Combining Progressive Hedging with a Frank-Wolfe Method to Compute Lagrangian Dual Bounds in Stochastic Mixed-Integer Programming

JIM LUEDTKE

Collaborators: Natasha Boland, Jeffrey Christansen, Brian Dandurand, Andrew Eberhard, Jeff Linderoth, Fabricio Oliveira

Georgia Tech, RMIT, UW-Madison

DIMACS Workshop on ADMM and Proximal Splitting Methods in Optimization

June 11, 2018
Two-Stage SMIP with Recourse

Stochastic Mixed Integer Program (SMIP)

\[
\begin{align*}
\min & \quad c^\top x + \mathbb{E}[Q(x, \xi)] \\
\text{s.t.} & \quad Ax \geq b \\
& \quad x \in \mathbb{R}_{+}^{n_1} \times \mathbb{Z}_{+}^{p_1}
\end{align*}
\]

where \( \xi = (q, h, T, W) \) and

\[
Q(x, \xi) = \min \quad q^\top y \\
\text{s.t.} & \quad Wy = h - Tx \\
& \quad y \in \mathbb{R}_{+}^{n_2} \times \mathbb{Z}_{+}^{p_2}
\]

- \( x \): first-stage decision variables
- \( y \): second-stage decision variables
Stochastic vehicle routing

- Random customer demands
- First-stage decisions: planned vehicle routes (binary)
- Second-stage decisions: recourse actions when capacity violated (binary or continuous)

Stochastic unit commitment

- Random electricity loads and wind/solar production
- First-stage decisions: units to commit and when (binary)
- Second-stage decisions: production amounts, line switching (binary or continuous)
SMIP $\equiv$ MIP

- Assume finite set of scenarios: $s \in S$

### Extensive form (deterministic equivalent) of SMIP

$$\min \ c^\top x + \frac{1}{|S|} \sum_{s \in S} q_s^\top y_s$$

s.t. $Ax \geq b$
$$T_s x + W_s y_s = h_s \quad \forall s \in S$$

$x \in \mathbb{R}^{n_1}_+ \times \mathbb{Z}^{p_1}_+$

$y_s \in \mathbb{R}^{n_2}_+ \times \mathbb{Z}^{p_2}_+, \quad \forall s \in S$

- Goal: Decompose this large MIP into smaller subproblems
Variable Splitting/Dual Decomposition \((\text{Carøe and Schultz, 1999})\)

- Create copies of the first-stage decision variables for each scenario: \(x_1, x_2, \ldots x_S\), and a “master” copy \(z\).
- Add constraints \(x_s = z \ \forall s \in S\).
- SMIP is reformulated to

\[
\min_{x,y,z} \left\{ \frac{1}{|S|} \sum_{s \in S} (c^\top x_s + q_s^\top y_s) : (x_s, y_s) \in K_s, x_s = z \ \forall s \in S \right\}
\]

where for \(s \in S\)

\[
K_s := \{(x, y) : Ax \geq b, T_s x + W_s y = h_s, x \in \mathbb{R}^{n_1}_+ \times \mathbb{Z}^{p_1}_+, y \in \mathbb{R}^{n_2}_+ \times \mathbb{Z}^{p_2}_+\}.
\]

- Applying Lagrangian relaxation to the constraints \(x_s = z\) yields the Lagrangian dual function

\[
\mathcal{L}(\lambda) := \min_{x,y,z} \left\{ \frac{1}{|S|} \sum_{s \in S} \left[ (c^\top x_s + q_s^\top y_s) + \lambda_s^\top (x_s - z) \right] : (x_s, y_s) \in K_s, \ \forall s \in S \right\}
\]
Lagrangian - Decomposes

- **Bonus:** The evaluation of $\mathcal{L}(\lambda)$ decomposes by scenario:

$$\mathcal{L}(\lambda) = \frac{1}{|S|} \sum_{s \in S} \mathcal{L}_s(\lambda_s) =$$

$$\frac{1}{|S|} \sum_{s \in S} \min_{x_s, y_s, z, \lambda_s} \{L_s(x_s, y_s, z, \lambda_s) : (x_s, y_s) \in K_s\}$$

where

$$L_s(x_s, y_s, z, \lambda_s) := c^T x_s + q_s^T y_s + \lambda_s^T (x_s - z).$$

- Since $z$ is unconstrained in definition of $\mathcal{L}(\lambda)$, in order for $\mathcal{L}(\lambda) > -\infty$, we require “dual feasibility”: $\sum_{s \in S} \lambda_s = 0$. 

Solving the Lagrangian dual

\[ \phi^{LD} := \max \left\{ \mathcal{L}(\lambda) : \sum_{s \in S} \lambda_s = 0 \right\} \]

- \( \lambda \): High-dimensional \( (S \times (n_1 + p_1)) \)
- \( \mathcal{L}(\lambda) \): Non-smooth, concave function of \( \lambda \)
- At points \( \lambda \) for which \( \sum_{s \in S} \lambda_s = 0 \), subgradients are of the form
  \[ \left[ x_s - z \right]_{s \in S} \]
  where \( x_s \) optimizes \( \mathcal{L}_s(\lambda_s) \)
Strength of Lagrangian dual

**Theorem**

The Lagrangian dual bound satisfies

\[
\phi^{LD} = \min \left\{ c^\top x + \frac{1}{|S|} \sum_{s \in S} q_s^\top y_s : (x, y_s) \in \text{conv}(K_s), \forall s \in S \right\}
\]

where for \( s \in S \)

\[
K_s := \{(x, y) : Ax \geq b, \ T_s x + W_s y = h_s \\
x \in \mathbb{R}^{n_1}_+ \times \mathbb{Z}^{p_1}_+, \ y \in \mathbb{R}^{n_2}_+ \times \mathbb{Z}^{p_2}_+ \}
\]

- In general \( \phi^{LD} < z^{SMIP} \)
- But \( \phi^{LD} \geq z^{SLP} \) (the usual LP relaxation)
- In many test instances, \( \phi^{LD} \) is very close to \( z^{SMIP} \)
**Progressive Hedging** (Rockafellar and Wets, 1991)

- Elegant (primal-dual) algorithm for solving **convex** stochastic programs
- Application of alternating direction method of multipliers (ADMM)

For $k = 1, 2, \ldots$
- Solve augmented (convex!) scenario problems and obtain solution $x^k_s$, $s \in S$:

$$
\mathcal{L}_s^\rho(\lambda_s; z^{k-1}) := \min (c + \lambda_s)^\top x + q_s^\top y + \left(\frac{\rho}{2}\right)\|x - z^{k-1}\|^2_2
$$

s.t. $Ax \geq b, T_s x + W_s y = h_s$

$$
x \in \mathbb{R}_{+}^{n_1}, y \in \mathbb{R}_{+}^{n_2}
$$

- Calculate $z^k = \frac{1}{|S|} \sum_s x^k_s$
- Update $\lambda_s$, $s \in S$ ($\rho$ is a fixed step size)

$$
\lambda_{s}^{k+1} \leftarrow \lambda_{s}^{k} + \rho (x_{s}^{k} - z^{k})
$$
MIQP-Based Progressive hedging for SMIP?

- What if we just solve quadratic MIP subproblems?

\[
D^\rho_s(\lambda_s; z^{k-1}) := \min \ (c + \lambda_s)^\top x + q_s^\top y + (\rho/2)\|x - z^{k-1}\|_2^2
\]

\[
\text{s.t. } (x, y) \in K_s
\]

where (recall that) \( K_s \) has integer restrictions:

\[
K_s = \{(x, y) : Ax \geq b, T_s x + W_s y = h_s \}
\]

\[
x \in \mathbb{R}^{n_1} \times \mathbb{Z}^{p_1}, \ y \in \mathbb{R}^{n_2} \times \mathbb{Z}^{p_2}
\]

- Basis of effective heuristic for primal feasible solutions: (Watson et al., 2010)
- Can take \( \lambda^k \) from algorithm, evaluate \( \mathcal{L}(\lambda^k) \) and get a lower bound (Guo et al., 2015)
Solving Lagrangian Dual

Subgradient Method

- Solve $|S|$ MILPs, take a step.
- Slow convergence to $\phi^{LD}$
- Step size requires significant tuning

MIQP-Based Progressive Hedging

- Solve $|S|$ MIQPs, take a step.
- Faster, but only “works” when $K_s$ is convex—May not converge to $\phi^{LD}$.
- Can get valid lower bound by also solving $|S|$ MILPs at a step.

Goal: Version of Progressive Hedging/ADMM for SMIP that converges to $\phi^{LD}$
Progressive Hedging for SMIP?

- Suppose we could solve this subproblem:

\[
\min \{(c + \lambda_s)^\top x + q_s^\top y + (\rho/2)\|x - z^{k-1}\|_2^2 : (x, y) \in \text{conv}(K_s)\}
\]

- Problem is convex \(\Rightarrow\) Progressive hedging works!
  - Dual solution \(\lambda\) converges to optimum of Lagrangian dual
  - Challenge: \(\text{conv}(K_s)\) not known explicitly

**Idea: Use Frank-Wolfe/Simplicial Decomposition**

- Solve a sequence of subproblems with *linearized* objective
  - With linear objective, optimizing over \(\text{conv}(K_s)\) same as over \(K_s\)
  \(\Rightarrow\) can solve with MILP solver
Frank-Wolfe/Simplicial Decomposition

Key subproblem:

$$\min\{ (c + \lambda_s)^\top x + q_s^\top y + (\rho/2)\|x - z^{k-1}\|_2^2 : (x, y) \in \text{conv}(K_s) \}$$

Simplicial Decomposition (improvement on Frank/Wolfe)

- Initialize $V_s^0 = \emptyset$, $x_s^0 = z^{k-1}$,
- For $t = 1, 2, \ldots$
  1. Solve MILP:
     $$(\hat{x}^t, \hat{y}^t) \in \arg\min_{x, y} \{(c + \lambda_s + \rho(x_s^{t-1} - z^{k-1}))^\top x + q_s^\top y : (x, y) \in K_s \}$$
  2. $V_s^t \leftarrow V_s^{t-1} \cup \{(\hat{x}^t, \hat{y}^t)\}$
  3. Solve convex QP:
     $$\min\{ (c + \lambda_s)^\top x + q_s^\top y + (\rho/2)\|x - z^{k-1}\|_2^2 : (x, y) \in \text{conv}(V_s^t) \}$$

Solutions to QP subproblem converge finitely to optimal solution of key subproblem.
Limitations/Improvements

Subproblem must be solved for all scenarios at every PH iteration
  • May require many Frank-Wolfe/simplicial decomposition iterations in every PH iteration
  • Improvement: Initialize sets $V_s^0$, $s \in S$ with final set in previous PH iteration
    • May lead to faster convergence of simplicial decomposition method in each PH iteration
    • At expense of growing list of points (but MILP time still likely to dominate)

Even better: We get away with doing just one simplicial decomposition iteration per PH iteration!
Frank-Wolfe/SD + Progressive Hedging

- Given: $V_s \subset \text{conv}(K_s)$, $z = \frac{1}{|S|} \sum_{s \in S} x_s^0$, $\lambda_s^0$ with $\sum_{s \in S} \lambda_s^0 = 0$

For $k = 1, 2, \ldots$

- For $s \in S$
  1. Evaluate $\mathcal{L}_s(\lambda_s^k) := \min_{x,y} \{(c + \lambda_s^k)^\top x + q_s^\top y : (x, y) \in K_s\}$
  2. Update $V_s$ with an optimal solution $(\hat{x}_s, \hat{y}_s)$.
  3. Solve convex QP: let $(x_s^k, y_s^k)$ be optimal solution to
     $$\min_{x,y} \{(c + \lambda_s)^\top x + q_s^\top y + (\rho/2)\|x - z^{k-1}\|_2^2 : (x, y) \in \text{conv}(V_s^k)\}$$
  4. $\phi^k \leftarrow \frac{1}{|S|} \sum_{s \in S} \mathcal{L}_s(\lambda_s^k)$
  5. $z^k \leftarrow \frac{1}{|S|} \sum_{s \in S} x_s^k$
  6. $\lambda_s^{k+1} \leftarrow \lambda_s^k + \rho(x_s^k - z_s^k)$, for $s \in S$
Convergence Proof

### Inner Approximation Expansion Lemma

If \((x^k_s, y^k_s)\) is not an optimal solution to

\[
\min_{x,y} \left\{ (c + \lambda_s)^T x + q_s^T y + (\rho/2)\|x - z^{k-1}\|_2^2 : (x, y) \in \text{conv}(K_s) \right\},
\]

then optimal solution of MILP at iteration \(k + 1\) is not in \(V_s\), so \(V_s\) grows.

- Proof follows from geometry of optimality conditions.
- Finitely many extreme points of \(\text{conv}(K_s)\) ⇒ Beyond some iteration, \((x^k_s, y^k_s)\) are optimal to true PH QP subproblem.
- At that point, convergence (and rate) follows from PH convergence theory.
Algorithm Properties

It Works Theorem

If initialized “appropriately”, \( \lim_{k \to \infty} \phi_k = \phi^{LD} \).

It Gives Lower Bounds Theorem

If \( \sum_{s \in S} \lambda^0_s = 0 \), then at each iteration \( \phi^k = \mathcal{L}(\lambda^k) \), and \( \sum_{s \in S} \lambda^k_s = 0 \).
So we have a valid lower bound at each iteration.
Numerical Experience

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>New “Frank-Wolfe/SDM + PH” Method</td>
</tr>
<tr>
<td>2</td>
<td>“MIQP”-Based PH Method</td>
</tr>
</tbody>
</table>

- Compare for various penalty parameters $\rho$
- Compare on three SMIP test problems
Stochastic server location problem from Ntaiimo and Sen (2005) (100 scenarios, 10 first-stage binary, 500 second-stage binary)
Stochastic capacitated facility location from Bodur et al. (2017) (250 scenarios, 25 first-stage binary, 1250 second-stage continuous)
Dynamic capacity allocation problem from Ahmed and Garcia (2003) (500 scenarios, 6 first-stage continuous, 27 second-stage integer)
Observations

Frank-Wolfe/SDM + PH
- Consistent convergence
- Speed depends on choice of $\rho$ (typically relatively large)

MIQP-PH
- Larger $\rho$: Converges to suboptimal value
- Smaller $\rho$: Slow convergence
Enhancements/Open Question

Simple computational enhancements
- Do not solve all $|S|$ MILP at every iteration (use heuristics to find “important” scenarios)
  - Just don’t update $V_s$ set for skipped scenarios
- When evaluate upper bound: Add solutions to $V_s$ sets
- Synchronous parallel: solve (up to) $|S|$ separate MILP and convex QP subproblems per iteration
- Semi-asynchronous parallel: Let MILP subproblems solve asynchronously and update $V_s$ sets when solved
  - But wait for convex QP subproblems to solve before taking a PH step

Open Question
- Convergence rate?

Thank you!

jim.luedtke@wisc.edu


