Predict+Optimize for Packing and Covering LPs with Unknown Parameters in Constraints

Paper ID: 6846

A Proofs

A.1 Proofs for Packing LPs

Lemma 1 *Let $x^*(\hat{\theta})$ denote the estimated optimal solution of the packing LP shown in (4), $x^*_{corr}(\hat{\theta},\theta) = \lambda x^*(\hat{\theta})$ be the correction function shown in (5). Suppose that at the optimal λ of (5), the i^{th} inequality constraint G_i is tight, namely $G_i^\top(\lambda x^*(\hat{\theta})) = h_i$. Then, we have*

$$\frac{\partial \lambda}{\partial x^*(\hat{\theta})} = -\frac{\lambda}{G_i^\top x^*(\hat{\theta})} G_i^\top.$$

As a corollary, we have

$$\frac{\partial x^*_{corr}(\hat{\theta},\theta)}{\partial x^*(\hat{\theta})} = \frac{-\lambda}{G_i^\top x^*(\hat{\theta})} x^*(\hat{\theta}) G_i^\top + \lambda I.$$

Proof. Since the i^{th} inequality constraint G_i is tight, we have:

$$\lambda \sum_{j=1}^{n} G_{ij} x^*(\hat{\theta})_j = h_i \tag{1}$$

The implicit differentiation of Equation 2 with respect to $x^*(\hat{\theta})$ is:

$$\frac{\partial}{\partial x^*(\hat{\theta})}(\lambda \sum_{j=1}^{n} G_{ij} x^*(\hat{\theta})_j) = \frac{\partial h_i}{\partial x^*(\hat{\theta})}$$

Since $x^*(\hat{\theta})$ is a vector, differentiation on the l^{th} variable is:

$$\frac{\partial}{\partial x^*(\hat{\theta})_l}(\lambda \sum_{j=1}^{n} G_{ij} x^*(\hat{\theta})_j) = \frac{\partial h_i}{\partial x^*(\hat{\theta})_l}$$

where

$$\frac{\partial}{\partial x^*(\hat{\theta})_l}(\lambda \sum_{j=1}^{n} G_{ij} x^*(\hat{\theta})_j) = \frac{\partial \lambda}{\partial x^*(\hat{\theta})_l} G_i^\top x^*(\hat{\theta}) + \lambda G_{il}$$

Since $\frac{\partial h_i}{\partial x^*(\hat{\theta})_l} = 0$, we can obtain:

$$\frac{\partial \lambda}{\partial x^*(\hat{\theta})} = -\frac{\lambda}{G_i^\top x^*(\hat{\theta})} G_i^\top.$$

Since $\frac{\partial x^*_{corr}(\hat{\theta},\theta)}{\partial \lambda} = x^*(\hat{\theta})$, $\frac{\partial x^*_{corr}(\hat{\theta},\theta)}{\partial x^*(\hat{\theta})}\Big|_\lambda = \lambda I$, the gradient of the corrected optimal solution with respect to the predicted optimal solution is:

$$\begin{aligned}
\frac{\partial x^*_{corr}(\hat{\theta},\theta)}{\partial x^*(\hat{\theta})} &= \frac{\partial x^*_{corr}(\hat{\theta},\theta)}{\partial \lambda}\frac{\partial \lambda}{\partial x^*(\hat{\theta})} + \frac{\partial x^*_{corr}(\hat{\theta},\theta)}{\partial x^*(\hat{\theta})}\Big|_\lambda \\
&= \frac{-\lambda}{G_i^\top x^*(\hat{\theta})} x^*(\hat{\theta}) G_i^\top + \lambda I.
\end{aligned}$$

□

Lemma 2 *Consider the LP relaxation (8), defining x^* as a function of c, G and h. Then, under this definition of x^*,*

$$\frac{\partial x^*}{\partial h} = -f_{xx}(x^*)^{-1} f_{hx}(x^*)$$

where f_{xx} denotes the matrix of second derivatives of f with respect to different coordinates of x, and similarly for other subscripts, and explicitly:

$$f_{x_k x_j}(x) = \begin{cases} -\mu x_j^{-2} - \mu \sum_{i=1}^p G_{ij}^2/(h_i - G_i^\top x)^2, & j = k \\ -\mu \sum_{i=1}^p G_{ij} G_{ik}/(h_i - G_i^\top x)^2, & j \neq k \end{cases}$$

and

$$f_{h_\ell x_j}(x) = \mu G_{\ell j}/(h_\ell - G_\ell^\top x)^2$$

Proof. Since $x^* = \arg\max_x f(x, c, G, h)$ is an optimum, $f_x(x^*) = \frac{\partial f(x)}{\partial x}\Big|_{x=x^*} = 0$. Thus,

$$\frac{\partial}{\partial h} f_x(x^*) = 0$$

By the chain rule,

$$\frac{\partial}{\partial h} f_x(x^*) = f_{hx}(x^*) + f_{xx}(x^*) \frac{\partial x^*}{\partial h}$$

Rearranging the aboved equation, we can obtain:

$$\frac{\partial x^*}{\partial h} = -f_{xx}(x^*)^{-1} f_{hx}(x^*)$$

where

$$f_{x_k x_j}(x) = \begin{cases} -\mu x_j^{-2} - \mu \sum_{i=1}^p G_{ij}^2/(h_i - G_i^\top x)^2, & j = k \\ -\mu \sum_{i=1}^p G_{ij} G_{ik}/(h_i - G_i^\top x)^2, & j \neq k \end{cases}$$

and

$$f_{h_\ell x_j}(x) = \mu G_{\ell j}/(h_\ell - G_\ell^\top x)^2$$

□

Lemma 3 *Consider the LP relaxation (8), defining x^* as a function of c, G and h. Then, under this definition of x^*,*

$$\frac{\partial x^*}{\partial G} = -f_{xx}(x^*)^{-1} f_{Gx}(x^*)$$

where f_{xx} denotes the matrix of second derivatives of f with respect to different coordinates of x, and similarly for other subscripts, and explicitly:

$$f_{x_k x_j}(x) = \begin{cases} -\mu x_j^{-2} - \mu \sum_{i=1}^p G_{ij}^2/(h_i - G_i^\top x)^2, & j = k \\ -\mu \sum_{i=1}^p G_{ij} G_{ik}/(h_i - G_i^\top x)^2, & j \neq k \end{cases}$$

and

$$f_{G_{\ell q} x_j}(x) = \begin{cases} -\mu G_{\ell j} x_q/(h_\ell - G_\ell^\top x)^2 - \mu/(h_\ell - G_\ell^\top x), & q = j \\ -\mu G_{\ell j} x_q/(h_\ell - G_\ell^\top x)^2, & q \neq j. \end{cases}$$

Proof. Since $x^* = \arg\max_x f(x, c, G, h)$ is an optimum, $f_x(x^*) = \frac{\partial f(x)}{\partial x}\Big|_{x=x^*} = 0$. Thus,

$$\frac{\partial}{\partial G} f_x(x^*) = 0$$

By the chain rule,

$$\frac{\partial}{\partial G} f_x(x^*) = f_{Gx}(x^*) + f_{xx}(x^*) \frac{\partial x^*}{\partial G}$$

Rearranging the aboved equation, we can obtain:

$$\frac{\partial x^*}{\partial G} = -f_{xx}(x^*)^{-1} f_{Gx}(x^*)$$

where

$$f_{x_k x_j}(x) = \begin{cases} -\mu x_j^{-2} - \mu \sum_{i=1}^p G_{ij}^2/(h_i - G_i^\top x)^2, & j = k \\ -\mu \sum_{i=1}^p G_{ij} G_{ik}/(h_i - G_i^\top x)^2, & j \neq k \end{cases}$$

and

$$f_{G_{\ell q} x_j}(x) = \begin{cases} -\mu G_{\ell j} x_q/(h_\ell - G_\ell^\top x)^2 - \mu/(h_\ell - G_\ell^\top x), & q = j \\ -\mu G_{\ell j} x_q/(h_\ell - G_\ell^\top x)^2, & q \neq j. \end{cases}$$

□

A.2 Proofs for Covering LPs

Lemma 4 *Let $x^*(\hat{\theta})$ denote the estimated optimal solution of the covering LP shown in (9), $x^*_{corr}(\hat{\theta}, \theta) = \lambda x^*(\hat{\theta})$ be the correction function shown in (10). Suppose that at the optimal λ of (10), the i^{th} inequality constraint G_i is tight, namely $G_i^\top(\lambda x^*(\hat{\theta})) = h_i$. Then, we have*

$$\frac{\partial \lambda}{\partial x^*(\hat{\theta})} = -\frac{\lambda}{G_i^\top x^*(\hat{\theta})} G_i^\top.$$

As a corollary, we have

$$\frac{\partial x^*_{corr}(\hat{\theta}, \theta)}{\partial x^*(\hat{\theta})} = \frac{-\lambda}{G_i^\top x^*(\hat{\theta})} x^*(\hat{\theta}) G_i^\top + \lambda I.$$

Proof. Since the i^{th} inequality constraint G_i is tight, we have:

$$\lambda \sum_{j=1}^{n} G_{ij} x^*(\hat{\theta})_j = h_i \tag{2}$$

The implicit differentiation of Equation 2 with respect to $x^*(\hat{\theta})$ is:

$$\frac{\partial}{\partial x^*(\hat{\theta})}(\lambda \sum_{j=1}^{n} G_{ij} x^*(\hat{\theta})_j) = \frac{\partial h_i}{\partial x^*(\hat{\theta})}$$

Since $x^*(\hat{\theta})$ is a vector, differentiation on the l^{th} variable is:

$$\frac{\partial}{\partial x^*(\hat{\theta})_l}(\lambda \sum_{j=1}^{n} G_{ij} x^*(\hat{\theta})_j) = \frac{\partial h_i}{\partial x^*(\hat{\theta})_l}$$

where

$$\frac{\partial}{\partial x^*(\hat{\theta})_l}(\lambda \sum_{j=1}^{n} G_{ij} x^*(\hat{\theta})_j) = \frac{\partial \lambda}{\partial x^*(\hat{\theta})_l} G_i^\top x^*(\hat{\theta}) + \lambda G_{il}$$

Since $\frac{\partial h_i}{\partial x^*(\hat{\theta})_l} = 0$, we can obtain:

$$\frac{\partial \lambda}{\partial x^*(\hat{\theta})} = -\frac{\lambda}{G_i^\top x^*(\hat{\theta})} G_i^\top.$$

Since $\frac{\partial x^*_{corr}(\hat{\theta},\theta)}{\partial \lambda} = x^*(\hat{\theta})$, $\left.\frac{\partial x^*_{corr}(\hat{\theta},\theta)}{\partial x^*(\hat{\theta})}\right|_\lambda = \lambda I$, the gradient of the corrected optimal solution with respect to the predicted optimal solution is:

$$\begin{aligned}
\frac{\partial x^*_{corr}(\hat{\theta},\theta)}{\partial x^*(\hat{\theta})} &= \frac{\partial x^*_{corr}(\hat{\theta},\theta)}{\partial \lambda} \frac{\partial \lambda}{\partial x^*(\hat{\theta})} + \left.\frac{\partial x^*_{corr}(\hat{\theta},\theta)}{\partial x^*(\hat{\theta})}\right|_\lambda \\
&= -\frac{\lambda}{G_i^\top x^*(\hat{\theta})} x^*(\hat{\theta}) G_i^\top + \lambda I.
\end{aligned}$$

□

Lemma 5 *In the context of covering LP, consider the LP relaxation in the following form:*

$$x^* = \arg\min_x c^\top x - \mu \left[\sum_{i=1}^{d} \ln(x_i) - \sum_{i=1}^{p} \ln(G_i^\top x - h_i) \right] \tag{3}$$

Defining x^* as a function of c, G and h. Then, under this definition of x^*,

$$\frac{\partial x^*}{\partial h} = -f_{xx}(x^*)^{-1} f_{hx}(x^*)$$

where f_{xx} denotes the matrix of second derivatives of f with respect to different coordinates of x, and similarly for other subscripts, and explicitly:

$$f_{x_k x_j}(x) = \begin{cases} \mu x_j^{-2} + \mu \sum_{i=1}^{p} G_{ij}^2/(h_i - G_i^\top x)^2, & j = k \\ \mu \sum_{i=1}^{p} G_{ij} G_{ik}/(h_i - G_i^\top x)^2, & j \neq k \end{cases}$$

and

$$f_{h_\ell x_j}(x) = -\mu G_{\ell j}/(h_\ell - G_\ell^\top x)^2$$

Proof. Since $x^* = \arg\min_x f(x, c, G, h)$ is an optimum, $f_x(x^*) = \left.\frac{\partial f(x)}{\partial x}\right|_{x=x^*} = 0$. Thus,

$$\frac{\partial}{\partial h} f_x(x^*) = 0$$

By the chain rule,

$$\frac{\partial}{\partial h} f_x(x^*) = f_{hx}(x^*) + f_{xx}(x^*) \frac{\partial x^*}{\partial h}$$

Rearranging the aboved equation, we can obtain:

$$\frac{\partial x^*}{\partial h} = -f_{xx}(x^*)^{-1} f_{hx}(x^*)$$

where

$$f_{x_k x_j}(x) = \begin{cases} \mu x_j^{-2} + \mu \sum_{i=1}^{p} G_{ij}^2/(h_i - G_i^\top x)^2, & j = k \\ \mu \sum_{i=1}^{p} G_{ij} G_{ik}/(h_i - G_i^\top x)^2, & j \neq k \end{cases}$$

and

$$f_{h_\ell x_j}(x) = -\mu G_{\ell j}/(h_\ell - G_\ell^\top x)^2$$

□

Lemma 6 *In the context of covering LP, consider the LP relaxation in the following form:*

$$x^* = \arg\min_x c^\top x - \mu \left[\sum_{i=1}^{d} \ln(x_i) - \sum_{i=1}^{p} \ln(G_i^\top x - h_i) \right] \tag{4}$$

Defining x^* as a function of c, G and h. Then, under this definition of x^*,

$$\frac{\partial x^*}{\partial G} = -f_{xx}(x^*)^{-1} f_{Gx}(x^*)$$

where f_{xx} denotes the matrix of second derivatives of f with respect to different coordinates of x, and similarly for other subscripts, and explicitly:

$$f_{x_k x_j}(x) = \begin{cases} \mu x_j^{-2} + \mu \sum_{i=1}^{p} G_{ij}^2/(h_i - G_i^\top x)^2, & j = k \\ \mu \sum_{i=1}^{p} G_{ij} G_{ik}/(h_i - G_i^\top x)^2, & j \neq k \end{cases}$$

and

$$f_{G_{\ell q} x_j}(x) = \begin{cases} \mu G_{\ell j} x_q/(h_\ell - G_\ell^\top x)^2 + \mu/(h_\ell - G_\ell^\top x), & q = j \\ \mu G_{\ell j} x_q/(h_\ell - G_\ell^\top x)^2, & q \neq j. \end{cases}$$

Proof. Since $x^* = \arg\max_x f(x, c, G, h)$ is an optimum, $f_x(x^*) = \left.\frac{\partial f(x)}{\partial x}\right|_{x=x^*} = 0$. Thus,

$$\frac{\partial}{\partial G} f_x(x^*) = 0$$

By the chain rule,

$$\frac{\partial}{\partial G} f_x(x^*) = f_{Gx}(x^*) + f_{xx}(x^*) \frac{\partial x^*}{\partial G}$$

Rearranging the aboved equation, we can obtain:

$$\frac{\partial x^*}{\partial G} = -f_{xx}(x^*)^{-1} f_{Gx}(x^*)$$

where

$$f_{x_k x_j}(x) = \begin{cases} \mu x_j^{-2} + \mu \sum_{i=1}^{p} G_{ij}^2/(h_i - G_i^\top x)^2, & j = k \\ \mu \sum_{i=1}^{p} G_{ij} G_{ik}/(h_i - G_i^\top x)^2, & j \neq k \end{cases}$$

and

$$f_{G_{\ell q} x_j}(x) = \begin{cases} \mu G_{\ell j} x_q/(h_\ell - G_\ell^\top x)^2 + \mu/(h_\ell - G_\ell^\top x), & q = j \\ \mu G_{\ell j} x_q/(h_\ell - G_\ell^\top x)^2, & q \neq j. \end{cases}$$

□

B Hyperparameters of the Experiments

Here are the final hyperparameter choices in the three problems: the maximum flow transportation problem, the alloy production problem, and the fractional knapsack problem.

Model	Hyperaprameters
Proposed	optimizer: optim.Adam; learning rate: 10^{-5}; $\mu = 10^{-3}$; epochs=6
k-NN	k=5
RF	n_estimator=100
NN	optimizer: optim.Adam; learning rate: 10^{-3}; epochs=6

Table 1: Hyperparameters of the experiments on the maximum flow transportation problem.

Model	Hyperaprameters
Proposed	optimizer: optim.Adam; learning rate: 5^{-6}; $\mu = 10^{-3}$; epochs=8
k-NN	k=5
RF	n_estimator=100
NN	optimizer: optim.Adam; learning rate: 10^{-3}; epochs=8

Table 2: Hyperparameters of the experiments on the alloy production problem.

C Detailed Experimental Results

This section shows the detailed experimental results on our three benchmarks: the maximum flow transportation problem, the alloy production problem, and the fractional knapsack problem.

Model	Hyperaprameters
Proposed	optimizer: optim.Adam; learning rate: 10^{-7}; $\mu = 10^{-3}$; epochs=8
k-NN	k=5
RF	n_estimator=100
NN	optimizer: optim.Adam; learning rate: 10^{-3}; epochs=8

Table 3: Hyperparameters of the experiments on the fractional knapsack problem.

C.1 Maximum Flow Transportation Problem

This section shows the experiment results of the maximum flow transportation problem.

Tables 4 and 5 report the mean post-hoc regrets and standard deviations across 10 runs, and the mean square errors (MSE) and standard deviations across 10 runs for each approach on the maximum flow transportation problem with unknown capacities respectively. We showed these as box plots in the main paper, but here in the appendix we present also the numerical values.

Table 4 shows that the proposed method achieves the best performance in all cases. We also report the average True Optimal Values (TOV) in the last column of Table 4 for reference. The proposed method achieves 11.49% relative error on POLSKA, 16.23% relative error on USANet, and 10.28% relative error on GÉANT.

Table 5 shows the numerical values of the MSE of the various methods in the max-flow experiment. Unsurprisingly, ridge regression achieves the best performance in all of the cases since it is explicitly designed to learn in ℓ_2 error, while RF always achieves the second best performance. We mentioned in the main paper that the MSE for our method is drastically higher than the other methods, and we gave justification as to why it is related to the fact that the penalty factor is zero in this experiment. Here, we give a scatterplot (Figure 1) of the norm of the predicted parameters versus the true parameters, across all the methods.

As we can see in Figure 1, the predicted parameters values of the proposed method are several orders of magnitude higher than the true parameters values. The explanation of this phenomenon is given in the main paper.

PReg	Proposed	Ridge	k-NN	CART	RF	NN	TOV
POLSKA	**10.00±0.67**	11.20±0.73	14.39±0.83	16.65±1.06	12.30±0.90	12.18±1.08	88.66±1.10
USANet	**16.64±1.34**	19.52±1.16	22.89±1.58	24.15±1.51	22.27±1.34	18.62±1.23	96.22±1.38
GÉANT	**10.84±1.10**	12.47±1.14	15.13±1.08	17.01±1.59	12.52±1.19	12.05±1.13	98.71±1.98

Table 4: Mean post-hoc regrets and standard deviations for the maximum flow transportation problem.

MSE	Proposed	Ridge	k-NN	CART	RF	NN
POLSKA	1.45E+04±2.63E+04	**290.75±127.31**	363.13±120.51	474.00±145.07	309.94±123.44	324.38±132.49
USANet	1.76E+04±2.20E+04	**755.54±90.39**	913.79±91.48	1626.40±195.31	779.04±83.86	903.86±105.96
GÉANT	1.62E+04±2.58E+04	**700.35±72.66**	842.45±75.78	1484.84±203.11	704.96±76.64	828.18±95.18

Table 5: Mean square errors and standard deviations for the maximum flow transportation problem.

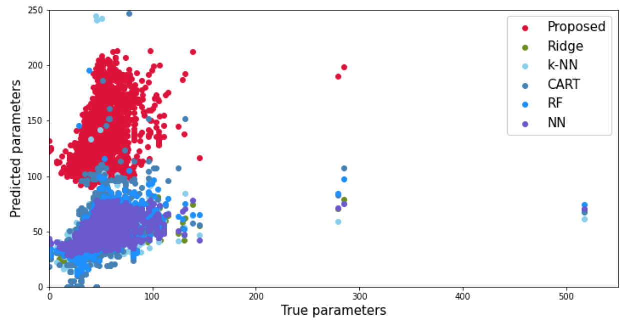

Figure 1: Groundtruth vs Predictions.

C.2 Detailed Experiment Results of the Alloy Production Problem

This section shows the experiment results of the alloy production problem, including the brass and the titanium-alloy.

Tables 6 and 7 report the mean post-hoc regrets and standard deviations across 10 runs, and the mean square errors (MSE) and standard deviations across 10 runs for each approach on the alloy production problem with unknown metal concentrations.

Tables 6 shows that, when the penalty factor is 0, our method improves the solution quality substantially in both of the two settings, obtaining at least 38.67% smaller post-hoc regret than the other methods in brass production, and at least 30.73% smaller post-hoc regret in titanium-alloy production. When the penalty factor is non-zero as given in the main paper, our method obtains at least 7.80%, 3.99%, 3.24%, and 6.56% smaller post-hoc regret respectively in brass production, and at least 9.65%, 7.30%, 3.14%, and 12.82% smaller post-hoc regret respectively in titanium-alloy production. The results in both of the two settings suggest that the advantages of the proposed method on solution quality first decreases and then increases as the penalty factor σ grows. The average True Optimal Values (TOV) are reported in the last column of Table 6. The relative errors of all the methods grow larger when the penalty factor grows larger. For example, the relative errors of the proposed method are 11.77%, 20.77%, 26.57%, 34.34%, and 47.03% on brass production when the penalty factors are all zero, or are sampled from $[0.25\pm0.015]$, $[0.5\pm0.015]$, $[1.0\pm0.015]$, $[2.0\pm0.015]$ respectively.

MSE of the predicted parameters across different methods are reported for reference in Table 7. The analysis of Table 7 is given in the main paper.

C.3 Detailed Experiment Results of the Fractional Knapsack Problem

This section shows the experiment results of the capacity of 50, 100, 150, and 200 in the fractional knapsack problem.

Tables 8 and 9 report the mean post-hoc regrets and standard deviations across 10

PReg		Proposed	Ridge	k-NN	CART	RF	NN	TOV
Alloy	Penalty factor							
Brass	0	**37.66±4.52**	61.93±3.17	65.68±5.76	87.57±8.83	61.40±2.96	61.46±6.69	312.02±6.94
	0.25±0.015	**68.16±6.26**	75.16±4.48	80.11±7.85	109.94±10.04	74.11±4.14	73.93±6.07	
	0.5±0.015	**82.91±5.45**	88.36±6.24	94.52±10.19	132.24±11.59	86.77±5.81	86.36±6.16	
	1±0.015	**107.64±6.85**	114.80±10.30	123.37±15.08	176.91±15.55	112.16±9.69	111.25±8.31	
	2±0.015	**150.47±12.99**	167.64±18.69	181.05±25.29	266.19±24.29	162.91±17.65	161.03±15.46	
Titanium-alloy	0	**4.07±0.75**	6.15±0.67	6.51±0.50	7.95±0.64	5.93±0.63	5.87±0.66	30.27±0.54
	0.25±0.015	**6.45±0.81**	7.54±0.81	8.03±0.59	10.05±0.67	7.22±0.75	7.14±0.79	
	0.5±0.015	**7.90±0.561**	8.92±0.96	9.56±0.69	12.15±0.73	8.53±0.88	8.52±0.90	
	1±0.015	**10.73±0.81**	11.69±1.28	12.59±0.92	16.34±0.87	11.12±1.16	11.08±1.19	
	2±0.015	**14.17±1.31**	17.23±1.92	18.69±1.41	24.72±1.24	16.32±1.75	16.25±1.72	

Table 6: Mean post-hoc regrets and standard deviations for the alloy production problem.

MSE		Proposed	Ridge	k-NN	CART	RF	NN
Alloy	Penalty factor						
Brass	0	395.81±331.56	39.33±0.64	43.68±0.92	73.98±1.74	**37.43±0.40**	37.80±0.47
	0.25±0.015	168.27±38.07					
	0.5±0.015	37.33±0.58					
	1±0.015	**36.97±0.56**					
	2±0.015	38.22±2.37					
Titanium-alloy	0	301.41±213.73	38.93±0.32	43.92±0.53	73.82±0.47	37.51±0.33	**36.60±0.26**
	0.25±0.015	48.23±7.95					
	0.5±0.015	44.69±5.74					
	1±0.015	39.00±2.63					
	2±0.015	45.28±4.28					

Table 7: Mean square errors and standard deviations for the alloy production problem.

runs, and the mean square errors (MSE) and standard deviations across 10 runs for each approach on the fractional knapsack problem with unknown prices and weights.

Observing Table 8, the performance of the proposed method is at least as good as other classical approaches when the capacity is 50, 100, or 150, and is consistent better than others when the capacity is 200. The relative errors of all the methods grow smaller when the capacity grows larger, for example, the relative errors of the proposed method are around 38-49%, 29-37%, 20-28%, 10-18% when the capacity is 50, 100, 150, and 200 respectively.

MSE of the predicted parameters across different methods are reported for reference in Table 9. The analysis of Table 9 is given in the main paper.

D Runtime Analysis

Table 10 shows the average runtime across 10 simulations for different optimization problems. In the alloy production problem and the fractional knapsack problem, the runtimes of the proposed method are comparable to NN, and are much better than RF. In the maximum flow transportation problem, the runtimes of the proposed method are comparable to NN in POLSKA and GÉANT, but the runtime of the proposed method is large in USANet. The reason is that we use the formulation where the decision variables each correspond to a simple path from the source to the sink. Thus, when the number of paths is large (the number of paths in USANet is 242), the number of the decision variables of the LP is large and the LP requires more time to be solved.

PReg		Proposed	Ridge	k-NN	CART	RF	NN	TOV
Capacity	Penalty factor							
50	0	**35.36±0.51**	38.00±0.89	36.95±1.04	**35.53±0.71**	37.90±0.65	39.75±1.18	90.79±0.46
	0.25±0.015	**38.17±0.76**	39.17±0.86	**38.46±0.96**	38.85±0.75	38.87±0.58	40.51±1.03	
	0.5±0.015	**39.57±0.85**	40.33±0.83	39.97±0.90	42.16±0.82	**39.85±0.53**	41.26±0.90	
	1.0±0.015	**41.90±0.85**	42.65±0.82	42.99±0.84	48.80±1.04	**41.99±0.47**	42.77±0.71	
	2.0±0.015	**44.92±0.91**	47.30±0.90	49.03±1.00	62.08±1.63	45.71±0.63	45.79±0.86	
100	0	**45.66±0.66**	49.52±1.29	48.20±1.31	48.08±0.75	49.85±1.31	52.19±1.84	156.46±0.79
	0.25±0.015	**49.97±0.86**	51.12±1.22	**50.38±1.14**	51.88±0.71	51.19±1.23	53.25±1.56	
	0.5±0.015	**52.27±0.66**	52.73±1.17	**52.36±1.08**	55.66±0.75	52.53±1.15	54.31±1.32	
	1.0±0.015	**55.71±1.12**	55.93±1.15	56.23±0.98	63.25±1.01	**55.74±0.63**	56.44±1.05	
	2.0±0.015	**58.88±0.79**	62.35±1.36	64.25±0.97	78.42±1.82	60.57±0.93	60.69±1.66	
150	0	**42.01±0.37**	47.56±1.08	46.16±1.13	46.91±0.67	48.09±0.97	49.78±2.02	207.92±0.99
	0.25±0.015	**46.59±0.23**	49.37±1.02	48.37±1.04	50.49±0.66	49.68±0.87	51.08±1.58	
	0.5±0.015	**50.25±0.59**	51.20±0.98	**50.58±0.97**	54.07±0.74	51.27±0.79	52.38±1.19	
	1.0±0.015	**54.07±0.66**	54.83±1.01	54.99±0.95	61.23±1.07	54.44±0.69	**54.97±0.86**	
	2.0±0.015	**58.40±0.63**	62.11±1.38	63.81±1.31	75.55±1.96	60.78±0.84	60.54±2.15	
200	0	**25.70±0.36**	33.07±0.98	32.73±0.92	33.18±0.88	33.63±0.84	34.67±2.13	246.86±1.20
	0.25±0.015	**31.50±0.50**	34.91±0.92	34.91±0.89	36.36±0.83	35.33±0.80	36.19±1.55	
	0.5±0.015	**35.08±0.69**	36.76±0.90	37.10±0.91	39.55±0.89	37.03±0.81	37.71±1.09	
	1.0±0.015	**39.54±0.45**	40.45±0.98	41.47±1.06	45.92±1.22	40.42±0.92	40.76±1.20	
	2.0±0.015	**44.59±0.55**	47.83±1.44	50.22±1.66	58.65±2.22	47.20±1.39	46.85±3.58	

Table 8: Mean post-hoc regrets and standard deviations for the fractional knapsack problem.

MSE		Proposed	Ridge	k-NN	CART	RF	NN
Capacity	Penalty factor						
50	0	190.05±3.89	75.40±0.65	83.47±0.77	140.51±1.75	72.04±0.73	**71.66±0.58**
	0.25±0.015	79.75±2.33					
	0.5±0.015	79.52±2.69					
	1.0±0.015	75.30±1.20					
	2.0±0.015	72.88±1.31					
100	0	162.95±22.35					
	0.25±0.015	83.28±3.32					
	0.5±0.015	77.57±1.62					
	1.0±0.015	72.28±1.08					
	2.0±0.015	**71.57±0.58**					
150	0	161.92±22.60					
	0.25±0.015	80.33±2.05					
	0.5±0.015	80.96±1.98					
	1.0±0.015	75.39±1.15					
	2.0±0.015	**71.35±0.66**					
200	0	151.91±35.18					
	0.25±0.015	79.12±3.26					
	0.5±0.015	75.40±1.50					
	1.0±0.015	**70.62±1.03**					
	2.0±0.015	**71.58±0.66**					

Table 9: Mean square errors and standard deviations for the fractional knapsack problem.

Runtime(s)	Maximum flow transportation			Alloy production		Fractional knapsack			
	POLSKA	USANet	GÉANT	Brass	Titanium-alloy	Capacity=50	Capacity=100	Capacity=150	Capacity=200
Proposed	18.65	132.22	15.48	228.00	331.38	131.49	132.89	139.44	132.37
Ridge	<1	<1	<1	20.22	56.89	22.33			
k-NN	<1	<1	<1	25.14	70.22	26.00			
CART	<1	<1	<1	30.33	94.89	34.83			
RF	4.11	11.00	11.89	959.50	2552.25	1034.07			
NN	10.33	12.82	13.89	212.22	321.11	135.80			

Table 10: Average runtime (in seconds) for the maximum flow transportation, alloy production, and fractional knapsack problems.