Constrained optimization

- General theoretical and practical aspects
- A quick intro to linear programming

Constrained optimization

- General theoretical and practical aspects
- A quick intro to linear programming

Motivation

⋆ all algorithms presented before dealt with unconstrained optimization
⋆ Advantage in the unconstrained case: when looking for the next iterate you can search in any direction you want!
⋆ In practice it may not be possible to include all information in the objective function!
⋆ Sometimes, a minimization problem does not have non-trivial examples if no constraints are imposed!

⋆ constraints are necessary and useful in practice: what are the implications from the theoretical point of view?
⋆ how to deduce what are the relevant optimality conditions and how to solve practically optimization problems under constraints?

Example 1

Source: `http://people.brunel.ac.uk/~mastjjb/jeb/or/morelp.html`

A company makes two products (X and Y) using two machines (A and B). Each unit of X that is produced requires 50 minutes processing time on machine A and 30 minutes processing time on machine B. Each unit of Y that is produced requires 24 minutes processing time on machine A and 33 minutes processing time on machine B.

At the start of the current week there are 30 units of X and 90 units of Y in stock. Available processing time on machine A is forecast to be 40 hours and on machine B is forecast to be 35 hours.

The demand for X in the current week is forecast to be 75 units and for Y is forecast to be 95 units. Company policy is to maximise the combined sum of the units of X and the units of Y in stock at the end of the week.

Getting the constraints and objective function...

- $50x + 24y \leq 40 \times 60$
- $30x + 33y \leq 35 \times 60$
- $x \geq 45$
- $y \geq 5$

Maximize: $x + y - 50$

Example 2

Optimal can

For an aluminum can one can infer that its production cost may be proportional to its surface area. On the other hand, the can must hold a certain volume c of juice. Supposing that the can has a cylindrical shape, what are its optimal dimensions?

- ⋆ we have two parameters: the height h and the radius r.
- ⋆ Area of the can (to be minimized): $A(h, r) = 2\pi r^2 + 2\pi rh$
- ⋆ Volume of the can (constraint): $V(h, r) = \pi r^2 h$
- ⋆ finally we obtain the problem

$$\min_{V(h,r)\geq c} A(h, r).$$

The milkmaid problem

Suppose a person (M) in a large field trying to get to a cow (C) as fast as possible. Before milking the cow the bucket needs to be cleaned in a river nearby defined by the equation $g(x, y) = 0$. What is the optimal point P on the river such that the total distance traveled $MP + PC$ is minimal?

If $M(x_0, y_0)$ is the initial position and $C(x_C, y_C)$ is the position of the cow then the problem becomes

$$\min_{g(P)=0} MP + PC.$$

General formulation

$\star$ given functions $f, h_1, ..., h_m, g_1, ..., g_k : \mathbb{R}^n \to \mathbb{R}$ we may consider problems like

$$\begin{array}{rl} (P) & \min f(x) \\ \text{s.t} & h_i(x) = 0, i = 1, ..., m \\ & g_j(x) \leq 0, j = 1, ..., k \end{array}$$

$\star$ in the following we assume that functions f, h_i, g_j are at least C^1 (even more regular if necessary)

$\star$ the cases where the constraints define a convex set are nice!

$\star$ we are interested in finding necessary and sufficient (when possible) optimality conditions

Some terminology

⋆ a feasible solution to (P) is any point which verifies all the constraints
⋆ the feasible set is the family of all feasible solutions
⋆ if among feasible solutions of (P) there exists one x^* such that $f(x^*) \leq f(x)$ for all x which are feasible then we found an optimal solution of (P)

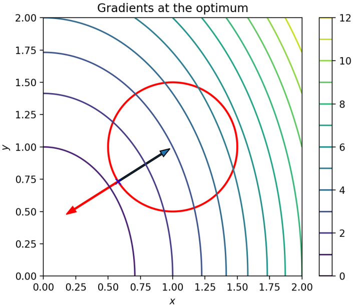
⋆ inequality constraints can be turned into equality constraints by introducing some slack variables: this increases the dimension of the problem...
⋆ keeping the inequality constraints is good in the convex case!

⋆ is good to picture the geometry given by the constraints and only then go to the analysis results

Intuitive Example

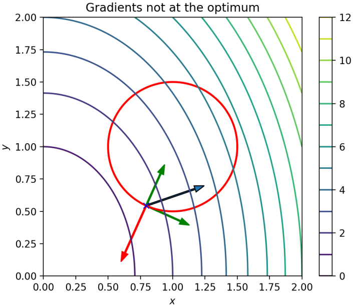
⋆ Minimize $f(x, y) = 2x^2 + y^2$ under the constraint $h(x, y) = \sqrt{(x-1)^2 + (y-1)^2} - 0.5 = 0$

⋆ Do the optimization and trace the gradients of f and h at the minimum:



⋆ Looks like the gradients are colinear! Why?

What happens if the gradients are not collinear?



⋆ the gradient ∇f has a non-zero component along the tangent line to the constraint

⋆ **Consequence:** it should be possible to further decrease the value of f by moving tangentially to the constraint!

Optimality condition: equality constraints

⋆ the gradient $\nabla f(x^*)$ should be orthogonal to the tangent plane to the constraint set $h(x^*) = 0$, otherwise following the non-zero tangential part we could still decrease the value of f

Questions:

⋆ definition of tangent space: look at the first order Taylor expansion!

The linearization of the constraint h_i around x s.t. $h_i(x) = 0$ is given by

$$\ell_i(y) = h_i(x) + \nabla h_i(x) \cdot (y - x) = \nabla h_i(x) \cdot (y - x)$$

If $h(x) = 0$ then the tangent plane at x is defined by

$$T_x = \{y : (y - x) \cdot \nabla h_i(x) = 0, i = 1, ..., m\}.$$

⋆ existence of well-defined tangent spaces: the function h should be regular around the minimizer

Examples

⋆ $h(x) = x_1^2 + x_2^2 - 1$ around the point $p = (\sqrt{2}/2, \sqrt{2}/2)$: we have $\nabla h(p) = 2(x_1, x_2)$ so the tangent plane is

$$T_p = \{y : (y - p) \cdot (x_1, x_2) = 0\},$$

which a well defined 1-dimensional line

⋆ $h(x) = x_1^2 - x_2^2$ at the point $p = (0, 0)$: we have $\nabla f(x) = (2x_1, -2x_2)$ so $\nabla f(p) = 0$. Using the same definition we have

$$T_p = \{y : (y - p) \cdot 0 = 0\} = \mathbb{R}^2,$$

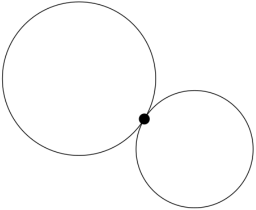
which is weird.

Goal: m equality constraints should give rise to a tangent space of dimension $k = n - m$! The gradient should be in the orthogonal to the tangent plane at the optimum: this has dimension equal to the rank of $Dh(x^*)$. Two situations occur:

- rank of $Dh(x^*)$ is strictly less than m: $\nabla f(x^*)$ might not be representable as a linear combination of $\nabla h_i(x^*)$!
- rank of $Dh(x^*)$ is exactly equal to m

Further Examples

⋆ intersect two spheres in $\mathbb{R}^3$: you may end up with a point which is not a set of dimension 1



⋆ intersect a sphere and a right cylinder: $h_1(x) = x_1^2 + x_2^2 + x_3^2 - 1$, $h_2(x) = x_1^2 + x_2^2 - x_2$. The gradients are $\nabla h_1(x) = 2(x_1, x_2, x_3)$ and $\nabla h_2(x) = (2x_1, 2x_2 - 1, 0)$ and they are linearly dependent at $(0, 1, 0)$.

We expect an intersection made of a 1D curve, but there are points where the tangent is not unique!

Regular points

Definition 1 (Regular points)

Given a family $h_1, ..., h_m$ of C^1 functions, $m \leq n$, a solution x_0 of the system

$$h_i(x) = 0, i = 1, ..., m$$

is called regular if the gradient vectors $(\nabla h_i(x_0))_{i=1}^m$ are linearly independent. Equivalently, the $m \times n$ matrix having $\nabla h_i(x_0)$ as rows has full rank m.

⋆ the implicit function theorem implies that around regular points the system $h_i(x) = 0$ defines a C^1 surface of dimension $k = n - m$!

⋆ moreover, you can pick some $k = n - m$ coordinates and express the set $h_i(x)$ in parametric form in terms of these coordinates

⋆ at regular points we can define the notion of tangent space which coincides with the one given by linearizing the constraints.

Tangent plane property

Proposition 2

Let S be given by $h_i(x) = 0, i = 1, ..., m$ where h_i are C^2 functions and $x \in S$ be a regular solutions. Then the plane T_x defined by

$$T_x = \{(y - x)Dh(x) = 0\}$$

is the tangent plane to S at x. Furthermore, there exists a constant C such that

$$\text{(1)} \quad \text{for every } x' \in S \text{ there exists } y' \in T_x \text{ s.t. } |x' - y'| \leq C|x' - x|^2$$

and

$$\text{(2)} \quad \text{for every } y' \in T_x \text{ there exists } x' \in S \text{ s.t. } |x' - y'| \leq C|y' - x|^2$$

⋆ Just look at the Taylor expansion of h_i and the linearization ℓ_i around x! They coincide up to the second order.

⋆ the statement (2) is false if x is not a regular point: the tangent space defined by T_x is larger than the real tangent space!

More details: tangent plane

$\star$ if $Dh(x)$ is of rank m then the linear system $Dh(x)y = 0$ can be solved in terms of $k = n - m$ parameters: e.g. $y_{m+1}, ..., y_n$:

$$\overline{y}_i = \ell_i(y_{m+1}, ..., y_n), \; i = 1, ..., m.$$

$\star$ implicit function theorem: there exist $k = n - m$ coordinates (say $y_{m+1}, ..., y_n$) such that there exist C^1 functions φ_j s.t.

$$y_i = \varphi_i(y_{m+1}, ..., y_n), \; i = 1, ..., m$$

$\star$ The gradients of φ_i are given by ℓ_i!

$\star$ Finally, the difference between the surface $h(x) = 0$ and the linearization contains only second order terms!

$$y_i - \overline{y}_i = O(|x - y|^2).$$

First order optimality conditions

⋆ suppose that x^* is a local minimum of f under the constraints $h(x) = 0$
⋆ suppose also that x^* is regular so that the tangent space T_x to the constraint gives a good approximation of $h(x) = 0$.
⋆ it is reasonable to assume that x^* also minimizes the linearization of f: $\overline{f}(y) = f(x^*) + (y - x^*)\nabla f(x^*)$ on this tangent plane defined by $Dh(x^*)(y - x^*) = 0$.
⋆ this would imply that $\nabla f(x^*)$ is orthogonal to $(y - x^*)$ for every y such that $Dh(x)(y - x^*) = 0$.
⋆ in usual notations we have $\nabla f(x^*) \in (\ker Dh(x^*))^\perp$
⋆ recall an important linear algebra result:

$$(\ker A)^\perp = \mathrm{Im} A^T.$$

⋆ finally, we obtain that there exists some $\lambda \in \mathbb{R}^m$ s.t.

$$\nabla f(x^*) = Dh(x^*)\lambda$$

which translates to the classical relation

$$\nabla f(x^*) = \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*).$$

Main result: Lagrange multipliers

Theorem 3

Let x^* be a local minimizer for the equality constrained problem

$$\min_{h(x)=0} f(x)$$

and suppose that x^* is a regular point for the system of equality constraints. Then the following two equivalent facts take place

- *The directional derivative of f in every direction along the space $\{y : Dh(x^*)(y - x^*) = 0\}$ tangent to the constraint at x^* is zero:*
$$Dh(x^*)d = 0 \Longrightarrow \nabla f(x^*) \cdot d = 0$$
- *There exist a uniquely defined vector of Lagrange multipliers $\lambda_i^*,\ i = 1, ..., m$ such that*
$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) = 0.$$

Proof:

- S denotes the set $h(x) = 0$.
- suppose that there exist a direction parallel to the tangent plane $Dh(x^*)\delta = 0$ which is not orthogonal to $\nabla f(x^*)$
- by eventually replacing it with $-\delta$ we may assume $\delta \cdot \nabla f(x^*) = -\alpha < 0$.
- denote $y_t = x^* + t\delta$. For small enough t we have $f(y_t) \leq f(x^*) - t\alpha/2$
- since x^* is regular, for every t small there exists a point $x_t \in S$ such that
$$|y_t - x_t| \leq C|y_t - x^*|^2 = C_1 t^2$$
- f is C^1 and therefore Lipschitz around x^* so
$$|f(x_t) - f(y_t)| \leq C_2|x_t - y_t| \leq C_1 C_2 t^2.$$
- Finally we get that $f(x_t) \leq f(x^*) - \alpha t/2 + C_1 C_2 t^2 < f(x^*)$ for $t > 0$ small enough, contradicting the optimality of x^*

$\star$ the second points comes from $(\ker A)^\perp = \text{Im} A^T$!

The result may be false at irregular points

Counterexample: Minimize the function $f(x_1, x_2, x_3) = x_2$ under the constraints

$$0 = h_1(x) = x_1^6 - x_3, \quad 0 = h_2(x) = x_2^3 - x_3.$$

- $\star$ the constraints define the curve $\gamma(x) = (x, x^2, x^6)$.
- $\star$ the minimum of f is attained at $(0, 0, 0)$
- $\star$ We have $\nabla f(0) = (0, 1, 0)$
- $\star$ on the other hand $\nabla h_1(0) = \nabla h_2(0) = (0, 0, -1)$
- $\star$ it is clear that $\nabla f(0)$ is not a linear combination of $\nabla h_1(0)$ and $\nabla h_2(0)$

Another counterexample

⋆ come back to the intersection between the sphere and the cylinder: $h_1(x) = x_1^2 + x_2^2 + x_3^2 - 1$, $h_2(x) = x_1^2 + x_2^2 - x_2$. The gradients are $\nabla h_1(x) = 2(x_1, x_2, x_3)$ and $\nabla h_2(x) = (2x_1, 2x_2 - 1, 0)$ and they are linearly dependent at $(0, 1, 0)$.

⋆ we can obtain that $x_1^2 = x_3^2 - x_3^4$ and $x_2 = 1 - x_3^2$ so the curve representing the intersection between h_1 and h_2 has the parametrization

$$(\pm\sqrt{x_3^2 - x_3^4}, 1 - x_3^2, x_3)$$

⋆ choose now the function $f(x_1, x_2, x_3) = x_1 + x_3 = x_3 \pm \sqrt{x_3^2 - x_3^4}$. This function has the minimum value 0 for $x_3 = 0$ associated to the point $(0, 1, 0)$.

⋆ the gradient of f at the minimum is $\nabla f(0, 1, 0) = (1, 0, 1)$

⋆ again, the conclusion of the theorem is not satisfied since the gradients of the constraints are not linearly independent at the optimum.

Example of usage

⋆ $\min(3x + 2y + 6z)$ such that $x^2 + y^2 + z^2 = 1$
⋆ obviously, there exists a solution, since $x^2 + y^2 + z^2 = 1$ is closed and bounded
⋆ write the optimality conditions: there exists λ such that
$\nabla f(x^*) + \lambda \nabla h(x^*) = 0$

$$(3, 2, 6) = \lambda(2x, 2y, 2z).$$

⋆ this immediately gives x, y, z in terms of λ
⋆ plug these expression in the constraint to get λ, and therefore x, y, z
⋆ in this case we get two values of λ: one corresponding to the minimum, the other corresponding to the maximum!

Order one optimality conditions do not indicate whether we are at a minimum or at a maximum!

The milkmaid problem

$$\min_{g(x)=0} d(P,x) + d(x,Q).$$

⋆ suppose that g is a nice curve in the plane with non-zero gradient
⋆ the gradient of the distance function:

$$\nabla_x d(P,x) = \frac{x-P}{d(P,x)},$$

is the unit vector that points from P to the variable point x.
⋆ the optimality condition says that there exists λ such that

$$\nabla_x d(P,x) + \nabla_x d(Q,x) + \lambda \nabla g(x) = 0$$

⋆ what does this mean geometrically? The normal vector $\nabla g(x)$ to $g(x) = 0$ cuts the angle PxQ in half
⋆ we obtain the classical reflection condition using Lagrange multipliers!

The isoperimetric inequality

What is the curve which has the maximum area for a given perimeter?

$\star$ suppose we have a 2D curve parametrized by $(x(t), y(t))$ in a counter-clockwise direction.

- the perimeter is $L = \int \sqrt{\dot{x}(t)^2 + \dot{y}(t)^2}$
- the area is $A = \int \frac{1}{2}(x(t)\dot{y}(t) - y(t)\dot{x}(t))$

Problem

Maximize A with the constraint $L = p$.

$\star$ $L = L(x, y),\ A = A(x, y)$ are functions for which variables are other functions. Sometimes the term functionals is employed!

$\star$ how to compute the gradient in such cases? when in doubt just come back to the one dimensional case using directional derivatives

$\star$ the integrals are taken over a whole period of the parametrization

Derivatives of A and L

- $\star$ pick two directions u and v and $t \in \mathbb{R}$. Then compute the derivative of
$$t \mapsto L(x+tu, y+tv) \text{ at } t=0.$$
- $\star$ it is useful to take all derivatives off u and v to get the linear form
$$L'(x,y)(u,v) = -\int \left[\left(\frac{\dot x}{\sqrt{\dot x^2+\dot y^2}}\right)' u + \left(\frac{\dot y}{\sqrt{\dot x^2+\dot y^2}}\right)' v\right]$$
- $\star$ do the same for $A(x,y)$ to get
$$A'(x,y)(u,v) = \int (\dot y u - \dot x v)$$
- $\star$ in the end we get
$$\nabla L(x,y) = \left(\left(\frac{\dot x}{\sqrt{\dot x^2+\dot y^2}}\right)', \left(\frac{\dot y}{\sqrt{\dot x^2+\dot y^2}}\right)'\right), \nabla A(x,y) = (\dot y, -\dot x).$$

Optimality condition and conclusion

$\star$ when maximizing A under the constraint $L = p$ the solution should verify the optimality condition

$$\nabla A(x,y) + \lambda \nabla P(x,y) = 0,\ \ \lambda \in \mathbb{R}$$

$\star$ plugging the derivatives found previously we get

$$\begin{cases} \dot y - \lambda \left(\frac{\dot x}{\sqrt{\dot x^2 + \dot y^2}} \right)' = 0 \\ -\dot x - \lambda \left(\frac{\dot y}{\dot x^2 + \dot y^2} \right)' = 0 \end{cases}$$

$\star$ integrating we obtain

$$\begin{cases} y - \lambda \frac{\dot x}{\sqrt{\dot x^2 + \dot y^2}} = b \\ x + \lambda \frac{\dot y}{\sqrt{\dot x^2 + \dot y^2}} = a \end{cases}$$

$\star$ in the end we have

$$(x-a)^2 + (y-b)^2 = \lambda^2,$$

so the solution should be a circle.

The Lagrangian

$\star$ the optimality conditions obtained involve the gradient of the objective function and the constraints.
$\star$ the optimality condition can be written as the gradient of a function combining the objective and the constraints called the Lagrangian: $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m$

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i=1}^m \lambda_i h_i(x) = f(x) + \lambda \cdot h(x).$$

$\star$ if x^* is a local minimum of f on the set $\{h(x) = 0\}$ then the optimality condition tells us that there exists $\lambda^* \in \mathbb{R}^m$ such that

$$\frac{\partial \mathcal{L}}{\partial x}(x^*, \lambda^*) = 0 \text{ and } \frac{\partial \mathcal{L}}{\partial \lambda}(x^*, \lambda^*) = 0$$

$\star$ moreover, $\displaystyle\sup_{\lambda \in \mathbb{R}^n} \mathcal{L}(x, \lambda) = \begin{cases} f(x) & \text{if } h(x) = 0 \\ +\infty & \text{if } h(x) \neq 0 \end{cases}$ which gives

$$\min_{h(x)=0} f(x) = \min_{x \in \mathbb{R}^n} \sup_{\lambda \in \mathbb{R}^m} \mathcal{L}(x, \lambda).$$

$\star$ the minimizer of f becomes a saddle point for the Lagrangian

Another point of view

⋆ for $c_i \in \mathbb{R}, i = 1, ..., m$ consider the problem

$$\min_{h_i(x)=c_i} f(x)$$

⋆ considering the Lagrangian

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i (c_i - h_i(x))$$

we see that $\frac{\partial L}{\partial c_i} = \lambda_i$ so the Lagrange multipliers represent the rate of change of the quantity being optimized as a function of the constraint parameter.

⋆ denote by $x^*(c), \lambda^*(c)$ the optimizer and the Lagrange multipliers as a function of c. Then

$$\begin{aligned}
\frac{\partial f(x^*(c))}{\partial c_i} &= \frac{\partial \mathcal{L}(x^*(c), \lambda^*)}{\partial c_i} \\
&= \frac{\partial \mathcal{L}}{\partial x}(x^*(c), \lambda^*) \frac{\partial x^*(c)}{c_i} + \frac{\partial \mathcal{L}}{\partial c_i}(x^*(c), \lambda^*) \\
&= \lambda_i^*
\end{aligned}$$

Another application: compute derivatives

⋆ how to compute derivatives under constraints?

Example: Compute the derivative of $x \mapsto f$ under the constraint $f^2 = x$.

⋆ write the Lagrangian: $L(x, f, p) = f + (f^2 - x)p$

⋆ if $f = \sqrt{x}$ then $L(x, f, p) = f$.

⋆ compute the derivative of f directly from above:

$$f'(x) = \frac{\partial L}{\partial x}(x, f, p) + \frac{\partial L}{\partial f}(x, f, p)\frac{df}{dx} + \frac{\partial L}{\partial p}(x, f, p)\frac{dp}{dx}$$

⋆ cancel the terms which you don't know using the Lagrangian:

$$\frac{\partial L}{\partial p} = f^2 - x = 0, \frac{\partial L}{\partial f} = 1 + 2fp = 0.$$

⋆ what remains is $f'(x) = \dfrac{\partial L}{\partial x}(x, f, -1/(2f)) = \dfrac{1}{2f} = \dfrac{1}{2\sqrt{x}}$.

⋆ we recover the classical result. This technique is known as the adjoint method and is useful for computing derivatives in complicated spaces: shape derivatives, control theory, etc.

What happens for inequality constraints?

⋆ minimize $f(x)$ such that $g_1(x) \leq 0, ..., g_k(x) \leq 0$.
⋆ not all inequality constraints play the same role: at the point x the constraint i is said to be active if $g_i(x) = 0$.
⋆ if a constraint g_i (where g_i is C^1) is inactive at a minimizer x^* then $g_i(x) < 0$ in a neighborhood of x^*

⋆ if x^* is a minimizer of $f(x)$ under the constraints g_i and $g_i(x^*) < 0$ then g_i does not impose any restriction on f locally: ignoring it produces the same result (locally)
⋆ equality constraints generally produced surfaces while inequality constraints can just give bunded regions of $\mathbb{R}^n$.

Qualification of constraints

⋆ denote by $I(x) = \{i \in \{1, ..., k\} : g_i(x) = 0\}$ be the indices of active constraints at x

⋆ we say that the constraints are qualified at x if the gradients $(\nabla g_i(x))_{i \in I(x)}$ are linearly independent!

⋆ geometrically, as in the equality constraints case, if the constraints are qualified at x then we may define a proper tangent space using the family $(\nabla g_i(x))_{i \in I(x)}$

⋆ **Special case:** if all g_i are affine constraints then they are automatically qualified. Why?

- in this case the constraints also define the tangent space themselves
- the linear independence of the gradients at a point x is equivalent to the removal of redundant constraints

Optimality conditions: inequalities

Theorem 4

Let x^* be a local minimizer for the inequality constrained problem

$$\min_{g(x)\leq 0} f(x)$$

and suppose that the constraints are qualified at x^*. Then the following affirmations are true:

- *There exists a uniquely defined vector of Lagrange multipliers $\lambda_i^* \geq 0,\ i = 1, ..., k$ such that*

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0.$$

- *Moreover, if $g_i(x^*) < 0$ then $\lambda_i = 0$, also called the complementary slackness relations. Equivalent formulation: $\lambda_i g_i(x^*) = 0$.*

⋆ why are Lagrange multipliers non-negative in this case? x^* would like to "get out of the constraints" to increase the value of f

⋆ if x^* is an interior point for $g(x) \leq 0$ then simply $\nabla f(x^*) = 0$

Example: qualification of constraints

Consider the set

$$k = \{x = (x_1, x_2) \in \mathbb{R}^2 : -x_1 \leq 0,\ -x_2 \leq 0,\ -(1 - x_1)^3 + x_2 \leq 0\}.$$

⋆ Maximize $J(x) = x_1 + x_2$ for $x \in K$.

⋆ making a drawing we find that immediately that the solutions are $(0, 1)$ and $(1, 0)$.

⋆ let's check if we can write the optimality condition at the two points:

- $(1, 0)$: constraints not qualified: unable to write the opt. cond
- $(0, 1)$: constraints qualified: the optimality condition can be written!

The Lagrangian - inequality case

⋆ the optimality conditions obtained involve the gradient of the objective function and the constraints.

⋆ the optimality condition can be written as the gradient of a function combining the objective and the constraints called the Lagrangian: $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m_+$

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i=1}^k \lambda_i g_i(x) = f(x) + \lambda \cdot g(x).$$

⋆ if x^* is a local minimum of f on the set $\{g(x) \leq 0\}$ then the optimality condition tells us that there exists $\lambda^* \in \mathbb{R}^m_+$ such that

$$\frac{\partial \mathcal{L}}{\partial x}(x^*, \lambda^*) = 0 \text{ and } \frac{\partial \mathcal{L}}{\partial \lambda}(x^*, \lambda^*) = 0$$

⋆ moreover, $\sup_{\lambda \in \mathbb{R}^m_+} \mathcal{L}(x, \lambda) = \begin{cases} f(x) & \text{if } g(x) \leq 0 \\ +\infty & \text{otherwise} \end{cases}$ which gives

$$\min_{g(x) \leq 0} f(x) = \min_{x \in \mathbb{R}^n} \sup_{\lambda \in \mathbb{R}^m_+} \mathcal{L}(x, \lambda).$$

⋆ the minimizer of f becomes a saddle point for the Lagrangian

Come back to the optimal can problem

⋆ Area of the can (to be minimized): $A(h, r) = 2\pi r^2 + 2\pi rh$
⋆ Volume of the can (constraint): $V(h, r) = \pi r^2 h$
⋆ finally we obtain the problem

$$\min_{V(h,r)\geq c} A(h, r).$$

⋆ the constraint will be active!
⋆ write the optimality condition: find r and h in terms of λ and finish!

⋆ in the end we find that the optimal can will have the height h equal to two times its radius r.

⋆ find now the optimal cup: only one of the two ends is filled with material!

Saddle points

Definition 5

We say that $(u,p) \in U \times P$ is a saddle point of $\mathcal{L}$ on $U \times P$ if

$$\forall q \in P \quad \mathcal{L}(u,q) \le \mathcal{L}(u,p) \le \mathcal{L}(v,p) \quad \forall v \in U$$

⋆ when fixing p: $v \mapsto \mathcal{L}(b,p)$ is minimal for $v = u$
⋆ when fixing u: $q \mapsto \mathcal{L}(u,q)$ is minimal for $q = p$

⋆ If J is the objective function and F defines the constraint set K (equality or inequality) then a saddle point (u,p) for the Lagrangian

$$\mathcal{L}(v,q) = J(v) + q \cdot F(v)$$

verifies that u is a minimum of J on K.
⋆ moreover, if the Lagrangian is defined on an open neighborhood U of the constraint set K then we also recover the optimality condition

$$\nabla J(u) + \sum_{i=1}^m p_i \nabla F_i(u) = 0.$$

Sufficient conditions

⋆ two options: go to the second order or use convexity
⋆ it is not enough to look at the second order approximation of f on the tangent space! The curvature of the constraint can also play a role.
⋆ the correct way is to look at the Hessian of the Lagrangian with respect to x, reduced to the tangent space!

⋆ in the convex case, for inequality constraints things are a little bit easier!

⋆ why only for inequality constraints? Imagine that equality constraints can produce curved surfaces and the only way to have convexity there is if they are flat!
⋆ why the choice $g_i(x) \leq 0$ as the definition of inequality constraints? Because if all g_i are convex functions then

$$K = \{x : g_i(x) \leq 0, i = 1, ..., k\} \text{ is a convex set.}$$

Sufficient conditions - convex case

Theorem 6 (Kuhn-Tucker)

Suppose that the functions $f, g_i,\ i = 1, ..., k$ are C^1 and convex. Define K as the set $K = \{x : g_i(x) \leq 0\}$ and introduce the Lagrangian

$$\mathcal{L}(v, q) = f(v) + q \cdot g(v),\ v \in \mathbb{R}^n, q \in \mathbb{R}^k_+.$$

Let x^* be a point of K where the constraints are qualified. Then the following are equivalent:

- *x^* is a global minimum of f on K*
- *there exists $\lambda^* \in \mathbb{R}^m$ such that (x^*, λ^*) is a saddle point for the Lagrangian*
- $g(x^*) \leq 0$, $\lambda^* \geq 0$, $\lambda^* \cdot F(x^*) = 0$, $\nabla f(x^*) + \sum_{i=1}^k \lambda_i^* \nabla g_i(x^*) = 0$.

⋆ why the reverse implication works? When $q \geq 0$ the Lagrangian

$$\mathcal{L}(v, q) = f(v) + q \cdot g(v),\ v \in \mathbb{R}^n, q \in \mathbb{R}^k_+$$

is convex when f and $g = (g_i)$ are convex!

⋆ particular case: affine equalities! convex and qualified!

Handle the constraints numerically

$\star$ we already saw two methods:

- projected gradient algorithm:

$$x_{i+1} = \text{Proj}_K(x_i - t\nabla f(x_i))$$

- penalization: include the constraint $\{g = 0\}$ in the objective

$$\min f(x) + \frac{1}{\varepsilon} g(x)^2$$

$\star$ we saw that the projection is not explicit in most cases! In the meantime we learned how to solve non-linear equations. Imagine the following algorithm:

- Compute x_i and the projection d_i of $-\nabla f(x_i)$ on the tangent space (orthogonal of $(\nabla g_j(x_i))$)
- advance in the direction of d_i: $x_{i+1} = x_i + \gamma_i d_i$
- project x_{i+1} on the set of constraints using the Newton method

Conclusion on Lagrange multipliers

- we may obtain necessary optimality conditions involving equality and inequality constraints: the gradient of f is a linear combination of the gradients of the constraints
- the gradients of the constraints need to be linearly independent at the optimum: proper definition of the tangent space!
- for inequality constraints only the active constraints come into play in the optimality condition
- sufficient conditions can be found in the convex case: Kuhn-Tucker theorem
- the theory gives new ways to handle constraints numerically

Constrained optimization

- General theoretical and practical aspects
- A quick intro to linear programming

Linear programming

⋆ maximizing or minimizing a linear function subject to linear constraints!

⋆ Example:

$$\max(x_1 + x_2)$$

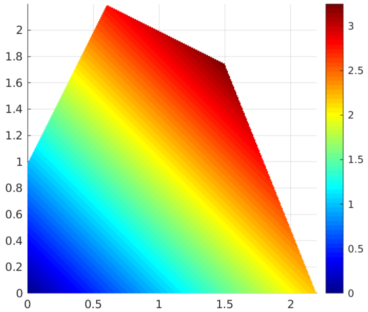
such that $x_1 \geq 0,\ x_2 \geq 0$ and

$$\begin{array}{rcl} x_1 + 2x_2 & \leq & 5 \\ 5x_1 + 2x_2 & \leq & 11 \\ -2x_1 + x_2 & \leq & 1 \end{array}$$

⋆ we have some non-negativity constraints and the main constraints

Geometric solution

⋆ in dimension 2 we can solve the problem by plotting the objective function on the admissible set determined by the constraints!



⋆ observe that in this case the solution is situated at the intersection of the lines

$$5x_1 + 2x_2 = 11 \text{ and } x_1 + 2x_2 = 5.$$

Theoretical observations

⋆ the gradient of $f(x_1, x_2) = x_1 + x_2$ is $(1, 1)$: it is constant and never zero!
⋆ the set K determined by the linear constraints is convex
⋆ the minimum or maximum cannot be attained in the interior of K, since $\nabla f(x) \neq 0$!
⋆ the optimal value is on the boundary of K. Moreover there exists a vertex of the polygon where it can be found! Why?

- start at a point x_0 inside K go against the gradient till you meet an edge
- if the function is constant along an edge then the gradient of the function and the constraint are collinear at that point: Kuhn-Tucker Theorem says that we reached the solution!
- otherwise, follow the direction where the function decreases till reaching a vertex. Then go to the next edge and repeat the previous reasoning.
- the process will finish: finite number of edges!

⋆ same reasoning can be applied in higher dimensions: follow the anti-gradient direction till it is collinear to the gradient of the constraint or no further decrease is possible along further facets!

Standard formulations

$\star$ **The Standard Maximum Problem:** Maximize $\mathbf{c}^t\mathbf{x} = c_1x_1 + ... + c_nx_n$ subject to the constraints

$$\begin{aligned} a_{11}x_1 + ... + a_{1n}x_n &\leq b_1 \\ &\vdots \\ a_{m1}x_1 + ... + a_{mn}x_n &\leq b_m \end{aligned} \qquad \text{or } A\mathbf{x} \leq \mathbf{b}$$

and $x_1 \geq 0, x_2 \geq 0, ..., x_n \geq 0$ or $\mathbf{x} \geq 0$

$\star$ **The Standard Minimum Problem:** Minimize $\mathbf{y}^t\mathbf{b} = y_1b_1 + ... + y_mb_m$ subject to the constraints

$$\begin{aligned} a_{11}y_1 + ... + a_{1m}y_m &\geq c_1 \\ &\vdots \\ a_{1n}y_1 + ... + a_{mn}y_m &\geq c_n \end{aligned} \qquad \text{or } y^TA \geq \mathbf{c}^T$$

and $y_1 \geq 0, y_2 \geq 0, ..., y_m \geq 0$ or $\mathbf{y} \geq 0$

⋆ There are I production sites $P_1, ..., P_I$ which supply a product and J markets $M_1, ..., M_J$ to which the product is shipped.
⋆ the site P_i contains s_i products and the market M_j must recieve r_j products.
⋆ the cost of transportation from P_i to M_j is b_{ij}
⋆ the objective is to minimize the transportation cost while meeting the market requirements!

⋆ denote by y_{ij} the quantity transported from P_i to M_j. Then the cost is

$$\sum_{i=1}^{I} \sum_{j=1}^{J} y_{ij} b_{ij}$$

⋆ the constraints are

$$\sum_{j=1}^{J} y_{ij} \leq s_i \text{ and } \sum_{i=1}^{I} y_{ij} \geq r_j.$$

Example 2

The Optimal Assignment Problem

⋆ There are I persons available for J jobs. The "value" of person i working 1 day at job j is a_{ij}.

⋆ **Objective:** Maximize the total "value"

⋆ the variables are x_{ij}: the proportion of person i's time spent on job j

⋆ the constraints are $x_{ij} \geq 0$

$$\sum_{j=1}^{J} x_{ij} \leq 1, i = 1, ..., I \text{ and } \sum_{i=1}^{I} x_{ij} \leq 1, \ j = 1, ..., J \leq 1$$

- can't spend a negative amount of time at a job
- a person can't spend more than 100% of its time
- no more than one person working on a job

Some Terminology

⋆ a point is said to be feasible if it verifies all the constraints
⋆ the set of feasible points is the constraint set
⋆ a linear programming problem is feasible if the constraint set is non-empty. If this is not the case then the problem is infeasible

⋆ every problem involving the minimization of a linear function under linear constraints can be put into standard form

- you can change a "$\geq$" inequality into "$\leq$" by changing the signs of the coefficients
- if a variable x_i has no sign restriction, write it as the difference of two new positive variables $x_i = u_i - v_i,\ u_i, v_i \geq 0$

⋆ it is possible to pass from inequality constraints to equality constraints (and the other way around)

- $Ax = b$ is equivalent to $Ax \leq b$ and $Ax \geq b$
- If $Ax \leq b$ then add some slack variables $\mathbf{u} \geq 0$ such that $Ax + u = b$

Duality in LP

Definition 7

The dual of the standard maximum problem

$$\begin{cases} \max \mathbf{c}^T\mathbf{x} \\ \text{s.t. } A\mathbf{x} \leq \mathbf{b} \text{ and } \mathbf{x} \geq 0 \end{cases}$$

is the standard minimum problem

$$\begin{cases} \min \mathbf{y}^T\mathbf{b} \\ \text{s.t. } \mathbf{y}^T A \geq \mathbf{c}^T \text{ and } \mathbf{y} \geq 0 \end{cases}.$$

Example

⋆ consider the problem

$$\begin{aligned} \text{maximize} \quad & x_1 + x_2 \\ \text{such that} \quad & \mathbf{x} \geq 0 \\ & \begin{array}{rcl} x_1 + 2x_2 & \leq & 5 \\ 5x_1 + 2x_2 & \leq & 11 \\ -2x_1 + x_2 & \leq & 1 \end{array} \end{aligned}$$

⋆ the dual problem is

$$\begin{aligned} \text{minimize} \quad & 5y_1 + 11y_2 + y_3 \\ \text{such that} \quad & \mathbf{y} \geq 0 \\ & \begin{array}{rcl} y_1 + 5y_2 - 2y_3 & \geq & 1 \\ 2y_1 + 2y_2 + y_3 & \geq & 1 \end{array} \end{aligned}$$

Relation between dual problems

Theorem 8

If $\mathbf{x}$ is feasible for the standard maximum problem and $\mathbf{y}$ is feasible for the dual problem then

$$\mathbf{c}^T\mathbf{x} \le \mathbf{y}^T\mathbf{b}.$$

⋆ The proof is straightforward:

$$\mathbf{c}^T\mathbf{x} \le \mathbf{y}^T A\mathbf{x} \le \mathbf{y}^T\mathbf{b}.$$

⋆ important consequences:

- if the standard maximum problem and its dual are both feasible, they are bounded feasible: the optimal values are finite!
- If there exist feasible $\mathbf{x}^*$ and $\mathbf{y}^*$ for the standard maximum problem and its dual such that $\mathbf{c}^T\mathbf{x}^* = \mathbf{y}^{*T}\mathbf{b}$ then both are optimal for their respective problems!

Theorem 9 (Duality)

If a standard linear programming problem is bounded feasible then so is its dual, their optimal values are equal and there exist optimal solutions for both problems.

Solve LP problems numerically

⋆ the simplex algorithm: travel along vertices of the set defined by the constraints until no decrease is possible

⋆ work with the matrix A and with vectors $\mathbf{b}$ and $\mathbf{c}$ and modify them using pivot rules: similar to the ones used when solving linear systems

⋆ exploit the connection between the standard formulation and its dual

⋆ things get more complicated when we restrict the variables to be integers. This gives rise to integer programming!

⋆ algorithms solving the main types of LP problems are implemented in various Python packages: `scipy.optimize.linprog`, `pulp`.

The simplex algorithm

⋆ bring the problem to the case of equality constraints using slack variables

$$\sum_{j=1}^{n} a_{ij}x_j \leq b_i \iff \sum_{j=1}^{n} a_{ij}x_j + s_i = b_i, s_i \geq 0$$

⋆ any free variable $x_j \in \mathbb{R}$ should be replaced with $u_j - v_j$ with $u_j, v_j \geq 0$

⋆ now we can solve

$$\begin{aligned} \text{maximize} \quad & \mathbf{c}^T\mathbf{x} \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq 0 \end{aligned}$$

⋆ start from the origin $\mathbf{x} = 0$ and go through the vertices of the polytype $A\mathbf{x} = \mathbf{b}$

⋆ at each step perform an operation similar to the Gauss elimination

⋆ Possible issues: cycling, numerical instabilities.

Practical Example 1

⋆ the first example of a standard maximum problem

$$\max(x_1 + x_2)$$

such that $x_1 \geq 0, \; x_2 \geq 0$ and

$$\begin{array}{rcc} x_1 + 2x_2 & \leq & 5 \\ 5x_1 + 2x_2 & \leq & 11 \\ -2x_1 + x_2 & \leq & 1 \end{array}$$

⋆ we saw geometrically that the solution should be the intersection of $x_1 + 2x_2 = 5$ and $5x_1 + 2x_2 = 11$

```
scipy.optimize.linprog(c, A_ub=None, b_ub=None, A_eq=None,
             b_eq=None,bounds=None, method='simplex',
             callback=None, options=None)
```

Practical Example 2

⋆ An optimal assignment problem: n

	Job 1	Job 2	Job 3
Person 1	100€	120€	80€
Person 2	150€	110€	120€
Person 3	90€	80€	110€

⋆ assign Person i to Job j in order to minimize the total cost!

⋆ we can model the situation as an LP problem with 9 variables: $x_{ij} = 1$ if and only if Person i has job j, $1 \leq i, j \leq 3$

⋆ the constraints are as follows:

- $\sum_{i=1}^{3} x_{ij} = 1$: exactly one Person for Job j
- $\sum_{j=1}^{3} x_{ij} = 1$: exactly one Job for Person i

⋆ we should also impose that $x_i \in \{0, 1\}$: no fractional jobs, but we'll neglect this condition and just suppose $x_i \geq 0$.

⋆ the cost is just

$$\sum_{1 \leq i,j \leq 3} c_{ij} x_{ij}$$

Find the LP parameters

⋆ let's look at the matrix of the problem: 9 variables and 6 constraints!

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

⋆ the matrix c_{ij} is given by the table shown previously: the cost of every person per function ⋆ the vector b is equal to 1 on every component

⋆ the solution is made of zeros and ones, without imposing this...
⋆ this phenomenon always happens: if A is a totally unimodular matrix and b is made of integers then $Ax \leq b$ has all its vertices at points with integer coordinates

A matrix is totally unimodular if every square submatrix has determinant in the set $\{0, 1, -1\}$.

Practical Example 3

⋆ solving a Sudoku with LP

					3		8	5
		1		2				
			5		7			
		4				1		
	9							
5							7	3
		2		1				
				4				9

⋆ Remember the rules: $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ should be found on every line, column and 3×3 square

⋆ in order to make this solvable via LP a different formulation should be used!

⋆ classical idea: use binary variables

Sudoku in Binary variables

⋆ how to represent a Sudoku puzzle using 0s and 1s?

⋆ build a 3D array $X = (x_{ijk})$ of size $9 \times 9 \times 9$ such that

$x_{ijk} = 1$ if and only if on position (i, j) we have the digit k; else $x_{ijk} = 0$

⋆ what are the constraints in this new formulation?

- $x_{ijk} \in \{0, 1\}$: again to be relaxed to $x_{ijk} \geq 0$ - 729 constraints
- fixing i, j: $\sum_{k=1}^{9} x_{ijk} = 1$ - one number per cell - 81 constraints
- fixing i, k: $\sum_{j=1}^{9} x_{ijk} = 1$ - k appears exactly once on line i - 81 constraints
- fixing j, k: $\sum_{i=1}^{9} x_{ijk} = 1$ - k appears exactly once on column j - 81 constraints
- small 3×3 squares condition: for $u, v \in \{0, 3, 6\}$

$$\sum_{i=1}^{3} \sum_{j=1}^{3} x_{i+u,j+v,k} = 1, \ k = 1, ..., 9 \text{ - 81 constraints}$$

- the initial given information for the puzzle may be written in the form $s_{ij} = k$ for some i, j, k. This gives the constraints $x_{i,j,s_{ij}} = 1$.

⋆ we are interested in finding a feasible solution: no objective function is needed!

Solving the Sudoku

⋆ a feasible solution can be found using the simplex algorithm
⋆ sometimes we may get non-integer results: apparently, the constraint matrix is not always a Totally Unimodular matrix
⋆ there are LP algorithms which will return integer solutions: integer programming

⋆ before solving we should check that the constraint matrix should be of maximal rank: eliminate redundant constraints
⋆ we could also eliminate fixed variables: the data $s_{ij} = k$ should eliminate all unknowns with first index i, second index j or third index k!

⋆ if the solution is unique: the algorithm will find it
⋆ if the solution is not unique: the algorithm will find one of the solutions. We may repeat with the constraint that the solution should be different than the previous one, until no other solutions are found!
⋆ check out the PuLP Python library: an example of Sudoku solver is given!

Conclusions on LP

- minimize/maximize linear functions under linear constraints
- many practical applications from an industrial point of view!
- there exist optimizers which are vertices of the constraint set
- simplex algorithm: travel along vertices decreasing the objective function
- computational complexity: worst case is exponential: Klee-Minty cube
- polynomial-time average case complexity: most of the LP problems will be solved very fast!

Conclusion of the course

⋆ numerical optimization (unconstrained case):

- derivatives-free algorithms: no-regularity needed, slow convergence
- gradient descent algorithms: linear convergence, sensitive to the condition number
- Newton, quasi-Newton: super-linear convergence in certain cases
- when dealing with large problems use L-BFGS
- Conjugate Gradient: solve linear systems, better than GD
- Gauss-Newton: useful when minimizing a non-linear least squares function

⋆ constrained case

- for simple constraints: use the projected gradient algorithm
- general smooth constraints: use the tangential part of the gradient and come back to the constraint set using the Newton method
- other options available: SQP, etc...
- Linear Programming: use specific techniques: the simplex algorithm $\longrightarrow$ to be continued next year in the course dealing with Convex Optimization!

Conclusion of the course

- know your options when looking at an optimization problem: choose the right algorithm depending on: the size of the problem, the number of variables, the regularity, the conditioning, etc.
- learn how to use existing solutions: `scipy.optimize` is a good starting point
- know how to code your own optimization algorithm if necessary: use gradients when possible, limit the number of function evaluations, choose a good stopping criterion, limit the number of iterations, etc.