

# End-to-End Constrained Optimization Learning: A Survey

James Kotary<sup>1</sup>, Ferdinando Fioretto<sup>1</sup>, Pascal Van Hentenryck<sup>2</sup> and Bryan Wilder<sup>3</sup>

<sup>1</sup>Syracuse University

<sup>2</sup>Georgia Institute of Technology

<sup>3</sup>Harvard University

{jkotary, ffiorett}@syr.edu, pvh@gatech.edu, bwilder@g.harvard.edu

## Abstract

This paper surveys the recent attempts at leveraging machine learning to solve constrained optimization problems. It focuses on surveying the work on integrating combinatorial solvers and optimization methods with machine learning architectures. These approaches hold the promise to develop new hybrid machine learning and optimization methods to predict fast, approximate, solutions to combinatorial problems and to enable structural logical inference. This paper presents a conceptual review of the recent advancements in this emerging area.

## 1 Introduction

*Constrained optimization* (CO) has made a profound impact in industrial and societal applications in numerous fields, including transportation, supply chains, energy, scheduling, and the allocation of critical resources. The availability of algorithms to solve CO problems is highly dependent on their form, and they range from problems that are efficiently and reliably solvable, to problems that provably have no efficient method for their resolution. Of particular interest in many fields are *combinatorial optimization* problems, which are characterized by discrete state spaces, and whose solutions are often combinatorial in nature, involving the selection of subsets, permutations, paths through a network, or other discrete structures to compose a set of optimal decisions. They are known for their difficulty, and are often NP-Hard.

Despite their complexity, many CO problems are solved routinely and the AI and Operational Research communities have devised a wide spectrum of techniques and algorithms to effectively leverage the problem structure and solve many hard CO problems instances within a reasonable time and with high accuracy. While this success has made possible the deployment of CO solutions in many real-world contexts, the complexity of these problems often prevent them to be adopted in contexts of repeated (e.g., involving expensive simulations, multi-year planning studies) or real-time nature, or when they depend in nontrivial ways on empirical data. However, in many practical cases, one is interested in solving problem instances sharing similar patterns. Therefore, machine learning (ML) methods appear to be a natural candidate

to aid CO decisions and have recently gained traction in the nascent area at the intersection between CO and ML.

Current research areas in the intersection of CO and ML can be categorized into two main directions: *ML-augmented CO* and *End-to-End CO learning*. The former focuses on using ML to aid the decisions performed within an optimization algorithm used to solve CO problems. The latter involves the combination of ML and CO techniques to form integrated models which predict solutions to optimization problems. The survey subdivides this context into two main application domains: **(1)** the fast, approximate prediction of solutions to CO problems and **(2)** the integration of data-driven inference with CO solvers for structure logical inference.

While there exists work surveying ML-augmented CO methods [Bengio *et al.*, 2020; Lodi and Zarpellon, 2017], the more modern end-to-end CO learning methods lack of a cohesive and critical analysis. The goal of this survey is to address this gap and provide a focused overview on the work to-date on end-to-end CO learning, provide a critical analysis, and pose a set of open questions and directions.

## 2 Preliminaries: Constrained Optimization

A constrained optimization (CO) problem poses the task of minimizing an *objective function*  $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$  of one or more variables  $z \in \mathbb{R}^n$ , subject to the condition that a set of *constraints*  $\mathcal{C}$  are satisfied between the variables:

$$O = \underset{z}{\operatorname{argmin}} f(z) \quad \text{subject to } z \in \mathcal{C}. \quad (1)$$

An assignment of values  $z$  which satisfies  $\mathcal{C}$  is called a *feasible solution*; if, additionally  $f(z) \leq f(w)$  for all feasible  $w$ , it is called an *optimal solution*.

A well-understood class of optimization problems are *convex* problems, those in which the constrained set  $\mathcal{C}$  is a convex set, and the objective  $f$  is a convex function. Convex problems have the favorable properties of being efficiently solvable with strong theoretical guarantees on the existence and uniqueness of solutions [Boyd *et al.*, 2004].

A particularly common constraint set arising in practical problems takes the form  $\mathcal{C} = \{z : Az \leq b\}$ , where  $A \in \mathbb{R}^{m \times n}$  and  $b \in \mathbb{R}^m$ . In this case,  $\mathcal{C}$  is a convex set. If the objective  $f$  is an affine function, the problem is referred to as *linear program* (LP). When a linearly constrained problem includes a quadratic objective rather than a linear one, the result is called a *quadratic program* (QP). If, in addition, some

subset of a problem’s variables are required to take integer values, it is called *mixed integer program* (MIP). While LP and QP with convex objectives belong to the class of convex problems, the introduction of integral constraints ( $\mathbf{x} \in \mathbb{N}^n$ ) results in a much more difficult problem. The feasible set in MIP consists of distinct points in  $\mathbf{x} \in \mathbb{R}^n$ , not only nonconvex but also disjoint, and the resulting problem is, in general, NP-Hard. Finally, nonlinear programs (NLPs) are optimization problems where some of the constraints or the objective function are nonlinear. Many NLPs are nonconvex and can not be efficiently solved [Nocedal and Wright, 2006].

Of particular interest are the *mixed integer linear programs* (MILPs), linear programs in which a subset of variables required to take integral values. This survey is primarily concerned with optimization problems involving linear constraints, linear or quadratic objective, and either continuous or integral variables, or a combination thereof.

**Optimization Solving Methods** A well-developed theory exists for solving convex problems. Methods for solving LP problems include *simplex* methods [Dantzig, 1951], *interior point* methods [Boyd et al., 2004] and *Augmented Lagrangian* methods [Hestenes, 1969; Powell, 1969]. Each of these methods has a variant that applies when the objective function is convex. Convex problems [Boyd et al., 2004] are in the class of  $\mathcal{P}$ , and can be assumed to be reliably and efficiently solvable in most contexts. For a review on convex optimization and Lagrangian duality, the interested reader is referred to [Boyd et al., 2004].

MILPs require a fundamentally different approach, as their integrality constraints put them in the class of NP-Hard problems. *Branch and bound* is a framework combining optimization and search, representable by a search tree in which a LP *relaxation* of the MILP is formed at each node by dropping integrality constraints, and efficiently solved using solution methods for linear programming. Subsequently, a variable  $z_i$  with (fractional) value  $a_i$  in the relaxation is selected for branching to two further nodes. In the right node, the constraint  $z_i \geq a_i$  is imposed and in the left,  $z_i < a_i$ ; bisecting the search space and discarding fractional values in between the bounds. The solution of each LP relaxation provides a lower bound on the MILP’s optimal objective value. When an LP relaxation happens to admit an integral solution, an upper bound is obtained. The minimal upper bound found thus far is used, at each node, for pruning.

Finally, *Constraint Programming* [Rossi et al., 2006] is an additional effective paradigm adopted to solve industrial-sized MI(L)P and discrete combinatorial programs.

### 3 Preliminaries: Deep Learning

Supervised deep learning can be viewed as the task of approximating a complex non-linear mapping from targeted data. Deep Neural Networks (DNNs) are deep learning architectures composed of a sequence of layers, each typically taking as inputs the results of the previous layer [LeCun et al., 2015]. Feed-forward neural networks are basic DNNs where the layers are fully connected and the function connecting the layer is given by  $\mathbf{o} = \pi(\mathbf{W}\mathbf{x} + \mathbf{b})$ , where  $\mathbf{x} \in \mathbb{R}^n$  and is the input vector,  $\mathbf{o} \in \mathbb{R}^m$  the output vector,  $\mathbf{W} \in \mathbb{R}^{m \times n}$  a matrix of

weights, and  $\mathbf{b} \in \mathbb{R}^m$  a bias vector. The function  $\pi(\cdot)$  is often non-linear (e.g., a rectified linear unit (ReLU)).

Supervised learning tasks consider datasets  $\mathcal{X} = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^N$  consisting of  $N$  data points with  $\mathbf{x}_i \in \mathcal{X}$  being a feature vector and  $\mathbf{y}_i \in \mathcal{Y}$  the associated targets. The goal is to learn a model  $\mathcal{M}_\theta : \mathcal{X} \rightarrow \mathcal{Y}$ , where  $\theta$  is a vector of real-valued parameters, and whose quality is measured in terms of a nonnegative, and assumed differentiable, *loss function*  $\mathcal{L} : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}_+$ . The learning task minimizes the empirical risk function (ERM):

$$\min_{\theta} J(\mathcal{M}_\theta, \mathcal{X}) = \frac{1}{n} \sum_{i=1}^n \mathcal{L}(\mathcal{M}_\theta(\mathbf{x}_i), \mathbf{y}_i). \quad (2)$$

Most of the techniques surveyed in this work use (variants of) DNNs whose training conforms to the objective above. Other notable classes of deep learning methods used to solve CO problems are Sequence Models and Graph Neural Networks (GNNs), which are reviewed below, and Reinforcement Learning (RL), which differs from supervised learning in not requiring labeled input/output pairs and concerns with learning a policy that maximizes some expected reward function. We refer the reader to [Sutton and Barto, 2018] for an extensive overview of RL.

**Sequence Models** Recurrent neural networks (RNNs) have proven highly effective for tasks requiring pattern recognition in sequential data. A basic RNN uses a *recurrent layer* to compute a sequence of output arrays  $\mathbf{y}_1, \dots, \mathbf{y}_N$  from a sequence of input arrays  $\mathbf{x}_1, \dots, \mathbf{x}_N$ . Each of several hidden units  $h_t$  depends on its corresponding input array  $\mathbf{x}_t$  along with the previous hidden unit  $h_{t-1}$  in the following manner:

$$h_t = \pi(\mathbf{W}^x \mathbf{x}_t + \mathbf{W}^h h_{t-1} + \mathbf{b})$$

where  $\pi$  is a nonlinear activation function,  $\mathbf{W}^x, \mathbf{W}^h$  are weight matrices, and  $\mathbf{b}$  is an additive bias.  $\mathbf{W}$  and  $\mathbf{b}$  are learned parameters, applied at each unit of the recurrent layer.

In their basic form, RNNs are known to suffer from vanishing gradient issues: the propagation of gradients through many hidden units by the chain rule can lead to loss of information as gradient magnitudes diminish. Modern RNNs use gated recurrent units (GRU) to form long-short term memory models (LSTM) which can better preserve relevant information across many hidden layers.

*Sequence-to-Sequence* models build on recurrent neural networks with the aim of extending the framework to handle input and output sequences of variable size. This is accomplished by the introduction of an *encoder-decoder* architecture in which an encoder LSTM maps an input sequence  $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$  to a fixed-length context vector  $\mathbf{c}$ , by way of the hidden encoder states  $(e_1, \dots, e_N)$  (typically,  $\mathbf{c} = e_N$ ). A subsequent decoder LSTM is used to map the context vector to a sequence of hidden decoder states  $(d_1, \dots, d_M)$ , which are used to target a sequence of labeled vectors  $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_M)$ . That is, some function  $g$  is trained to approximate the likelihood of predicting  $\mathbf{y}_t$  given the context vector  $\mathbf{c}$ , the current decoder state and all previous  $\mathbf{y}_i$ :

$$\Pr(\mathbf{y}_t | \{\mathbf{y}_1, \dots, \mathbf{y}_{t-1}\}, \mathbf{c}) = g(\mathbf{y}_{t-1}, d_t, \mathbf{c}),$$

where the overall training goal is to model the probability of predicting sequence  $\mathbf{y}$ . This joint probability is decomposed into a product of ordered conditionals:

$$\Pr(\mathbf{y}) = \prod_{t=1}^M \Pr(\mathbf{y}_t | \{\mathbf{y}_1, \dots, \mathbf{y}_{t-1}\}, \mathbf{c}).$$

The primary downside to this approach is the reliance on a single fixed-length vector to encode the input sequence, resulting in limited expressivity, particularly when the input sequences observed at test time are of a different length than those used during training. *Attention* mechanisms [Bahdanau et al., 2016] are used to address this shortcoming. They use the same encoder design, but allow the decoder to adaptively use any subset of the encoder’s hidden units during decoding. In place of a single context vector, a *sequence* of context vectors  $\mathbf{c}_r$  (one per decoder unit) is computed as

$$\mathbf{c}_r = \sum_{t=1}^N \mathbf{a}_{rt} e_t \quad \forall r \leq M \quad (3a)$$

$$\mathbf{a}_{rt} = \frac{\exp(\mathbf{u}_{rt})}{\sum_{k=1}^N \exp(\mathbf{u}_{rk})} \quad (3b)$$

$$\mathbf{u}_{rt} = \mathbf{v}^T \pi(\mathbf{W}_1 e_t + \mathbf{W}_2 d_r), \quad (3c)$$

where  $\mathbf{a}$  above is computed as the softmax of  $\mathbf{u}$ , the result of *alignment* model (3c) in which  $\mathbf{v}$ ,  $\mathbf{W}_1$ , and  $\mathbf{W}_2$  are learnable matrices and  $\pi$  is a nonlinear activation function. The values  $\mathbf{a}_{rt}$  constitute a sequence of *attention* vectors which measure the alignment between decoder state  $d_r$  and encoder state  $e_t$ . Each attention vector is used as a set of weights to form a linear combination of encoder states, used as a context vector to inform subsequent predictions.

*Pointer networks* [Vinyals et al., 2017] are a special use case of attention mechanisms, used when the intended output of a network is a permutation of its input. Rather than utilizing the attention vector to assist output sequence prediction pointer networks treat the attention vector itself as the prediction. That is,

$$\mathbf{u}_{rt} = \mathbf{v}^T \pi(\mathbf{W}_1 e_t + \mathbf{W}_2 d_r) \quad \forall 1 \leq t \leq N,$$

$$\Pr(\mathbf{y}_t | \{\mathbf{y}_1, \dots, \mathbf{y}_{t-1}\}, \mathbf{x}) = \text{softmax}(\mathbf{u}_r),$$

where  $\mathbf{x}$  is again the input sequence and  $\mathbf{y}$  is a permutation of  $\mathbf{x}$ . The attention vector is viewed as a pointer to a particular encoder state, and can be used in subsequent predictions.

**Graph Neural Networks** Graph Neural Networks (GNNs) [Scarselli et al., 2008], learn to represent each node  $v$  of a graph with a *state embedding*  $x_v$ , which depends on information from its neighboring nodes. The goal of these embeddings is to provide useful latent information about each node to subsequent components of a prediction model. Let  $x_v, x_{co[v]}, h_{ne[v]}, x_{ne[v]}$  represent, respectively, the features of node  $v$ , the features of its edges, the state embeddings, and features of its neighbors. Then consider two functions:

$$h_v = f(x_v, x_{co[v]}, h_{ne[v]}, x_{ne[v]}) \quad (4a)$$

$$o_v = g(h_v, x_v) \quad (4b)$$

Equivalently to (4a) and (4b),

$$H = F(H, X) \quad (5a)$$

$$O = G(H, X_N) \quad (5b)$$

where  $X$  and  $H$  represent the concatenation of all features and all state embeddings, and  $X_N$  contains only the node-wise features. When  $F$  is a contractive mapping, the equation has a fixed point solution  $H^*$  which can be found by repeated application of (5a). When  $F$  and  $G$  are interpreted as feed-forward neural networks, their parameters can be trained to produce state embeddings that are optimized to assist a particular learning task.

Several variants and enhancements to the basic GNN have been proposed; Zhou et al. [2018] provides a thorough survey of techniques and applications.

## 4 Overview of ML and CO

Current research areas in the synthesis of constrained optimization and machine learning can be categorized into two main directions: *ML-augmented CO*, which focuses on using ML to aid the performance of CO problem solvers, and *End-to-End CO learning (E2E-COL)*, which focuses on integrating combinatorial solvers or optimization methods into deep learning architectures. A diagram illustrating the main research branches within this area is provided in Figure 1.

The area related with End-to-End CO learning is the focus of this survey and is concerned with the data-driven prediction of solutions to CO problems. This paper divides this area into: (1) approaches that develop ML architectures to predict fast, approximate solutions to predefined CO problems, further categorized into *learning with constraints* and *learning on graphs*, and (2) approaches that exploit CO solvers as neural network layers for the purpose of structured logical inference, referred to here as the *Predict-and-Optimize* paradigm.

## 5 ML-augmented CO

ML-augmented CO involves the augmentation of already highly-optimized CO solvers with ML inference models. Techniques in this area draw on both supervised and RL frameworks to develop more efficient approaches to various aspects of CO solving for both continuous and discrete combinatorial problems.

In the context of combinatorial optimization, these are broadly categorized into methods that learn to guide the search decisions in branch and bound solvers, and methods that guide the application of primal heuristics within branch and bound. The former include low-cost emulation of expensive branching rules in mixed integer programming [Khalil et al., 2016; Gasse et al., 2019; Gupta et al., 2020], prediction of optimal combinations of low-cost variable scoring rules to derive more powerful ones [Balcan et al., 2018], and learning to cut [Tang et al., 2020] in cutting plane methods within MILP solvers. The latter include prediction of the most effective nodes at which to apply primal heuristics [Khalil et al., 2017], and specification of primal heuristics such as the optimal choice of variable partitions in large neighborhood search [Song et al., 2020]. The reader is referred to the excellent surveys [Lodi and Zarpellon, 2017; Bengio et al., 2020] for a thorough account of techniques developed within ML-augmented combinatorial optimization.

Techniques in this area have also used ML to improve decisions in continuous CO problems and include learning restart

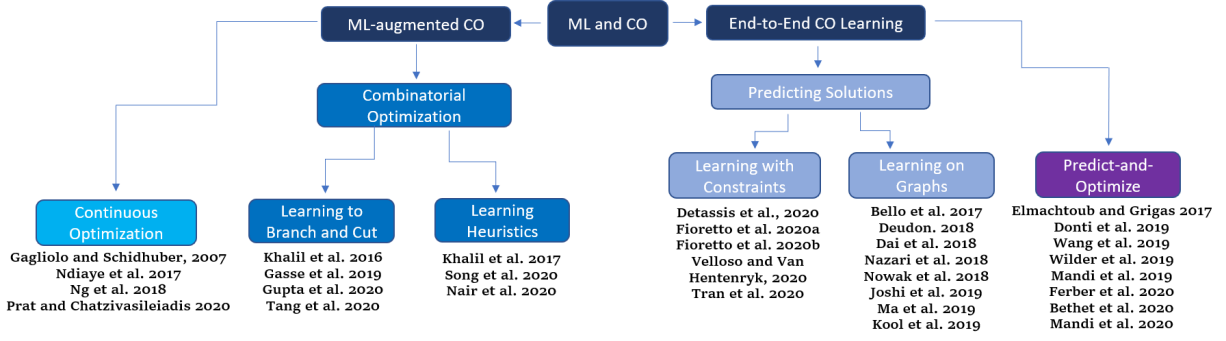


Figure 1: Machine Learning and Constrained Optimization.

strategies [Gagliolo and Schmidhuber, 2007], learn rules to ignore some optimization variables leveraging the expected sparsity of the solutions and consequently leading to faster solvers, [Ndiaye et al., 2017], and learning active constraints [Ng et al., 2018; Prat and Chatzivasileiadis, 2020] to reduce the size of the problem to be fed into a CO solver.

## 6 E2E-COL: Predicting CO Solutions

A diverse body of work within the end-to-end CO learning literature has focused on developing ML architectures to predict fast, approximate solutions to predefined CO problems *end-to-end* without the use of CO solvers at the time of inference, by observing a set of solved instances or execution traces. These approaches contrast with those that use ML to augment search-based CO solvers and configure their sub-routines to direct the solver to find solutions efficiently. This survey categorizes the literature on *predicting CO solutions* into (1) methods that incorporate constraints into end-to-end learning, for the prediction of feasible or near-feasible solutions to both continuous and discrete constrained optimization problems, and (2) methods that learn combinatorial solutions on graphs, with the goal of producing outputs as combinatorial structures from variable-sized inputs. These two categories, referred to as *learning with constraints* and *learning CO solutions*, respectively, are reviewed next.

### 6.1 Learning with Constraints

The methods below consider datasets  $\chi = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^N$  whose inputs  $\mathbf{x}_i$  describe some problem instance specification, such as matrix  $\mathbf{A}$  and vector  $\mathbf{b}$  describing linear constraints in MILPs, and the labels  $\mathbf{y}_i$  describe complete solutions to problem  $\mathcal{O}$  with input  $\mathbf{x}_i$ . Notably, each sample may specify a different problem instance (with different objective function coefficients and constraints).

An early approach to the use of ML for predicting CO problem solutions was presented by Hopfield and Tank [1985], which used Hopfield Networks (Hopfield [1988]) with modified energy functions to emulate the objective of a traveling salesman problem (TSP), and applied Lagrange multipliers to enforce feasibility to the problem’s constraints. It was shown in Wilson and Pawley [1988] however, that this approach suffers from several weakness, notably sensitivity to parameter initialization and hyperparameters. As noted in

Bello et al. [2017], similar approaches have largely fallen out of favor with the introduction of practically superior methods.

Frameworks that exploit Lagrangian duality to guide the prediction of approximate solutions to satisfy the problem’s constraints have found success in the context of continuous NLPs including energy optimization [Fioretto et al., 2020b; Velloso and Van Hentenryck, 2020] as well as constrained prediction on tasks such as transprecision computing and fair classification [Fioretto et al., 2020a; Tran et al., 2020].

Other end-to-end learning approaches have demonstrated success on the prediction of solutions to constrained problems by injecting information about constraints from targeted feasible solutions. Recently, Detassis et al. [2020] presented an iterative process of using an external solver for discrete or continuous optimization to adjust targeted solutions to more closely match model predictions while maintaining feasibility, reducing the degree of constraint violation in the model predictions in subsequent iterations.

### 6.2 Learning Solutions on Graphs

By contrast to approaches learning solutions to unstructured CO problems, a variety of methods learn to solve CO cast on graphs. The development of deep learning architectures, such as sequence models and attention mechanisms, as well as GNNs, has provided a natural toolset for these tasks.

The survey categorizes these modern approaches broadly based on whether they rely supervised learning or reinforcement learning. It observes that: (1) Supervised learning concerns training models that take as input a representation of a problem instance and output a solution. Target solutions must be provided for this training task; this is often problematic as it entails solving, typically, NP-hard problems. (2) Optimization problems have a native objective function, which in principle can be used in lieu of a loss function in supervised learning (possibly obviating the need for training labels) and can serve as a natural reward function in reinforcement learning.

#### Supervised Learning

Vinyals et al. [2017] introduced the *pointer network*, in which a sequence-to-sequence model uses an encoder-decoder architecture paired with an attention mechanism to produce permutations over inputs of variable size. Noting that certain CO problems provide a compelling use case for their architecture,



the authors test it for solving the Traveling Salesman (TSP), Delaunay Triangulation, and Convex Hull problem variants. In each, the solution to a problem instance can be expressed as a single permutation. They develop a pointer network model to predict approximately optimal solutions by learning from previously solved instances in a supervised manner, and demonstrate some ability to generalize over variable-sized problem instances. For example, in the 2D Euclidean TSP, the pointer network’s inputs are the 2D coordinates of each city that must be visited. A predicted permutation represents a tour over all cities, and each target label is a permutation representing the pre-computed minimum-length tour over its corresponding input coordinates. Any city is directly reachable from any other—by a straight line whose length is equal to the euclidean distance. In general, the approach introduced in this work applies only to problems whose solutions take the form of a single permutation and where all permutations are feasible (e.g., the problem does not feature constraints on which tours are allowed in the TSP).

Despite the subsequent trend toward RL-oriented frameworks, as discussed in the next section, [Nowak et al. \[2018\]](#) studied a purely supervised learning method for the general Quadratic Assignment Problem (QAP). The proposed solution was based on the use of simple graph neural networks trained on representations of individual problem instances and their targeted solutions. The TSP is chosen as a test problem, along with two graph matching problems (all instances of the QAP). Inferences from the model take the form of approximate permutation matrices, which are converted into feasible tours by a beam search. The authors note promising accuracy results on small instances, but the ability for trained models to generalize their performance to larger-sized instances was not shown.

[Joshi et al. \[2019\]](#) built on this concept by applying a Graph Convolutional Network model to the 2D Euclidean TSP. Aside from the enhanced deep learning architecture, most aspects of the approach remain similar. However, accurate results are reported on problem instances much larger than in [Nowak et al. \[2018\]](#). The authors observe that compared to contemporary reinforcement learning approaches, their training method is more sample-efficient, but results in models that do not generalize nearly as well to problem instances larger than those used in training.

## Reinforcement Learning

The pointer network architecture of [Vinyals et al. \[2017\]](#) was adopted by [Bello et al. \[2017\]](#), who proposed to train an approximate TSP solver with reinforcement learning rather than supervised learning. The transition to RL was motivated partly by the difficulties associated with obtaining target solutions that are optimal, and the existence of nonunique optimal solutions to TSP instances. Rather than generating and targeting precomputed solutions, the authors present an *actor-critic* RL framework using expected tour length  $L(\pi|g)$  as the reward signal, where  $g$  and  $\pi$  represent a graph (problem instance) and a permutation (a tour over the graph):

$$J(\theta|g) = \mathbb{E}_{\pi \sim p_{\theta}(\pi|g)} L(\pi|g). \quad (6)$$

A policy, represented by the underlying pointer network with parameters  $\theta$ , is optimized using the well-known RE-

INFORCE rule for computing policy gradients [[Williams, 1992](#)]:

$$\nabla_{\theta} J(\theta|g) = \mathbb{E}_{\pi \sim p_{\theta}(\pi|g)} [(L(\pi|g) - b(g)) \nabla_{\theta} \log p_{\theta}(\pi|g)] \quad (7)$$

The total training objective is then the expectation of  $J(\theta|g)$  over a distribution of graphs. The policy gradient calculation requires a baseline function  $b$  which estimates the expected reward. In this work, an auxiliary *critic* network with its own set of parameters is trained to predict the expected tour length for any graph in supervised fashion, using empirically observed tours from the most recent policy as training labels. At the time of inference, two methods are available for producing tours from a trained model: (1) A set of tours can be sampled from the stochastic policy by varying the softmax temperature parameter, from which the shortest is selected or (2) an active search method which modifies the stochastic policy based on solutions sampled during inference.

[Nazari et al. \[2018\]](#) departed from the RNN encoder scheme used in [Bello et al. \[2017\]](#) and [Vinyals et al. \[2017\]](#) arguing that when a problem’s defining data do not adhere to a natural order (e.g. city coordinates in the TSP), it need not be modeled sequentially. This provides motivation to leave out the encoder RNN and consider combinatorial problems whose state changes over time, replacing the role of encoder hidden states with representations of the problem in each time frame. The authors consider a capacitated vehicle routing problem (VRP/CVRP) in which demands are associated to points in space and change over time; a vehicle with limited capacity must satisfy all the demands by delivering supply from a central depot while minimizing total tour length. This setting also diverges from those previously discussed in that solutions no longer take the form of permutations, since an optimal tour may now include several trips to the supply depot or a demand sink. The static (location coordinates) and dynamic (demand value) data defining the problem are concatenated into a single embedding vector  $x_t$  for each time frame  $t$ . In each decoder step, the attention layer computes a context vector  $c_t$  based only on  $x_t$  and the previous decoder state  $d_t$ . A final prediction at time  $t$  indicates the next visited location based on  $x_t$  and  $c_t$ , and the embeddings  $x_t$  can then be updated based on the resulting state-changes. The training consists of an actor-critic method similar to that of [Bello et al. \[2017\]](#).

Several improvements of these approaches exists. Notably, [Kool et al. \[2019\]](#) developed a general reinforcement learning framework based on a Graph Attention Network architecture [[Veličković et al., 2017](#)] and trained with the REINFORCE policy gradients, and present significant improvements in accuracy on 2D euclidean TSP over [Vinyals et al. \[2017\]](#), [Bello et al. \[2017\]](#), [Dai et al. \[2018\]](#), and [Nowak et al. \[2018\]](#).

Differently from previous approaches, [Ma et al. \[2019\]](#) focused on CO instances with hard constraints and consider TSP variants in which certain tours are considered infeasible. A multi-level RL framework is proposed in which each layer of a hierarchy learns a different policy, from which actions can be sampled. Lower layers learn policies that result in feasible solutions by using reward functions that bias feasibility, and provide latent variables to higher layers which learn to optimize a given objective while sampling actions

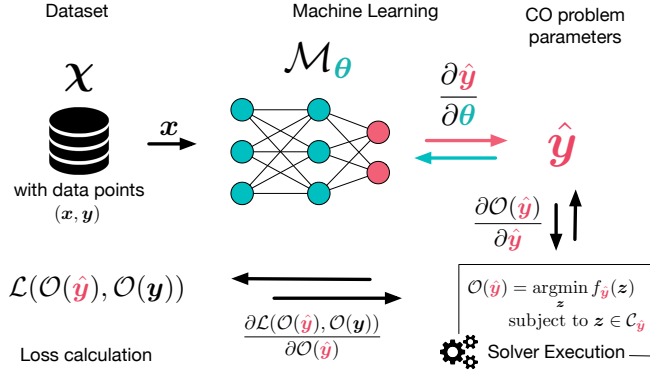


Figure 2: Predict-and-optimize framework; gradients of a solver output (solution) must be computed with respect to its input (problem parameters) in order to maximize empirical model performance.

from lower layers. Layer-wise policy optimization uses the REINFORCE algorithm, training each layer in turn before fixing its weights and sampling actions to train the next layer. Each policy is represented by a graph neural network to produce context embeddings that capture inter-node relationships from node-wise inputs, and an attention-based decoder to predict distributions over candidate actions as in prior works. When the direct output  $u_i$  from layer  $i$  is predicted at a particular decoding step, it is combined with the output of the lower layer  $i - 1$  to ensure mutual compatibility in the policy distribution

$$p_i = \operatorname{softmax}(u_i + \alpha u_{i-1})$$

where  $\alpha$  is a trainable parameter.

Finally, Dai *et al.* [2018] adopted a different RL approach based on a greedy heuristic framework which determines approximate solutions by the sequential selection of graph nodes. The selection policy is learned using a combination of Q-learning and fitted Q-iteration, by optimizing the optimization objective directly. The graph embedding network `structure2vec` is used to represent the policy of the learned greedy algorithm. It is used to produce feature embeddings for each node of the graph, which are updated between actions and capture relationships between neighboring nodes, given the current state of a partial solution. It is noted that this framework excels in combinatorial problems that have a true graphical structure, as opposed to most previous previously studied applications (primarily the TSP and VRP variants) in the underlying ‘graph’ is typically fully connected. The Minimum Vertex Cover, Max Cut, and TSP are used as test problems, and a strong ability to generalize over various problem sizes is advertised in comparison to prior approaches.

## 7 E2E-COL: Predict-and-Optimize

A burgeoning topic in the intersection of ML and CO is the fusion of prediction (ML) and decision (CO) models, in which decision models are represented by partially defined optimization problems, whose specification is completed by parameters that are predicted from data. The resulting composite models employ constrained optimization as a neural

network layer and are trained *end-to-end*, based on the optimality of their decisions. This setting is altogether different in motivation to those previously discussed, in which the goal was to solve predefined CO instances with increased efficiency. The goal here is the synthesis of predictive and prescriptive techniques to create ML systems that learn to make decisions based on empirical data.

The following constrained optimization problem is posed, in which the objective function  $f_y$  and feasible region  $\mathcal{C}_y$  depend on a parameter vector  $y$ :

$$\mathcal{O}(y) = \underset{z}{\operatorname{argmin}} f_y(z) \quad \text{subject to } z \in \mathcal{C}_y. \quad (8)$$

The goal here is to use supervised learning to predict  $\hat{y}$  the unspecified parameters from empirical data. The learning task is performed so that the optimal solution  $\mathcal{O}(\hat{y})$  best matches a targeted optimal solution  $\mathcal{O}(y)$ , relative to some appropriately chosen loss function. The empirical data in this setting is defined abstractly as belonging to a dataset  $\mathcal{X}$ , and can represent any empirical observations correlating with targeted solutions to (8) for some  $y$ . See Figure 2 for an illustration.

This framework aims to improve on simpler two-stage approaches, in which a conventional loss function (e.g. MSE) is used to target labeled parameter vectors  $y$  that are provided in advance, before solving the associated CO problem to obtain a decision. Such approaches are suboptimal in the sense that predictions of  $y$  do not take into account the accuracy of the resulting solution  $\mathcal{O}(y)$  during training.

We note that there are two ways to view the predictions that result from these integrated models. If  $\hat{y}$  is viewed as the prediction, then the calculation of  $\mathcal{O}(\hat{y})$  is absorbed into the loss function  $\mathcal{L}(\hat{y}, y)$ , which targets the provided parameter vectors. Otherwise, the loss function  $\mathcal{L}(\mathcal{O}(\hat{y}), \mathcal{O}(y))$  is considered separately from the decision model and aims to match computed optimal solutions to targeted ones. One advantage sought in either case is the opportunity to minimize during training the ultimate error in the computed optimal objective values  $f_{\hat{y}}(\mathcal{O}(\hat{y}))$ , relative to those of the target data. This notion of training loss is known as *regret*:

$$\operatorname{regret}(\hat{y}, y) = f_{\hat{y}}(\mathcal{O}(\hat{y})) - f_y(\mathcal{O}(y)).$$

Otherwise the optimal solution  $\mathcal{O}(y)$  is targeted and one can use  $\operatorname{regret}(\mathcal{O}(\hat{y}), \mathcal{O}(y))$ , regardless of whether the corresponding  $y$  is available. Depending on the techniques used, it may be possible to minimize the regret without access to ground-truth solutions, as in Wilder *et al.* [2019], since the targeted solutions  $\mathcal{O}(y)$  contribute only a constant term to the overall loss. It is worth mentioning that because the optimization problem in (8) is viewed as a function, the existence of nonunique optimal solutions is typically not considered. The implication then is that  $\mathcal{O}(y)$  is directly determined by  $y$ .

Training these end-to-end models involves the introduction of external CO solvers into the training loop of a ML model, often a DNN. Note that combinatorial problems with discrete state spaces do not offer useful gradients; viewed as a function, the *argmin* of a discrete problem is piecewise constant. The challenge of forming useful approximations to  $\frac{\partial \mathcal{L}}{\partial y}$  is central in this context and must be addressed in order to perform backpropagation. It may be approximated directly, but

a more prevalent strategy is to model  $\frac{\partial \mathcal{O}(\mathbf{y})}{\partial \mathbf{y}}$  and  $\frac{\partial \mathcal{L}}{\partial \mathcal{O}}$  separately, in which case the challenge is to compute the former term by *differentiation through argmin*. Figure 2 shows the role of this gradient calculation in context.

## 7.1 Quadratic Programming

One catalyst for the development of this topic was the introduction of *differentiable optimization layers*, beginning with Amos and Kolter [2017] which introduced a GPU-ready QP solver that offers exact gradients for backpropagation by differentiating the KKT optimality conditions of a quadratic program at the time of solving, and utilizing information from the forward pass to solve a linear system for incoming gradients, once the outgoing gradients are known. Subsequently, Donti et al. [2019] proposed a *predict-and-optimize* model in which QPs with stochastic constraints were integrated in-the-loop to provide accurate solutions to inventory and power generator scheduling problems specified by empirical data.

## 7.2 Linear Programming

Concurrent with Donti et al. [2019], an alternative methodology for end-to-end learning with decision models, called *smart predict-and-optimize* (SPO), was introduced by Elmachetoub and Grigas [2020], which focused on prediction with optimization of constrained problems with linear objectives, in which the cost vector is predicted from data and the feasible region  $\mathcal{C}$  is invariant to the parameter  $\mathbf{y}$ :

$$\mathcal{O}(\mathbf{y}) = \underset{\mathbf{z}}{\operatorname{argmin}} \mathbf{y}^T \mathbf{z} \text{ subject to } \mathbf{z} \in \mathcal{C}. \quad (9)$$

The target data in this work are the true cost vectors  $\mathbf{y}$ , and an inexact subgradient calculation is used for the backpropagation of regret loss  $\mathcal{L}(\hat{\mathbf{y}}, \mathbf{y}) = \hat{\mathbf{y}}^T (\mathcal{O}(\hat{\mathbf{y}}) - \mathcal{O}(\mathbf{y}))$  on the decision task, by first defining a convex surrogate upper bound on regret called the *SPO+loss*, for which it is shown that  $\mathcal{O}(\mathbf{y}) - \mathcal{O}(2\hat{\mathbf{y}} - \mathbf{y})$  is a useful subgradient. Since this work is limited to the development of surrogate loss functions on regret from the optimization task, it does not apply to learning tasks in which the full solution to an optimization problem is targeted. The paper includes a discussion justifying the method’s use on problems with discrete constraints in  $\mathcal{C}$ , as in combinatorial optimization, but experimental results are not provided on that topic. It is, however, demonstrated that the approach succeeds in a case where  $\mathcal{C}$  is convex but nonlinear.

Wilder et al. [2019] introduced an alternative framework to predict-and-optimize linear programming problems, based on exact differentiation of a smoothed surrogate model. While LPs are special cases of QPs, the gradient calculation of Amos and Kolter [2017] does not directly apply due to singularity of the KKT conditions when the objective function is purely linear. This is addressed by introducing a small quadratic regularization term to the LP objective  $f_{\mathbf{y}}(\mathbf{z}) = \mathbf{y}^T \mathbf{z}$  so that the problem in (8) becomes

$$\mathcal{O}(\mathbf{y}) = \underset{\mathbf{z}}{\operatorname{argmin}} \mathbf{y}^T \mathbf{z} + \epsilon \|\mathbf{z}\|^2 \text{ subject to } \mathbf{A}\mathbf{z} \leq \mathbf{b}. \quad (10)$$

The resulting problems approximate the desired LP, but have unique solutions that vary smoothly as a function of their parameters, allowing for accurate backpropagation of the result.

The integrated model is trained to minimize the expected optimal objective value across all training samples, which is equivalent to minimizing the regret loss but without the need for a target dataset. This work demonstrated success on problems such as the knapsack (using LP relaxation) and bipartite matching problems where a cost vector is predicted from empirical data (e.g., historical cost data for knapsack items), and is shown to outperform two-stage models which lack integration of the LP problem. We note that although the differentiable QP solving framework of Amos and Kolter [2017] is capable of handling differentiation with respect to any objective or constraint coefficient, this work only report results on tasks in which the *cost* vector is parameterized within the learning architecture, and constraints are held constant across each sample. *This limitation is common to all of the works described below, as well.*

Next, Mandi and Guns [2020] introduced an altogether different approach to obtaining approximate gradients for the *argmin* of a linear program. An interior point method is used to compute the solution of a homogeneous self-dual embedding with early stopping, and the method’s log-barrier term is recovered and used to solve for gradients in the backward pass. Equivalently, this can be viewed as the introduction of a log-barrier regularization term, by analogy to the QP-based approach of Wilder et al. [2019]:

$$\mathcal{O}(\mathbf{y}) = \underset{\mathbf{z}}{\operatorname{argmin}} \mathbf{y}^T \mathbf{z} + \lambda \left( \sum_i \ln(z_i) \right) \text{ subject to } \mathbf{A}\mathbf{z} \leq \mathbf{b}.$$

Further, the method’s performance on end-to-end learning tasks is evaluated against the QP approach of Wilder et al. [2019] on LP-based predict-and-optimize tasks, citing stronger accuracy results on energy scheduling and knapsack problems with costs predicted from data.

Berthet et al. [2020] detailed an approach based on stochastic perturbation to differentiate the output of linear programs with respect to their cost vectors. The output space of the LP problem is smoothed by adding low-amplitude random noise to the cost vector and computing the expectation of the resulting solution in each forward pass. This can be done in Monte Carlo fashion and in parallel across the noise samples. The gradient calculation views the solver as a black box in this approach, and does not require the explicit solving of LP for operations that can be mathematically formulated as LP, but are simple to perform (e.g., sorting and ranking). Results include a replication of the shortest path experiments presented in [Pogančić et al., 2019], in which a model integrated with convolutional neural networks is used to approximate the shortest path through stages in a computer game, solely from images.

## 7.3 Combinatorial Optimization

Ferber et al. [2020] extended the approach of Wilder et al. [2019] to integrate MILP within the end-to-end training loop, with the aim of utilizing more expressive NP-Hard combinatorial problems with parameters predicted from data. This is done by reducing the MILP with integer constraints to a LP by a method of cutting planes. In the ideal case, the LP that results from the addition of cutting planes has the same optimal solution as its mixed-integer parent. Exact gradients can



then be computed for its regularized QP approximation as in [Wilder et al. \[2019\]](#). Although the LP approximation to MILP improves with solving time, practical concerns arise when the MILP problem cannot be solved to completion. Each instance of the NP-Hard problem must be solved in each forward pass of the training loop, which poses clear obstructions in terms of runtime. A disadvantage of the approach is that cutting-plane methods are generally considered to be inferior in efficiency to staple methods like branch and bound. Improved results were obtained on portfolio optimization and diverse bipartite matching problems, when compared to LP-relaxation models following the approach of [Wilder et al. \[2019\]](#).

[Mandi et al. \[2020\]](#) investigates the application of the same SPO approach to NP-Hard combinatorial problems. Primary among the challenges faced in this context is, as in [Ferber et al., 2020](#), the computational cost of solving hard problems within every iteration of the training. The authors found that continuous relaxations of common MILP problems (e.g. knapsack) often offer subgradients of comparable quality to the full mixed-integer problem with respect to the SPO loss, so that training end-to-end systems with hard CO problems can be simplified in such cases by replacing the full problem with an efficiently solvable relaxation, an approach termed *SPO-relax*. The authors put continuous relaxations into the broader category of “weaker oracles” for the CO solver, which also includes approximation algorithms (e.g. greedy approximation for knapsack). The main results showed that SPO-relax achieves accuracy competitive with the full SPO approach but with shorter training times on a handful of discrete problems. The SPO-relax approach was compared also against the formulation of [Wilder et al. \[2019\]](#) on equivalent relaxations, but no clear winner was determined.

[Pogančić et al. \[2019\]](#) introduced a new idea for approximating gradients over discrete optimization problems for end-to-end training, which relies on viewing a discrete optimization problem as a function of its defining parameters (in this context coming from previous layers), whose range is piecewise constant. The only requirement is that the objective be linear. The gradient calculation combines the outputs of two calls to an optimization solver; one in the forward pass on initial parameters  $\mathbf{y}$ , and one in the backward pass on perturbed parameters  $\bar{\mathbf{y}}$ . The results are used to construct a piecewise linear function which approximates the original solver’s output space, but has readily available gradients. Because the gradient calculation is agnostic to the implementation of the solver, it is termed “black-box differentiation”. As such, input parameters to the solver do not correspond explicitly to coefficients in the underlying optimization problem. Results on end-to-end learning for the shortest path problem, TSP and min-cost perfect matching are shown. In each case, the discrete problem’s specification is implicitly defined in terms of images, which are used to predict parameters of the appropriate discrete problem through deep networks. The optimal solutions coming from blackbox solvers are expressed as binary indicator matrices in each case and matched to targeted optimal solutions using a Hamming distance loss function.

Finally, [Wang et al. \[2019\]](#) presented a differentiable solver for the MAXSAT problem, another problem form capable of representing NP-Hard combinatorial problems. Approximate

gradients are formed by first approximating the MAXSAT problem as a related semidefinite program (SPD), then differentiating its solution analytically during a specialized coordinate descent method [\[Wang et al., 2018\]](#) which solves the SDP. The successful integration of MAXSAT into deep learning is demonstrated with a model trained to solve sudoku puzzles represented only by handwritten images.

## 8 Challenges and Research Directions

The current state of the art in integrating combinatorial optimization with end-to-end machine learning shows promise on challenging tasks for which there was previously no viable approach. Further, it has been demonstrated that a variety of non-overlapping approaches can be effective. Despite these encouraging results, a number of challenges remain that must be addressed to allow an integration that lives up to its full potential. (1) Despite the variety of approaches, the success of the predict-and-optimize paradigm has been demonstrated on a relatively limited set of optimization problems and a majority of reported experimental results focus on linear programming formulations. Challenges posed by the parametrization of constraints stand in the way of broader applications, but have not been yet been addressed. (2) Issues associated with the runtime of combinatorial solvers in-the-loop still make some potential applications impractical. (3) Additionally, despite being possible in theory, the role of the CO model in-the-loop has not been generalized successfully beyond being applied as the final layer of a deep model. The use of additional layers beyond the solver, or even compositions of CO solving layers, could potentially lead to new applications if the practical challenges were to be overcome. (4) In predicting solutions to CO problems, the current methods cannot reliably guarantee the problem constraints to be satisfied. This critical shortcoming may be addressed by integrating ML approaches with methods from the robust optimization literature or by developing ad-hoc layers to project the predictions onto the feasible space. (5) While it has been observed in limited contexts [Demirović et al. \[2019\]](#) that predict-and-optimize frameworks based on optimization layers are competitive only when the underlying constrained problem is convex, this area still lacks theoretical results providing guarantees on their viability or performance. Finally, (6) the need for uniform benchmark experiments and systematic comparisons between each predict-and-optimization framework is apparent. [Demirović et al. \[2019\]](#) provided a study comparing the approaches of [Wilder et al. \[2019\]](#) and [Elmachtoub and Grigas \[2020\]](#), along with problem-specific approaches, on knapsack problems but did not conclude strongly as to which method should be preferred. Further, [Mandi et al. \[2020\]](#) reports that for knapsack problems, SPO performs comparably on the knapsack problem whether the LP relaxation or the full problem is used, but does not show that this effect generalize to other NP-Hard problems. This signals a need for comprehensive studies that test performance on a variety of hard CO problems.

Although the approaches surveyed are still in an early stage of their development, and have been adopted mainly for academic purposes, we strongly believe that the integration be-



tween combinatorial optimization and machine learning is a promising direction for the development of new, transformative, tools in combinatorial optimization and learning.

## Acknowledgments

This research is partially supported by NSF grant 2007164. Its views and conclusions are those of the authors only and should not be interpreted as representing the official policies, either expressed or implied, of the sponsoring organizations, agencies, or the U.S. government.

## References

- Brandon Amos and J Zico Kolter. Optnet: Differentiable optimization as a layer in neural networks. In *International Conference on Machine Learning*, pages 136–145. PMLR, 2017.
- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate, 2016.
- Maria-Florina Balcan, Travis Dick, Tuomas Sandholm, and Ellen Vitercik. Learning to branch. *arXiv:1803.10150*, 2018.
- Irwan Bello, Hieu Pham, Quoc V. Le, Mohammad Norouzi, and Samy Bengio. Neural combinatorial optimization with reinforcement learning. *arXiv:1611.09940*, 2017.
- Yoshua Bengio, Andrea Lodi, and Antoine Prouvost. Machine learning for combinatorial optimization: a methodological tour d’horizon. *European Journal of Operational Research*, 2020.
- Quentin Berthet, Mathieu Blondel, Olivier Teboul, Marco Cuturi, Jean-Philippe Vert, and Francis Bach. Learning with differentiable perturbed optimizers. *arXiv:2002.08676*, 2020.
- Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- Hanjun Dai, Elias B. Khalil, Yuyu Zhang, Bistra Dilkina, and Le Song. Learning combinatorial optimization algorithms over graphs. *arXiv:1704.01665*, 2018.
- George B Dantzig. Maximization of a linear function of variables subject to linear inequalities. *Activity analysis of production and allocation*, 13:339–347, 1951.
- Emir Demirović, Peter J Stuckey, James Bailey, Jeffrey Chan, Chris Leckie, Kotagiri Ramamohanarao, and Tias Guns. An investigation into prediction+ optimisation for the knapsack problem. In *International Conference on Integration of Constraint Programming, Artificial Intelligence, and Operations Research*, pages 241–257. Springer, 2019.
- Fabrizio Detassis, Michele Lombardi, and Michela Milano. Teaching the old dog new tricks: Supervised learning with constraints. *arXiv:2002.10766*, 2020.
- Priya L. Donti, Brandon Amos, and J. Zico Kolter. Task-based end-to-end model learning in stochastic optimization. *arXiv:1703.04529*, 2019.
- Adam N. Elmachtoub and Paul Grigas. Smart “predict, then optimize”. *arXiv:1710.08005*, 2020.
- Aaron Ferber, Bryan Wilder, Bistra Dilkina, and Milind Tambe. Mipaal: Mixed integer program as a layer. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2020.
- Ferdinando Fioretto, Pascal Van Hentenryck, Terrence W. K. Mak, Cuong Tran, Federico Baldo, and Michele Lombardi. Lagrangian Duality for Constrained Deep Learning. In *European Conference on Machine Learning and Knowledge Discovery in Databases ECML PKDD*, volume 12461 of *Lecture Notes in Computer Science*, pages 118–135. Springer, 2020.
- Ferdinando Fioretto, Terrence W.K. Mak, and Pascal Van Hentenryck. Predicting ac optimal power flows: Combining deep learning and lagrangian dual methods. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2020.
- Matteo Gagliolo and Jürgen Schmidhuber. Learning restart strategies. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 792–797, 2007.
- Maxime Gasse, Didier Chételat, Nicola Ferroni, Laurent Charlin, and Andrea Lodi. Exact combinatorial optimization with graph convolutional neural networks. *arXiv:1906.01629*, 2019.
- Prateek Gupta, Maxime Gasse, Elias B. Khalil, M. Pawan Kumar, Andrea Lodi, and Yoshua Bengio. Hybrid models for learning to branch. *arXiv:2006.15212*, 2020.
- Magnus R Hestenes. Multiplier and gradient methods. *Journal of optimization theory and applications*, 4(5):303–320, 1969.
- John Hopfield and D Tank. Neural computation of decisions in optimization problems. *Biological cybernetics*, 52:141–52, 02 1985.
- J.J. Hopfield. *Neural networks and physical systems with emergent collective computational abilities*, pages 7–19. 01 2018.
- Chaitanya K Joshi, Thomas Laurent, and Xavier Bresson. An efficient graph convolutional network technique for the travelling salesman problem. *arXiv preprint arXiv:1906.01227*, 2019.
- Elias Khalil, Pierre Le Bodic, Le Song, George Nemhauser, and Bistra Dilkina. Learning to branch in mixed integer programming. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.
- Elias B. Khalil, Bistra Dilkina, George L. Nemhauser, Shabbir Ahmed, and Yufen Shao. Learning to run heuristics in tree search. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 659–666, 2017.
- Wouter Kool, Herke van Hoof, and Max Welling. Attention, learn to solve routing problems! *arXiv:1803.08475*, 2019.

- Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521:436–444, 2015.
- Andrea Lodi and Giulia Zarpellon. On learning and branching: a survey. *Top*, 25(2):207–236, Jul 2017.
- Qiang Ma, Suwen Ge, Danyang He, Darshan Thaker, and Iddo Drori. Combinatorial optimization by graph pointer networks and hierarchical reinforcement learning. *arXiv preprint arXiv:1911.04936*, 2019.
- Jayanta Mandi and Tias Guns. Interior point solving for lp-based prediction-optimisation. *arXiv:2010.13943*, 2020.
- Jayanta Mandi, Peter J Stuckey, Tias Guns, et al. Smart predict-and-optimize for hard combinatorial optimization problems. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 34, pages 1603–1610, 2020.
- Mohammadreza Nazari, Afshin Oroojlooy, Lawrence V Snyder, and Martin Takáč. Reinforcement learning for solving the vehicle routing problem. *arXiv preprint arXiv:1802.04240*, 2018.
- Eugene Ndiaye, Olivier Fercoq, Alexandre Gramfort, and Joseph Salmon. Gap safe screening rules for sparsity enforcing penalties. *The Journal of Machine Learning Research*, 18(1):4671–4703, 2017.
- Yeesian Ng, Sidhant Misra, Line A Roald, and Scott Backhaus. Statistical learning for dc optimal power flow. In *2018 Power Systems Computation Conference (PSCC)*, pages 1–7. IEEE, 2018.
- Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.
- Alex Nowak, Soledad Villar, Afonso S. Bandeira, and Joan Bruna. Revised note on learning algorithms for quadratic assignment with graph neural networks, 2018.
- Marin Vlastelica Pogančić, Anselm Paulus, Vit Musil, Georg Martius, and Michal Rolínek. Differentiation of black-box combinatorial solvers. In *International Conference on Learning Representations*, 2019.
- Michael JD Powell. A method for nonlinear constraints in minimization problems. *Optimization*, pages 283–298, 1969.
- Eléa Prat and Spyros Chatzivasileiadis. Learning active constraints to efficiently solve bilevel problems. *arXiv preprint arXiv:2010.06344*, 2020.
- Francesca Rossi, Peter Van Beek, and Toby Walsh. *Handbook of constraint programming*. Elsevier, 2006.
- Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE transactions on neural networks*, 20(1):61–80, 2008.
- Jialin Song, ravi lanka, Yisong Yue, and Bistra Dilkina. A general large neighborhood search framework for solving integer linear programs. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 20012–20023. Curran Associates, Inc., 2020.
- Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.
- Yunhao Tang, Shipra Agrawal, and Yuri Faenza. Reinforcement learning for integer programming: Learning to cut. In *International Conference on Machine Learning*, pages 9367–9376. PMLR, 2020.
- Cuong Tran, Ferdinando Fioretto, and Pascal Van Hentenryck. Differentially private and fair deep learning: A lagrangian dual approach. *arXiv:2009.12562*, 2020.
- Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.
- Alexandre Velloso and Pascal Van Hentenryck. Combining deep learning and optimization for security-constrained optimal power flow. *arXiv:2007.2007.07002*, 2020.
- Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. Pointer networks. *arXiv:1506.03134*, 2017.
- Po-Wei Wang, Wei-Cheng Chang, and J. Zico Kolter. The mixing method: low-rank coordinate descent for semidefinite programming with diagonal constraints. *arXiv:1706.00476*, 2018.
- Po-Wei Wang, Priya Donti, Bryan Wilder, and Zico Kolter. Satnet: Bridging deep learning and logical reasoning using a differentiable satisfiability solver. In *International Conference on Machine Learning*, pages 6545–6554. PMLR, 2019.
- Bryan Wilder, Bistra Dilkina, and Milind Tambe. Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 33, pages 1658–1665, 2019.
- Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256, 1992.
- GV Wilson and GS Pawley. On the stability of the travelling salesman problem algorithm of hopfield and tank. *Biological Cybernetics*, 58(1):63–70, 1988.
- Jie Zhou, Ganqu Cui, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications. *arXiv preprint arXiv:1812.08434*, 2018.