The sufficient second-order condition to have a strict local minimiser

$$\nabla^2 f(w^*) \succ 0$$

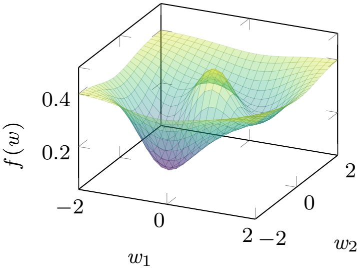
□

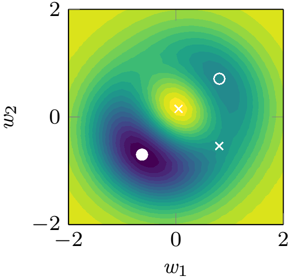
Optimality conditions | Unconstrained problems (cont.)

Example

Consider the unconstrained optimisation problem

$$\min_{w \in \mathcal{R}^2} \quad \frac{2}{5} - \frac{1}{10}\left(5w_1^2 + 5w_2^2 + 3w_1w_2 - w_1 - 2w_2\right) e^{\left(-\left(w_1^2+w_2^2\right)\right)}$$





Equality constraints

Optimality conditions

Optimality conditions | Equality constraints

Consider the equality constrained optimisation problem in the general form

$$\min_{w \in \mathcal{R}^N} \quad f(w)$$
$$\text{subject to} \quad g(w) = 0$$

- We assume that $f : \mathcal{R}^N \rightarrow \mathcal{R}$ and $g : \mathcal{R}^N \rightarrow \mathcal{R}^{N_g}$ are smooth functions
- The feasible set is $\Omega = \{w \in \mathcal{R}^N | g(w) = 0\}$, a differentiable manifold

---

We are interested in the optimality conditions for this class of optimisation problems

- To have a condition $\nabla f(w) = 0$ (or $\nabla f(w) = 0$ and $\nabla^2 f(w) \succeq 0$) is not enough
- Variations in other feasible directions must not improve the objective function

Optimality conditions | Equality constraints (cont.)

To formulate the optimality conditions, we need two notions from differential geometry

- The **tangent vector** to the feasible set Ω
- The **tangent cone** to the feasible set Ω

These notions will allow for a local characterisation of the feasible set

For (standard, well-behaved) equality constrained optimisation problems, the set of all the tangent vectors to the feasibility set Ω at a feasible point w^* form a vector space

- The **tangent space**

Optimality conditions | Equality constraints (cont.)

Remember the equality constraint function, each component function need be smooth

$$g(w) = \underbrace{\begin{bmatrix} g_1(w) \\ \vdots \\ g_{n_g}(w) \\ \vdots \\ g_{N_g}(w) \end{bmatrix}}_{N_g \times 1}$$

Each function is required to be at least differentiable once, to compute the Jacobian

---

Jacobian of the equality constraints

The Jacobian of the equality constraint functions is a rectangular $(N_g \times N)$ matrix

- It collects (transposed) gradients $\nabla g_{n_g}(w)$ of component functions $g_{n_g}(w)$

Optimality conditions | Equality constraints (cont.)

$$g(w) = \underbrace{\begin{bmatrix} g_1(w) \\ \vdots \\ g_{n_g}(w) \\ \vdots \\ g_{N_g}(w) \end{bmatrix}}_{N_g \times 1}$$

More explicitly, the gradient vector of an equality constraint function $g_{n_g}(w)$

$$\nabla g_{n_g}(w) = \underbrace{\begin{bmatrix} \partial g_{n_g}(w_1, \dots, w_N)/\partial w_1 \\ \vdots \\ \partial g_{n_g}(w_1, \dots, w_N)/\partial w_n \\ \vdots \\ \partial g_{n_g}(w_1, \dots, w_N)/\partial w_N \end{bmatrix}}_{N \times 1}$$

Each gradient $\nabla g_{n_g}(w)$ is a column-vector of size $(N \times 1)$

Optimality conditions | Equality constraints (cont.)

In the Jacobian of $g(w)$, the gradients are transposed and arranged along the rows

That is,

$$\nabla g(w)^T = \begin{bmatrix} \nabla g_1(w)^T \\ \nabla g_2(w)^T \\ \vdots \\ \nabla g_{n_g}(w)^T \\ \vdots \\ \nabla g_{N_g}(w^*)^T \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} \begin{bmatrix} \partial g_1(w)/\partial w_1 & \cdots & \partial g_1(w)/\partial w_n & \cdots & \partial g_1(w)/\partial w_N \end{bmatrix} \\ \begin{bmatrix} \partial g_2(w)/\partial w_1 & \cdots & \partial g_2(w)/\partial w_n & \cdots & \partial g_2(w)/\partial w_N \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \partial g_{n_g}(w)/\partial w_1 & \cdots & \partial g_{n_g}(w)/\partial w_n & \cdots & \partial g_{n_g}(w)/\partial w_N \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \partial g_{N_g}(w)/\partial w_1 & \cdots & \partial g_{N_g}(w)/\partial w_n & \cdots & \partial g_{N_g}(w)/\partial w_N \end{bmatrix} \end{bmatrix}}_{N_g \times N}$$

Optimality conditions | Equality constraints (cont.)

$$\nabla g(w)^T = \underbrace{\begin{bmatrix} \nabla g_1(w)^T \\ \nabla g_2(w)^T \\ \vdots \\ \nabla g_{n_g}(w)^T \\ \vdots \\ \nabla g_{N_g}(w^*)^T \end{bmatrix}}_{N_g \times N}$$

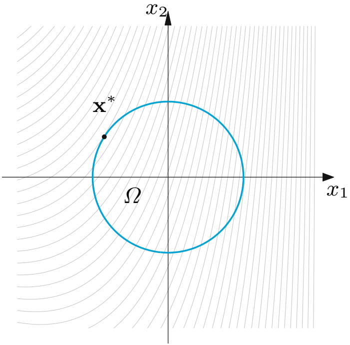
We denote the Jacobian matrix of vector-valued multivariate function $g(w)$ as $\nabla g(w)^T$

- Alternative notation used for the Jacobian, $J_g(w)$ and $\dfrac{\partial g(w)}{\partial w}$

Optimality conditions | Equality constraints (cont.)

Example

Consider the minimisation of some function $f(w)$ under some equality constraint $g(w)$



Let $f : \mathcal{R}^2 \rightarrow \mathcal{R}$

$$f(x) = \frac{3}{5} x_1^2 + \frac{1}{2} x_1 x_2 - x_2 + 3 x_1$$

Let $g : \mathcal{R}^2 \rightarrow \mathcal{R}$

$$g(x) = x_1^2 + x_2^2 - 1$$

The feasible set

$$\Omega = \{x \in \mathcal{R}^2 : g(x) = 0\}$$

When on the constraint(s), feasibility is satisfied when moving along tangent directions

- Optimality conditions must be verified along these directions

Optimality conditions | Equality constraints (cont.)

Tangent vector

A vector $p \in \mathcal{R}^N$ is a tangent vector to the feasible set Ω at point $w^* \in \Omega \subset \mathcal{R}^N$ if there exists a smooth curve $\overline{w}(t) : [0, \varepsilon) \rightarrow \mathcal{R}^N$ such that the following is true

- The curve for $t = 0$ starts at the feasible point w^*

$$\overline{w}(0) = w^*$$

- The curve is in feasible set for all $t \in [0, \varepsilon)$

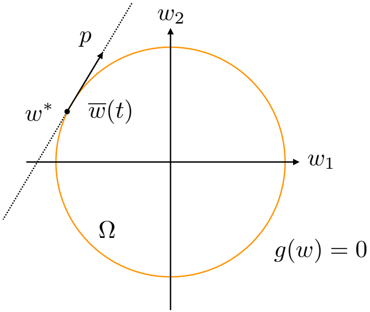
$$\overline{w}(t) \in \Omega, \quad \forall t$$

Vector p is derivative of curve $\overline{w}$, at $t = 0$

$$p = \frac{d\overline{w}(t)}{dt}\Big|_{t=0}$$

Optimality conditions | Equality constraints (cont.)

Curve $\overline{w}(t)$ is parameterised by t, t varies over the infinitesimally small interval $[0, \varepsilon)$



$$\overline{w}(t) = \begin{bmatrix} \overline{w}_1(t) \\ \vdots \\ \overline{w}_N(t) \end{bmatrix}$$

$$t \in [0, \varepsilon)$$

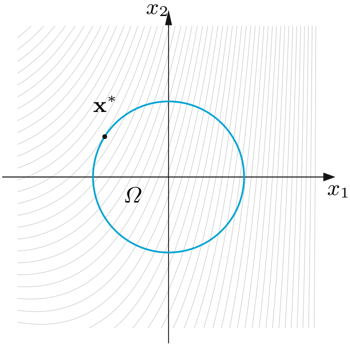
- $w^* \in \Omega$ is where the curve starts, $\overline{w}(t = 0) = w^*$ and ε is small enough
- Thus, the curve $\overline{w}(t)$ remains inside Ω (surely in the limit $\varepsilon \to 0$)

$$p(t) = \frac{d\overline{w}(t)}{dt} = \begin{bmatrix} d\overline{w}_1(t)/dt \\ \vdots \\ d\overline{w}_N(t)/dt \end{bmatrix} = \begin{bmatrix} p_1(t) \\ \vdots \\ p_N(t) \end{bmatrix}$$

Tangent vector p defines a direction along which it is possible move without leaving Ω

Optimality conditions | Equality constraints (cont.)

Example



Consider the problem with feasibility set

$$\Omega = \{x \in \mathcal{R}^2 : x_1^2 + x_2^2 - 1 = 0\}$$

The points x^* on the unit circle

$$x^* = \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix}$$

An alternative characterisation of a feasible point x^*, for some fixed $\alpha^* \in [0, 2\pi]$

$$x^*(\alpha) = \begin{bmatrix} \cos(\alpha^*) \\ \sin(\alpha^*) \end{bmatrix}$$

Optimality conditions | Equality constraints (cont.)

For a fixed α^* (fixed x^*) and some $\omega \in \mathcal{R}$, we construct a feasible curve $\overline{x}(t)$ from x^*

$$\overline{x}(t|\alpha^*,\omega) = \begin{bmatrix} \cos(\alpha^* + \omega t) \\ \sin(\alpha^* + \omega t) \end{bmatrix}$$

We can also determine the tangent vectors $p(t)$ to the curve $\overline{x}(t)$, along the curve

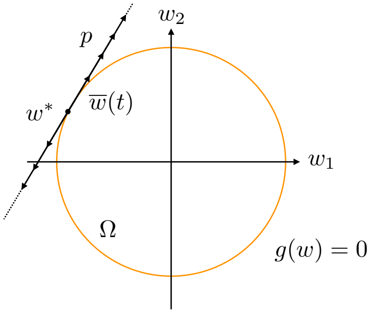
$$\begin{aligned} p_{\alpha^*,\omega}(t) &= \frac{d\overline{x}(t|\alpha^*,\omega)}{dt} \\ &= \begin{bmatrix} -\omega \sin(\alpha^* + \omega t) \\ \omega \cos(\alpha^* + \omega t) \end{bmatrix} \\ &= \omega \begin{bmatrix} -\sin(\alpha^* + \omega t) \\ \cos(\alpha^* + \omega t) \end{bmatrix} \end{aligned}$$

The tangent vector at $t = 0$ (or, at x^*),

$$\begin{aligned} p_\omega &= \frac{d\overline{x}(t)}{dt}\Big|_{t=0} \\ &= \omega \begin{bmatrix} -\sin(\alpha^*) \\ \cos(\alpha^*) \end{bmatrix} \end{aligned}$$

□

Optimality conditions | Equality constraints (cont.)



Tangent cone

The tangent cone $T_\Omega(w^*)$ of the feasible set Ω at some feasible point $w^* \in \Omega \subset \mathcal{R}^N$ is the set of all the tangent vectors at w^*

- 'If p is a tangent vector, then also $2p$ is a tangent vector, ...'

Sometimes the set of elements of the tangent cone define a space, the **tangent space**

Optimality conditions | Equality constraints (cont.)

To construct a smooth curve $\overline{w}(t)$ that satisfies the conditions needed to define tangent vectors, we can consider the equality constraint $g(w)$ and its Taylor's expansion at w^*

Consider the first-order Taylor's series expansion of function g at point w^*

$$g(w) = \underbrace{g(w^*)}_{=0} + \nabla g(w^*)^T (w - w^*) + \mathcal{O}\left((w - w^*)^2\right)$$

- We know g and we can compute its gradients ($\rightsquigarrow$ Jacobian)

Similarly, we construct the approximated curve and at $t = 0$ (at point w^*) we have

$$\begin{aligned}\overline{w}(t) &= \underbrace{w(0)}_{w^*} + \underbrace{\left.\frac{d\overline{w}(t)}{dt}\right|_{t=0}}_{p} (t - 0) + \mathcal{O}\left((t - 0)^2\right)\\ &\approx w^* + tp\end{aligned}$$

We can then construct a direction such that from w^* it is feasible, up to the first-order

$$g(w) = \underbrace{\underbrace{g(w^*)}_{=0} + \underbrace{\nabla g(w^*)^T (w - w^*)}_{=0} + \mathcal{O}\left((w - w^*)^2\right)}_{=0}$$

Optimality conditions | Equality constraints (cont.)

$$g(w) \approx \underbrace{g(w^*)}_{=0} + \underbrace{\nabla g(w^*)^T (w - w^*)}_{=0}$$

We consider the tangent vectors p that projected by the Jacobian $\nabla g(w^*)^T$ are zero

$$\nabla g(w^*)^T p = 0$$

Tangent directions p that satisfy the orthogonality condition are feasible, $g(\overline{w}(t)) = 0$

- If the constraints at w^* are zero, along p they will remain zero (up to first-order)

The feasible tangent directions are in the null-space of the Jacobian $J_g(w) = \nabla g(w^*)^T$

This suggests a criterion for building a possible tangent cone $T_\Omega(w^*)$

$$T_\Omega(w^*) = \{p \in \mathcal{R}^N : \nabla g(w^*)^T p = 0\}$$

Optimality conditions | Equality constraints (cont.)

The collection of tangent directions to Ω that are orthogonal to the equality constraints

$$\mathcal{F}_{\Omega}\left(w^*\right) = \{p \in \mathcal{R}^N : \nabla g_{n_g}\left(w^*\right)^T p = 0, \text{ with } n_g = 1, 2, \dots, N_g\}$$

The collection in this set is denoted as the **linearised feasible cone for equality constraints**

- For equality constrained problems that are smooth, $\mathcal{F}_{\Omega}\left(w^*\right)$ is a space
- More generally, the set of all tangent vectors to Ω is just a cone

---

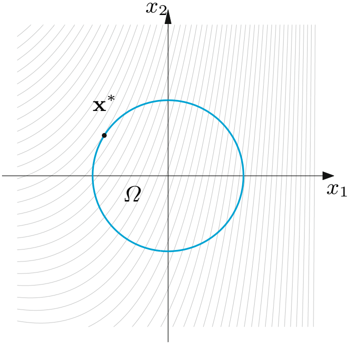
In general (with inequality constraints), it is difficult to characterise the tangent cone

- The linearised feasible cone for equality constraints is a good proxy to it

Though, in general, we have

$$\mathcal{F}_{\Omega}(w) \neq T_{\Omega}\left(w\right)$$

Example



Consider the problem with feasibility set

$$\Omega = \{x \in \mathcal{R}^2 | x_1^2 + x_2^2 - 1 = 0\}$$

A possible tangent vector $p_\omega(x^*)$

$$p_\omega(x^*) = \omega \begin{bmatrix} -\sin(\alpha^*) \\ \cos(\alpha^*) \end{bmatrix}$$

The vector space is mono-dimensional

The vector space corresponds to the tangent cone, it is constructed by choosing $\omega \in \mathcal{R}$

$$T_\Omega(x^*) = \{p \in \mathcal{R}^2 : p = \omega \begin{bmatrix} -\sin(\alpha^*) \\ \cos(\alpha^*) \end{bmatrix}, \text{ with } \omega \in \mathcal{R}\}$$

The tangent vectors are orthogonal to the gradient vector of the constraint function

$$\nabla g(x^*) = 2 \begin{bmatrix} \cos(\alpha^*) \\ \sin(\alpha^*) \end{bmatrix}$$

□

Optimality conditions | Equality constraints (cont.)

First-order necessary optimality conditions (I)

Consider the equality constrained optimisation problem

$$\min_{w \in \mathcal{R}^N} \quad f(w)$$
$$\text{subject to} \quad g(w) = 0$$

A point w^* is a local minimiser, if $w^* \in \Omega$ and for all tangents $p \in T_\Omega(w^*)$, we have

$$\nabla f(w^*)^T p \geq 0$$

When we consider the directions that are in the tangent cone $T_\Omega(w^*)$ of point w^* in the feasible set Ω, we must only have directions along which the objective worsens

---

If $\nabla f(w^*)^T p < 0$, then there would also exist some feasible curve $\overline{w}(t)$ such that

$$\left.\frac{df(\overline{w}(t))}{dt}\right|_{t=0} = \nabla f(w^*)^T p$$
$$< 0$$

There would exist a feasible descent direction, along which the objective improves

Optimality conditions | Equality constraints (cont.)

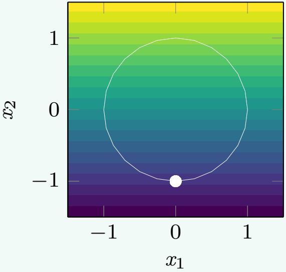
Example

Consider the constrained optimisation

$$\begin{aligned} \min_{w \in \mathcal{R}^2} \quad & w_2 \\ \text{subject to} \quad & w_1^2 + w_2^2 - 1 = 0 \end{aligned}$$

The minimiser w^*

$$w^* = (0, -1)$$



The gradient vector of the objective function at the minimiser

$$\nabla f\left(w^*\right) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

The gradient at w^* is orthogonal to the tangent space at w^*

- Not true for (most of the) other feasible points

Optimality conditions | Equality constraints (cont.)

We are interested in the conditions under which the identity $\mathcal{F}_{\Omega}(w^*) = T_{\Omega}(w^*)$ holds

- (When the tangent cone is also a tangent (vector) space?)

We say that the **linear independence constraint qualification** (**LICQ**) holds at point w^* if and only if the vectors $\nabla g_{n_g}(w^*)$ are linearly independent, $n_g = 1, \dots, N_g$

- $\{\nabla g_{n_g}(w^*)^T\}$ are the rows of the Jacobian, gradients of the equality constraints

$$\nabla g(w)^T = \underbrace{\begin{bmatrix} \nabla g_1(w)^T \\ \nabla g_2(w)^T \\ \vdots \\ \nabla g_{n_g}(w)^T \\ \vdots \\ \nabla g_{N_g}(w^*)^T \end{bmatrix}}_{N_g \times N}$$

The linear independence qualification is equivalent to requiring $\text{rank}\Big(\nabla g(w^*)^T\Big) = N_g$

- This condition can be satisfied if and only if $N_g \leq N$

Optimality conditions | Equality constraints (cont.)

It can be shown that, in general, the following holds

$$T_\Omega(w^*) \subseteq \mathcal{F}_\Omega(w^*)$$

When LICQ holds, we have

$$T_\Omega(w^*) = \mathcal{F}_\Omega(w^*)$$

We can restate the **first-order optimality conditions (II)**

$$\min_{w \in \mathcal{R}^N} \quad f(w)$$
$$\text{subject to} \quad g(w) = 0$$

Point w^* is a local minimiser, if $w^* \in \Omega$, LICQ holds at w^*, and for all $p \in \mathcal{F}_\Omega(w^*)$

$$\rightsquigarrow \quad \nabla f(w^*)^T p = 0$$

Optimality conditions | Equality constraints (cont.)

We can further restate the **first-order optimality conditions (III)**

$$\begin{aligned} \min_{w \in \mathcal{R}^N} \quad & f(w) \\ \text{subject to} \quad & g(w) = 0 \end{aligned}$$

Point w^* is a local minimiser, if $w^* \in \Omega$, LICQ holds at w^*, and there is a $\lambda^* \in \mathcal{R}^{N_g}$

$$\rightsquigarrow \quad \nabla f(w^*) = \nabla g(w^*) \lambda^*$$

Remember the Lagrangian function for equality constrained problems, we have

$$\mathcal{L}(w, \lambda) = f(w) - \lambda^T g(w)$$

We retrieve the optimality condition, by differentiating

$$\begin{aligned} \nabla_w \mathcal{L}(w^*, \lambda^*) &= \nabla f(w^*) - \nabla g(w^*) \lambda^* \\ &= 0 \end{aligned}$$

This result is important, because we can optimise simultaneously for both w^* and λ^*

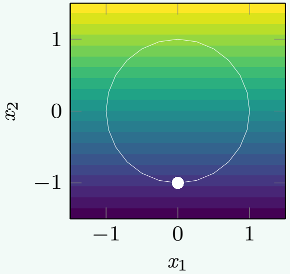
Example

Consider the constrained optimisation

$$\min_{w \in \mathcal{R}^2} \quad w_2$$
$$\text{subject to} \quad w_1^2 + w_2^2 - 1 = 0$$

The Lagrangian function

$$\mathcal{L}\left(w, \lambda\right) = w_2 - \lambda(w_1^2 + w_2^2 - 1)$$



The gradient of $\mathcal{L}\left(w, \lambda\right) = w_2 - \lambda(w_1^2 + w_2^2 - 1)$ with respect to the primal variables w

$$\nabla_w \mathcal{L}\left(w, \lambda\right) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \lambda \begin{bmatrix} 2w_1 \\ 2w_2 \end{bmatrix}$$

The first-order optimality conditions, $g\left(w^*\right)$ and $\nabla_w \mathcal{L}\left(w, \lambda\right) = 0$

$$\begin{aligned} w_1^2 + w_2^2 - 1 &= 0 \\ -2\lambda w_1 &= 0 \\ -2\lambda w_2 + 1 &= 0 \end{aligned}$$

Optimality conditions | Equality constraints (cont.)

Some remarkable facts about first-order optimality conditions and Lagrangian functions

$$\mathcal{L}(w, \lambda) = f(w) - \lambda^T g(w)$$

The gradient of the Lagrangian function with respect to the dual λ equals $-g(w)$

$$\nabla_\lambda \mathcal{L}(w, \lambda) = -g(w)$$

At a minimiser $w^* \in \Omega$, we have $g(w^*) = 0$ and $\nabla_w \mathcal{L}(w^*, \lambda^*) = 0$, or

$$\begin{aligned} \begin{bmatrix} \nabla_w \mathcal{L}(w^*, \lambda^*) \\ \nabla_\lambda \mathcal{L}(w^*, \lambda^*) \end{bmatrix} &= \nabla_{w,\lambda} \mathcal{L}(w^*, \lambda^*) \\ &= 0 \end{aligned}$$

The LICQ condition led to define the **Karhush-Kuhn-Tucker (KKT) conditions**

$$\begin{aligned} \nabla_{w,\lambda} \mathcal{L}(w^*, \lambda^*) &= 0 \\ g(w^*) &= 0 \end{aligned}$$

Optimality conditions | Equality constraints (cont.)

Second-order necessary optimality conditions

$$\begin{aligned} \min_{w \in \mathcal{R}^N} \quad & f(w) \\ \text{subject to} \quad & g(w) = 0 \end{aligned}$$

Point w^* is a local minimiser if $w^* \in \Omega$, LICQ holds at w^*, there exists a $\lambda^* \in \mathcal{R}^{N_g}$ such that $\nabla f(w^*) = \nabla g(w^*)\lambda^*$, and for all tangent vectors $p \in \mathcal{F}_\Omega(w^*)$ we also have

$$p^T \nabla_w^2 \mathcal{L}(w^*, \lambda^*) p \geq 0$$

Second-order sufficient optimality conditions

$$p^T \nabla_w^2 \mathcal{L}(w^*, \lambda^*) p > 0$$

Equality and inequality constraints

Optimality conditions

Optimality conditions | Constrained problems

Consider the equality and inequality constrained optimisation problem in general form

$$\begin{aligned} \min_{x \in \mathcal{R}^N} \quad & f(x) \\ \text{subject to} \quad & g(x) = 0 \\ & h(x) \geq 0 \end{aligned}$$

We assume smooth functions $f : \mathcal{R}^N \to \mathcal{R}$, $g : \mathcal{R}^N \to \mathcal{R}^{N_g}$, and $h : \mathcal{R}^N \to \mathcal{R}^{N_h}$

$$g(w) = \begin{bmatrix} g_1(w) \\ \vdots \\ g_{N_g}(w) \end{bmatrix}$$

$$h(w) = \begin{bmatrix} h_1(w) \\ \vdots \\ h_{N_h}(w) \end{bmatrix}$$

⇝ We have the set of feasible points $\Omega = \{w \in \mathcal{R}^N : g(w) = 0, h(w) \geq 0\}$

To formulate the optimality conditions for these problems, we extend previous notions

Optimality conditions | Constrained problems (cont.)

Tangent vector

A vector $p \in \mathcal{R}^N$ is a tangent vector to the feasible set Ω at point $w^* \in \Omega \subset \mathcal{R}^N$ if there exists a smooth curve $\overline{w}(t) : [0, \varepsilon) \rightarrow \mathcal{R}^N$ such that the following is verified

- The curve for $t = 0$ starts at the feasible point w^*

$$\overline{w}(0) = w^*$$

- The curve is in feasible set for all $t \in [0, \varepsilon)$

$$\overline{w}(t) \in \Omega$$

- Vector p is the derivative of $\overline{w}$ at $t = 0$

$$\frac{d\overline{w}(t)}{dt}\bigg|_{t=0} = p$$

Tangent cone

The tangent cone $T_\Omega(w^*)$ of the feasible set Ω at point $w^* \in \Omega \subset \mathcal{R}^N$ is the set of all the tangent vectors at w^* (same definition, now it requires a different characterisation)

Optimality conditions | Constrained problems (cont.)

With equality constrained problems, we defined the linearised feasible cone $\mathcal{F}_\Omega(w^*)$

- For feasible points w^*, we have first-order necessary optimality conditions

$$\nabla f\left(w^*\right)^T p \geq 0, \quad \text{for all } p \in \mathcal{T}_\Omega(w^*)$$

- Under linear independence constraint qualification (LICQ) conditions

$$T_\Omega(w^*) = \mathcal{F}_\Omega(w^*)$$

To characterise the tangent cone with inequality constrains, we introduce new concepts

Optimality conditions | Constrained problems (cont.)

We need to describe the feasibility set in the neighbourhood of a local minimiser $w^* \in \Omega$

Earlier, we mentioned the notion of **active constraints** and **active set**

Consider the inequality constraint functions

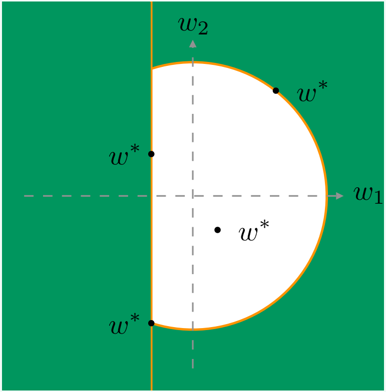
$$h(w) = \begin{bmatrix} h_1(w) \\ \vdots \\ h_{N_h}(w) \end{bmatrix}$$

An inequality constraint $h_{n_g}(w^*) \leq 0$ is said to be an **active inequality constraint** at $w^* \in \Omega$ if and only if $h_{n_g}(w^*) = 0$, otherwise it is an **inactive inequality constraint**

- The index set of active inequality constraints is $\mathcal{A}(w^*) \subset \{1, 2, \dots, N_h\}$
- The index set $\mathcal{A}(w^*)$ of active inequality constraints is the **active set**
- The cardinality of the active set, $N_{\mathcal{A}} = |\mathcal{A}(w^*)|$

Optimality conditions | Constrained problems (cont.)

Example



Determine the active set for the different feasible points w^*

Optimality conditions | Constrained problems (cont.)

The **linearised feasible cone for equality and inequality constraints**

The linearised feasible cone $\mathcal{F}_\Omega(w^*)$ at point $w^* \in \Omega$ is the set of all tangent directions to Ω that are orthogonal to the equality constraints and the active inequality constraints

$$\mathcal{F}_\Omega(w^*) = \{p \in \mathcal{R}^N : \nabla g_{n_g}(w^*)^T p = 0 \underbrace{\text{ with } n_g = 1, \dots, N_g}_{\text{all equalities}}$$
$$\nabla h_{n_h}(w^*)^T p \geq 0 \underbrace{\text{ with } n_h \in \mathcal{A}(w^*)}_{\text{active inequalitites}}\}$$

We require that tangent directions remain inside the feasible set, up to the first order

Optimality conditions | Constrained problems (cont.)

Consider point $w^* \in \Omega$ and the gradient vectors $\{\nabla g_{n_g}(w^*)\}_{n_g=1}^{N_g}$ and $\{\nabla h_{n_h}(w^*)\}_{n_h=1}^{N_h}$

The gradient vectors are the rows of the respective Jacobians, evaluated at point w^*

$$\underbrace{\begin{bmatrix} \nabla g_1(w^*) \\ \vdots \\ \nabla g_{N_g}(w^*) \end{bmatrix}}_{\nabla g(w^*)^T} = \begin{bmatrix} \begin{bmatrix} \partial g_1(w)/\partial w_1 & \partial g_1(w)/\partial w_2 & \cdots & \partial g_1(w)/\partial w_N \end{bmatrix}^T \\ \vdots \\ \begin{bmatrix} \partial g_{N_g}(w)/\partial w_1 & \partial g_{N_g}(w)/\partial w_2 & \cdots & \partial g_{N_g}(w)/\partial w_N \end{bmatrix}^T \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} \nabla h_1(w^*) \\ \vdots \\ \nabla h_{N_h}(w^*) \end{bmatrix}}_{\nabla h(w^*)^T} = \begin{bmatrix} \begin{bmatrix} \partial h_1(w)/\partial w_1 & \partial h_1(w)/\partial w_2 & \cdots & \partial h_1(w)/\partial w_N \end{bmatrix}^T \\ \vdots \\ \begin{bmatrix} \partial h_{N_\mathcal{A}}(w)/\partial w_1 & \partial h_{N_\mathcal{A}}(w)/\partial w_2 & \cdots & \partial h_{N_h}(w)/\partial w_N \end{bmatrix}^T \end{bmatrix}$$

Optimality conditions | Constrained problems (cont.)

At any point $w^* \in \Omega$ in the feasible set, we have that all constraints must be satisfied

$$g(w) = 0$$
$$h(w) \geq 0$$

Moreover, at each active inequality constraint $n_g \in \mathcal{A}(w^*)$ we have

$$\begin{bmatrix} \vdots \\ h_{n_g \in \mathcal{A}}(w^*) \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ 0 \\ \vdots \end{bmatrix}$$

For points w^* on the equality and active inequality constraint, we define

$$\overline{g}(w^*) = \underbrace{\left[\begin{array}{c} g_1(w^*) \\ \vdots \\ g_{N_g}(w^*) \\ \hline \vdots \\ h_{n_g \in \mathcal{A}}(w^*) \\ \vdots \end{array}\right]}_{(N_g + N_{\mathcal{A}}) \times 1}$$

Optimality conditions | Constrained problems (cont.)

We say that the **linear independence constraint qualitification** (**LICQ**) holds at point w^* is and only if vectors $\{\nabla g_{n_g}(w^*)\}$ and $\{h_{n_h \in \mathcal{A}}(w^*)\}$ are linearly independent

That is, when the rank condition on the Jacobian of function $\overline{g}$ holds

$$\text{rank}\left(\frac{\partial \overline{g}(w^*)}{\partial w}\right) = N_g + N_{\mathcal{A}}$$

Importantly, note that inactive inequality constraint do not affect the LICQ coinditions

---

For feasible points $w^* \in \Omega$, we have

$$\mathcal{T}_{\Omega}(w^*) \subset \mathcal{F}_{\Omega}(w^*)$$

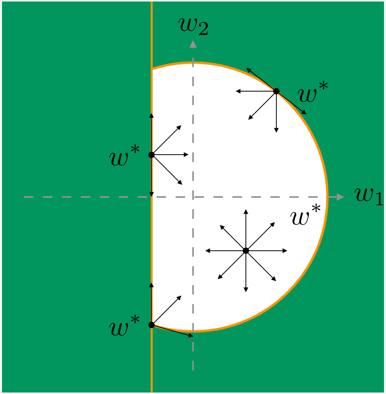
If LICQ holds at w^*, we also have

$$\mathcal{T}_{\Omega}(w^*) = \mathcal{F}_{\Omega}(w^*)$$

Inactive constraints do not affect the tangent cone

Optimality conditions | Constrained problems (cont.)

Example



Determine the tangent cone for the different feasible points w^*

Optimality conditions | Constrained problems (cont.)

First-order necessary optimality conditions (I)

$$\begin{aligned} \min_{w \in \mathcal{R}^N} \quad & f(w) \\ \text{subject to} \quad & g(w) = 0 \\ & h(w) \leq 0 \end{aligned}$$

Point w^* is a local minimiser, if $w^* \in \Omega$, LICQ holds at w^*, and for all $p \in \mathcal{F}_\Omega(w^*)$

$$\rightsquigarrow \quad \nabla f(w^*)^T p \geq 0$$

Optimality conditions | Constrained problems (cont.)

$$\begin{aligned} \min_{w \in \mathcal{R}^N} \quad & f(w) \\ \text{subject to} \quad & g(w) = 0 \\ & h(w) \leq 0 \end{aligned}$$

The LICQ condition leads to define the **Karhush-Kuhn-Tucker (KKT) conditions**

Let w^* be a minimiser of objective function f, given constraint functions g and h

If LICQ holds at w^*, then there exists vectors $\lambda^* \in \mathcal{R}^{N_g}$ and $\mu^* \in \mathcal{R}^{N_h}$ such that

$$\begin{aligned} \nabla f(w^*) - \nabla g(w^*)\lambda^* - \nabla h(w^*)\mu^* &= 0 \\ g(w^*) &= 0 \\ h(w^*) &\geq 0 \\ \mu^* &\geq 0 \\ \mu^*_{n_h} h_{n_h}(w^*) &= 0, \quad n_h = 1, \dots, N_h \end{aligned}$$

First-order necessary optimality conditions (II)

Optimality conditions | Constrained problems (cont.)

$$\underbrace{\nabla f\left(w^*\right)}_{N\times 1} - \underbrace{\nabla g\left(w^*\right)}_{N\times N_g} \underbrace{\lambda^*}_{N_g\times 1} - \underbrace{\nabla h\left(w^*\right)}_{N\times N_h} \underbrace{\mu^*}_{N_h\times 1} = 0$$

$$\underbrace{g\left(w^*\right)}_{N_g\times 1} = 0$$

$$\underbrace{h\left(w^*\right)}_{N_h\times 1} \geq 0$$

$$\underbrace{\mu^*}_{N_h\times 1} \geq 0$$

$$\underbrace{\mu^*_{n_h}}_{1\times 1} \underbrace{h_{n_h}\left(w^*\right)}_{1\times 1} = 0, \quad n_h = 1,\dots,N_h$$

We defined the following terms,

$$\nabla f\left(w^*\right) = \left(\frac{\partial f\left(w^*\right)}{\partial w}\right)^T$$

$$\nabla g\left(w^*\right) = \left(\frac{\partial g\left(w^*\right)}{\partial w}\right)^T$$

$$\nabla h\left(w^*\right) = \left(\frac{\partial h\left(w^*\right)}{\partial w}\right)^T$$

Optimality conditions | Constrained problems (cont.)

$$
\begin{aligned}
\nabla f\left(w^*\right)-\nabla g\left(w^*\right)\lambda^*-\nabla h\left(w^*\right)\mu^* &= 0\\
g\left(w^*\right) &= 0\\
h\left(w^*\right) &\geq 0\\
\mu^* &\geq 0\\
\mu_{n_h}^* h_{n_h}\left(w^*\right) &= 0, \quad n_h = 1,\dots,N_h
\end{aligned}
$$

The KKT conditions are first-order necessary optimality conditions for arbitrarily constrained problems, and thus correspond to $\nabla f\left(w^*\right)=0$ for unconstrained problems

- For convex problems, the KKT conditions are sufficient for globality

---

The last three KKT conditions are often denoted as **complementarity conditions**