Risk-averse multi-stage stochastic equilibria: models and algorithms

Michael C. Ferris

Computer Sciences Department and Wisconsin Institute for Discovery,
University of Wisconsin, Madison

(Joint work with Olivier Huber and Jiajie Shen)

Erice, Italy
May 20, 2022

The MOPEC problem (GNE)

Assume there are N agents, find $(x_1^*, \dots, x_N^*, \pi^*)$ such that for each agent:

$$\begin{aligned} x_a^* \in \arg\min \quad & f_a(x_a; x_{-a}^*, \pi^*) \\ \text{s.t.} \;\; & x_a \in X_a(x_{-a}^*, \pi^*) \end{aligned}$$

and a market equilibrium constraint:

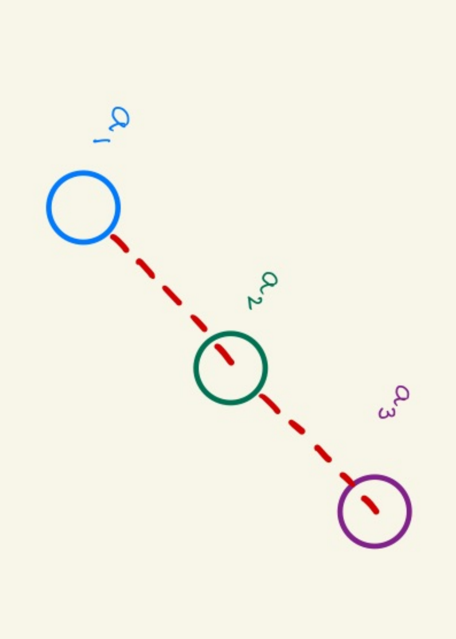
$$0 \in H(\pi^*; x^*) + N_P(\pi^*)$$

Variables:

- x_a: variable controlled by each agent a
- $x_{-a} = (x_1, x_2, \dots, x_{a-1}, x_{a+1}, \dots, x_N)$: action of other agents
- price variable π, set by the market equilibrium constraint

- Optimizations might be large LP or QP models of particular sectors
- Extensive literature, hard problems (non-monotone) even if f_a strongly convex

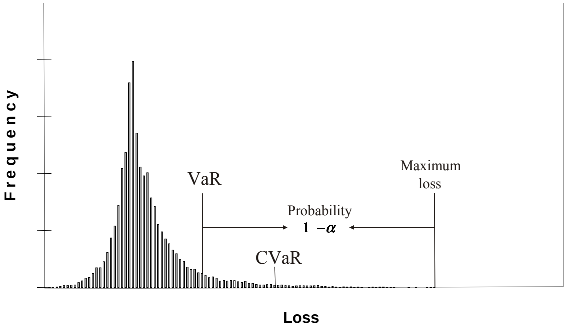
Structure



- Agent optimization problems at nodes
- Complementarity links across agents

Risk modeling

- Modern approach to modeling risk aversion uses concept of risk measures
- Considers not only the expected value of the uncertain quantities, but also more "extreme events"



- $\overline{CVaR}_\alpha$: mean of upper tail beyond α-quantile (e.g. $\alpha = 0.95$)
- Dual representation (of coherent r.m.) in terms of risk sets: $\mathcal{D}$ [4]

$$\rho(Z) = \sup_{\mu \in \mathcal{D}} \mathbb{E}_\mu[Z]$$

- Different agents have different risk profiles

One example: MOPEC equilibrium

Agents (e.g): 'fos', 'ren', 'trns', 'dem':

$$
\begin{array}{llll}
\mathrm{S}(a): & \min \ \rho_a(\psi_a) & \text{s.t.} & (z_a, y_a, q_a, r_a) \in \mathcal{X}_a \\
& \quad \psi_a(\omega) & = & \mathcal{C}_a(z_a) + \mathcal{Z}_a(y_a, q_a, r_a, \omega) \\
& & & +\pi_e(\omega)\left(d_a(\omega) - q_a(\omega) - r_a(\omega)\right) \\
& & & +\pi_c(\omega)\mathcal{E}(y_a, \omega)
\end{array}
$$

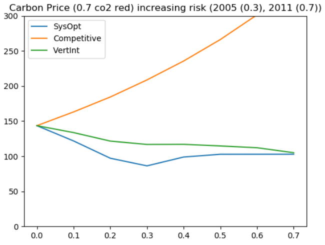
and the prices, production and purchases satisfy the market clearing conditions

$$
\begin{aligned}
0 &\le \sum_a \left(q_a(\omega) + r_a(\omega) - d_a(\omega)\right) \perp \pi_e(\omega) \ge 0, \\
0 &\le E - \sum_a \mathcal{E}(y_a, \omega) \perp \pi_c(\omega) \ge 0.
\end{aligned}
$$

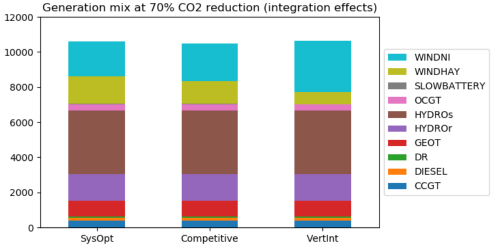
[2] provides theory to show when system optimization is equivalent

Increasing risk aversion: carbon price and investment

- $\rho(Z) = (1-\lambda)\mathbb{E}[Z] + \lambda \text{AVaR}_{0.90}(Z)$
- Same price risk neutral
- Competitive equilibrium: increased price
- VertInt: co-ownership of wind/thermal results in more wind closer to existing thermal



(a) Carbon prices with increasing λ



(b) Ownership at $\lambda = 0.3$

These problems are computationally challenging

Standard methods to solve the MOPEC problem

- Convert the MOPEC problem to mixed complementarity problem (EMP does this) and solve it using PATH solver
- Or traditional decomposition method: splitting, prox-gradient
- EMP/PATH fails to solve large-scale MOPEC problems
- Decompositions usually fail to solve the problem without helpful problem properties, and slow convergence

Solution method: Primal penalty and dual method

- Agent based decomposition (prox gradient)
- Penalty (Augmented Lagrangian) of the constraint $H(x, \pi) \geq 0$ in the primal agents' problems and updating dual in the major iterations.
- Able to solve the problem in situation without having an implicit function $\pi = h(x)$ from the constraint $0 \leq H(x, \pi) \perp \pi \geq 0$.
- Performance mainly depends on the choice of γ. γ small enough enables algorithm to converge to the true solution, but too small γ may cause slow convergence.

Algorithm 1 Gauss-Seidel Primal penalty and dual method

set $k = 0$, define initial point π^0.
while stopping criterion not met **do**
  for $a = 1, 2, \ldots, N$ **do**
    get solution (x_a^{k+1}, y_a^{k+1}) from solving

$$\begin{aligned} \min \quad & f_a(x_a, \bar{x}_{-a}^{k+1}, \pi^k) + y_a^T \pi^k + 0.5\gamma \cdot (y_a)^2 \\ \text{s.t.} \quad & x_a \in X_a(\bar{x}_{-a}^{k+1}, \pi^k) \\ & y_a \geq -H(x_a, \bar{x}_{-a}^{k+1}, \pi^k) \\ & y_a \geq 0 \end{aligned}$$

    here $\bar{x}_{-a}^{k+1} = (x_1^{k+1}, \ldots, x_{a-1}^{k+1}, x_{a+1}^k, \ldots, x_N^k)$.
  end for
  $\pi^{k+1} = \max\{0, \pi^k - \gamma \cdot H(x^{k+1}, \pi^k)\}$
  $k = k + 1$
end while

Comparison between PATH and Primal-Dual method

risk neutral

size	PATH	Primal-Dual		
	time(secs)	γ	# Iter	time(secs)
62K $\times$ 22K	1795.79	0.005	75	333.21

risk averse

size	residual	PATH	Primal-Dual		
		time(secs)	γ	# Iter	time(secs)
114 $\times$ 62	1e-6	-	0.05	264	35.87
114 $\times$ 62	1e-6	-	0.1	162	20.97
114 $\times$ 62	1e-6	-	0.5	334	45.45
21K $\times$ 8.5K	< 1	-	0.005	32	165.76

- The stopping criterion is **1e-6**
- In risk-averse setting, PATH fails to find a solution without good initial point even in small cases

Risk Measures

Problem type

Objective function or Constraint

$$\min_{x \in X} \theta(x) + \rho(F(x)) \qquad\qquad \min_{x \in X} \theta(x) \text{ s.t. } \rho(F(x)) \leq \alpha$$

- If $\mathcal{D} = \{p\}$ then $\rho(Z) = \mathbb{E}[Z]$
- If $\mathcal{D}_{\alpha,p} = \{\lambda \in [0, p/(1-\alpha)] : \langle \mathbf{1}, \lambda \rangle = 1\}$, then $\rho(Z) = \overline{CVaR}_\alpha(Z)$
- Popular examples include: expectation, Conditional Value at Risk, also known as expected shortfall, Average Value at Risk (AVaR), and expected tail loss (ETL), and mean-upper-absolute semideviation.

Using the algebra of support function, we can create new risk measures from existing ones: for instance

$$(1-\lambda)\mathbb{E} + \lambda \overline{CVaR}_\alpha$$

captures more realistic risk-averse behavior. For $\lambda < 1$, it is strictly monotone (desirable for time-consistency)

The transformation to complementarity

$$\min_{x \in X} \theta(x) + \rho(F(x))$$

$$\rho(y) = \sup_{u \in U} \left\{ \langle u, y \rangle - \frac{1}{2} \langle u, Mu \rangle \right\}$$

conjugate composite function:

$$0 \in \partial\theta(x) + \nabla F(x)^T \partial\rho(F(x)) + N_X(x)$$

calculus:

$$0 \in \partial\theta(x) + \nabla F(x)^T u + N_X(x)$$

$$0 \in -u + \partial\rho(F(x)) \iff 0 \in -F(x) + Mu + N_U(u)$$

- This is a complementarity problem (solvable by PATH)
- Equilibrium formulation
- (Fenchel) duality formulation
- Extreme point formulation

Conjugate composite function (CCF)

$$\rho(y) := \sup_{u \in U} \langle G(y), u \rangle - k(u) \tag{1}$$

$G(y) := By + b$, k is convex, U polyhedral [1]

Conjugate function $\quad G \equiv Id$

ρ is the conjugate function of $\delta_U + k$

Support function $\quad G \equiv Id,\ k \equiv 0$

ρ is the support function of U.

Conversion of constraint to objective

Can extend the conjugacy result to a nested version. Suppose that each component of F has the form $F_i = f_i + \hat{\rho}_i \circ \hat{F}_i$ and consider the CCF composition $\rho \circ F$.
Then, for any $\bar{x} \in \mathrm{dom}(\rho \circ F)$ we have

$$\partial(\rho \circ F)(\bar{x}) = \{\partial\langle v, F\rangle(\bar{x}) \mid v \in \partial\rho(F(\bar{x}))\}.$$

and

$$\partial\langle v, F\rangle(\bar{x}) = \{\langle v, \nabla f\rangle(\bar{x}) + \langle v, w\rangle \quad \text{where } v \in \partial\rho(F(\bar{x})) \\ \text{and } w_i \in \{\partial\langle \hat{v}_i, \hat{F}_i\rangle(\bar{x}) \mid \hat{v}_i \in \partial\hat{\rho}_i(\hat{F}_i(\bar{x}))\} \quad \text{for } i \in \{1, \ldots, q\}\},$$

where f collects all f_i.
So

$$\min_{x \in X} \theta(x) + \delta_{\mathbb{R}_-}(\rho(F(x)) - \alpha) = \min_{x \in X} \theta(x) + \sigma_{\mathbb{R}_+}(\rho(F(x)) - \alpha)$$

and we can apply the nested conjugacy result.

When is $\rho \circ F$ convex?

Uses the concept of K-convexity.

Lemma

Let $F : \mathbb{R}^p \to \mathbb{R}^l_\bullet$ with $F_i : \mathbb{R}^p \to \overline{\mathbb{R}}$ lsc, proper, convex for all $i \in \{1, \ldots, l\}$. Then, for any coherent risk measure ρ, the composition $\rho \circ F$ is lsc, proper, convex and $\mathrm{dom}(F) \subseteq \mathrm{dom}(\rho \circ F)$.

Reformulation via duality

Dualization [3]

$$\max_u \langle u, G(F(x))\rangle - \langle u, Mu\rangle$$
$$Au - b \in K_c$$
$$u \in K_u$$

$$\min_{z,w} \langle b, z\rangle + \tfrac{1}{2}\langle w, Jw\rangle$$
$$G(F(x)) - A^T z - Dw \in K_u^\circ$$
$$z \in K_c^\circ \quad w \text{ free}$$

K_u and K_c convex cones with polar K_u° and K_c°

Improvement to dual QP reformulation

- "The larger K_u, the smaller K_u° is"
- If u is free, then K_u is the whole space and $K_u^\circ = \{0\}$
- Try to use simple bounds to reduce K_u
- Look for $\tilde{u}$ such that $u - \tilde{u} \in \mathbb{R}^n_+$
- $G(F(x)) - A^T z - Dw - M^T \tilde{u} \in \mathbb{R}^n_-$: F convex $\Rightarrow$ convex constraints

Reformulation via conjugacy

ρ as a conjugate function

- ρ is the (Fenchel) conjugate of $k + \delta_U$:

$$\rho(u) = \inf_{u=u_1+u_2} k^*(u_1) + \sigma_U(u_2)$$

- $k(u) = u^T M u = \|L^T u\|_2$ (M psd)

$$\rho(F(x)) = \inf_s \ \tfrac{1}{2}\|s\|_2^2 + \sigma_U(G(F(x)) - Ls) \tag{2}$$

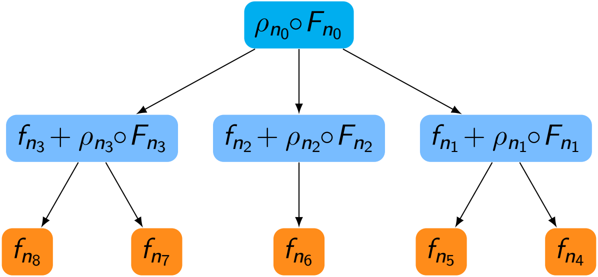
⊕ Problem (2) may be convex if all F_i are convex ($U \subset \mathbb{R}_m^+$)

⊕ Equivalent minimization problem (can use broad range of solvers)

⊖ Need closed-loop expression for σ_U

- Replace σ_U by t and compute vertices V of U and add constraints $\langle v, G(F(x)) - Ls \rangle \leq t \quad \forall v \in V$
- If U is a convex cone, replace σ_U by δ_{U°

Scenario tree with nodes $\mathcal{N} = \{0, 1, \ldots, 8\}$, and $T = 3$



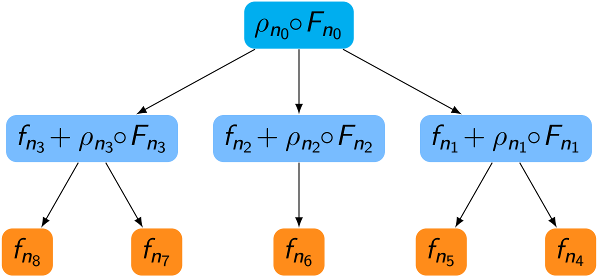
At leaf nodes:

$$\min_{x_{a\ell}} f_{a\ell}(x_{a\ell}; x_{-a\ell}, x_{\cdot\ell\_}, \pi_\ell) \quad \forall a \in \mathcal{A},$$

$$0 \in H_\ell(\pi_\ell; x_{\cdot\ell}) + N_{P_\ell}(\pi_\ell)$$

";" separates variables from parameters in function definition

Stochastic equilibrium (nested definition)

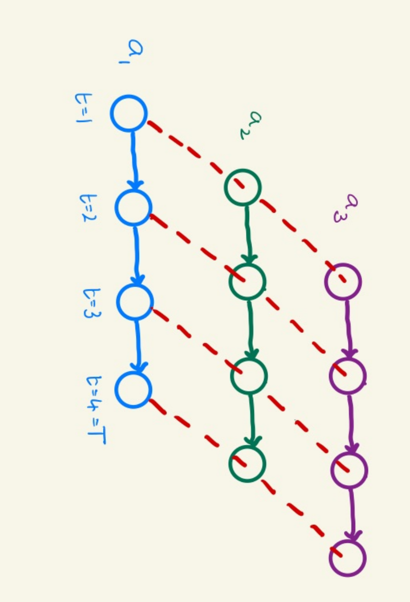


Recursing back to the root node:

$$\begin{aligned}
\min_{x_{a\mathcal{S}(n_0)}} \; & f_{an_0}(x_{an_0}; x_{-an_0}, x_{\cdot n_{0-}}, \pi_{n_0}) \\
& + \mathcal{R}_{an_0}([f_{aj}(x_{aj}; x_{-aj}, x_{\cdot n_0}, \pi_j) \\
& \quad + \mathcal{R}_{aj}([f_{a\ell}(x_{a\ell}; x_{-a\ell}, x_{\cdot \ell_-}, \pi_\ell)]_{\ell \in j_+})]_{j \in n_{0+}}) \qquad \forall a \in \mathcal{A}, \\
0 \in & H_j(\pi_j; x_{\cdot j}) + N_{P_j}(\pi_j), \qquad \forall j \in \mathcal{S}(n_0).
\end{aligned}$$

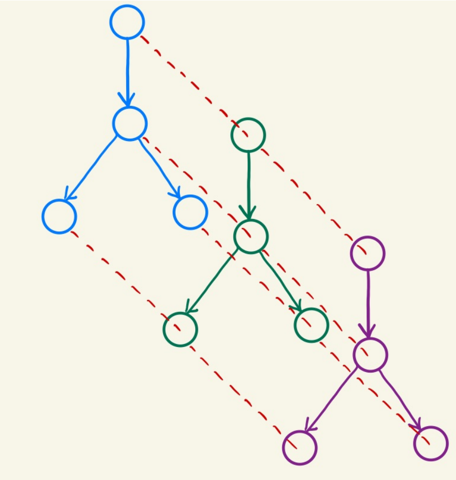
$\mathcal{S}(n)$ is the set of successor nodes of n, including n

Simple dynamics (discrete time, finite horizon)



- Complementarity links nodes across agents
- Dynamics link over time

Scenario trees linked across agents



- Complementarity links nodes of scenario tree across agents
- Dynamics link over time

Example: risk-averse stochastic equilibria

- market equilibrium: price defined by equilibrium constraints
- producers have a random upper bound on their production capacities and their ability to store goods from one stage to the other induces a coupling across stages
- objective function: revenue minus cost of production
- A, the scenario tree has 3 stages with 13 nodes, and there are 5 players in the market with 2 goods.
- B, the scenario tree has 4 stages with 30 nodes, and we have 2 players with 1 good.
- C has 5 stages, 121 nodes, 2 players and 1 good.

	Equilibrium			Duality			Conjugate		
	T (s)	vars	nnz	T (s)	vars	nnz	T (s)	vars	nnz
A	1.6	584	2775	5.2	644	2990	3.8	584	3530
B	9.0	455	2382	3.0	533	2774	Fail	455	2498
C	2.2	1400	8700	Fail	1640	10280	Fail	1400	7736

Different reformulations via option file

Multistage deterministic equivalent

$P(y)$ $\quad \min_{x^t_{an} \in X_{at}} \quad f_{a1}(x^1_{a1}, x^1_{-a1}, \pi^1_1) + \sum_{n \in 1+} y^2_{an} \cdot \left[f_{a2}(x^2_{an}, x^2_{-an}, \pi^2_n, \xi^2_n) + \sum_{m \in n+} y^3_{am} [\ldots] \right]$

$$\text{s.t} \quad h_{a1}(x_{a0}, x^1_{a1}) = 0, \quad g_{a1}(x^1_{a1}, x^1_{-a1}, \pi^1_1) \leq 0,$$

$$h_{at}(x^{t-1}_{an-}, x^t_{an}, \xi^t_n) = 0, \quad g_{at}(x^t_{an}, x^t_{-an}, \pi^t_n, \xi^t_n) \leq 0, \quad \forall t = 2, \ldots, T, \quad \forall n \in \mathcal{N}(t)$$

with the VI constraints

$$0 \leq H_1(\boldsymbol{x^1_1}, \pi^1_1) \quad \perp \pi^1_1 \geq 0$$

$$0 \leq H_t(\boldsymbol{x^t_n}, \pi^t_n, \xi^t_n) \quad \perp \pi^t_n \geq 0, \quad \forall t = 2, \ldots, T, \quad \forall n \in \mathcal{N}(t)$$

For any $t = 1, \ldots, T-2, n \in \mathcal{N}(t)$ the dual maximization problem

$D^t_{an}(\boldsymbol{x}, \boldsymbol{\pi}, \boldsymbol{y}_{n++})$: $\quad \max_{\{y^{t+1}_{am}\}_{m \in n+}} \quad \sum_{m \in n+} y^{t+1}_{am} \cdot \left[f_{at+1}(x^{t+1}_{am}, x^{t+1}_{-am}, \pi^{t+1}_m, \xi^{t+1}_m) + \sum_{r \in m+} y^{t+2}_{ar} [\ldots] \right]$

$$\text{s.t} \quad y^{t+1}_a \in \mathcal{D}^{t+1}_a$$

For any $t = T-1, n \in \mathcal{N}(t)$ the dual maximization problem

$D^t_{an}(\boldsymbol{x}, \boldsymbol{\pi})$: $\quad \max_{\{y^{t+1}_{am}\}_{m \in n+}} \quad \sum_{m \in n+} y^{t+1}_{am} \cdot \left[f_{at+1}(x^{t+1}_{am}, x^{t+1}_{-am}, \pi^{t+1}_m, \xi^{t+1}_m) \right]$

$$\text{s.t} \quad y^{t+1}_a \in \mathcal{D}^{t+1}_a$$

Forward backward algorithm

Define $y \in SOL(D(x, \pi)) \iff$

$$\{y_{am}^{t+1}\}_{m \in n+} \in D_{an}^{t}(\boldsymbol{x^k}, \boldsymbol{\pi^k}), \forall t = T-1, n \in \mathcal{N}(t)$$
$$\{y_{am}^{t+1}\}_{m \in n+} \in D_{an}^{t}(\boldsymbol{x^k}, \boldsymbol{\pi^k}, \boldsymbol{y_{n++}}), \forall t = 1, \ldots, T-2, n \in \mathcal{N}(t)$$

Finding a solution of the stochastic MOPEC with risk-averse agents is equivalent to find the solution (x^*, π^*, y^*) of the system

$$\begin{aligned}(x^*, \pi^*) &\in SOL(P(y^*)) \\ y^* &\in SOL(D(x^*, \pi^*))\end{aligned}$$

Detail of Forward backward algorithm

Algorithm 2 Forward-backward algorithm

set $k = 1$, set starting y^0 equal to the probability of risk-neutral case.
while stopping criterion not met **do**
  Solve the MOPEC with fixed risk probabilities $P(y^{k-1})$ to get $(x^k, \pi^k) \in SOL(P(y^{k-1}))$
  for $t = T-1, \ldots, 1$ **do**
    for $n \in \mathcal{N}(t)$ **do**
      if $t = T-1$ **then**
        $\{y_{am}^{k,t+1}\}_{m \in n+} \in D_{an}^t(\boldsymbol{x^k}, \boldsymbol{\pi^k})$
      else
        $\{y_{am}^{k,t+1}\}_{m \in n+} \in D_{an}^t(\boldsymbol{x^k}, \boldsymbol{\pi^k}, \boldsymbol{y_{n++}^k})$
      end if
    end for
  end for
  $k = k + 1$
end while

Numerical experiments

Test problem cases:

- MOPEC properites:
  - Type I: $f_a(x_a, x_{-a}, \pi) = \frac{1}{2}\epsilon||x_a||^2 + c^T x_a - \pi^T x_a + d$, $H(\boldsymbol{x}, \pi) = A\boldsymbol{x} - b$
  - Type II: $f_a(x_a, x_{-a}, \pi) = \frac{1}{2}\epsilon||x_a||^2 + c^T x_a - \pi^T x_a + d$, $H(\boldsymbol{x}, \pi) = A\boldsymbol{x} + B\pi - b$
  - Type III: $f_a(x_a, x_{-a}, \pi) = \frac{1}{2}\epsilon||x_a||^2 + c^T x_a - (B^{-1}(b - A\boldsymbol{x}))^T x_a + d$, no VI constraint and market price variable π
- Coherent risk measure:
  - $\rho(v) = (1-\lambda)\mathbb{E}[v] + \lambda CVaR_{1-\alpha}(v)$, where $CVaR_{1-\alpha}(\cdot)$ is the upper tail risk measure.
- Initial point strategy for PATH solver:
  - Strategy 1: Initial point (x, π, y) is uniformly randomly picked in the feasible region
  - Strategy 2: (x, π) of the initial point is the solution of risk-neutral problem and y is generated so initial basis matrix of PATH is nonsingular.
  - Strategy 3: Run several sweep forward-backward algorithms and use the point achieved as the initial point

Numerical results: performance of different strategies in choosing initial point

MOPEC Type	Ini Stra	total #	success #	success ratio
I	1	1000	375	37.5%
I	2	1000	555	55.5%
I	3(2)	1000	865	86.5%
II	1	1000	539	53.9%
II	2	1000	711	71.1%
II	3(2)	1000	870	87%
III	1	1000	813	81.3%
III	2	1000	892	89.2%
III	3(2)	1000	921	92.1%

- test problem size:
  - agent #: 2
  - scenario tree node size: 39
  - time stage size: 4
  - Corresponding MCP size: 455

Numeral results: changing ϵ and λ with fixed $\alpha = 0.75$

$\lvert\mathcal{N}\rvert$	T	MOPEC type	ϵ	λ	Ini Stra	total #	succ #	succ_r	FB_s #	FB_s_r
39	4	I	0	0.1	3(5)	16	16	100.00%	0	0.00%
39	4	I	0	0.3	3(5)	16	16	100.00%	0	0.00%
39	4	I	0	0.5	3(5)	16	8	50.00%	0	0.00%
39	4	I	0	0.7	3(5)	16	2	12.50%	0	0.00%
39	4	I	0	0.9	3(5)	16	0	0.00%	0	0.00%
39	4	I	1e-2	0.1	3(5)	16	16	100.00%	7	43.75%
39	4	I	1e-2	0.3	3(5)	16	16	100.00%	1	6.25%
39	4	I	1e-2	0.5	3(5)	16	16	100.00%	0	0.00%
39	4	I	1e-2	0.7	3(5)	16	8	50.00%	0	0.00%
39	4	I	1e-2	0.9	3(5)	16	4	25.00%	0	0.00%
39	4	I	1e-1	0.1	3(5)	16	16	100.00%	12	75.00%
39	4	I	1e-1	0.3	3(5)	16	16	100.00%	11	68.75%
39	4	I	1e-1	0.5	3(5)	16	16	100.00%	7	43.75%
39	4	I	1e-1	0.7	3(5)	16	16	100.00%	5	31.25%
39	4	I	1e-1	0.9	3(5)	16	16	100.00%	7	43.75%
39	4	I	1	0.1	3(5)	16	16	100.00%	16	100.00%
39	4	I	1	0.3	3(5)	16	16	100.00%	16	100.00%
39	4	I	1	0.5	3(5)	16	16	100.00%	16	100.00%
39	4	I	1	0.7	3(5)	16	16	100.00%	15	93.75%
39	4	I	1	0.9	3(5)	16	16	100.00%	16	100.00%
39	4	I	10	0.1	3(5)	16	16	100.00%	16	100.00%
39	4	I	10	0.3	3(5)	16	16	100.00%	16	100.00%
39	4	I	10	0.5	3(5)	16	16	100.00%	15	93.75%
39	4	I	10	0.7	3(5)	16	16	100.00%	16	100.00%
39	4	I	10	0.9	3(5)	16	16	100.00%	15	93.75%

Conclusions

- Markets naturally modeled via complementarity
- Solvers exist for medium to large scale problems
- Frameworks (EMP) exist to streamline model transformations
- empinfo: dualvar, bilevel, equilibrium, vi, CCF
- Very large scale models (many agents with many instruments acting strategically) with risk are hard
- Decomposition/diagonalization methods are effective when sensitivity information is exploited
- New algorithms enable solution of more detailed, authentic problems and address underlying policy questions

A. Aravkin, J. V. Burke, and G. Pillonetto.
Sparse / robust estimation and kalman smoothing with nonsmooth log-concave densities: Modeling, computation, and theory.
Journal of Machine Learning Research, 14(1):2689–2728, 2013.

M. Ferris and A. Philpott.
Dynamic risked equilibrium.
Operations Research, sep 2021.

R. T. Rockafellar and R. J. B. Wets.
Variational Analysis.
Springer-Verlag, 1998.

A. Ruszczyński and A. Shapiro.
Optimization of convex risk functions.
Mathematics of Operations Research, 31(3):433–452, Aug. 2006.