Interior Point Crashstarts for Stochastic Programming

Andreas Grothey
(joint work with Marco Colombo and Jacek Gondzio)

School of Mathematics, University of Edinburgh

Lancaster, 12 April 2011

Decomposition Based IPM-Warmstarts: Overview

- (Anotated) whistle-stop tour of
  - Stochastic Programming
  - Interior Point Methods
  - Warmstarting Interior Point Methods
- Reduced tree based SP warmstarts
- Multilevel reduced tree warmstart
- Decomposition based SP warmstarts
- Numerical Results

Stochastic Programming:

A Stochastic Programming Problem

$$\begin{array}{rl} \min & c^T x + \boldsymbol{E}_\xi[Q(x,\xi)] \\ \text{s.t.} & Ax = b, \\ & x \geq 0 \end{array} \qquad \text{(SP)}$$

where $Q(x,\xi) = \min\{q(\xi)^T y(\xi) : T(\xi)x + W(\xi)y(\xi) = h(\xi)\}$

This models a decision process: $x \longrightarrow \xi \longrightarrow y(\xi)$

Stochastic Programming:

A Stochastic Programming Problem

$$\begin{array}{ll} \min & c^T x + \boldsymbol{E}_\xi[Q(x,\xi)] \\ \text{s.t.} & Ax = b, \\ & x \geq 0 \end{array} \qquad \text{(SP)}$$

where $Q(x,\xi) = \min\{q(\xi)^T y(\xi) : T(\xi)x + W(\xi)y(\xi) = h(\xi)\}$

This models a decision process: $x \rightarrow \xi \rightarrow y(\xi)$

For a discrete distribution $\xi = \{\xi_i : P(\xi = \xi_i) = p_i\}_i$ we get (with $\mathcal{T} = \{(p_i, \xi_i)\}_i$):

Deterministic equivalent:

$$\begin{array}{ll} \min & c_0^T x + \sum_i p_i c_i^T y_i \\ \text{s.t.} & Ax = b, \\ & T_i x + W_i y_i = h_i \quad \forall i \\ & x \geq 0, y_i \geq 0 \end{array} \qquad (P(\mathcal{T}))$$

Structure of Deterministic equivalent

The constraint matrix of the deterministic equivalent

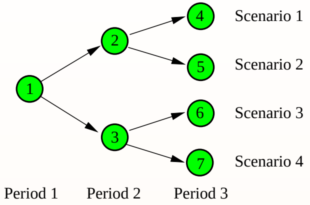
$$
\begin{array}{ll}
\min & c_0^T x + \sum_i p_i c_i^T y_i \\
\text{s.t.} & \begin{bmatrix} W_1 & & & & T_1 \\ & W_2 & & & T_2 \\ & & \ddots & & \vdots \\ & & & W_n & T_n \\ & & & & A \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \\ x \end{pmatrix} = \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \\ b \end{pmatrix}
\end{array}
$$

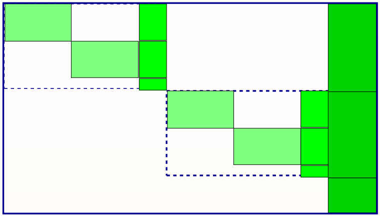
is a **column bordered block-diagonal** matrix.

Structure can be exploited

Multistage Stochastic Programming

(models decision process $x_1 \to \xi_2 \to x_2(\xi_2) \to \ldots \to \xi_T \to x_T(\xi_T)$)





Scenario Tree

Constraint Matrix

$\Rightarrow$ **nested** column bordered block-diagonal constraint matrix
Symmetrical event tree with K realizations/node and T periods corresponds to

$$K^{T-1} \text{ scenarios} \qquad \frac{K^T - 1}{K - 1} \text{ nodes (blocks)}$$

Applications

(Multistage) Stochastic Programming has many applications

- Portfolio Optimization
- Network Design with Uncertain Demand
- Electricity Generation Planning (involving hydro or wind)
- etc

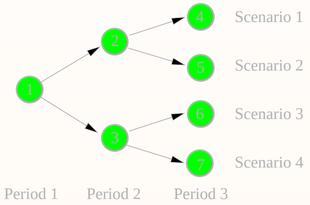
Solving Stochastic Programming Problems

Size of deterministic equivalent quickly becomes very large

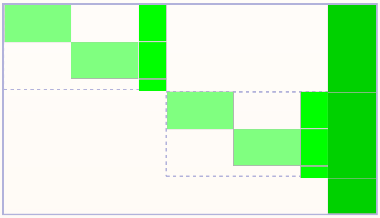
⇒Difficult for standard solvers. Suitable approaches are:

- Decomposition (Benders, L-shaped method)
- Interior Point Method
  (*Gondzio, G. (2005): solved multistage SP problem with* 10^9 *variables on 1280 processors in under 2h*)

- **Idea:** speed up solution process by crash-starting from a smaller tree
- **But:** IPMs are notoriously bad at exploiting a known starting point



Scenario Tree



Constraint Matrix

Interior Point Methods (for LP)

Linear Program

$$\min c^\top x \qquad \text{s.t.} \quad \begin{aligned} Ax &= b \\ x &\geq 0 \end{aligned} \tag{LP}$$

KKT Conditions

$$\begin{aligned} c - A^\top \lambda - s &= 0 \\ Ax &= b \\ XSe &= 0 \\ x, s &\geq 0 \end{aligned} \tag{KKT}$$

$X = \mathrm{diag}(x), S = \mathrm{diag}(s)$

Interior Point Methods (for LP)

Barrier Problem

$$\min c^\top x - \mu \sum \ln x_i \quad \text{s.t.} \quad \begin{array}{rcl} Ax & = & b \\ x & \geq & 0 \end{array} \qquad (\mathrm{LP}_\mu)$$

KKT Conditions

$$\begin{array}{rcl} c - A^\top \lambda - s & = & 0 \\ Ax & = & b \\ XSe & = & \mu e \\ x, s & \geq & 0 \end{array} \qquad (\mathrm{KKT}_\mu)$$

- Introduce logarithmic barriers for $x \geq 0$

Interior Point Methods (for LP)

Barrier Problem

$$\min c^\top x - \mu \sum \ln x_i \quad \text{s.t.} \quad \begin{aligned} Ax &= b \\ x &\geq 0 \end{aligned} \qquad (\mathrm{LP}_\mu)$$

KKT Conditions

$$\begin{aligned} c - A^\top \lambda - s &= 0 \\ Ax &= b \\ XSe &= \mu e \\ x, s &\geq 0 \end{aligned} \qquad (\mathrm{KKT}_\mu)$$

- Introduce logarithmic barriers for $x \geq 0$
- (LP_μ) ist striclty convex
- System (KKT_μ) can be solved per Newton-Method

Interior Point Methods (for LP)

Barrier Problem

$$\min c^\top x - \mu \sum \ln x_i \quad \text{s.t.} \quad \begin{array}{rcl} Ax & = & b \\ x & \geq & 0 \end{array} \qquad (\text{LP}_\mu)$$

KKT Conditions

$$\begin{array}{rcl} c - A^\top \lambda - s & = & 0 \\ Ax & = & b \\ XSe & = & \mu e \\ x, s & \geq & 0 \end{array} \qquad (\text{KKT}_\mu)$$

- Introduce logarithmic barriers for $x \geq 0$
- (LP_μ) ist striclty convex
- System (KKT_μ) can be solved per Newton-Method
- For $\mu \to 0$ solution of (LP_μ) converges to solution of (LP)

Interior Point Methods (for LP)

KKT conditions

$$\begin{aligned} c - A^\top \lambda - s &= 0 \\ Ax &= b \\ XSe &= \mu e \\ x, s &\geq 0 \end{aligned} \qquad (\mathrm{KKT}_\mu)$$

Interior Point Methods (for LP)

KKT conditions

$$\begin{array}{rcl} c - A^\top \lambda - s & = & 0 \\ Ax & = & b \\ XSe & = & \mu e \\ x, s & \geq & 0 \end{array} \qquad (\mathrm{KKT}_\mu)$$

Central Path

The set of all solutions to (KKT_μ) for all $\mu > 0$.
Central Path joins the **analytical center** (for $\mu{=}\infty$) with the LP solution (for $\mu = 0$).

Neighbourhoods (of the central path)

$$\begin{array}{rcl} \mathcal{N}_2(\theta) & := & \{(x, \lambda, s) \in \mathcal{F}^0 : \|XSe - \mu e\|_2 \leq \theta\mu\} \\ \mathcal{N}_{-\infty}(\gamma) & := & \{(x, \lambda, s) \in \mathcal{F}^0 : x_i s_i \geq \gamma\mu\} \end{array}$$

where $\mathcal{F}^0 := \{(x, \lambda, s) : c - A^\top\lambda - s = 0, Ax = b, x > 0, s > 0\}$.

Warmstarting Interior Point Methods

Aim: Use information from solution process of

$$\min c^\top x \quad \text{s.t.} \quad \begin{array}{rcl} Ax & = & b \\ x & \geq & 0 \end{array} \qquad \text{(LP)}$$

to construct a starting point for (nearby problem)

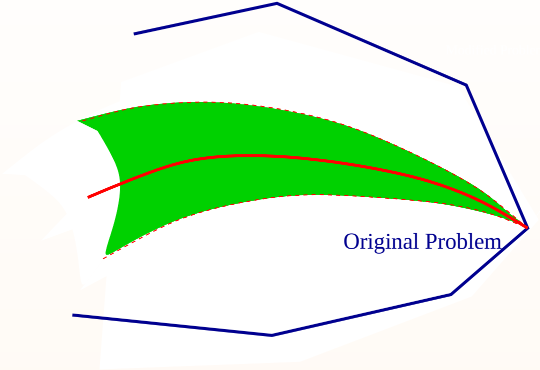
$$\min \tilde{c}^\top x \quad \text{s.t.} \quad \begin{array}{rcl} \tilde{A}x & = & \tilde{b} \\ x & \geq & 0 \end{array} \qquad (\widetilde{\mathrm{LP}})$$

where $\tilde{A} \approx A, \tilde{b} \approx b, \tilde{c} \approx c$

- It is **not** a good idea to use the solution of (LP) to start $(\widetilde{\mathrm{LP}})$.
- *Unlike for the Simplex/Active Set Method!*

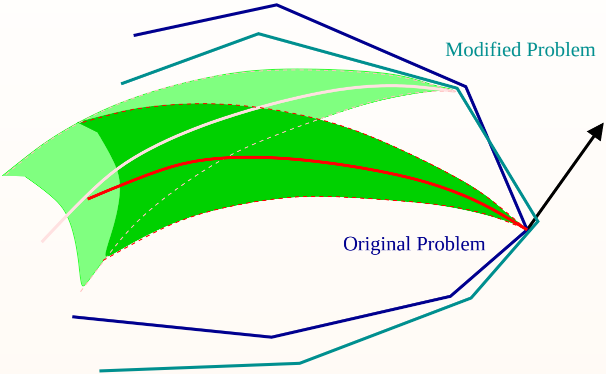
Why?

Hippolito (1993): Search direction is parallel to nearby constraints



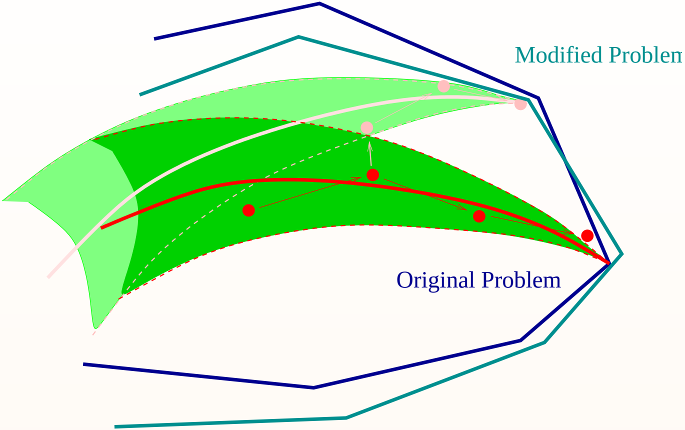
Why?

Hippolito (1993): Search direction is parallel to nearby constraints



$\Rightarrow$ only small step in search direction can be taken

Idea: Start close to the (new) central path, not close to the (old) solution



$\Rightarrow$ Start from a previous iterate and do additional *modification* step.

- Yıldırım/Wright ('02): Weighted Least Squares (WLS), Newton correction
- Gondzio/G. ('03/'08): Splitting Directions, Unblocking

Interior Point Warmstarts: Theoretical Results

A typical warmstart results is (Assume $\tilde{A} = A$):

Lemma (based on Gondzio/G. '03)

Let $(x, \lambda, s) \in \mathcal{N}_{-\infty}(\gamma_0)$ for problem (LP) then a full recentering step $(\Delta x, \Delta\lambda, \Delta s)$ in the perturbed problem $(\widetilde{LP})$ is feasible and

$$(x + \Delta x, \lambda + \Delta\lambda, s + \Delta s) \in \tilde{\mathcal{N}}_{-\infty}(\gamma)$$

provided that

$$\delta_{bc} \le \frac{\sqrt{\gamma_0 - \gamma}}{2B_\infty^2}\gamma_0\mu^{3/2}$$

$\Rightarrow$ **"small δ_{bc}, large μ"**

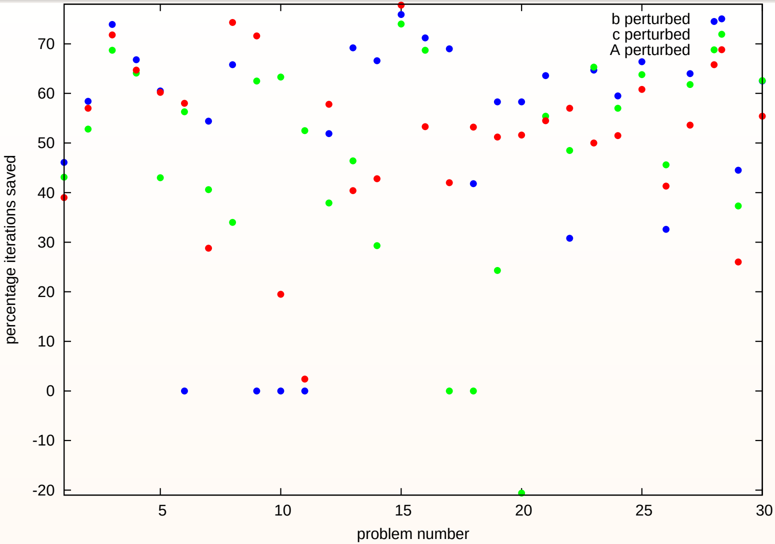
---

where

$$\begin{aligned} \delta_{bc} &:= \|\tilde{c} - \tilde{A}^T y - s\|_2 + \|\tilde{A}^T(\tilde{A}\tilde{A}^T)^{-1}(\tilde{b} - \tilde{A}x)\|_2 \\ &:= P_{\tilde{c} - \tilde{A}^T y - s}(s) + P_{\tilde{b} - \tilde{A}x}(x) \end{aligned}$$

is the **orthogonal distance** of the warmstart point from primal-dual feasibility in the perturbed problem

Results for LP problems (NETLIB) (Gondzio, G. '08)



IPM warmstart can save 50%-60% of iterations

Stochastic Programming Crashstarts

- **Idea:** speed up solution process by crash-starting from a smaller tree
- **But:** IPMs are notoriously bad at exploiting a known starting point

The aim of a crash-start is to construct (cheaply) a point on (or near) the central path.

Stochastic Programming Crashstarts

- **Idea:** speed up solution process by crash-starting from a smaller tree
- **But:** IPMs are notoriously bad at exploiting a known starting point

The aim of a crash-start is to construct (cheaply) a point on (or near) the central path.

Central Path conditions for full problem

primal feasibility	dual feasibility	centrality
$Ax = b$	$\sum_i T^T \lambda_i + s = c_0$	$XSe = \mu e$
$Tx + W_i y_i = h_i$	$W_i^T \lambda_i + z_i = p_i c_i$	$Y_i Z_i e = \mu e$
		$x, s, y, z \geq 0$

The (unique) solution is denoted by

$$(x_\mu(\mathcal{T}), y_\mu(\mathcal{T}), \lambda_\mu(\mathcal{T}), s_\mu(\mathcal{T}), z_\mu(\mathcal{T}))$$

Stochastic Programming Crashstarts

- **Idea:** speed up solution process by crash-starting from a smaller tree
- **But:** IPMs are notoriously bad at exploiting a known starting point

The aim of a crash-start is to construct (cheaply) a point on (or near) the central path.

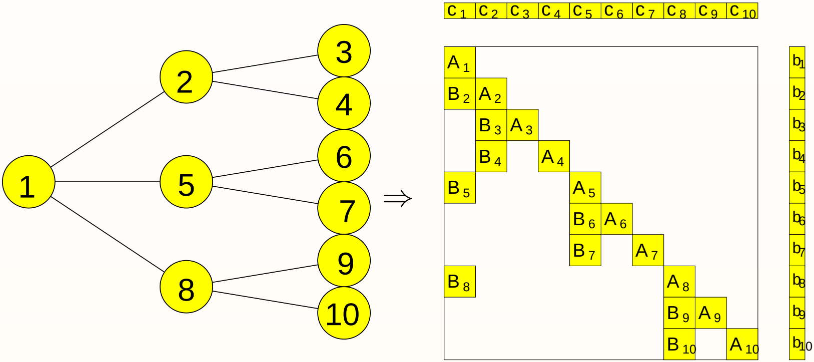
Central Path conditions for full problem

primal feasibility	dual feasibility	centrality
$Ax = b$	$\sum_i T^T \lambda_i + s = c_0$	$XSe = \mu e$
$Tx + W_i y_i = h_i$	$W_i^T \lambda_i + z_i = p_i c_i$	$Y_i Z_i e = \mu e$
		$x, s, y, z \geq 0$

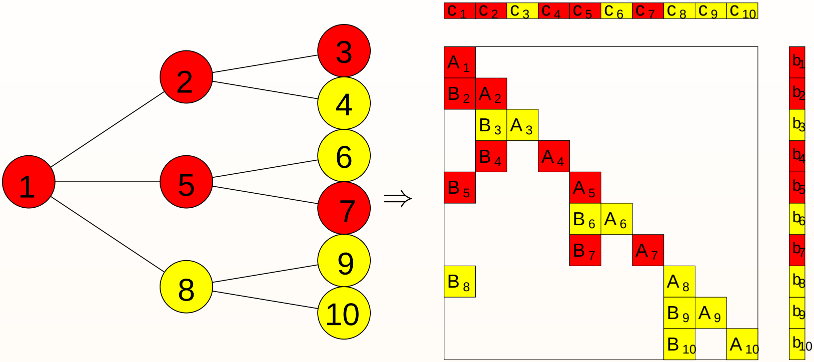
The (unique) solution is denoted by

$$(x_\mu(\mathcal{T}), y_\mu(\mathcal{T}), \lambda_\mu(\mathcal{T}), s_\mu(\mathcal{T}), z_\mu(\mathcal{T}))$$

- The stochastic programming crashstart schemes attempt to (cheaply) find an approximation to $(x_\mu(\mathcal{T}), y_\mu(\mathcal{T}), \lambda_\mu(\mathcal{T}), s_\mu(\mathcal{T}), z_\mu(\mathcal{T}))$
- **Reduced tree based warmstart** (Colombo, Gondzio, G. '09)

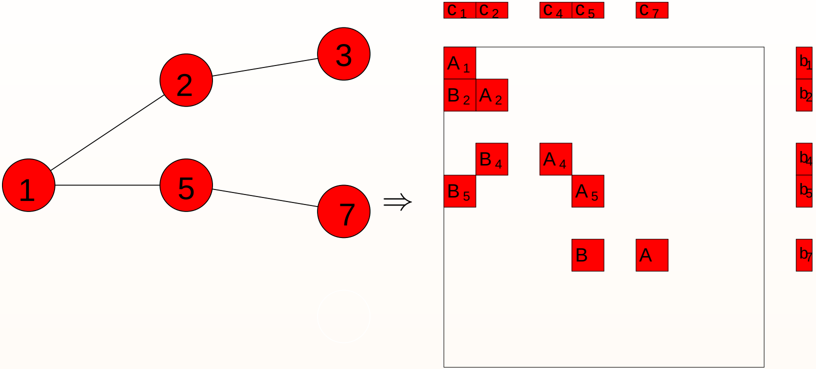

1
2
3
4
5
6
7
8
9
10
⇒
c 1 c 2 c 3 c 4 c 5 c 6 c 7 c 8 c 9 c 10
A 1
B 2 A 2
B 3 A 3
B 4 A 4
B 5 A 5
B 6 A 6
B 7 A 7
B 8 A 8
B 9 A 9
B 10 A 10
b1 b2 b3 b4 b5 b6 b7 b8 b9 b10

Stochastic Programming Crashstart



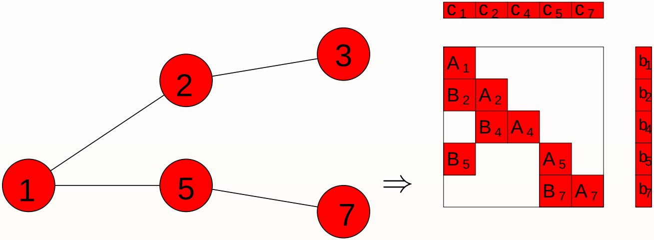
- Select sample scenarios

Stochastic Programming Crashstart



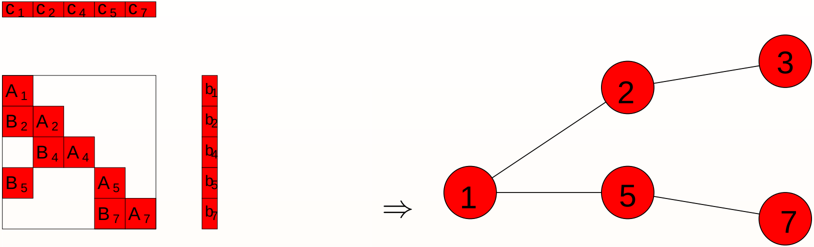
- Select sample scenarios
- Aggregate Scenarios/Reduce Problem

Stochastic Programming Crashstart

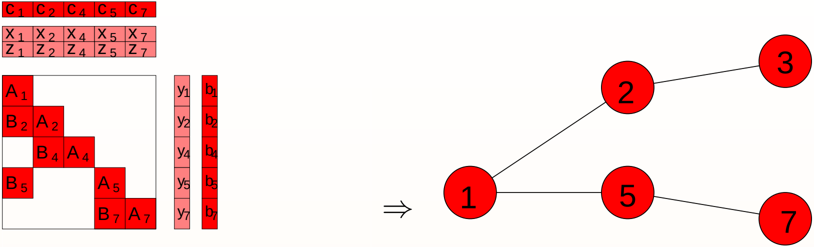


- Select sample scenarios
- Aggregate Scenarios/Reduce Problem

Stochastic Programming Warmstarts

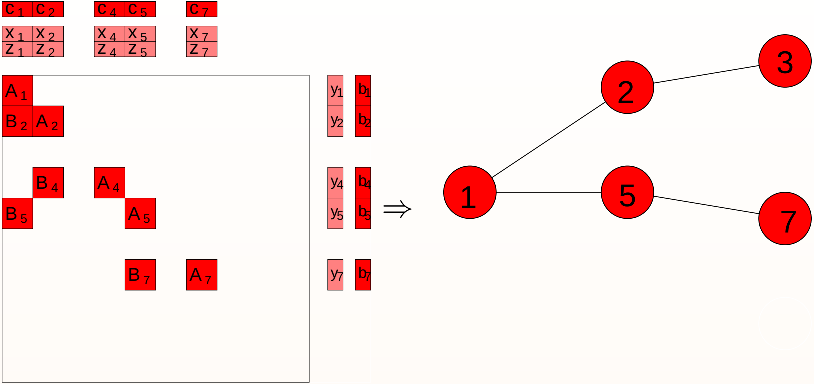


To solve the problem by warmstarting, reverse the process



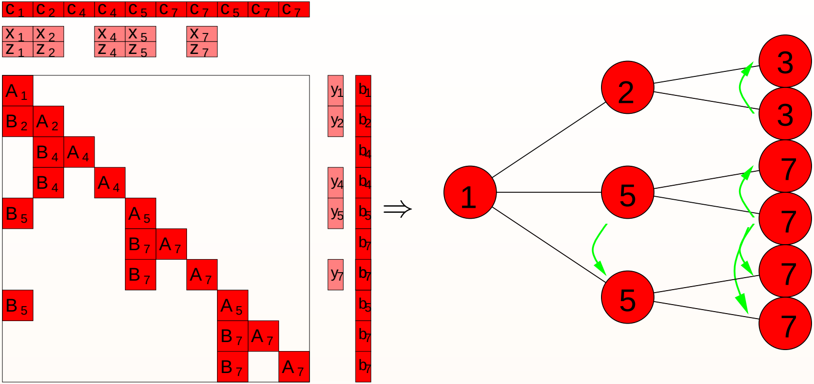
To solve the problem by warmstarting, reverse the process

- Find central point for reduced problem



To solve the problem by warmstarting, reverse the process

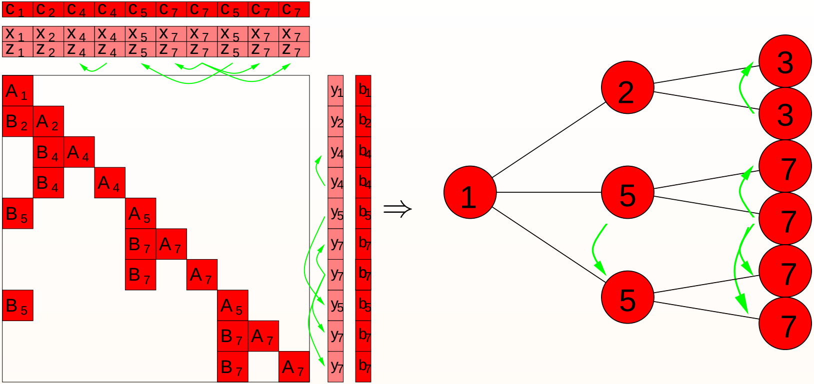
- Find central point for reduced problem
- Expand the problem to original size



To solve the problem by warmstarting, reverse the process

- Find central point for reduced problem
- Expand the problem to original size (by duplicating scenarios)

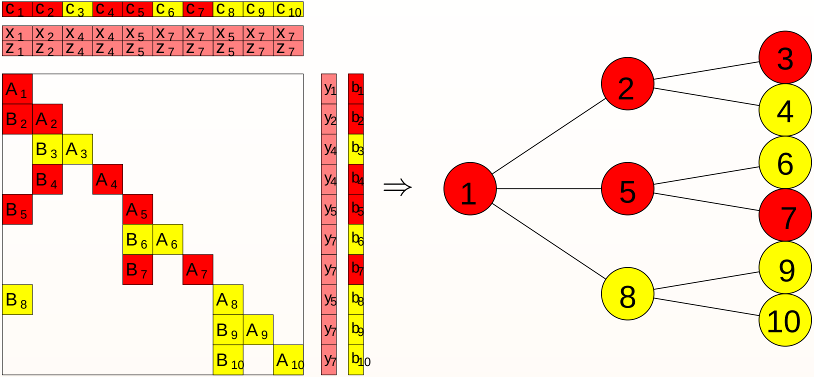
Stochastic Programming Warmstarts



To solve the problem by warmstarting, reverse the process

- Find central point for reduced problem
- Expand the problem to original size (by duplicating scenarios)
- Expand solution to primal/dual feasible point for expanded problem

Stochastic Programming Warmstarts

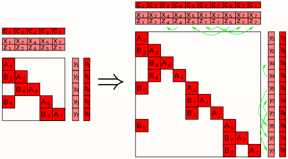


To solve the problem by warmstarting, reverse the process

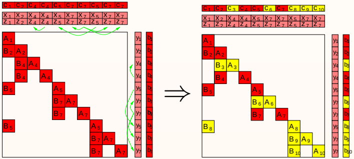
- Find central point for reduced problem
- Expand the problem to original size (by duplicating scenarios)
- Expand solution to primal/dual feasible point for expanded problem
- Use this to warmstart full problem

The propsed warmstart procedure is a two stage procedure:

- Reduced Problem ⇒ Expanded Problem
  - *Can construct primal/dual feasible starting point*
  - *Although point is not central (We are duplicating constraints!)*

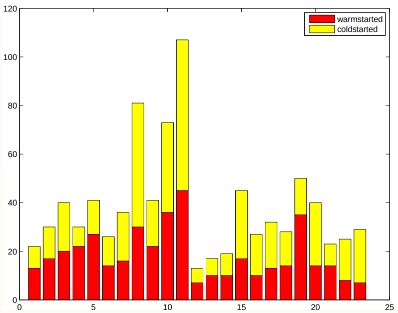


- Expanded Problem ⇒ Full Problem
  - *Can bound changes in problem data (differences in scenarios)*

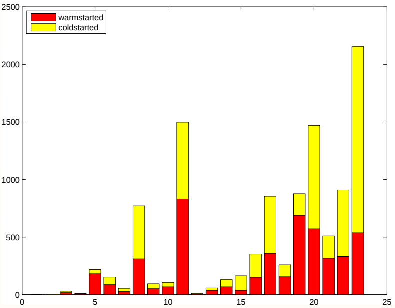


Reduced Tree Warmstart: Results

- SP test problems & Capacity assignment problems
- Ranging from 1,000 - 100,000 variables



Number of IPM iterations
-50.5%



Total solution time (s)
-42.1%

→ Colombo, Gondzio, G. (2009)

Extensions of Basic Idea

Possible refinements of the idea

- Multilevel Reduced Tree Warmstart
  (uses sequence of trees $\mathcal{T}_1 \subset \mathcal{T}_2 \subset \cdots \subset \overline{\mathcal{T}}$)
- Decomposition Based Warmstart
  (Find better approximation to central path on full problem by an additional half iteration of Benders Decomposition)

Multilevel Reduced Tree Warmstart

Algorithm: Multiple-tree stochastic programming warm-start

Given: A sequence of trees $\mathcal{T}^{(1)}, \mathcal{T}^{(2)}, \cdots, \mathcal{T}^{(K)} = \overline{\mathcal{T}}$ with resulting problem sizes n_k and closeness measures $d^{(k)} = d(\mathcal{T}^{(k)}, \mathcal{T}^{(k+1)})$. Further a sequence of μ-values $\{\mu_k\}_k$.

for $k = 1$ to K do

1. Solve problem $P(\mathcal{T}^{(k)})$ to accuracy μ_k, to obtain a point $(x^{(k)}, y^{(k)}, s^{(k)}) \in \mathcal{N}_{-\infty}^{(k)}$ with $(x^{(k)})^T s^{(k)}/n_k = \mu_k$.
2. Perform additional centering steps to obtain a point in a tighter neighbourhood.
3. Expand this point to a solution $(\hat{x}^{(k)}, \hat{y}^{(k)}, \hat{s}^{(k)})$ of the expanded problem.
4. Use $(\hat{x}^{(k)}, \hat{y}^{(k)}, \hat{s}^{(k)})$ to warm-start the problem on the next tree in the sequence $\mathcal{T}^{(k+1)}$.

end for

Scenario tree distance

Scenario: $\eta_i = (T_i, W_i, h_i, q_i)$. **Tree:** $\mathcal{T} = \{\eta_i : i \in I\}$.

Distance between two scenarios

Let $\eta_i, \eta_j \in \mathcal{T}$. Then

$$d(\eta_i, \eta_j) := \|T_i - T_j\|_2 + \|W_i - W_j\|_2 + \|h_i - h_j\|_2 + \|q_i - q_j\|_2.$$

is the distance between two scenarios

Distance of a **scenario** η **from a tree** $\mathcal{T}$

$$d(\eta, \mathcal{T}) := \min_{\eta_i \in \mathcal{T}} d(\eta, \eta_i).$$

Distance between two trees

Let $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ be two trees

$$d(\mathcal{T}^{(1)}, \mathcal{T}^{(2)}) := \max_{\eta_i \in \mathcal{T}^{(1)}} d(\eta_i, \mathcal{T}^{(2)}).$$

Well-balancedness measure

Given the reduced tree $\mathcal{T}_R \subset \mathcal{T}$. We have a mapping

$$
\begin{aligned}
r : \mathcal{T} &\rightarrow \mathcal{T}_R \\
r(\eta) &= \arg\min_{\eta_R \in \mathcal{T}_R} d(\eta, \eta_R), \qquad \textit{(closest node in } \mathcal{T}_R\textit{)}
\end{aligned}
$$

And its inverse

$$\mathcal{I}_k = \mathcal{I}(\eta_R^k) := \{\eta^l \in \mathcal{T} : r(\eta^l) = \eta_R^k\}$$

then

$$p_R^k = \sum_{\eta^l \in \mathcal{I}_k} p^l$$

$\Rightarrow$ Scenario η_R^k *aggregates* the scenarios in $\mathcal{I}(\eta_R^k)$.

Well-balancedness measure ρ

$$\rho = \min_{\eta^l \in \mathcal{T}} \left\{ \frac{p^l}{p_R^{r(l)}} \frac{n}{n_R} \right\}, \qquad \Rightarrow \rho \leq 1$$

For a well balanced selection of nodes $\mathcal{T}_R \subset \mathcal{T}$ we expect $\rho \approx 1$

Boundedness assumptions

Lemma.

The stochastic programming warmstart is succesful if

$$d(\mathcal{T}_R, \mathcal{T})\sqrt{|\mathcal{T}|} \le \frac{\rho\gamma_0\sqrt{\rho\gamma_0 - \gamma}}{2B_\infty^2 \|W^T(WW^T)^{-1}\|_2} \left(\frac{n}{n_R}\right)^{3/2} \mu_0^{3/2}$$

Can we get a globally valid bound B_∞ for all iterates encountered in the process?

Boundedness assumptions

Lemma.

The stochastic programming warmstart is succesful if

$$d(\mathcal{T}_R, \mathcal{T})\sqrt{|\mathcal{T}|} \leq \frac{\rho\gamma_0\sqrt{\rho\gamma_0 - \gamma}}{2B_\infty^2\|W^T(WW^T)^{-1}\|_2}\left(\frac{n}{n_R}\right)^{3/2}\mu_0^{3/2}$$

Can we get a globally valid bound B_∞ for all iterates encountered in the process?

Boundedness of iterates

All primal-dual feasible iterates (x, y, s) satisfy

$$\|x\|_\infty, \|s\|_\infty \leq \overline{B}_\infty := 2\|\mathfrak{d}\|_1(\overline{C(\mathfrak{d})}^2 + n\mu/\rho(\mathfrak{d}))$$

where $C(\mathfrak{d})$ is the Renegar condition number of the problem $\mathfrak{d} = (A, b, c)$:

$$C(\mathfrak{d}) = \frac{\|\mathfrak{d}\|}{\rho(\mathfrak{d})}, \quad \rho(\mathfrak{d}) = \textit{“distance to infeasibility”}$$

(Nunez, Freund '98)

Boundedness assumptions

Lemma.

The stochastic programming warmstart is succesful if

$$d(\mathcal{T}_R, \mathcal{T})\sqrt{|\mathcal{T}|} \le \frac{\rho\gamma_0\sqrt{\rho\gamma_0 - \gamma}}{2B_\infty^2 \|W^T(WW^T)^{-1}\|_2} \left(\frac{n}{n_R}\right)^{3/2} \mu_0^{3/2}$$

Can we get a globally valid bound B_∞ for all iterates encountered in the process?

Boundedness of ρ and $C(\mathfrak{d}^{(k)})$

Given a sequence of trees $\mathcal{T}^{(1)}, \mathcal{T}^{(2)}, \cdots, \mathcal{T}^{(K)} := \overline{\mathcal{T}}$ where $\mathcal{T}^{(k-1)}$ is obtained from $\mathcal{T}^{(k)}$ by scenario aggregation, then

- $\rho(\mathfrak{d}(\mathcal{T}^{(k)})) \ge \overline{\rho} := \rho(\mathfrak{d}(\overline{\mathcal{T}}))$,
- $C(\mathfrak{d}^{(k)}) \le \overline{C(\mathfrak{d})} := C(\mathfrak{d}^{(K)}), \quad \forall k.$

Multilevel Reduced Tree Warmstart

$\mathcal{N}_{-\infty}$-neighbourhood

Given is a sequence of trees $\mathcal{T}^{(1)}, \mathcal{T}^{(2)}, \cdots, \mathcal{T}^{(K)}$ with resulting problem sizes $n^{(k)}$ and closeness measures $d^{(k)} = d(\mathcal{T}^{(k)}, \mathcal{T}^{(k+1)})$. If

$$\gamma^{(k+1)} = \frac{1}{2}(\gamma^{(k)} + \frac{1}{10})$$

and satisfy the conditions

$$\begin{aligned} 1 - \rho^{(k,k+1)} &\leq \frac{1}{4}(\gamma^{(k)} - \frac{1}{10}) \\ d^{(k)}\sqrt{|\mathcal{T}^{(k+1)}|} &\leq \frac{\sqrt{\gamma^{(k)} - \frac{1}{10}}}{60\overline{B}_\infty^2 \|W^T(WW^T)^{-1}\|_2}\left(\frac{(n^{(k)})}{(n^{(k+1)})}\mu\right)^{3/2} \end{aligned}$$

then the algorithm can be performed successfully, that is all warm-starts are successful and lead to points

$$(x^{(k)}, y^{(k)}, s^{(k)}) \in \mathcal{N}_{-\infty}^{(k)}(\gamma^{(k)}).$$

Complexity Analysis

- Computational cost of an interior point iteration is $\mathcal{O}(n^\alpha)$ (typically $\alpha > 1$).

$\Rightarrow$ The computational cost will be dominated by the number of iterations that have to be performed on the full problem $P(\mathcal{T}^{(K)})$

- From an iterate in $N_{-\infty}(\gamma)$ corresponding to μ need

$$\mathcal{O}\left(n_K \log \frac{\mu}{\epsilon}\right)$$

iterations to convergence in the full problem

Complexity Analysis

2-step crash-start

- Smallest acceptable μ_1-value (in the reduced problem) that leads to a successful warm-start

$$\mu_1 \geq C d(\mathcal{T}^{(1)}, \mathcal{T}^{(K)})^{2/3} |\mathcal{T}^{(K)}|^{1/3} \frac{n_K}{n_1}$$

- Using $|\mathcal{T}^{(K)}| \sim n_K$ we have

$$\mu_1 \geq C d(\mathcal{T}^{(1)}, \mathcal{T}^{(K)})^{2/3} n_K^{4/3} / n_1$$

- this yields a warm-start point in the full problem with a corresponding μ-value

$$\underline{\mu}^{(1)} = \mu_1 \frac{n_1}{n_K} \geq C d(\mathcal{T}^{(1)}, \mathcal{T}^{(K)})^{2/3} n_K^{1/3} \qquad (1)$$

Complexity Analysis

Use a sequence of trees $\mathcal{T}^{(1)}, \mathcal{T}^{(2)}, \mathcal{T}^{(K)}$

- final warm-start will be performed from $\mathcal{T}^{(2)}$ and by a similar argument we obtain the bound

$$\underline{\mu}^{(2)} \geq Cd(\mathcal{T}^{(2)}, \mathcal{T}^{(K)})^{2/3} n_K^{1/3}. \tag{2}$$

- Comparing (1) and (2) the number of full problem iterations saved by performing the multi-step warm-start is

$$\mathcal{O}\left(n_K \log \frac{\underline{\mu}^{(2)}}{\underline{\mu}^{(1)}}\right) = \mathcal{O}\left(n_K \cdot a \log \frac{d(\mathcal{T}^{(2)}, \mathcal{T}^{(K)})}{d(\mathcal{T}^{(1)}, \mathcal{T}^{(K)})}\right).$$

Multilevel Reduced Tree Warmstart: Numerical Results

Test Problem Sizes

		core size		problem size	
Problem	scenarios	constraints	variables	constraints	variables
Minoux	10000	118	239	1.160.001	3.050.051
Jll_gva	4000	350	781	1.392.001	4.180.085
T1B3	10000	50	109	1.070.001	1.350.024
r4c	10000	177	272	2.700.001	2.970.151
ex1	20000	50	70	960.001	1.780.006
	40000	50	70	1.920.001	3.560.006
ex3	10000	72	112	700.001	1.380.009
	20000	72	112	1.400.001	2.760.009
s97	10000	147	157	1.450.001	2.390.010
s98	10000	149	141	1.470.001	2.060.006
j99	10000	158	148	1.560.001	2.170.007

The first set of problems are stochastic capacity assignment problems, the second set are stochastic blending problems.

Multilevel Warmstart Results

Stochastic Programming Testproblems

Problem	scenarios	cold	2-step	multistep
Minoux	10000	2644	1212	1176
JII_gva	4000	4981	2523	2251
T1B3	10000	995	663	637
r4c	10000	2098	835	749
ex1	20000	580	327	302
	40000	1559	766	701
ex3	10000	563	346	316
	20000	1793	626	586
s97	10000	498	389	307
s98	10000	3189	826	481
j99	10000	1796	375	295

(Solution time in seconds)

Decomposition Based Warmstart

Recall:

The aim of a crash-start is to construct (cheaply) a point on (or near) the central path.

Central Path conditions for full problem

primal feasibility	dual feasibility	centrality
$Ax = b$	$\sum_i T^T\lambda_i + s = c_0$	$XSe = \mu e$
$Tx + W_i y_i = h_i$	$W_i^T\lambda_i + z_i = p_i c_i$	$Y_i Z_i e = \mu e$
		$x, s, y, z \geq 0$

The (unique) solution is denoted by

$$(x_\mu(\mathcal{T}), y_\mu(\mathcal{T}), \lambda_\mu(\mathcal{T}), s_\mu(\mathcal{T}), z_\mu(\mathcal{T}))$$

- The stochastic programming warmstart schemes attempt to (cheaply) find an approximation to $(x_\mu(\mathcal{T}), y_\mu(\mathcal{T}), \lambda_\mu(\mathcal{T}), s_\mu(\mathcal{T}), z_\mu(\mathcal{T}))$

Decomposition Based Warmstart

The aim of a crash-start is to construct (cheaply) a point on (or near) the central path.

primal feasibility	dual feasibility	centrality
$Ax = b$	$\sum_i T^T \lambda_i + s = c_0$	$XSe = \mu e$
$Tx + W_i y_i = h_i$	$W_i^T \lambda_i + z_i = p_i c_i$	$Y_i Z_i e = \mu e$
		$x, s, y, z \geq 0$

Analysis of reduced tree warmstart scheme:

- The tree expansion constructs a primal/dual feasible (not central) point
- Scenario changes destroy primal/dual feasibility.

Is it possible to regain primal/dual feasibility & centrality?

Decomposition based Warmstarting Scheme

Idea

- Use estimate of first stage decisions x, s from reduced tree $\mathcal{T}_R$.
- Given these, solve scenario subproblems to get $(y_i, z_i, \lambda_i) \forall i \in \mathcal{T}$.

Decomposition based Warmstarting Scheme

Idea

- Use estimate of first stage decisions x, s from reduced tree $\mathcal{T}_R$.
- Given these, solve scenario subproblems to get $(y_i, z_i, \lambda_i) \forall i \in \mathcal{T}$.

Given x, s the conditions for a point on the central path (almost) decompose into

$$\begin{aligned} W_i y_i &= h_i - Tx \\ W_i^T \lambda_i + z_i &= p_i c_i \\ Y_i Z_i e &= \mu e \end{aligned}$$

which are the optimality conditions for the barrier scenario subproblem

$$\begin{array}{ll} \min_{y_i} & p_i c_i^T y_i - \mu \sum_j \ln y_{ij} \\ \text{s.t.} & W_i y_i = h_i - Tx \end{array} \qquad (P_i(x))$$

Decomposition based Warmstarting Scheme

Idea

- Use estimate of first stage decisions x, s from reduced tree $\mathcal{T}_R$.
- Given these, solve scenario subproblems to get $(y_i, z_i, \lambda_i) \forall i \in \mathcal{T}$.

Given x, s the conditions for a point on the central path (almost) decompose into

$$\begin{aligned} W_i y_i &= h_i - Tx \\ W_i^T \lambda_i + z_i &= p_i c_i \\ Y_i Z_i e &= \mu e \end{aligned}$$

which are the optimality conditions for the barrier scenario subproblem

$$\begin{array}{ll} \min_{y_i} & p_i c_i^T y_i - \mu \sum_j \ln y_{ij} \\ \text{s.t.} & W_i y_i = h_i - Tx \end{array} \qquad (P_i(x))$$

Note

- Disregard the global dual constraint: $\sum_i T^T \lambda_i + s = c_0$.
- Satisfied (approximately) if $\mathcal{T}_R \approx \mathcal{T}$.

Decomposition based Warmstarting Scheme: Algorithm

Algorithm

- Input: Reduced tree $\mathcal{T}_R$, target $\mu : \mu_T$.

1. Solve $P(\mathcal{T}_R)$ to find $(\hat{x}, \hat{s}) := (x_{\mu_T}(\mathcal{T}_R), s_{\mu_T}(\mathcal{T}_R))$.
   as approximation to $(x_{\mu_T}(\mathcal{T}), s_{\mu_T}(\mathcal{T}))$
2. $\forall i$:Solve scenario subproblems $P_i(\hat{x})$
   to get μ_T-centers $(\hat{y}_i, \hat{z}_i, \hat{\lambda}_i) := (y_{\mu,i}, z_{\mu,i}, \lambda_{\mu,i})$.
   as approximation to $(y_{\mu_T}(\mathcal{T}), z_{\mu_T}(\mathcal{T}), \lambda_{\mu_T}(\mathcal{T}))$

- Output: $(\hat{x}, \hat{y}, \hat{\lambda}, \hat{s}, \hat{z})$ as the warmstart point for $P(\mathcal{T})$.

Decomposition based Warmstarting Scheme: Algorithm

Algorithm

- Input: Reduced tree $\mathcal{T}_R$, target $\mu : \mu_T$.

1. Solve $P(\mathcal{T}_R)$ to find $(\hat{x}, \hat{s}) := (x_{\mu_T}(\mathcal{T}_R), s_{\mu_T}(\mathcal{T}_R))$.
   as approximation to $(x_{\mu_T}(\mathcal{T}), s_{\mu_T}(\mathcal{T}))$
2. $\forall i$:Solve scenario subproblems $P_i(\hat{x})$
   to get μ_T-centers $(\hat{y}_i, \hat{z}_i, \hat{\lambda}_i) := (y_{\mu,i}, z_{\mu,i}, \lambda_{\mu,i})$.
   as approximation to $(y_{\mu_T}(\mathcal{T}), z_{\mu_T}(\mathcal{T}), \lambda_{\mu_T}(\mathcal{T}))$

- Output: $(\hat{x}, \hat{y}, \hat{\lambda}, \hat{s}, \hat{z})$ as the warmstart point for $P(\mathcal{T})$.

Rationale

- If $\mathcal{T}_R \approx \mathcal{T}$
- then $(\hat{x}, \hat{s}) := (x_{\mu_T}(\mathcal{T}_R), s_{\mu_T}(\mathcal{T}_R)) \approx (x_{\mu_T}(\mathcal{T}), s_{\mu_T}(\mathcal{T}))$
- and thus $(\hat{y}, \hat{\lambda}, \hat{z}) \approx (y_{\mu_T}(\mathcal{T}), \lambda_{\mu_T}(\mathcal{T}), z_{\mu_T}(\mathcal{T}))$

Decomposition based Warmstarting Scheme: Analysis

By construction $(\hat{x}, \hat{y}, \hat{\lambda}, \hat{s}, \hat{z})$ is

- *Central*
- *Primal feasible*

Decomposition based Warmstarting Scheme: Analysis

By construction $(\hat{x}, \hat{y}, \hat{\lambda}, \hat{s}, \hat{z})$ is

- *Central*
- *Primal feasible*

Dual feasibility depends on residual in $\sum_i T^T \lambda_i + s = c_0$:

$$= \mathcal{O}(\|\hat{\lambda} - \lambda_\mu(\mathcal{T}))\| \stackrel{?}{=} \mathcal{O}(d(\mathcal{T}, \mathcal{T}_R))$$

Decomposition based Warmstarting Scheme: Analysis

By construction $(\hat{x}, \hat{y}, \hat{\lambda}, \hat{s}, \hat{z})$ is

- *Central*
- *Primal feasible*

Dual feasibility depends on residual in $\sum_i T^T \lambda_i + s = c_0$:

$$= \mathcal{O}(\|\hat{\lambda} - \lambda_\mu(\mathcal{T}))\| \stackrel{?}{=} \mathcal{O}(d(\mathcal{T}, \mathcal{T}_R))$$

Wasserstein (Kantorovich) distance

A standard measure of distance between scenario sets

$$\mathcal{T} = \{(p_i, \xi_i)\}, \qquad \mathcal{T}_R = \{(\tilde{p}_i, \tilde{\xi}_i\}$$

is the *Wasserstein* or *Transportation* distance:

$$W_1(\mathcal{T}, \mathcal{T}_R) := \min_{\eta \geq 0} \left\{ \sum_{(\pi_i, \xi_i) \in \mathcal{T}} \sum_{(\tilde{\pi}_j, \tilde{\xi}_j) \in \mathcal{T}_R} \|\xi_i - \tilde{\xi}_j\| \eta_{ij} : \sum_i \eta_{ij} = \tilde{\pi}_j, \sum_j \eta_{ij} = \pi_i \right\}$$

Decomposition based Warmstarting: Theoretical Results

Given

- L_h : $\|h(\xi) - h(\tilde{\xi})\| \leq L_h \|\xi - \tilde{\xi}\|$
- $\overline{B}$: $\|x_{\mu_T}(\mathcal{T}_R)\|_\infty \leq B, \qquad \forall \mathcal{T}_R$
- $L_Q = 3\overline{C(d)}(\overline{C(d)\|d\|} + \mu_T n)\chi(W) L_h$
- $C_\lambda = 4\chi(W)\|T\|\overline{C(d)}^2[\overline{C(d)\|d\|} + \mu_T n]^2/\mu_T$

Lemma

We have

$$\|\lambda_\mu(\mathcal{T}) - \hat{\lambda}\|_\infty \quad \leq \quad \frac{2}{\sqrt{\mu_T}} C_\lambda \overline{B} \sqrt{L_Q} \sqrt{W_1(\mathcal{T}, \mathcal{T}_R)}$$

$\Rightarrow$ **"small $W_1(\mathcal{T}, \mathcal{T}_R)$, large μ"**

Decomposition based Warmstarting: Theoretical Results

Theorem

Let $\hat{w} = (\hat{x}, \hat{y}, \hat{\lambda}, \hat{s}, \hat{z})$ be the warm-start point for problem $P(\mathcal{T})$ obtained by following the above algorithm starting from the reduced tree $\mathcal{T}_R$. If

$$W_1(\mathcal{T}, \mathcal{T}_R) \leq \frac{\theta^2}{\overline{C(d)}^2 C_\lambda^2} \min\left\{ \frac{\overline{\|d\|}^2}{4(2n+1)^2}, \frac{\mu_T^2}{16\overline{C(d)}^2} \right\}$$

then the warmstart is successful, that is the full centering step is feasible and

$$(\hat{x} + \Delta x, \hat{y} + \Delta y, \hat{\lambda} + \Delta\lambda, \hat{s} + \Delta s, \hat{z} + \Delta z) \in \mathcal{N}_2^{\mathcal{T}}(\theta).$$

Theorem

Let $\hat{w} = (\hat{x}, \hat{y}, \hat{\lambda}, \hat{s}, \hat{z})$ be the warm-start point for problem $P(\mathcal{T})$ obtained by following the above algorithm starting from the reduced tree $\mathcal{T}_R$. If

$$W_1(\mathcal{T}, \mathcal{T}_R) \leq \frac{\theta^2}{\overline{C(d)}^2 C_\lambda^2} \min\left\{ \frac{\overline{\|d\|}^2}{4(2n+1)^2}, \frac{\mu_{\mathcal{T}}^2}{16\overline{C(d)}^2} \right\}$$

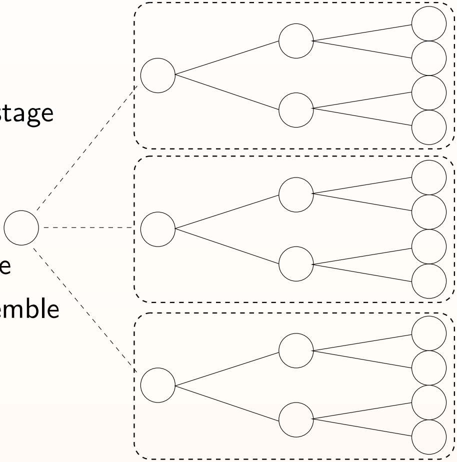
then the warmstart is successful, that is the full centering step is feasible and

$$(\hat{x} + \Delta x, \hat{y} + \Delta y, \hat{\lambda} + \Delta\lambda, \hat{s} + \Delta s, \hat{z} + \Delta z) \in \mathcal{N}_2^{\mathcal{T}}(\theta).$$

A similar result can be proven if we only require approximate $\mu_{\mathcal{T}}$-centers!

Extension to Multistage SP

- Easy to generalize to multistage SP
- Decompose problem at 2nd stage
  $\Rightarrow P_i(x)$ is based on subtree
- Solve each subtree and assemble warmstart point



2-Stage Problems: Dimensions

Problem	scenarios	rows	columns	nonzeros
AIRL1	25	152	306	706
AIRL2	25	152	306	706
AIRL3	676	4,058	8,118	18,934
cargo-4node32	32	2,382	6,396	15,656
cargo-4node64	64	4,750	12,732	31,016
cargo-4node128	128	9,486	25,404	61,736
cargo-4node256	256	18,958	50,748	123,176
cargo-4node512	512	37,902	101,436	246,056
cargo-4node1024	1024	75,790	202,812	491,816
cargo-4node2048	2048	151,566	405,564	983,336
cargo-4node4096	4096	303,118	811,068	1,966,376
cargo-4node8192	8196	606,222	1,622,076	3,932,456
cargo-4node16384	16384	1,212,430	3,244,092	7,864,616
asset1	100	505	1,313	2,621
asset2	37500	187,505	487,513	975,021
env.1200	1200	57,648	102,085	220,972
env.1875	1875	90,048	159,460	345,172
env.3780	3780	181,488	321,385	695,692
env.5292	5292	254,064	449,905	973,900
env.lrge	8232	395,184	699,805	1,514,860
stocfor2	64	6,543	9,237	29,985
sslp_10_50_100	100	6,001	52,011	101,911
sslp_10_50_500	500	30,001	260,011	509,511
sslp_10_50_1000	1000	60,001	520,011	1,019,011
dcap233_500	500	7,506	16,518	31,518
dcap243_500	500	9,006	21,018	39,018
storm27	27	14,441	37,485	94,274
storm125	125	66,185	172,431	433,256
storm1000	1000	528,185	1,377,306	3,459,881

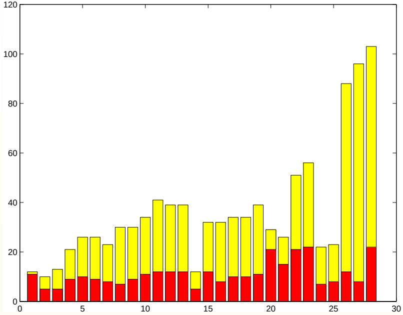
Table: Characteristics of 2-stage problems

2-stage Problems: Results

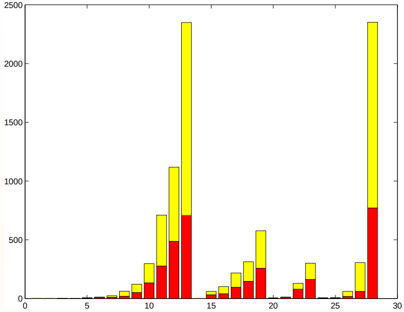
Problem	$\lvert\mathcal{T}\rvert$	$\lvert\mathcal{T}^R\rvert$	cold iter	cold time	dec iter	dec time	$\bar{\mu}$
AIRL1	25	10	12	0.08	11	0.16	1.0
AIRL2	25	10	10	0.07	5	0.16	1.0
AIRL3	676	60	13	2.8	5	3.5	1.0
cargo-4node32	32	6	21	2.6	9	2.5	1.0
cargo-4node64	64	10	26	6.4	10	4.9	1.0
cargo-4node128	128	10	26	13.2	9	8.2	0.1
cargo-4node256	256	10	23	23.8	8	11.2	0.01
cargo-4node512	512	10	30	62.9	7	20.6	0.01
cargo-4node1024	1024	50	30	123	9	52	0.01
cargo-4node2048	2048	200	34	297	11	134	0.01
cargo-4node4096	4096	200	41	710	12	277	0.01
cargo-4node8192	8192	400	39	1118	12	487	0.01
cargo-4node16384	16384	800	39	2349	12	708	0.01
asset1	100	10	12	0.3	5	0.3	0.01
env.1200	1200	50	32	61	12	32	0.01
env.1875	1875	50	32	102	8	41	0.001
env.3780	3780	50	34	218	10	97	0.001
env.5292	5292	200	34	313	10	147	0.001
env.lrge	8232	200	39	577	11	258	0.001
stocfor2	64	2	21	3.9	29	4.8	0.01
sslp_10_50_100	100	10	26	12.6	15	12.3	0.001
sslp_10_50_500	500	50	51	130	21	81	0.0001
sslp_10_50_1000	1000	50	56	301	22	163	0.0001
dcap233_500	500	50	22	6.5	7	4.3	0.01
dcap243_500	500	50	23	7.8	8	5.3	0.001
storm27	27	5	88	61	12	18.8	0.1
storm125	125	10	96	306	8	61	1.0
storm1000	1000	100	103	2351	22	771	1.0

Table: Results for 2-stage problems.

2-stage Problems: Results



Number of IPM iterations
-68.4%



Total solution time (s)
-62.8%

→ Colombo, G. (2010)

Test Problem Characteristics

Problem	stages	scenarios	rows	columns	nonzeros
fxm3-6	3	36	6,200	12,628	57,722
fxm3-16	3	256	41,340	85,575	392,252
pltexpA4-6	4	216	26,894	70,364	143,059
swing8-4	8	65,536	262,142	349,522	786,422
mmix-50	2	50	32,951	72,517	366,817
mmix-100	2	100	65,901	145,017	733,617
watson10-64	7	64	15,101	28,097	72,648
watson10-128	8	128	26,237	49,153	128,648
watson10-256	9	256	43,517	82,177	218,888

Numerical Results

Problem	cold		reduced		decomp	
	iter	time	iter	time	iter	time
fxm3-6	24	3.7	9	1.4	11	3.9
fxm3-16	62	46.2	21	18.7	19	37.0
pltexpA4-6	68	30.3	-	-	40	23.0
swing8-4	35	69.8	66	114.9	14	72.9
mmix-50	31	11.5	25	8.7	10	18.6
mmix-100	35	27.3	48	29.1	13	8.3
watson10-64	58	11.8	89	15.8	6	9.2
watson10-128	74	26.1	106	34.2	15	20.8
watson10-256	85	51.6	-	-	48	56.7

Conclusions & Outlook

Conclusions:

- IPM **can** be warmstarted
- Can exploit stochastic programming structure to significantly speed-up solution.
- Decomposition based crashstart/Multilevel crashstart shows potential for further time-savings.

Future Work:

- Warmstart scenario subproblems
- Dynamic Scenario Generation ($\rightarrow$ AC SCOPF Problems)
- Towards “proper” multilevel scheme for stochastic programming

Bibliography

- M. Colombo, J. Gondzio, A. Grothey: *A Warm-Start Approach for Large-Scale Stochastic Linear Programs*, Mathematical Programming 127/2 (2011), pp371-397.
- M. Colombo, A. Grothey: *A decomposition-based warm-start method for stochastic programming*, Technical Report MS-09-008, School of Mathematics, The University of Edinburgh, June 2009.
- M. Colombo, A. Grothey: *A multi-step interior point warm-start approach for large-scale stochastic linear programming*, Technical Report MS-09-007, School of Mathematics, The University of Edinburgh, June 2009, revised Dec 2010.

`http://www.maths.ed.ac.uk/ERGO/preprints.html`