# 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

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

$\star$ constraints are necessary and useful in practice: what are the implications from the theoretical point of view?
$\star$ 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

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

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

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

⋆ the cases where the constraints define a convex set are nice!

⋆ 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 $(P)$

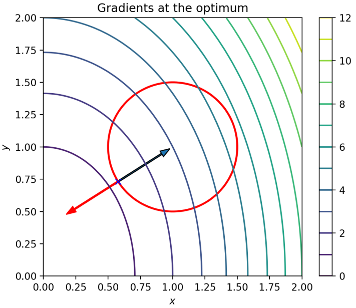
⋆ inequality constraints can be turned into equality constraints by introducing some slack variables: this increases the dimension of the problem...
⋆ keeping the inequality constraints is good in the convex case!

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

# Intuitive Example

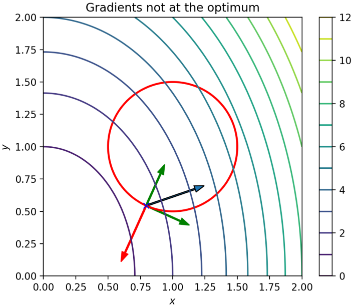
⋆ Minimize $f(x, y) = 2x^2 + y^2$ under the constraint $h(x, y) = \sqrt{(x-1)^2 + (y-1)^2} - 0.5 = 0$

⋆ 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 $\nabla f$ 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

$\star$ 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:**

$\star$ 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

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

$\star$ 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

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

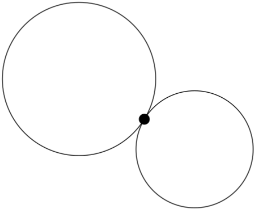
which is weird.

**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*

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

$\star$ Just look at the Taylor expansion of $h_i$ and the linearization $\ell_i$ around $x$! They coincide up to the second order.

$\star$ 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 $\varphi_j$ s.t.

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

$\star$ The gradients of $\varphi_i$ are given by $\ell_i$!

$\star$ 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

$\star$ suppose that $x^*$ is a local minimum of $f$ under the constraints $h(x) = 0$
$\star$ suppose also that $x^*$ is regular so that the tangent space $T_x$ to the constraint gives a good approximation of $h(x) = 0$.
$\star$ 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 $Dh(x^*)(y - x^*) = 0$.
$\star$ this would imply that $\nabla f(x^*)$ is orthogonal to $(y - x^*)$ for every $y$ such that $Dh(x)(y - x^*) = 0$.
$\star$ in usual notations we have $\nabla f(x^*) \in (\ker Dh(x^*))^\perp$
$\star$ recall an important linear algebra result:

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

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

# 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 $\lambda$ 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 $\lambda$
⋆ plug these expression in the constraint to get $\lambda$, and therefore $x,y,z$
⋆ in this case we get two values of $\lambda$: 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 $\lambda$ 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?

⋆ 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 $L = p$.

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

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

⋆ 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

⋆ 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: $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m$

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

⋆ 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

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

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

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

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

# Another point of view

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

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

$\star$ 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.

$\star$ 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

$\star$ how to compute derivatives under constraints?

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

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

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

$\star$ compute the derivative of $f$ directly from above:

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

$\star$ cancel the terms which you don't know using the Lagrangian:

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

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

$\star$ 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.