Convex Optimization in Machine Learning and Computational Finance Lecture 3: Duality

Dr. Michel Baes, Pr. Patrick Cheridito

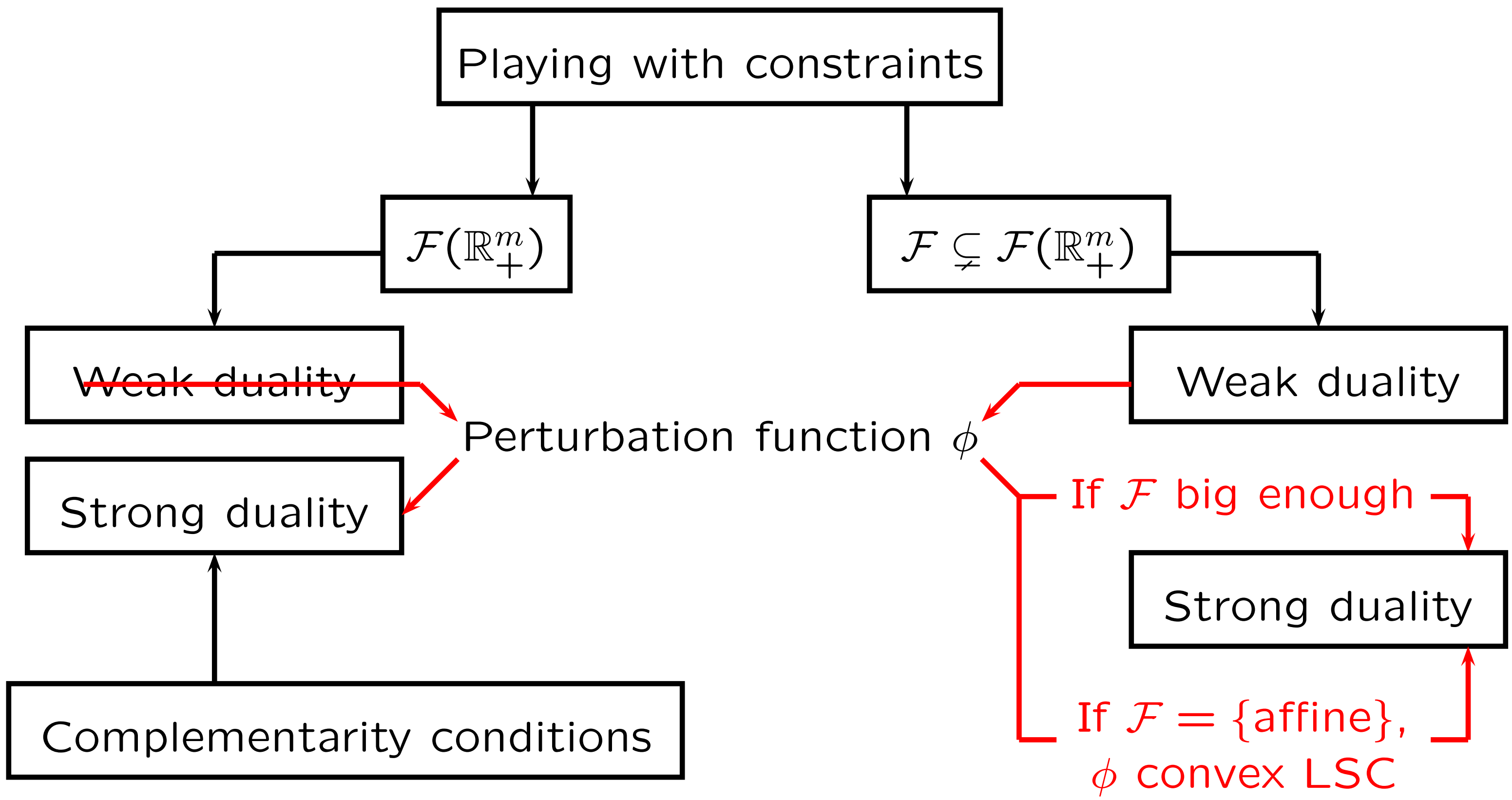
RiskLab / ETH Zürich

Quick recall of last week's lecture

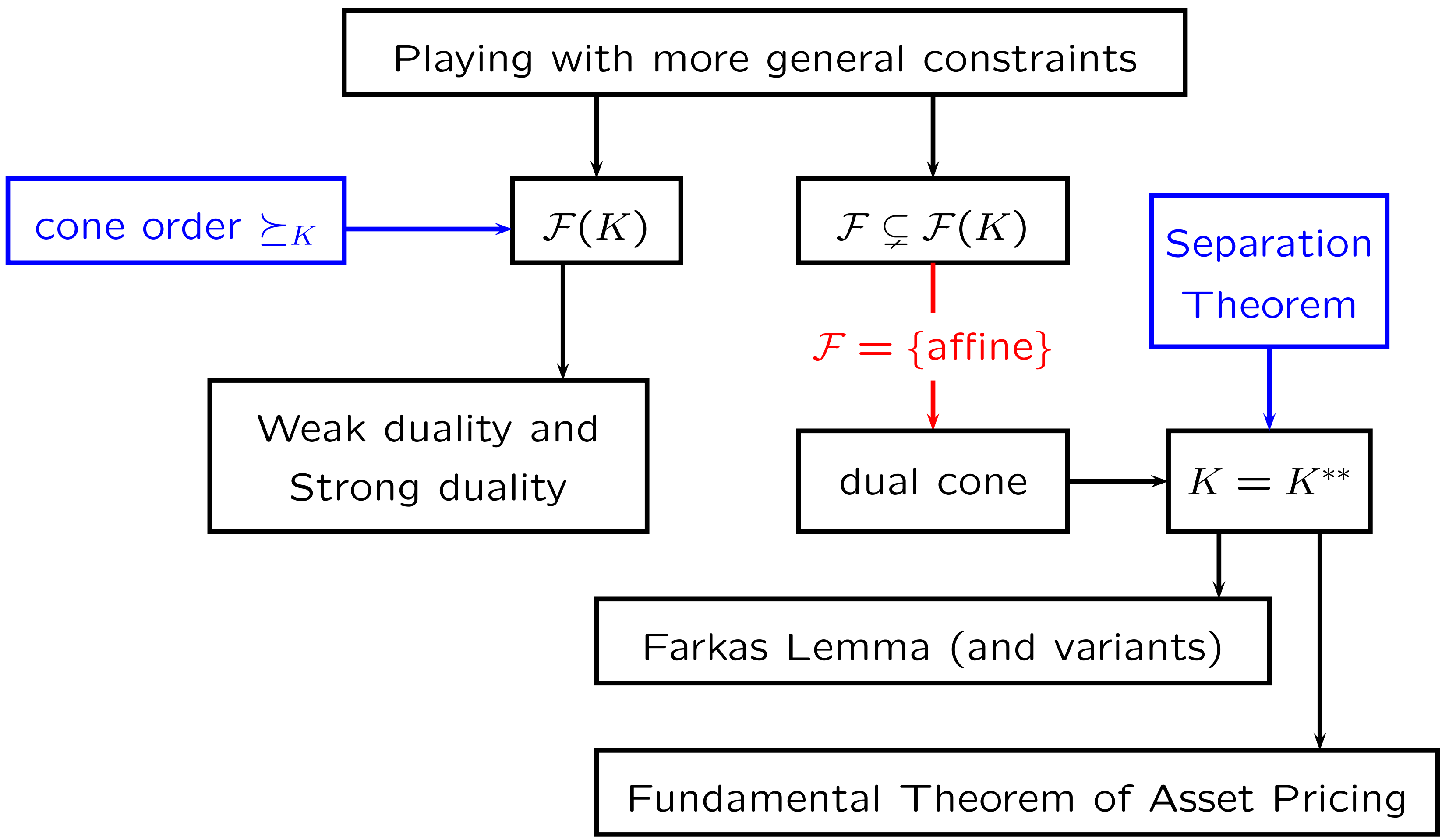
- Convexity also occurs with matrices:
  the SDP cone is convex.
- Convex functions:
  definition, supremum, Minkowski function.

- Hahn-Banach Extension Theorem
- Separation theorems:
  open convex vs. convex,
  closed convex vs. compact convex,
  affine vs. relatively open convex.
- Convex LSC functions are supremum of affine functions

Duality: Guideline of the lecture (1/2)



Duality: Guideline of the lecture (2/2)



How can you check the quality of your almost optimal solution?

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$.

- Suppose that we need to solve $f(x) = 0$.
  How good is an approximate solution $\hat{x}$?
  Just compute $f(\hat{x})$!

- Suppose that we need to solve $\min f(x)$.
  How good is an approximate solution $\hat{x}$?
  Ideally, you need to compute $f(\hat{x}) - f^*$.
  However, you don't know f^*!

Duality is an indirect way of checking the quality of $\hat{x}$

Duality: proving that things don't exist

$$\begin{array}{rrcrcccrl}
\text{min} & y_1 & + & y_2 & + & \cdots & + & y_{2999} & \\
\text{s.t.} & y_1 & + & 2y_2 & + & \cdots & + & 2999y_{2999} & \geq 1 \\
 & 2999y_1 & + & 2998y_2 & + & \cdots & + & y_{2999} & \geq 99
\end{array}$$

The optimal value is greater than 1/30

(sum up the two inequalities and divide by 3000)

Duality is **nothing else**
than a generalization of this trick

A first generalization avenue

$$
\begin{array}{rrcrcccrll}
\text{min} & y_1 & + & y_2 & + & \cdots & + & y_{2999} & & \\
\text{s.t.} & \ln(y_1 & + & 2y_2 & + & \cdots & + & 2999y_{2999}) & \geq & 0 \\
 & \exp(2999y_1 & + & 2998y_2 & + & \cdots & + & y_{2999}) & \geq & C
\end{array}
$$

- We can have nonlinear constraints
- We can even have a nonlinear objective

There will be an **extra** "generalization" soon!

Another use of duality: setting taxes

In the long run, electricity producers might be submitted to CO_2 emission constraints. Legislators have to set fines.

Let x_c be the energy to be produced in coal plants [MWh]
x_g be the energy to be produced in natural gas plants [MWh]
x_n be the energy to be produced in nuclear plants [MWh]
d be the demand to satisfy [MWh]
E be the maximal CO_2 emission allowed [Tons]
c_i be the capacity of production of mean i [MWh]

Another use of duality: setting taxes

In the long run, electricity producers might be submitted to CO_2 emission constraints. Legislators have to set fines.

$$\begin{array}{lrcrcrcll}
\text{min} & 22x_c & + & 75x_g & + & 17x_n & & & \text{[Euros]} \\
\text{s.t.} & x_c & + & x_g & + & x_n & \geq & d & \text{[MWh]} \\
 & 0.9x_c & + & 0.6x_g & & & \leq & E & \text{[Tons]} \\
 & 0 & \leq & x_i & \leq & c_i & \forall i & & \text{[MWh]}
\end{array}$$

Question: How much could the producer save if he were allowed to exceed E ? (i.e. how do we need to tax him ?)

Duality can tell you!

A very general optimization problem

A very general duality theory

Let $X \subseteq \mathbb{R}^n$ be the set of allowed activity levels x

$f(x) \in \mathbb{R}$ be the cost induced by the activity level x

$g(x) \in \mathbb{R}^m$ be the constraints caused by the activity level x

$b \in \mathbb{R}^m$ represent the level of these constraints

$$
\begin{aligned}
p^* := \inf \quad & f(x) \\
\text{s.t.} \quad & g(x) \succeq b \\
& x \in X
\end{aligned}
$$

1. $u \succeq v$ means here $u_i \geq v_i$ for all component i

2. Many such representations exist for the same problem.

Illustration: the problem $\min\{f(x) : g_1(x) \geq 0, g_2(x) \geq 0\}$
could also be written as $\min\{f(x) : g_1(x) \geq 0, x \in X\}$,
where $X := \{x : g_2(x) \geq 0\}$.

Playing with constraints

$$\begin{aligned} p^* := \text{inf} \quad & f(x) \\ \text{s.t.} \quad & g(x) \succeq b \\ & x \in X \end{aligned}$$

Let $\mathcal{F}(\mathbb{R}^m_+)$ be the set of *increasing functions* $F : \mathbb{R}^m \to \mathbb{R} \cup \{\pm\infty\}$:

$$F(u) \geq F(v) \text{ when } u \succeq v.$$

$g(x) \succeq b$ implies $F(g(x)) \geq F(b)$ when $F \in \mathcal{F}(\mathbb{R}^m_+)$.
If $f(x) \geq F(g(x))$ for all $x \in X$, then $p^* \geq F(b)$.
Looking for the **best** function in $\mathcal{F}(\mathbb{R}^m_+)$,
we get the *dual problem*:

$$\begin{aligned} p^* \geq d^* := \text{sup} \quad & F(b) \\ \text{s.t.} \quad & F(g(x)) \leq f(x),\ \forall x \in X \\ & F \in \mathcal{F}(\mathbb{R}^m_+). \end{aligned}$$

Weak duality comes easily

We have organized everything to get:

Theorem 1 (Weak Duality) *We have $p^* \geq d^*$, that is:*

$$\begin{array}{lll} \text{inf} & f(x) & \geq & \text{sup} & F(b) \\ \text{s.t.} & g(x) \succeq b & & \text{s.t.} & F(g(x)) \leq f(x) \ \forall x \in X \\ & x \in X & & & F \in \mathcal{F}(\mathbb{R}^m_+). \end{array}$$

(Recall $f(x) \geq F(g(x)) \geq F(b)$ for feasible x)

- As there are many ways of defining X vs. g, there are many possible duals for an optimization problem.
- Even better: in this setting, equality always holds.

Equality always holds
The perturbation function solves the dual

Theorem 2 (Strong Duality) *We have* $p^* = d^*$*, that is:*

$$\begin{array}{llcll} \inf & f(x) & = & \sup & F(b) \\ \text{s.t.} & g(x) \succeq b & & \text{s.t.} & F(g(x)) \le f(x) \ \forall x \in X \\ & x \in X & & & F \in \mathcal{F}(\mathbb{R}^m_+). \end{array}$$

Proof: Define the *perturbation function* ϕ as follows:

$$\begin{array}{rl} \phi(b) := \inf & f(x) \\ \text{s.t.} & g(x) \succeq b \\ & x \in X. \end{array}$$

Observe that $\phi \in \mathcal{F}(\mathbb{R}^m_+)$ and $\phi(g(x)) \le f(x) \ \forall x \in X$. Thus:

$$\begin{array}{llcllc} \phi(b) = \inf & f(x) & \ge & \sup & F(b) & \ge \phi(b) \\ \text{s.t.} & g(x) \succeq b & & \text{s.t.} & F(g(x)) \le f(x) \ \forall x \in X & \\ & x \in X & & & F \in \mathcal{F}(\mathbb{R}^m_+) & \end{array}$$

The cost for strong duality

This dual is too complicated

- it contains infinitely many constraints
- $\mathcal{F}(\mathbb{R}^m_+)$ is huge (every multivariate increasing function)

So, use a subclass $\mathcal{F}$ instead of the full $\mathcal{F}(\mathbb{R}^m_+)$

- Weak duality still holds
- Strong duality probably not.

Central issue of duality:
Given a problem class, choose
a simple $\mathcal{F}$ for which strong duality holds

Restricting the dual: When we don't loose anything

$$
\begin{aligned}
d^*(\mathcal{F}) = \sup \quad & F(b) \\
\text{s.t.} \quad & F(g(x)) \le f(x), \ \forall x \in X \\
& F \in \mathcal{F}.
\end{aligned}
$$

The quantity $p^* - d^*(\mathcal{F})$ is called the *duality gap* for $\mathcal{F}$

Recall the perturbation function:

$\phi(b) := \inf\{f(x) : g(x) \succeq b, x \in X\}$.

Theorem 3 (Strong duality)

If $\phi = \sup_{i \in I} F_i$ for some $F_i \in \mathcal{F}$, then $p^* = d^*(\mathcal{F})$.

If $\phi \in \mathcal{F}$, then ϕ is also optimal for the dual.

Proof: Since $F_i(g(x)) \le \phi(g(x)) \le f(x)$ when $x \in X$, $i \in I$,

$$
\begin{aligned}
\phi(b) = \inf \quad & f(x) \quad = \quad p^* \ \ge \ d^*(\mathcal{F}) \ \ge \sup_{i \in I} F_i(b) = \phi(b). \\
\text{s.t.} \quad & g(x) \succeq b \\
& x \in X
\end{aligned}
$$

A direct application:

Strong duality in Nonlinear Programming

Consider the class $\mathcal{F} = \{x \mapsto \langle u, x \rangle + u_0 : u \succeq 0\}$ of nondecreasing affine functions. We know that:

$$\tilde{\mathcal{F}} := \left\{ \sup_{i \in I} F_i : F_i \in \mathcal{F} \right\}$$

is the set of LSC, convex and nondecreasing functions.

Theorem 4 *If the perturbation function ϕ of a problem is LSC and convex then strong duality holds with $\mathcal{F}$ as above.*

Interestingly, this result, or some immediate variants of it, keeps being rediscovered in a myriad of engineering papers.

A direct application:

Strong duality in Nonlinear Programming

Theorem 4 *If the perturbation function ϕ of a problem is LSC and convex then strong duality holds with $\mathcal{F}$ as above.*

- Sometimes, proving "ϕ is LSC and convex" is easy.
- Perturbation functions are automatically nondecreasing.
- **Benefit**: The exact dual is a convex problem.

$$
\begin{aligned}
d^*(\mathcal{F}) = \sup \quad & \langle u, b \rangle + u_0 \\
\text{s.t.} \quad & \langle u, g(x) \rangle + u_0 \leq f(x), \ \forall x \in X \\
& u \succeq 0
\end{aligned}
$$

$$
\begin{aligned}
\text{or: } d^*(\mathcal{F}) = \sup \quad & \inf_{x \in X} \langle u, b - g(x) \rangle + f(x) \\
\text{s.t.} \quad & u \succeq 0.
\end{aligned}
$$

Retrieving info from zero duality gap

Theorem 5 (Complementarity conditions) *Suppose that x^* and F^* are feasible for their respective problems, and that $f(x^*) = F^*(b)$. Then:*

$$g(x^*) \succeq b, \quad F^*(g(x)) \le f(x) \; \forall x \in X.$$

$$p^* = f(x^*) = F^*(g(x^*)) = F^*(b) = d^*(\mathcal{F}).$$

Proof: First line comes from feasibility. Next,

$$f(x^*) \ge F^*(g(x^*)) \ge F^*(b) = f(x^*).$$

Also $f(x^*) \ge p^* \ge d^*(\mathcal{F}) \ge F^*(b) = f(x^*).$

Intuitive interpretation of the dual maximizer

$$
\begin{aligned}
d^* = \sup \quad & F(b) \\
\text{s.t.} \quad & F(g(x)) \leq f(x),\ \forall x \in X \\
& F \in \mathcal{F}.
\end{aligned}
$$

We know that the optimal F is the perturbation function

$$\phi(b) := \inf\{f(x) : g(x) \succeq b, x \in X\}.$$

ϕ describes the sensitivity of the optimal value wrt b
and is thus critical in robustness analysis,
for pricing constraints, etc...

Intuitive interpretation of the dual maximizer

$$\begin{aligned} d^* = \text{sup} \quad & F(b) \\ \text{s.t.} \quad & F(g(x)) \leq f(x),\ \forall x \in X \\ & F \in \mathcal{F}. \end{aligned}$$

We know that the optimal F is the perturbation function

$$\phi(b) := \inf\{f(x) : g(x) \succeq b, x \in X\}.$$

Suppose we strengthen the constraints by changing b by $b + h$ (with h small). If f is a pricing function, the minimal cost goes from $\phi(b)$ to $\phi(b + h)$. Its relative increase is $(\phi(b + h) - \phi(b))/||h||$.

The price of constraint i at b is $\partial\phi(b)/\partial b_i$ if ϕ is differentiable at b.

A superficial generalization which proves to be extremely useful

$$p^* := \inf \quad f(x)$$
$$\text{s.t.} \quad g(x) \succeq b$$
$$x \in X$$

The type of inequalities can be modified

We want $g(x)$ to be at least as good as b.

The desirable axioms for such an order would be:

- **Transitivity:** If $g(\hat{x}) \succeq g(x)$ and $g(x) \succeq b$, a fortiori $g(\hat{x}) \succeq b$.
- **Reflexivity:** $b \succeq b$ for all b.
- **Antisymmetry:** If $a \succeq b$ and $b \succeq a$, then $a = b$.
- **Translation invariance:** If $g(x) \succeq b$ then $g(x)+c \succeq b+c$.
- **Homogeneity:** If $a \succeq b$, $t > 0$ then $ta \succeq tb$
- **Closedness:** If $a_n \to a$ and $a_n \succeq b$, then $a \succeq b$.

Convexity appears in the notion of order

Let $\succeq$ be any order satisfying for all $a, b, c \in \mathbb{R}^m$:

- **Transitivity (T):** If $a \succeq b$ and $b \succeq c$, then $a \succeq c$.
- **Reflexivity (R):** $b \succeq b$ for all b.
- **Antisymmetry (A):** If $a \succeq b$ and $b \succeq a$, then $a = b$.
- **Translation invariance (TI):** If $a \succeq b$ then $a + c \succeq b + c$.
- **Homogeneity (H):** If $a \succeq b$, $t > 0$ then $ta \succeq tb$
- **Closedness (C):** If $a_n \to a$ and $a_n \succeq b$, then $a \succeq b$.

Let $K := \{x \in \mathbb{R}^m : x \succeq 0\}$.
By **(H - C)**, K is a closed cone. By **(R)**, $0 \in K$.
By **(A - TI)**, K does not contain any straight line.
By **(TI)**, the order is defined entirely by K:

$$a \succeq b \text{ iff } a - b \in K.$$

By **(T)**, K is a convex cone.

Notation: $\succeq_K$ is the order for which
$K := \{x \in \mathbb{R}^m : x \succeq_K 0\}$.

Some examples of "order" cones

Almost all the cones we have seen qualify as order-defining cones.

- $K = \{0\}$ yields the order $a \succeq_{\{0\}} b$ iff $a = b$.
- $K = \mathbb{R}^m_+$ gives the order we used: $a \succeq_{\mathbb{R}^m_+} b$ iff $a_i \geq b_i\ \forall i$.
- The Lorentz cone $\{(t,x) \in \mathbb{R} \times \mathbb{R}^n : ||x||_2 \leq t\}$ yields the order used in *second-order programming*.
- The SDP cone $\mathbb{S}^N_+$ is used in *semidefinite programming*.
- The light cone of every compact convex set induces an order-defining cone.

A new set of increasing functions

$$
\begin{array}{rl}
p^* := \inf & f(x) \\
\text{s.t.} & g(x) \succeq_K b \\
& x \in X.
\end{array}
$$

Define

$$\mathcal{F}(K) := \{F : \mathbb{R}^m \to \mathbb{R} \cup \{\pm\infty\} : F(u) \geq F(v) \text{ when } u \succeq_K v\}.$$

Of course, weak duality still holds:

$$
\begin{array}{rl}
\inf & f(x) \\
\text{s.t.} & g(x) \succeq_K b \\
& x \in X
\end{array}
\quad \geq \quad
\begin{array}{rl}
\sup & F(b) \\
\text{s.t.} & F(g(x)) \leq f(x) \; \forall x \in X \\
& F \in \mathcal{F}(K).
\end{array}
$$

Strong duality holds, the complementarity theorem too.

(Just replicate the proof with the perturbation function).

A closer look at the set $\mathcal{F}(K)$

The dual cone

$$\mathcal{F}(K) := \{F : \mathbb{R}^m \to \mathbb{R} \cup \{\pm\infty\} : F(u) \geq F(v) \text{ when } u \succeq_K v\}$$

What are the linear functions in $\mathcal{F}(K)$?

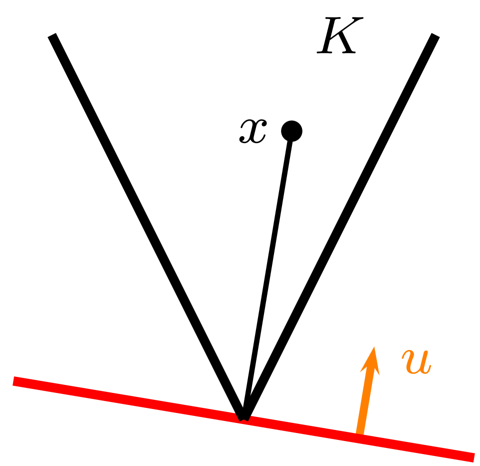
We write $\langle \cdot, \cdot \rangle$ for a scalar product in $\mathbb{R}^m$.
Let $x \mapsto \langle u, x \rangle$ be in $\mathcal{F}(\mathcal{K})$; then $x \succeq_K y \Rightarrow \langle u, x \rangle \geq \langle u, y \rangle$, or:

$$x - y \in K \;\Rightarrow\; \langle u, x - y \rangle \geq 0$$

That is $\langle u, z \rangle \geq 0$ for all $z = y - x \in K$.
The *dual cone* of K is $K^* := \{u \in \mathbb{R}^m : \langle u, z \rangle \geq 0 \; \forall z \in K\}$.

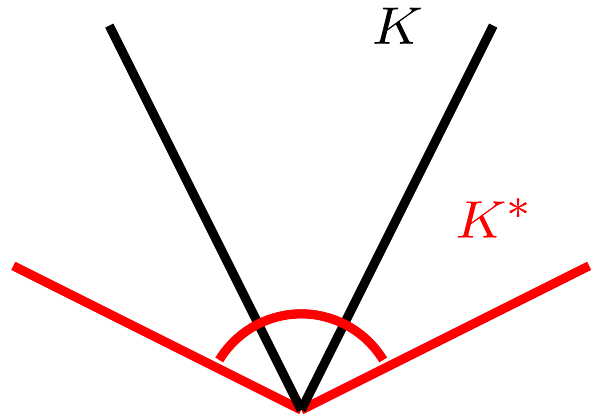
Everything you need to know about the dual cone

$$K^* := \{u \in \mathbb{R}^m : \langle u, x \rangle \geq 0 \ \forall x \in K\}$$



Everything you need to know about the dual cone

$$K^* := \{u \in \mathbb{R}^m : \langle u, x\rangle \geq 0 \ \forall x \in K\}$$



K^* is the intersection of all the half-spaces whose normal is in K

Everything you need to know about the dual cone

- K^* is closed and convex ($K^* = \cap_{x \in K}\{u : \langle u, x\rangle \geq 0\}$).
- If $K_1 \subseteq K_2$, then $K_1^* \supseteq K_2^*$.
- If K is closed and convex, then $K^{**} = K$.

Proof: $K^{**} \supseteq K$. Let $\hat{x} \in K$; as $\langle u, \hat{x}\rangle \geq 0\ \forall u \in K^*$, we deduce $\hat{x} \in K^{**}$. $K^{**} \subseteq K$. Let $\hat{x} \notin K$; we prove that $\hat{x} \notin K^{**}$, i.e. $\exists u$ such that $\langle u, \hat{x}\rangle < 0$ and $\langle u, x\rangle \geq 0\ \forall x \in K$. Since K is closed, there exists $\epsilon > 0$ so that $B := B(\hat{x}, \epsilon)$ does not intersect K. The Separation Theorem gives $\hat{f}$ linear (say, $\hat{f}(x) = \langle u, x\rangle$) such that $\hat{f}(y) < \hat{f}(z)$ for every $y \in B$, $z \in K$. In particular, as $\hat{x} \in B$ and $0 \in K$, we have $\hat{f}(\hat{x}) < 0$. Finally, $\hat{f}(x) \geq 0$ for $x \in K$: suppose on the contrary that $\hat{f}(z) < 0$ for a $z \in K$; as $\lambda z \in K$ when $\lambda > 0$, $\lambda \hat{f}(z)$ can be taken smaller than $\hat{f}(\hat{x})$, a contradiction.

- If K is a (possibly nonconvex) set, we can adapt the proof to show $K^{**} = \mathrm{cl}(\mathrm{cone}(K))$.

Some dual cones 1

- $(\mathbb{R}^m)^* = \{0\}$.

Proof: $(\mathbb{R}^m)^* = \{u : \langle u, x\rangle \geq 0 \ \forall x \in \mathbb{R}^m\} = \{0\}$.

- $\{0\}^* = \mathbb{R}^m$.

Proof: Use $\{0\}^* = ((\mathbb{R}^m)^*)^* = \mathbb{R}^m$.
Alternatively, $\{0\}^* = \{u : \langle u, 0\rangle \geq 0\} = \mathbb{R}^m$.

- $(\mathbb{R}^m_+)^* = \mathbb{R}^m_+$ with the *dot scalar product* $x^T y$.

Proof: Let $u \in (\mathbb{R}^m_+)^*$. Then $u_i \geq 0$ for every i,
for otherwise $u^T e_i = u_i < 0$ with $e_i := (0, \ldots, 0, 1, 0 \ldots 0)^T$
being the ith canonical unit vector.
Conversely, if $u \in \mathbb{R}^m_+$, then $\sum_i u_i x_i \geq 0$ when $x \in \mathbb{R}^m_+$.

Some dual cones 2

- $(\mathbb{S}_+^N)^* = \mathbb{S}_+^N$ with the Frobenius scalar product.

Proof: $(\mathbb{S}_+^N)^* \supseteq \mathbb{S}_+^N$. It comes immediately from $\text{Tr}(XY) \geq 0$ when $X, Y \in \mathbb{S}_+^N$.

$(\mathbb{S}_+^N)^* \subseteq \mathbb{S}_+^N$. Assume $U \in (\mathbb{S}_+^N)^*$, but $U \notin \mathbb{S}_+^N$. Then U has a negative eigenvalue $\lambda_i(U)$. We denote by u_i a corresponding normed eigenvector. Letting $X := u_i u_i^T$, we have $X \in \mathbb{S}_+^N$, and $\langle U, X \rangle_F = u_i^T U u_i = \lambda_i(U) < 0$, a contradiction.

An important application: Farkas Lemma

An intermediate result

Let $a_1, \ldots, a_m, b \in \mathbb{R}^n$.

There exist $t_1, \ldots, t_m \in \mathbb{R}_+$

such that $b = \sum_i t_i a_i$

iff $\forall z \in \mathbb{R}^n$ such that

$z^T a_1 \geq 0, \ldots, z^T a_m \geq 0$,

we have $z^T b \geq 0$.

Let (S_1) be the first statement and (S_2) the second one.

- Set $K = \text{cone}\{a_1, \ldots, a_m\}$. $(S_1) \Leftrightarrow b \in K$.
- Let $z \in \mathbb{R}^n$ such that $z^T a_1 \geq 0, \ldots, z^T a_m \geq 0$.

  Then $z \in K^*$ by definition of K. Thus $(S_2) \Leftrightarrow b \in K^{**}$
- $K = K^{**}$ because K is closed and convex.

Both statements are thus equivalent.

Another view on dual cones: Farkas Lemma

An intermediate result

Let $a_1, \ldots, a_m, b \in \mathbb{R}^n$.

There exist $t_1, \ldots, t_m \in \mathbb{R}_+$ such that $b = \sum_i t_i a_i$

iff $\forall z \in \mathbb{R}^n$ such that $z^T a_1 \geq 0, \ldots, z^T a_m \geq 0$,

we have $z^T b \geq 0$.

Farkas Lemma Let $A \in \mathbb{R}^{n \times m}$ and $b \in \mathbb{R}^n$. Then exactly one of the following system has a solution:

$$\textbf{a.} \quad Ax = b, \quad x \succeq_{\mathbb{R}^m_+} 0$$

$$\textbf{b.} \quad A^T y \succeq_{\mathbb{R}^m_+} 0, \quad b^T y < 0.$$

Proof: replace t by x and z by y in the intermediate result.

Other versions of Farkas Lemma 1

Farkas Lemma Let $A \in \mathbb{R}^{n\times m}$ and $b \in \mathbb{R}^n$. Then exactly one of the following system has a solution:

$$\textbf{a.} \quad Ax = b, \quad x \succeq_{\mathbb{R}^m_+} 0$$

$$\textbf{b.} \quad A^T y \succeq_{\mathbb{R}^m_+} 0, \quad b^T y < 0.$$

Variant 1: Checking that a polyhedron is empty

Exactly one of the following system has a solution:

$$\textbf{a.} \quad Ax \preceq_{\mathbb{R}^n_+} b, \quad x \succeq_{\mathbb{R}^m_+} 0$$

$$\textbf{b.} \quad A^T y \succeq_{\mathbb{R}^m_+} 0, \quad y \succeq_{\mathbb{R}^n_+} 0, \quad y^T b < 0$$

Verification of Farkas $\Leftrightarrow$ Variant 1:

$\Rightarrow$ [**a.** $\Leftrightarrow Ax + s = b, x \succeq 0, s \succeq 0$]. Use Farkas replacing A by $[A, I]$.

$\Leftarrow$ Replace A by $[A; -A]$ and b by $[b; -b]$ in Variant 1.

Other versions of Farkas Lemma 2

Farkas Lemma Let $A \in \mathbb{R}^{n \times m}$ and $b \in \mathbb{R}^n$. Then exactly one of the following system has a solution:

$$\textbf{a.} \quad Ax = b, \quad x \succeq_{\mathbb{R}^m_+} 0$$

$$\textbf{b.} \quad A^T y \succeq_{\mathbb{R}^m_+} 0, \quad b^T y < 0.$$

Variant 2: Tucker's Key Theorem

Let $B \in \mathbb{R}^{n \times m}$. Then there exist $x \in \mathbb{R}^m$, $y \in \mathbb{R}^m$ for which:

$$Bx = 0, \quad x \succeq_{\mathbb{R}^m_+} 0, \quad B^T y \succeq_{\mathbb{R}^n_+} 0, \quad [B^T y + x]_i > 0 \text{ for all } i.$$

This is an important theorem in Linear Programming
allowing us also to prove the "hard part"
of the Fundamental Theorem of Asset Pricing

Note: There is no assumption B (!)

Farkas implies Tucker:

A proof in two steps

1. Suppose that $M \in \mathbb{R}^{N\times N}$ is a *skew-symmetric* matrix: $M^T = -M$ (Thus $x^T M x = 0$ and $x^T M y = -y^T M x$ for all $x, y \in \mathbb{R}^N$).

Claim: $\exists z \in \mathbb{R}^N_+$ with $Mz \in \mathbb{R}^N_+$ and $[Mz + z]_i > 0$ for all i. Equivalently,

$$\exists y \text{ such that } A^T y := \begin{pmatrix} M & 0 \\ M+I & -\mathbf{1} \\ I & 0 \end{pmatrix} \begin{pmatrix} z \\ \epsilon \end{pmatrix} \succeq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \text{ and } b^T y := -\epsilon < 0.$$

By Farkas, we need to check that there is no $x \succeq_{\mathbb{R}^N_+} 0$ for which $Ax = b$.

Otherwise, with $x = [u; v; w]$: $M^T u + M^T v + v + w = 0$ and $\mathbf{1}^T v = 1$.

Left-multiply the first equation by u^T then by v^T, and use $u, v, w \succeq 0$:

we get $u^T M^T v + u^T v \le 0$ and $v^T M^T u + ||v||_2^2 \le 0$. But $u^T M^T v = -v^T M^T u$,

so $0 \le u^T v \le v^T M^T u \le -||v||_2^2$ and $v = 0$, contradicting $\mathbf{1}^T v = 1$.

2. Define $M := \begin{pmatrix} 0 & A^T & -A^T \\ -A & 0 & 0 \\ A & 0 & 0 \end{pmatrix}$ and $z = \begin{pmatrix} x \\ y_+ \\ y_- \end{pmatrix}$. We're done.

Farkas Lemma also works for convex cone ordering*

Let $\langle \cdot, \cdot \rangle$ be a scalar product on $\mathbb{R}^n$, let $A \in \mathbb{R}^{n \times m}$, let A^* be the *adjoint* of A w.r.t. $\langle \cdot, \cdot \rangle$, let $b \in \mathbb{R}^n$ and let K be a closed convex cone with no line. Then exactly one of the following system has a solution:

$$\textbf{a.} \quad Ax = b, \quad x \succeq_K 0$$

$$\textbf{b.} \quad A^* y \succeq_{K^*} 0, \quad b^T y < 0.$$

Proof: Observe first that the dual of $AK := \{Ax : x \in K\}$ is:

$$(AK)^* = \{u \in \mathbb{R}^m : u^T A x \geq 0 \; \forall x \in K\}$$

$$= \{u \in \mathbb{R}^m : \langle A^* u, x \rangle \geq 0 \; \forall x \in K\} = \{u \in \mathbb{R}^m : A^* u \in K^*\}.$$

Now, **a.** means that $b \in AK$, i.e. $u^T b \geq 0$ for all $u \in (AK)^*$, i.e. $A^* y \in K^*$ implies $b^T y \geq 0$.

An application in Math Finance 1/3: Arbitrage

We have: d risky assets $v_1, \ldots, v_d$,
one risk-free asset v_0, with risk-free rate r.
Their initial price ($t = 0$) is $\pi_0, \pi_1, \ldots, \pi_d$,
and at time $t = 1$ becomes $S_0, S_1, \ldots, S_d$.
a portfolio $\xi \in \mathbb{R}^{d+1}$ (i.e. ξ_i units of v_i).
$\xi_i < 0$ is allowed (we *short* v_i).

$S_0 = (1 + r)\pi_0$, S_i is non-deterministic for $i > 0$
(all random variables on a prob. space $(\Omega, \Sigma, \mathbb{P})$).

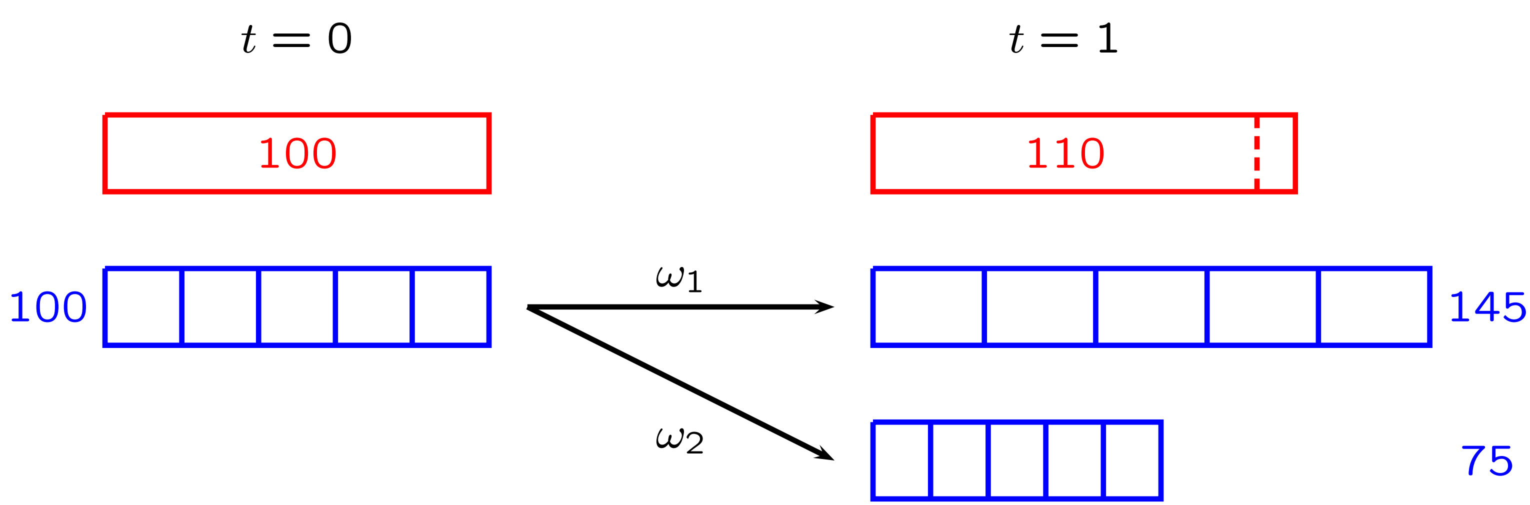
ξ is an *arbitrage* if:

- $\xi^T \pi \leq 0$, [possible initial cost]
- $\xi^T S \geq 0$ a.s., [no loss with probability 1]
- $\mathbb{P}[\xi^T S > 0] > 0$. [a chance of winning]

An application in Math Finance 1/3:

Arbitrage: an example

Assume $\Omega = \{\omega_1, \omega_2\}$, $d = 1$, $r = 0.1$.
Let $\pi_0 = 100$, $\pi_1 = 20$, $S_1(\omega_1) = 29$, $S_1(\omega_2) = 15$
If $\xi = (-1; 5)$, then $\xi^T \pi = 0$.



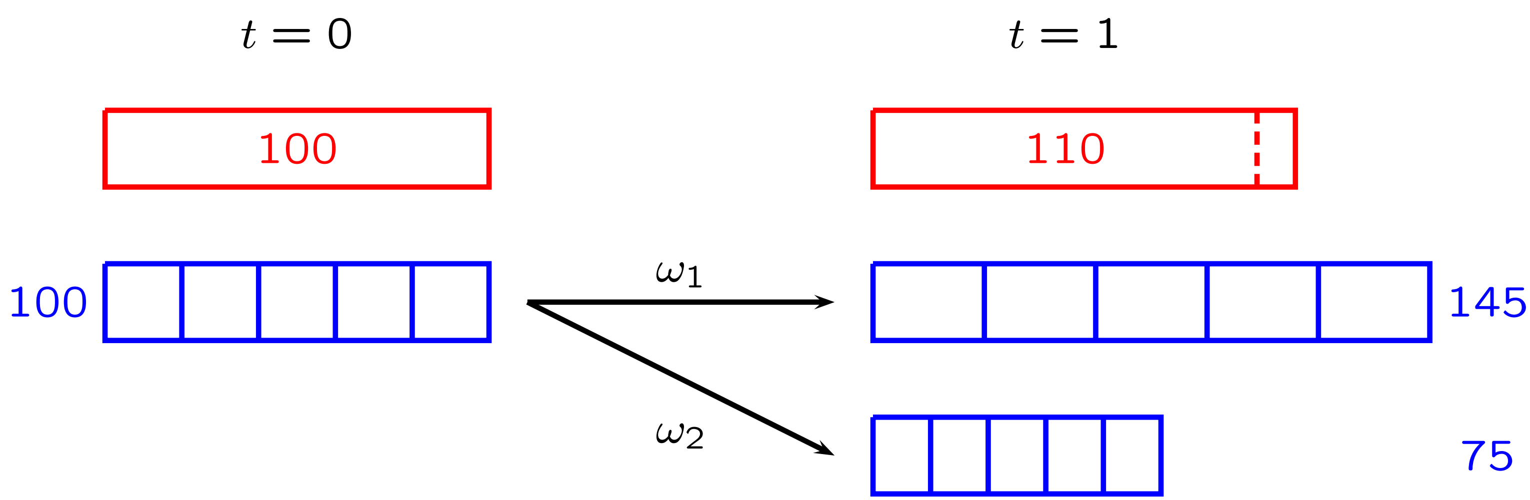
We do not know $\mathbb{P}[\omega_1]$.

If $\mathbb{P}[\omega_1] = 4/5$, then $\mathbb{E}[5 \cdot S_1] = 4/5 \cdot 145 + 1/5 \cdot 75 = 131 > 110$

An application in Math Finance 1/3:

Arbitrage: an example

Assume $\Omega = \{\omega_1, \omega_2\}$, $d = 1$, $r = 0.1$.
Let $\pi_0 = 100$, $\pi_1 = 20$, $S_1(\omega_1) = 29$, $S_1(\omega_2) = 15$
If $\xi = (-1; 5)$, then $\xi^T \pi = 0$.



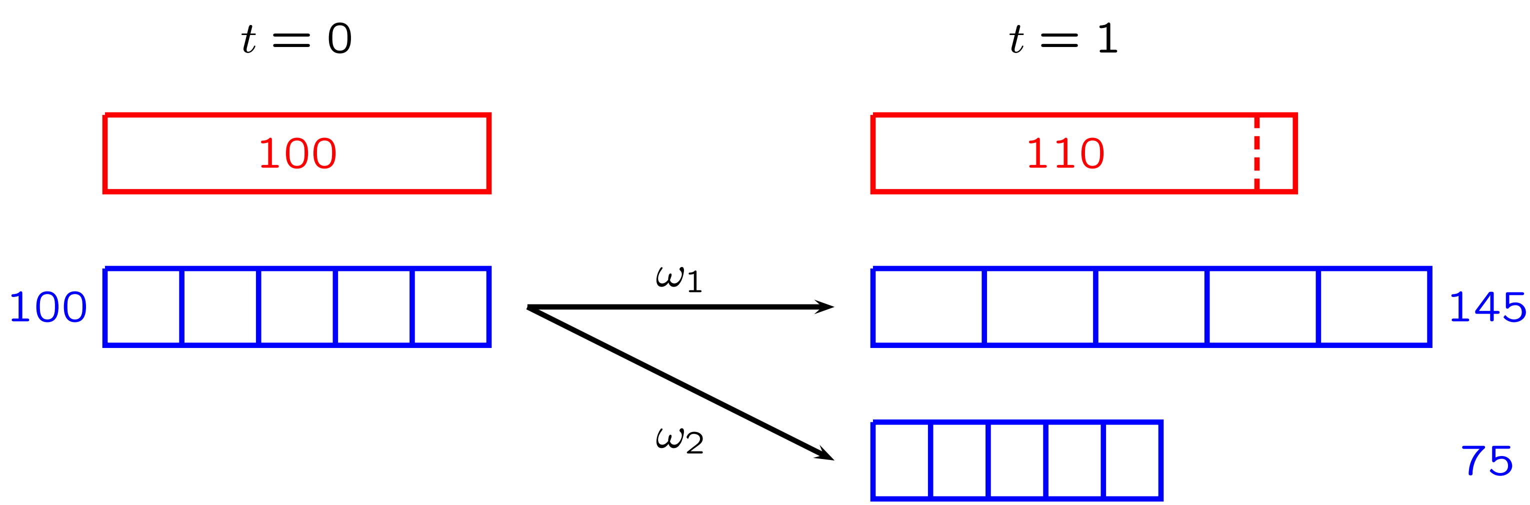
We do not know $\mathbb{P}[\omega_1]$.

If $\mathbb{P}[\omega_1] = 2/5$, then $\mathbb{E}[5 \cdot S_1] = 2/5 \cdot 145 + 3/5 \cdot 75 = 103 < 110$

An application in Math Finance 1/3:

Arbitrage: an example

Assume $\Omega = \{\omega_1, \omega_2\}$, $d = 1$, $r = 0.1$.
Let $\pi_0 = 100$, $\pi_1 = 20$, $S_1(\omega_1) = 29$, $S_1(\omega_2) = 15$
If $\xi = (-1; 5)$, then $\xi^T \pi = 0$.



We do not know $\mathbb{P}[\omega_1]$.

If $\mathbb{P}[\omega_1] = 1/2$, then $\mathbb{E}[5 \cdot S_1] = 1/2 \cdot 145 + 1/2 \cdot 75 = 110$

An application in Math Finance 2/3: Risk-neutral probability

ξ is an *arbitrage* if:

- $\xi^T \pi \leq 0$, [possible initial cost]
- $\xi^T S \geq 0$ a.s., [no loss with probability 1]
- $\mathbb{P}[\xi^T S > 0] > 0$. [a chance of winning]

Suppose that for all i, we would have $\mathbb{E}[S_i] = \pi_i(1+r)$.

- Advantage: Computing $\mathbb{E}[\xi^T S]$ is trivial for every ξ.
- Annoying issue: It is wrong. Otherwise
  - there would be no compensation for the risk of assets
  - why investing in risky assets when S_0 does as well?

However, if there is no arbitrage, there exists a probability $\mathbb{P}^*$ for which $\mathbb{E}_{\mathbb{P}^*}[S_i] := \int S_i(\omega) d\mathbb{P}^*[\omega] = \pi_i(1+r)$ for all i.

An application in Math Finance 2/3: Risk-neutral measure

A *risk-neutral measure* $\mathbb{P}^*$ is a probability equivalent to $\mathbb{P}$ for which $\mathbb{E}_{\mathbb{P}^*}[S_i] = \pi_i(1+r)$ for all i.
(in the example, $\mathbb{P}^*(\omega_1) = \mathbb{P}^*(\omega_2) = 1/2$.)

If there is no arbitrage, there exists a probability $\mathbb{P}^*$ for which $\mathbb{E}_{\mathbb{P}^*}[S_i] := \int S_i(\omega) d\mathbb{P}^*[\omega] = \pi_i(1+r)$ for all i.

This statement and its reciprocal constitutes the *Fundamental Theorem of Asset Pricing*.
(Proved here for finite Ω. More general versions exist, e.g. for infinite Ω and multiple time-steps.)

(Still) heavily used in practice to price financial products
Warning: ignores other sources of risk
and execution costs.

An application in Math Finance 3/3

Theorem 6 (Fundamental Theorem of Asset Pricing) *Suppose Ω is finite. A market S has no arbitrage iff there exists a risk-neutral measure.*

Proof: $\Leftarrow$: Let $\mathbb{P}^*$ be a risk-neural measure and ξ an arbitrage. Then $0 \geq \xi^T \pi = \mathbb{E}_{\mathbb{P}^*}[\xi^T S]/(1+r) > 0$. $\not\subset$

$\Rightarrow$: Define $S(\Omega') := \{S(\omega) : \omega \in \Omega'\} \subseteq \mathbb{R}^{d+1}$ for $\Omega' \subseteq \Omega$. Note that if $v \in S(\Omega')$, then $v_0 = \pi_0(1+r)$.

Assume no arbitrage. **Claim:** there is Ω' with $\mathbb{P}[\Omega'] = 1$ for which all $\xi \in S(\Omega')^*$ satisfy $\xi^T \pi \geq 0$. Otherwise, we would have $\xi^T S \geq 0$ a.s. with $\xi^T \pi < 0$. By no arbitrage, $\xi^T S = 0$ a.s. Take $\hat{\xi} := \xi + \epsilon e_0$ for $\epsilon > 0$ such that $\hat{\xi}^T \pi \leq 0$. Then $\hat{\xi}^T S = \epsilon \pi_0 (1+r) > 0$, a contradiction.

Thus $\pi(1+r) \in S(\Omega')^{**} = \mathrm{cl}(\mathrm{cone}(S(\Omega')))$ (**Why?**),

and $\pi(1+r) = \sum_{\omega \in \Omega'} p(\omega) S(\omega)$ for $p \geq 0$.

Also, $\sum_{\omega \in \Omega'} p(\omega) = 1$ (because $S_0(\omega) = \pi_0(1+r)$ for all ω).

Then, we complete p on Ω with null components.

What remains to prove:
we can take p equivalent to $\mathbb{P}$

This part is harder and relies essentially on an equivalent reformulation of Farkas'Lemma, *Tucker's Key Theorem*:

Theorem 7 *Let $A \in \mathbb{R}^{m\times n}$. If there exists $x \in \mathbb{R}^n_+$, $y \in \mathbb{R}^m$ such that $Ax = 0$ and $A^T y \in \mathbb{R}^n_+$, we can take them so that either $x_i > 0$ or $[A^T y]_i > 0$ for all i.*

Claim: suppose that $\pi \in \text{bd}(\text{cone}(S(\Omega')))$, where Ω' is the support of $\mathbb{P}$. Then there exists $u \in S(\Omega')^* \backslash \{0\}$ for which $u^T \pi = 0$. By no arbitrage, $u^T S = 0$ a.s. Define $A_{i,\omega} := S_i(\omega) - \pi_i(1+r)$, so that A is a $(d+1)$-by-$|\Omega|$ matrix. Then $A^T u = 0$ and $Ap = 0$. But then $p(\omega) > 0$ for all $\omega \in \Omega'$, contradicting $\pi \in \text{bd}(\text{cone}(S(\Omega')))$.

Hence π is in the interior of cone$(S(\Omega'))$ and $p(\omega) > 0$.

For next week

- Lagrangian duality: another view on strong duality
- Convex functions, (sub)differentiability