

Spectral Risk Sharing

Stephen J. Mildenhall

2026-03-23



This note presents a summary of the spectral risk sharing theorem from Jouini et al. (2008). The theorem describes how two agents with spectral risk cost functions optimally share risk. It provides a wonderful, intuitive, and constructive solution to the problem, involving a horizontal layering of the total loss, with each horizontal layer assumed by the agent with lowest cost of risk at the tail probability level. The proof can be extended to any finite number of agents. We provide a detailed sketch of the proof, summarize the extensions, and give some examples. Our presentation follows the original argument closely, but translates the notation to use cost functions rather than utilities and the actuarial loss sign convention rather than the payoff investment convention. Throughout, X models losses, and a large positive value of X is undesirable.

The result, once you are told it, seems eminently plausible, but the proof is quite involved and relies on several tricks, ranging from general results about convex functions to specific ones about distortions. These tricks include

- convex duality,
- common supporting subgradient,
- conditional-expectation reduction to functions of X ,
- Landsberger monotone rearrangement,
- Denneberg comonotonicity,
- Kusuoka quantile representation,
- lower-envelope rule $g_0 \wedge g_1$,
- singular dual mass for ess sup terms,
- subgradient = flatness where support is strict, and
- integrate flatness to get stop-loss layers.

The proof touches on some esoteric mathematics: finitely additive measures. These appear of necessity, not perversity; they allow the model to incorporate risk measures used in practice. The interplay of theory and practice illustrates the power of mathematical formalism, how mathematics follows reality, but also shows it enforces its own constraints. Each definition

begets implications that cannot be ignored, however much we might wish. It is a particular thorn that $f \mapsto f(x)$ defines a bounded linear functional on the space of bounded functions that cannot be expressed as an integral with respect to a countably additive measure. Said another way: although the dual of L^1 is L^∞ , and of L^p is L^q for $p, q > 1$, $p^{-1} + q^{-1} = 1$, the dual of L^∞ is bigger than L^1 .

1 Setup and notation

We work on an atomless standard probability space $(\Omega, \mathcal{F}, \mathbb{P})$. All losses lie in L^∞ and so are bounded. For a loss X , write F_X for its distribution function and

$$q_X(u) := \inf \{ x : F_X(x) \geq u \}, \quad 0 \leq u \leq 1,$$

for its quantile function. When $\mu \in (L^\infty)^*$ is a bounded finitely additive functional, the pairing with X is $\langle \mu, X \rangle$; if μ is countably additive with density $Z \in L^1$, then $\langle \mu, X \rangle = \mathbb{P}(ZX)$. The paper keeps the full dual $(L^\infty)^*$ in play because, in the comonotonic additive spectral case, subgradients need not live in L^1 when there is a jump at 0 in the distortion function.

Agent $i \in \{0, 1\}$ evaluates a bounded loss X using a risk-cost functional

$$g_i : L^\infty \rightarrow \mathbb{R}.$$

In this note we take the primitive objects to be these cost functionals, not utilities. The standing assumptions are the risk-measure analogs of the paper's monetary utility assumptions: that g_i are convex, monotone increasing, cash-invariant (aka translation-invariant), and normalized $g_i(0) = 0$. Thus larger losses cost more, diversification does not hurt, and adding a sure amount c to the loss simply adds c to the cost:

$$g_i(X + c) = g_i(X) + c.$$

Law-invariance means $g_i(X)$ depends only on the law of X . Coherence means, in addition, positive homogeneity:

$$g_i(\lambda X) = \lambda g_i(X), \quad \lambda \geq 0.$$

These assumptions are just the risk-measure side of Definition 2.1, Theorem 2.1, and Remark 2.1 in the paper.

Each agent owns (is endowed with) a loss X_i , and the aggregate loss is

$$X := X_0 + X_1.$$

An allocation of X is a pair

$$(\xi_0, \xi_1) \in \mathbb{A}(X) := \{ (\xi_0, \xi_1) \in L^\infty \times L^\infty : \xi_0 + \xi_1 = X \}.$$

Thus, ξ_i is the loss borne by agent i . In the law-invariant setting, the paper shows that one may restrict attention to allocations measurable with respect to X , and in fact to allocations that increase with X ; equivalently, to comonotone allocations. We therefore also introduce

$$\mathbb{A}^\uparrow(X) := \{ (\xi_0, \xi_1) \in \mathbb{A}(X) : \xi_0 \uparrow X, \xi_1 \uparrow X \}.$$

By Denneberg's lemma, this is the same as the set of comonotone allocations.

Since we are minimizing costs, the natural aggregate objective is the inf-convolution

$$(g_0 \square g_1)(X) := \inf_{\xi_0 + \xi_1 = X} (g_0(\xi_0) + g_1(\xi_1)).$$

An allocation $(\xi_0, \xi_1) \in \mathbb{A}(X)$ is efficient, or Pareto optimal in cost language, when there is no competing allocation $(\zeta_0, \zeta_1) \in \mathbb{A}(X)$ with

$$g_0(\zeta_0) \leq g_0(\xi_0), \quad g_1(\zeta_1) \leq g_1(\xi_1),$$

and at least one inequality strict. For these convex cash-invariant criteria, efficiency is equivalent to solving the inf-convolution problem:

$$(\xi_0, \xi_1) \text{ is efficient} \iff g_0(\xi_0) + g_1(\xi_1) = (g_0 \square g_1)(X).$$

This is the cost-version of Theorem 3.1(ii).

Next, we need to introduce conjugates and subgradients. Take a proper convex functional

$$g : L^\infty \rightarrow \mathbb{R}.$$

The Fenchel conjugate of g is

$$g^*(\mu) := \sup_{X \in L^\infty} (\langle \mu, X \rangle - g(X)), \quad \mu \in (L^\infty)^*.$$

At a point $X \in L^\infty$, a subgradient of g at X is a dual functional

$$\mu \in (L^\infty)^*$$

such that

$$g(Y) \geq g(X) + \langle \mu, Y - X \rangle \quad \text{for every } Y \in L^\infty.$$

The set of all such μ is the subdifferential

$$\partial g(X).$$

Thus μ is a linear functional that touches the graph of g at X and stays below it everywhere else. Equivalently,

$$Y \mapsto g(X) + \langle \mu, Y - X \rangle$$

is a supporting affine functional for g at X . If g were differentiable, then $\partial g(X)$ would just be the singleton containing the gradient. The point of the subgradient is that it still makes sense when g is not differentiable (has kinks).

We can interpret

$$\mu \in \partial g(X)$$

as the linearized cost rule $\langle \mu, \cdot \rangle$ touches g at X , or, in the coherent case, that μ is a worst-case dual scenario for X .

Next: convex duality. The Fenchel inequality follows immediately from the definition of the conjugate:

$$g(X) + g^*(\mu) \geq \langle \mu, X \rangle.$$

There is equality if and only if

$$\mu \in \partial g(X) \iff X \in \partial g^*(\mu).$$

When g is lower semicontinuous in the relevant topology, we also have the biconjugate formula

$$g(X) = \sup_{\mu \in (L^\infty)^*} \langle \mu, X \rangle - g^*(\mu).$$

This is the convex form of the paper's conjugacy between U and V .

When g is coherent, the conjugate function is a convex indicator function. More precisely, there is a convex weak*-closed set $\mathcal{Q}_i \subset (L^\infty)^*$ of positive normalized dual functionals such that

$$g(X) = \sup_{\mu \in \mathcal{Q}} \langle \mu, X \rangle, \quad g^*(\mu) = I_{\mathcal{Q}}(\mu),$$

where $I_{\mathcal{Q}}$ is 0 on \mathcal{Q} and $+\infty$ off it. In the coherent case the dual penalty does not grade scenarios, it simply says which dual functionals are admissible. The paper exploits this structure in the comonotonic additive case, where the utility-side conjugate becomes an indicator of a set C_φ and intersections of those sets produce the optimal aggregate criterion.

The simple form of the penalty function for coherent measures leads to simple duality:

$$\mu \in \partial g(X) \iff \mu \in \mathcal{Q} \text{ and } g(X) = \langle \mu, X \rangle.$$

A subgradient is just an admissible dual measure that attains the supremum for X , that is a worsening-case pricing rule, or stress scenario, exactly supporting the risk cost of X . If μ is countably additive with density Z , then this reads

$$g(X) = \mathbb{P}(ZX).$$

If μ has a singular part, that singular piece captures the $\text{ess sup } X$ part coming from a mass of the spectral profile at 0. In the coherent case, where \mathcal{Q} replaces the penalty, subgradients are exactly the contact functionals.

The common-subgradient condition is very powerful. If

$$\mu \in \partial g_0(\xi_0) \cap \partial g_1(\xi_1),$$

then the same linear pricing rule supports (prices) both agents' allocated losses. That means neither agent sees a profitable local reshuffle of an infinitesimal layer of loss, because both are already tangent to the same supporting functional. That is the first-order condition behind efficient risk sharing. On the utility side, this is exactly the role played by the common supergradient in Theorem 3.1 and in the proof of Proposition 3.1.

The duality identity behind risk sharing is

$$(g_0 \square g_1)^* = g_0^* + g_1^*.$$

In the coherent case, where $g_i^* = I_{\mathcal{Q}_i}$,

$$g_0^* + g_1^* = I_{\mathcal{Q}_0} + I_{\mathcal{Q}_1} = I_{\mathcal{Q}_0 \cap \mathcal{Q}_1}.$$

This is the coherent counterpart of Lemma 2.1 and Theorem 3.1 in the paper. In particular, for an allocation $(\xi_0, \xi_1) \in \mathbb{A}(X)$, the following conditions are the same:

1. $g_0(\xi_0) + g_1(\xi_1) = (g_0 \square g_1)(X)$,
2. $\exists \mu \in (L^\infty)^* : \mu \in \partial g_0(\xi_0) \cap \partial g_1(\xi_1)$, and
3. $\exists \mu \in (L^\infty)^* : g_i(\xi_i) + g_i^*(\mu) = \langle \mu, \xi_i \rangle$, $i = 0, 1$.

The proof of Proposition 3.1 later boils down to understanding what this common μ looks like in the spectral case.

For the spectral subclass, we use the same symbol g_i for both the functional and its generating distortion profile; the argument tells you which object is meant. Thus $g_i(X)$ is the risk-cost of a loss X , while $g_i(t)$ is a concave nondecreasing function on $[0, 1]$ with

$$g_i(0) = 0, \quad g_i(1) = 1.$$

The representation is NEED gS VERSION and INT BY PARTS

$$g_i(X) = \int_0^1 q_X(u) d\check{g}_i(u).$$

If the profile is absolutely continuous on $(0, 1]$, this becomes

$$g_i(X) = \int_0^1 q_X(u) g_i'(u) du.$$

A jump at $u = 0$ contributes an $\text{ess sup } X$ term; this is why singular dual pieces appear and why the full dual $(L^\infty)^*$ matters. TVaR is the basic example.

So the entire framework is already in place:

1. choose two convex law-invariant risk-costs g_0, g_1 ;
2. solve the inf-convolution problem on $\mathbb{A}(X)$, or on $\mathbb{A}^\uparrow(X)$ once law-invariance is used;
3. characterize optimality by a common subgradient μ ;
4. in the spectral/comonotonic additive case, rewrite that subgradient condition as a statement about quantiles and distortion profiles.

That last step is where the condition flat on the region where one agent is more risk-averse appears.

Paper notation, for cross-reference only: their U_i is our $-g_i$, their V_i corresponds to our $g_i^*(-\mu)$, and their $\bar{\varphi}_i$ is our distortion profile $g_i(\cdot)$.

2 Efficient allocations and the common subgradient

Let

$$(g_0 \square g_1)(X) := \inf_{\xi_0 + \xi_1 = X} (g_0(\xi_0) + g_1(\xi_1)).$$

This is the total cost of splitting the aggregate loss X in the best possible way between the two agents. An allocation $(\xi_0, \xi_1) \in \mathbb{A}(X)$ is efficient exactly when it attains this infimum:

$$g_0(\xi_0) + g_1(\xi_1) = (g_0 \square g_1)(X).$$

This is the cost-side version of Theorem 3.1(ii).

The conjugate of the inf-convolution is the sum of the conjugates:

$$(g_0 \square g_1)^*(\mu) = g_0^*(\mu) + g_1^*(\mu).$$

This is the dual backbone of the whole argument. The primal problem asks to split X ; the dual problem asks to find one common dual variable μ that supports both agents at once. On the utility side this is Lemma 2.1 together with Theorem 3.1.

Fenchel's inequality gives, for every $\mu \in (L^\infty)^*$,

$$g_i(\xi_i) + g_i^*(\mu) \geq \langle \mu, \xi_i \rangle, \quad i = 0, 1.$$

Adding and using $\xi_0 + \xi_1 = X$,

$$g_0(\xi_0) + g_1(\xi_1) + g_0^*(\mu) + g_1^*(\mu) \geq \langle \mu, X \rangle.$$

Taking the supremum over μ gives

$$g_0(\xi_0) + g_1(\xi_1) \geq (g_0 \square g_1)(X).$$

This is just the obvious statement that every candidate allocation costs at least the optimum, but the real point is what happens when equality holds: equality forces equality in Fenchel agent by agent. That is the first-order condition.

So the key criterion is that an allocation $(\xi_0, \xi_1) \in \mathbb{A}(X)$ is efficient if and only if there exists some

$$\mu \in (L^\infty)^*$$

such that

$$\mu \in \partial g_0(\xi_0) \cap \partial g_1(\xi_1).$$

Equivalently,

$$g_i(\xi_i) + g_i^*(\mu) = \langle \mu, \xi_i \rangle, \quad i = 0, 1.$$

This is the risk-cost version of the paper's common supergradient condition for utilities and it is a common subgradient here.

That common μ is the first big trick of the proof. Instead of solving directly for the whole functions ξ_0, ξ_1 , one looks for a single shadow price μ that makes both agents individually happy to hold what they hold. Once such a μ exists, the allocation is efficient. If no such common μ exists, the allocation is not efficient.

There is also a nice ELI5 way to read this. Think of μ as a system of state prices, or marginal penalties, attached to the final loss. Agent i chooses ξ_i optimally when its own marginal cost matches that same price system. Efficiency means both agents agree on the same marginal valuation of the last tiny piece of loss. If they disagree somewhere, you can move a little loss from the one who values it more to the one who values it less and reduce total cost.

Cash-invariance is what removes Lagrange multipliers. In a generic two-agent optimization problem you often get weights λ_0, λ_1 . Here cash-invariance forces them to be equal, so there is just one common μ . This is one of the clean structural simplifications of the monetary-risk setting. On the utility side this appears in the proof of Theorem 3.1, where Hahn–Banach gives multipliers and cash-invariance collapses them to one common one.

In the coherent case the picture becomes even simpler. Then each g_i is a support function:

$$g_i(X) = \sup_{\mu \in \mathcal{Q}_i} \langle \mu, X \rangle$$

for some convex weak*-closed admissible set \mathcal{Q}_i , and

$$g_i^*(\mu) = I_{\mathcal{Q}_i}(\mu).$$

So the common-subgradient condition becomes

$$\mu \in \mathcal{Q}_0 \cap \mathcal{Q}_1, \quad g_i(\xi_i) = \langle \mu, \xi_i \rangle, \quad i = 0, 1.$$

In words: an efficient allocation admits a single admissible dual measure that simultaneously attains both agents' worst-case costs. On the utility side this is exactly the indicator-function structure in Remark 2.1 and Lemma 2.2.

At this abstract level, that is already the whole theorem:

1. solve the primal problem by inf-convolution;

2. pass to the dual sum $g_0^* + g_1^*$;
3. characterize optimality by a common μ ;
4. then specialize that common μ in the law-invariant/comonotonic additive case.

The next section is where the second big trick enters. Once the functionals are law-invariant, one can replace an arbitrary efficient allocation by one depending only on the aggregate loss X , namely

$$(P[\xi_0 | X], P[\xi_1 | X]),$$

without increasing cost. Then, using Landsberger plus Denneberg, one moves from measurable with respect to X to increasing in X , hence comonotone, and the whole problem becomes one-dimensional in quantile space. That is Theorem 3.1(vi) and Theorem 3.2.

Paper notation, only for cross-reference: their condition

$$\partial U_0(\xi_0) \cap \partial U_1(\xi_1) \neq \emptyset$$

becomes here

$$\partial g_0(\xi_0) \cap \partial g_1(\xi_1) \neq \emptyset.$$

The sign flip is the only change in substance.

Next I would do section 3 as the law-invariant reduction: from arbitrary efficient allocations to allocations of the form

$$\xi_i = f_i(X),$$

then to increasing/comonotone allocations, and then to quantiles.

3 Reductions: function of total loss and then increasing functions

The first big trick is dual: efficient allocation \iff common supporting subgradient.

The second big trick is probabilistic: once the costs are law-invariant, we can simplify the shape of the allocation without making anyone worse off. This is the step that turns an infinite-dimensional problem into a one-dimensional quantile problem. It comes in two moves.

We can move from an arbitrary allocation to one measurable with respect to X . Start with any allocation

$$\xi_0 + \xi_1 = X.$$

At first ξ_0 and ξ_1 may depend on extra randomness beyond X itself. Economically, that means the agents may be using some private randomization to decide who bears which part of the loss, even after the total loss X is known.

The paper shows that, for law-invariant criteria, this extra randomness is useless for Pareto efficiency. Replace the allocation by

$$\bar{\xi}_i := P[\xi_i | X], \quad i = 0, 1.$$

Then still

$$\bar{\xi}_0 + \bar{\xi}_1 = X,$$

so this is again an allocation of the same total loss. Moreover, if (ξ_0, ξ_1) is Pareto optimal, then

$$(\bar{\xi}_0, \bar{\xi}_1) = (\mathbb{P}[\xi_0 | X], \mathbb{P}[\xi_1 | X])$$

is also Pareto optimal. This is Theorem 3.1(vi) in the paper. Why is this plausible? Because once X is fixed, replacing ξ_i by its conditional mean given X removes idiosyncratic noise. Convexity says randomization does not help when you measure cost. Law-invariance says only the conditional distribution matters, not the labels of states. It is beneficial to average out the useless extra randomness.

An ELI5 version is: after the total loss X is known, there is no benefit in flipping extra coins to decide who pays what. A clean deterministic rule based only on the size of X does at least as well.

This is exactly the first reduction you want to keep in mind: an efficient allocation must be of the form $\xi_i = f_i(X)$.

This reduction is not just a Jensen-style heuristic. It also appears on the dual side. Suppose

$$\mu \in \partial g(\xi).$$

On the utility side, Lemma 4.1 says one may replace the supporting dual object by its conditional expectation on any σ -algebra containing $\sigma(\xi)$, without losing optimality. In our cost language, this means that if μ supports ξ , then

$$\mu_1 := \mathbb{P}[\mu | \mathcal{G}]$$

also supports ξ whenever ξ is \mathcal{G} -measurable, and the dual cost does not increase.

So once the allocation is replaced by functions of X , the supporting dual object may also be replaced by one depending only on X . That is a major simplification. It means both primal and dual objects can be collapsed onto the one-dimensional sigma-algebra generated by the total loss.

This is one of the hidden structural tricks in the proof: law-invariance plus conditional expectation allows you to compress the whole problem onto $\sigma(X)$.

Next, we move from functions of X to increasing functions of X . Suppose

$$\xi_i = f_i(X).$$

This still allows wild measurable functions f_i . The question is: can we arrange that f_i is nondecreasing? The answer is yes. Theorem 3.2 says that for law-invariant monetary criteria one may optimize over the smaller class

$$\Delta^\uparrow(X) := \{ (\xi_0, \xi_1) : \xi_0 + \xi_1 = X, \xi_0 \uparrow X, \xi_1 \uparrow X \},$$

and Pareto optimal allocations exist there. So the second reduction is from an efficient allocation to an efficient allocation with $\xi_i = f_i(X)$ nondecreasing. This is where Landsberger and Denneberg enter.

Very roughly, Landsberger's idea is a monotone-rearrangement improvement. If an agent's final loss is not increasing in the total loss X , then there are states where the aggregate loss is larger but that agent pays less, and other states where the aggregate loss is smaller but that agent pays more. By swapping these pieces into the right order, one creates an allocation that is more aligned with X .

For law-invariant monotone-convex preferences, that monotone rearrangement is never worse and can be strictly better. The paper cites Landsberger for this improvement principle in the proof architecture around Theorem 3.2. The intuition is easy:

- bad aggregate states should not assign surprisingly small losses to one agent while better aggregate states assign that same agent larger losses;
- if such crossings occur, one can untangle them;
- untangling makes the allocation more ordered, and law-invariant convex criteria prefer that ordered version.

So Landsberger supplies the economic monotonicity principle: optimal sharing can be taken to be monotone in the total loss.

Denneberg's lemma says that increasing in X is the same as comonotone with X . Once the allocation is increasing in X , it is automatically comonotone with X , and therefore the two pieces are comonotone with each other. This is the observation the paper attributes to Denneberg:

$$\mathbb{A}^\uparrow(X) = \text{comonotone allocations of } X.$$

That equivalence is stated explicitly just before Theorem 3.2. This matters because comonotonicity turns the random-variable problem into a quantile problem. If ξ_0, ξ_1, X are comonotone and

$$X = \xi_0 + \xi_1,$$

then their quantiles add pointwise:

$$q_X(u) = q_{\xi_0}(u) + q_{\xi_1}(u), \quad 0 < u < 1.$$

That is the real prize. Once you have this, the sharing problem becomes an optimization over nondecreasing functions on $(0, 1)$ instead of over arbitrary random variables on Ω . This is the bridge from abstract convex duality to the explicit layer-cake formulas later in Proposition 3.1.

After these two moves, the problem has changed completely. Originally:

$$\inf_{\xi_0 + \xi_1 = X} (g_0(\xi_0) + g_1(\xi_1))$$

ranges over all bounded random variables on Ω . After the reduction:

$$\inf_{q_0 + q_1 = q_X} \left(\int_0^1 q_0(u), dg_0(u) + \int_0^1 q_1(u), dg_1(u) \right),$$

where q_0, q_1 are nondecreasing quantile functions. So the entire infinite-dimensional risk-sharing problem becomes a one-dimensional allocation of quantile layers. Each u -layer of the aggregate loss must be assigned to one agent or the other, and the distortion profiles g_0, g_1 determine which assignment is cheaper. That is the conceptual heart of the paper.

Now combine both big tricks. First big trick:

$$\text{efficiency} \iff \exists \mu \in \partial g_0(\xi_0) \cap \partial g_1(\xi_1).$$

Second big trick: we may assume

$$\xi_i = f_i(X), \quad f_i \text{ nondecreasing,}$$

hence everything is comonotone and can be read in quantiles.

Then Lemma 4.1 lets you choose the supporting μ comonotone with the allocation as well. On the utility side they write this as a comonotonicity statement for the pair $(-\mu, \xi)$ after replacing μ by a suitably rearranged version with the same law and no worse dual value. In

your cost language the same point is: the supporting dual functional can be arranged to be aligned with the loss.

That alignment is what lets Lemma 4.2 read the subgradient condition as a simple geometric statement about the distortion profiles and flat pieces of the quantile function. That is the next step.

The second big trick is:

1. average out all randomness not already contained in X :

$$\xi_i \mapsto \mathbb{P}[\xi_i \mid X];$$

2. reorder the resulting sharing rule so that it is nondecreasing in X ;
3. use Denneberg to identify this with comonotonicity;
4. pass to quantiles, where sums become pointwise sums.

From this point on, think of the optimal contract not as an arbitrary random variable, but as a monotone rule that assigns each quantile layer of the aggregate loss to one agent or the other.

Paper notation note only: their Theorem 3.1(vi) is the conditional-expectation reduction, their Theorem 3.2 is the existence of Pareto optima in the increasing/comonotone class, and the sentence before Theorem 3.2 invokes Denneberg for the identification of increasing allocations with comonotone ones.

Next comes the final step: once we are in quantile space, the common-subgradient condition becomes the flatness rule,

$$q_{\xi_i} \text{ is flat where agent } i \text{ is not the cheaper bearer of marginal loss.}$$

That is the core of Proposition 3.1.

4 The flatness rule in quantile space

At this point, the whole problem is one-dimensional. We take an efficient allocation in the increasing/comonotone class

$$(\xi_0, \xi_1) \in \mathbb{A}^\uparrow(X), \quad q_X = q_{\xi_0} + q_{\xi_1}.$$

Each agent has a comonotonic additive spectral risk cost

$$g_i(Y) = \int_0^1 q_Y(u) d\check{g}_i(u),$$

where now $g_i : [0, 1] \rightarrow [0, 1]$ is the concave distortion profile of agent i . The aggregate inf-convolution has distortion

$$g := g_0 \wedge g_1,$$

the lower envelope. So the cheapest possible total cost is

$$(g_0 \square g_1)(X) = \int_0^1 q_X(u) d\check{g}(u).$$

This is the risk-side analog of Lemma 2.2 and Proposition 3.1 in the paper.

The result we want is that for $(\xi_0, \xi_1) \in \mathbb{A}^\uparrow(X)$, (ξ_0, ξ_1) is efficient if and only if, for each $i = 0, 1$, q_{ξ_i} is flat on $\{g_i > g_{1-i}\} \cap dq_X > 0$. Here flat means

$$dq_{\xi_i} = 0$$

a.e. on that set, together with the analogous jump-at-zero condition if there is an atom of the distortion at 0. In words: wherever agent i is more expensive than the other agent, agent i does not absorb any moving layer of loss. In words: each marginal quantile layer of the aggregate loss goes to the cheaper agent.

By the first big trick, efficiency is equivalent to existence of one common supporting dual functional

$$\mu \in \partial g_0(\xi_0) \cap \partial g_1(\xi_1).$$

Because we are already in the law-invariant/comonotonic additive setting, Lemma 4.1 lets us choose that supporting object aligned with the loss, so its law is summarized by a single concave profile, call it g_μ . On the utility side the paper writes this as φ_μ ; here I write g_μ .

What does it mean that μ supports ξ in quantile terms? This is exactly the content of Lemma 4.2. In our language, if

$$\mu \in \partial g_i(\xi_i),$$

then

$$g_\mu \leq g_i$$

pointwise, and q_{ξ_i} is flat on $\{g_\mu < g_i\}$. So a supporting dual profile sits below the agent's distortion profile, and strict gap means flatness of that agent's quantile. This is the key local statement. It turns the abstract subgradient condition into a geometric condition in u -space. If the contact profile g_μ lies strictly below agent i 's profile at some quantile level u , then agent i cannot be carrying a genuinely varying loss there. If agent i were carrying a moving layer there, the support would fail to touch.

We now apply the same idea to the aggregate problem. The same μ also supports the inf-convolution itself at X . Since the aggregate distortion is

$$g = g_0 \wedge g_1,$$

Lemma 4.2 applied to the aggregate criterion gives $g_\mu \leq g$ and q_X is flat on $\{g_\mu < g\}$. Therefore, on every region where $dq_X > 0$, we must have $g_\mu = g = g_0 \wedge g_1$. This is the decisive move. Wherever the aggregate loss is actually changing, the supporting profile coincides with the cheaper of the two agents' profiles.

We can now deduce the flatness rule for each agent. Take a point u where $g_0(u) > g_1(u)$ and $dq_X(u) > 0$. Then from Step 3, $g_\mu(u) = g(u) = g_1(u) < g_0(u)$. Now apply Step 2 to agent 0: since $g_\mu < g_0$, the quantile q_{ξ_0} must be flat there. Therefore, on $g_0 > g_1 \cap dq_X > 0$, agent 0's allocated loss does not move. Likewise, on $\{g_1 > g_0\} \cap \{dq_X > 0\}$, agent 1's allocated loss does not move. That is the risk-side version of

$$q_{\xi_i} \text{ flat on } \{\bar{\varphi}_i > \bar{\varphi}_{1-i}\} \cap \{dq_X > 0\}$$

in the paper.

The converse also holds. Now suppose we start with a comonotone allocation satisfying the flatness rule: q_{ξ_i} flat on $\{g_i > g_{1-i}\} \cap dq_X > 0$, $i = 0, 1$. Set $g := g_0 \wedge g_1$. Then on the region where $g_i > g$, the rule says exactly that q_{ξ_i} is flat wherever q_X moves. Since $q_X = q_{\xi_0} + q_{\xi_1}$, the remaining moving part must belong to the other agent, the cheaper one. This is enough to

make the same dual profile g support both agents: for each i , $g \leq g_i$ and the required flatness holds on $\{g < g_i\}$, so Lemma 4.2 gives a common supporting subgradient $\mu \in \partial g_0(\xi_0) \cap \partial g_1(\xi_1)$. Then Theorem 3.1 gives efficiency. That is the converse implication.

Hence the optimal contract is built by carving the aggregate loss into layers and assigning each layer to the cheaper agent. When the profiles cross only finitely many times, these layers become stop-loss pieces, so the contract is a sum of options on X . That is exactly why Example 3.1 produces finite sums of calls / stop-loss contracts.

4.1 One-line summary

The proof of Proposition 3.1 is:

1. efficient allocation \iff common supporting subgradient;
2. law-invariance lets us reduce to increasing/comonotone allocations;
3. comonotonicity turns the problem into quantiles;
4. Lemma 4.2 says there is a strict gap between support profile and agent profile \iff flat quantile;
5. for the aggregate problem the support profile is the lower envelope $g_0 \wedge g_1$ on the moving part of q_X ;
6. therefore each moving loss layer goes to the cheaper agent.

Paper notation note only: their $\bar{\varphi}_i$ is your distortion profile g_i , and their utility-side condition

$$q_{\xi_i} \text{ flat on } \bar{\varphi}_i > \bar{\varphi}_{1-i}$$

becomes your cost-side condition

$$q_{\xi_i} \text{ flat on } g_i > g_{1-i}.$$

SETUP AGAIN?!

Let $X \in L^\infty$ be an aggregate loss on an atomless probability space.

For $i = 0, 1$, let

$$g_i : [0, 1] \rightarrow [0, 1]$$

be concave, nondecreasing, with

$$g_i(0) = 0, \quad g_i(1) = 1,$$

and let the corresponding spectral risk cost, again denoted by g_i , be

$$g_i(Y) := \int_0^1 q_Y(u) d\check{g}_i(u), \quad Y \in L^\infty.$$

Allow g_i to have a jump at 0, so an ess sup term is included when present. This is the risk-cost version of the paper's comonotonic additive law-invariant monetary utility setup.

Define the class of increasing, equivalently comonotone, allocations of X by

$$\mathbb{A}^\uparrow(X) := \{(\xi_0, \xi_1) \in L^\infty \times L^\infty : \xi_0 + \xi_1 = X, \xi_0 \uparrow X, \xi_1 \uparrow X\}.$$

By Denneberg's lemma and Theorem 3.2 in the paper, this is the right class in which to look for Pareto-optimal allocations.

Set

$$g := g_0 \wedge g_1, \quad g(u) = \min g_0(u), g_1(u), \quad 0 \leq u \leq 1.$$

For a nondecreasing function f on $[0, 1]$, say that f is flat on a set $A \subseteq [0, 1]$ if

$$df = 0 \quad \text{a.e. on } A,$$

and, if needed, the jump-at-zero term is also zero there:

$$(f(0+) - f(0)) 1_A(0+) = 0.$$

This is the loss-side version of Definition 3.2 in the paper.

Then:

Theorem 4.1 (Risk sharing theorem). *For an allocation $(\xi_0, \xi_1) \in \mathbb{A}^\uparrow(X)$, the following are equivalent.*

1. (ξ_0, ξ_1) is Pareto optimal.
2. For each $i = 0, 1$, the quantile function q_{ξ_i} is flat on

$$\{g_i > g_{1-i}\} \cap \{dq_X > 0\}.$$

Equivalently, at every quantile level where the aggregate loss X is actually increasing, agent i carries no layer of loss on the region where it is more expensive than the other agent. In that case,

$$(g_0 \square g_1)(X) = \inf_{\xi_0 + \xi_1 = X} g_0(\xi_0) + g_1(\xi_1) = \int_0^1 q_X(u) d\check{g}(u).$$

The aggregate optimal risk cost is obtained by the lower envelope $g = g_0 \wedge g_1$.

5 Magic Ingredients

1. convex duality and Fenchel contact: conjugates, biconjugates, Fenchel inequality, and subgradients as contact functionals.
2. inf-convolution dualizes to sum:

$$(g_0 \square g_1)^* = g_0^* + g_1^*.$$

This is the algebraic backbone of the Pareto problem. On the utility side this is Lemma 2.1.

3. Pareto optimality equals common supporting subgradient: efficient allocation \iff one common dual object supports both agents. This is Theorem 3.1.
4. Hahn–Banach separation, plus cash-invariance: the separation argument gives Pareto \implies weighted optimum; cash-invariance collapses the two multipliers to one. This is an important hidden simplification.
5. coherent case: the dual penalty is an indicator:

$$g^* = I_Q,$$

so the dual problem becomes set intersection rather than a penalized optimization.

6. Kusuoka / quantile representation for law-invariant criteria: general law-invariant costs have a variational quantile representation, and comonotone ones collapse to a single distortion profile. This is Theorems 2.2 and 2.3.
7. lower-envelope rule: in the comonotonic additive case the aggregate criterion is generated by the lower envelope

$$g = g_0 \wedge g_1.$$

On the utility side this is Lemma 2.2.

8. conditional-expectation reduction onto $\sigma(X)$: any Pareto allocation can be replaced by

$$(\mathbb{P}[\xi_0 | X], \mathbb{P}[\xi_1 | X])$$

without loss. This is one of the big structural moves in Theorem 3.1(vi).

9. Landsberger monotone-improvement idea: once only the law matters, unordered sharing can be rearranged into monotone sharing. This is not stated as a formal theorem in this paper, but it is part of the background logic behind the passage to increasing allocations.
10. Denneberg's lemma: increasing in X is the same as comonotone with X , hence with each other. This is what turns the allocation problem into a quantile problem.
11. quantile additivity under comonotonicity: for comonotone allocations,

$$q_X = q_{\xi_0} + q_{\xi_1}.$$

This is what makes the layer-by-layer assignment possible. This sits behind Theorem 3.2 and Proposition 3.1.

12. singular-vs-regular dual decomposition: the dual lives in $(L^\infty)^*$, not just L^1 , because a jump of the distortion at 0 creates an $\text{ess sup } X$ term in your loss convention. The regular/singular split is therefore not bookkeeping; it is essential.
13. rearrangement/comonotone-coupling trick for the dual variable: Lemma 4.1 lets one replace a supporting dual object by one aligned with the loss, and also push it down by conditional expectation to $\sigma(X)$. This is one of the real magic tricks in the proof.
14. explicit description of the subgradient in distortion language: Lemma 4.2 says, in effect,

$$\mu \in \partial g(X)$$

iff the dual profile lies below the distortion profile and the quantile is flat where the inequality is strict. This is the bridge from abstract duality to the concrete contract shape.

15. flatness criterion: in the optimal allocation, agent i 's quantile is flat where $g_i > g_{1-i}$ on the moving part of q_X . This is Proposition 3.1 itself.
16. stop-loss / option synthesis: when the profiles cross in finitely many places, the flatness rule integrates to a finite sum of layers, hence a sum of stop-loss contracts or options. That is Example 3.1.

6 Appendix: Reconciliation to Paper

This section ties our notation to Jouini et al. (2008).

Table 1: Notation reconciliation to Jouini et al. (2008).

Paper notation	Post notation	Meaning
U_i	$-g_i$	Utility / risk measure
V_i	g_i^*	Convex dual
$\bar{\varphi}$	g distortion function	

Table 2: Summary of results in Jouini et al. (2008).

Result	Statement	Tricks
asd		

References

Jouini, E., Schachermayer, W., & Touzi, N. (2008). Optimal risk sharing for law invariant monetary utility functions. *Mathematical Finance*, 18(2), 269–292. <https://doi.org/10.1111/1/j.1467-9965.2007.00332.x>