A Supplementary Material for Interior Point Solving for LP-based prediction+optimisation

A.1 Solution of Newton Equation System of Eq. (11)

Here we discuss how we solve an equation system of Eq (11), for more detail you can refer to [4]. Consider the following system with a generic R.H.S-

\[
\begin{bmatrix}
-X^{-1}T & A^\top & -c \\
-A & 0 & -b \\
-c^\top & b^\top & \kappa/\tau
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix}
=
\begin{bmatrix}
r_1 \\
r_2 \\
r_3
\end{bmatrix}
\]  

(13)

If we write:

\[
W = \begin{bmatrix}
-X^{-1}T & A^\top \\
A & 0
\end{bmatrix}
\]  

(14)

then, observe \( W \) is nonsingular provided \( A \) is full row rank. So it is possible to solve the following system of equations-

\[
W \begin{bmatrix}
p \\
q
\end{bmatrix}
=
\begin{bmatrix}
c \\
b
\end{bmatrix}
\]  

(15)

Once we find \( p, q, u, v \) finally we compute \( x_3 \) as:

\[
x_3 = \frac{r_3 + u^\top c - v^\top b}{-c^\top p + b^\top q + \frac{\kappa}{\tau}};
\]  

(16)

And finally

\[
\begin{align*}
x_1 &= u + px_3 \\
x_2 &= v + qx_3
\end{align*}
\]  

(17) \hspace{1cm} (18)

To solve equation of the form

\[
W \begin{bmatrix}
u \\
v
\end{bmatrix}
=
\begin{bmatrix}
-X^{-1}T & A^\top \\
A & 0
\end{bmatrix}
\begin{bmatrix}
u \\
v
\end{bmatrix}
=
\begin{bmatrix}
r_1 \\
r_2
\end{bmatrix}
\]  

Notice we can reduce it to

\[
Mv = AT^{-1}Xr_1 + r_2 \quad \text{(where } M = AT^{-1}XA^\top\text{). As } M \text{ is positive definite for a full row-rank } A, \text{ we obtain } v \text{ by Cholesky decomposition and finally } u = T^{-1}X(A^\top v - r_1).\]

A.2 Differentiation of HSD formulation in Eq. (9)

We differentiate Eq. (9) with respect to \( c \):

\[
\begin{align*}
\frac{\partial (Ax)}{\partial c} - \frac{\partial (b\tau)}{\partial c} &= 0 \\
\frac{\partial (A^\top y)}{\partial c} + \frac{\partial t}{\partial c} - \frac{\partial (cr)}{\partial c} &= 0 \\
-\frac{\partial (c^\top x)}{\partial c} + \frac{\partial (b^\top y)}{\partial c} - \frac{\partial \kappa}{\partial c} &= 0 \\
\frac{\partial t}{\partial c} &= \frac{\partial (\lambda X^{-1}e)}{\partial c} \\
\frac{\partial \kappa}{\partial c} &= \frac{\partial (\frac{1}{\tau})}{\partial c}
\end{align*}
\]  

(19)
Applying the product rule we can further rewrite this into:

\[ A \frac{\partial x}{\partial c} - b \frac{\partial \tau}{\partial c} = 0 \]

\[ A^\top \frac{\partial y}{\partial c} + \frac{\partial t}{\partial c} - (c \frac{\partial \tau}{\partial c} + \tau I) = 0 \]

\[-(c^\top \frac{\partial x}{\partial c} + x^\top) + b^\top \frac{\partial y}{\partial c} - \frac{\partial \kappa}{\partial c} = 0 \]

(20)

Using \( t = \lambda X^{-1} e \leftrightarrow \lambda e = X^T e \) we can rewrite the fourth equation to \( \frac{\partial t}{\partial c} = -X^{-1} T \frac{\partial x}{\partial c} \). Similarly we use \( \kappa = \lambda \tau \leftrightarrow \lambda = \kappa \times \tau \) and rewrite the fifth equation to \( \frac{\partial \kappa}{\partial c} = -\frac{\lambda \tau}{\tau^2} \frac{\partial \tau}{\partial c} \). Substituting these into the first three we obtain:

\[ A \frac{\partial x}{\partial c} - b \frac{\partial \tau}{\partial c} = 0 \]

\[ A^\top \frac{\partial y}{\partial c} - X^{-1} T \frac{\partial x}{\partial c} - c \frac{\partial \tau}{\partial c} - \tau I = 0 \]

\[-c^\top \frac{\partial x}{\partial c} - x^\top + b^\top \frac{\partial y}{\partial c} + \frac{\kappa \partial \tau}{\tau} \frac{\partial \tau}{\partial c} = 0 \]

(21)

This formulation is written in matrix form in Eq. (12).

### A.3 LP formulation of the Experiments

#### A.3.1 Details on Knapsack formulation of real estate investments

In this problem, \( H \) is the set of housings under consideration. For each housing \( h \), \( c_h \) is the known construction cost of the housing and \( p_h \) is the (predicted) sales price. With the limited budget \( B \), the constraint is

\[ \sum_{h \in H} c_h x_h = B, \quad x_h \in \{0, 1\} \]

where \( x_h \) is 1 only if the investor invests in housing \( h \). The objective function is to maximize the following profit function

\[ \max_{x_h} \sum_{h \in H} p_h x_h \]

#### A.3.2 Details on Energy-cost aware scheduling

In this problem, \( J \) is the set of tasks to be scheduled on \( M \) number of machines maintaining resource requirement of \( R \) resources. The tasks must be scheduled over \( T \) set of equal length time periods. Each task \( j \) is specified by its duration \( d_j \), earliest start time \( e_j \), latest end time \( l_j \), power usage \( p_j \), resource usage \( u_r \) is the resource usage of task \( j \) for resource \( r \) and \( c_{mr} \) is the capacity of machine \( m \) for resource \( r \). Let \( x_{jmt} \) be a binary variable which possesses 1 only if task \( j \) starts at time \( t \) on machine \( m \). The first constraint ensures each task is scheduled and only once.

\[ \sum_{m \in M} x_{jmt} = 1, \quad \forall j \in J \]

The next constraints ensure the task scheduling abides by earliest start time and latest end time constraints.

\[ x_{jmt} = 0 \quad \forall j \in J, \forall m \in M, \forall t < e_j \]

\[ x_{jmt} = 0 \quad \forall j \in J, \forall m \in M, \forall t + d_j > l_j \]

Finally the resource requirement constraint:

\[ \sum_{j \in J} \sum_{t-d_j < t' \leq t} x_{jmt' u_{jr}} \leq c_{mr}, \quad \forall m \in M \forall r \in R \forall t \in T \]
If $c_t$ is the (predicted) energy price at time $t$, the objective is to minimize the energy cost of running all tasks, given by:

$$\min_{x_{jmt}} \sum_{j \in J} \sum_{m \in M} \sum_{t \in T} x_{jmt} \left( \sum_{t' < t + d_j} p_j c_{t'} \right)$$

### A.3.3 Details on Shortest path problem

In this problem, we consider a directed graph specified by node-set $N$ and edge-set $E$. Let $A$ be the $|N| \times |E|$ incidence matrix, where for an edge $e$ that goes from $n_1$ to $n_2$, the $(n_1, e)$th entry is 1 and $(n_2, e)$th entry is -1 and the rest of entries in column $e$ are 0. In order to, traverse from source node $s$ to destination node $d$, the following constraint must be satisfied:

$$Ax = b$$

where $x$ is $|E|$ dimensional binary vector whose entries would be 1 only if corresponding edge is selected for traversal and $b$ is $|N|$ dimensional vector whose $s$th entry is 1 and $d$th entry is -1; and rest are 0. With respect to the (predicted) cost vector $c \in \mathbb{R}^{|E|}$, the objective is to minimize the cost

$$\min_{x} c^\top x$$

### A.4 Additional Knapsack Experiments

This knapsack experiment is taken from [18], where the knapsack instances are created from the energy price dataset [15]. The 48 half-hour slots are considered as 48 knapsack items and a random cost is assigned to each slot. The energy price of a slot is considered as the profit-value and the objective is to select a set of slots which maximizes the profit ensuring the total cost of the selected slots remains below a fixed budget. We also added the approach of Blackbox [25], which also deals with a combinatorial optimization problem with a linear objective.

<table>
<thead>
<tr>
<th>Budget</th>
<th>Two-stage QPTL SPO Blackbox IntOpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>1042 (3) 579 (3) 624 (3) 533 (40) 570 (58)</td>
</tr>
<tr>
<td>120</td>
<td>1098 (5) 380 (2) 425 (4) 383 (14) 406 (71)</td>
</tr>
</tbody>
</table>

### A.5 Hyperparameters of the experiments

#### A.5.1 Knapsack formulation of real estate investments

<table>
<thead>
<tr>
<th>Model</th>
<th>Hyperparameters*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two-stage</td>
<td>• optimizer: optim.Adam; learning rate: $10^{-3}$</td>
</tr>
<tr>
<td>SPO</td>
<td>• optimizer: optim.Adam; learning rate: $10^{-3}$</td>
</tr>
<tr>
<td>QPTL</td>
<td>• optimizer: optim.Adam; learning rate: $10^{-3}$; $\tau$ (quadratic regularizer): $10^{-3}$</td>
</tr>
<tr>
<td>IntOpt</td>
<td>• optimizer: optim.Adam; learning rate: $10^{-2}$; $\lambda$-cut-off: $10^{-3}$; damping factor $\alpha$: $10^{-3}$</td>
</tr>
</tbody>
</table>

* for all experiments embedding size: 7 number of layers:1,hidden layer size: 2

#### A.5.2 Energy-cost aware scheduling

<table>
<thead>
<tr>
<th>Model</th>
<th>Hyperparameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two-stage</td>
<td>• optimizer: optim.SGD; learning rate: 0.1</td>
</tr>
<tr>
<td>SPO</td>
<td>• optimizer: optim.Adam; learning rate: 0.7</td>
</tr>
<tr>
<td>QPTL</td>
<td>• optimizer: optim.Adam; learning rate: 0.1; $\tau$ (quadratic regularizer): $10^{-3}$</td>
</tr>
<tr>
<td>IntOpt</td>
<td>• optimizer: optim.Adam; learning rate: 0.7; $\lambda$-cut-off: 0.1; damping factor $\alpha$: $10^{-6}$</td>
</tr>
</tbody>
</table>

---

2For more details refer to https://github.com/JayMan91/NeurIPSIntopt
A.5.3 Shortest path problem

<table>
<thead>
<tr>
<th>Model</th>
<th>1-layer</th>
<th>2-layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two-stage</td>
<td>optimizer: optim.Adam; learning rate: 0.01</td>
<td>optimizer: optim.Adam; learning rate: $10^{-4}$</td>
</tr>
<tr>
<td>SPO</td>
<td>optimizer: optim.Adam; learning rate: $10^{-3}$</td>
<td>optimizer: optim.Adam; learning rate: $10^{-3}$</td>
</tr>
<tr>
<td>QPTL</td>
<td>optimizer: optim.Adam; learning rate: 0.7; $\tau$ (quadratic regularizer): $10^{-1}$</td>
<td>optimizer: optim.Adam; learning rate: 0.7; $\tau$ (quadratic regularizer): $10^{-1}$</td>
</tr>
<tr>
<td>IntOpt</td>
<td>optimizer: optim.Adam; learning rate: 0.7; $\lambda$-cut-off: 0.1; damping factor $\alpha$: $10^{-2}$</td>
<td>optimizer: optim.Adam; learning rate: 0.7; $\lambda$-cut-off: 0.1; damping factor $\alpha$: $10^{-2}$</td>
</tr>
</tbody>
</table>

* for all experiments hidden layer size: 100

A.6 Learning Curves

(a) Energy-cost aware scheduling

(b) Shortest path problem

Figure 2: IntOpt Learning Curve