Distributionally Robust Max Flow

Louis Chen[*] Will Ma[†] James Orlin[‡] David Simchi-Levi[§]

Abstract

In this paper, we extend the study of Chen et al. (2018) to the problem of distributionally robust network design. In this problem, the decision maker is to decide on the prepositioning of resources on arcs in a given s-t flow network in anticipation of an adversary's selection of a probability distribution for the arc capacities that seeks to minimize the expected max flow. The adversary's selection is limited to those distributions that are couplings of given arc capacity distributions, one for each arc. We find that modeling the uncertainty in this way is certainly more sensible than prescribing the stochastic behavior of the arc capacities across an entire network. Furthermore, while it is in fact #P-Hard to compute even the expectation with respect to the independent coupling of the stochastic arc capacities, we show that we can efficiently solve the distributionally robust network design problem. Indeed, this particular problem satisfies the sufficiency condition for tractability that we proposed in the previous work. But what's more, a highlight and extension in this work is to take advantage of the network setting to go even further and efficiently solve for the distribution the adversary responds with.

1 Max Flow with Random Arc Capacities

In the max flow problem (cf. Ahuja, Magnanti and Orlin (1993)), we are given a graph $G = (N, A)$. N is the node set consisting of at least two distinct elements—the *source* s and the *sink* t. A is the arc set consisting of ordered pairs of the form (i, j), which indicates that flow can be directed from i to j, for distinct nodes $i, j \in N$. For each arc $(i, j) \in A$, we let $\tilde{u}_{ij} \geq 0$ denote the random *capacity* of that arc (an upper bound on the flow that can be directed from node i to node j); we

[*]Operations Research Center, Massachusetts Institute of Technology, `llchen@mit.edu`.
[†]Operations Research Center, Massachusetts Institute of Technology, `willma@mit.edu`.
[‡]Sloan School of Management, Massachusetts Institute of Technology, `jorlin@mit.edu`.
[§]Institute for Data, Systems, and Society, Department of Civil and Environmental Engineering, and Operations Research Center, Massachusetts Institute of Technology, `dslevi@mit.edu`.

use $\tilde{u}$ to refer to the set $\{\tilde{u}_{ij} : (i,j) \in A\}$. We are given the marginal distribution μ_{ij} of $\tilde{u}_{ij}$ for each arc (i,j), and would like to evaluate the expected value of the maximum flow that can be directed from s to t under the worst-case correlation between $\{\mu_{ij} : (i,j) \in A\}$.

For a realization of the arc capacities $\tilde{u}$, the value of the max flow $Z(\tilde{u})$ is given by the optimal objective value of the following LP.

$$\max v \tag{MaxFlow-1}$$

$$\text{s.t.} \sum_{j:(i,j)\in A} x_{ij} - \sum_{j:(j,i)\in A} x_{ji} = \begin{cases} v, & i = s \\ 0, & i \in N \setminus \{s,t\} \\ -v, & i = t \end{cases} \quad \forall i \in N$$

$$0 \le x_{ij} \le \tilde{u}_{ij},\ (i,j) \in A$$

We are interested in studying the value of the worst-case expected max-flow,

$$\inf_{\theta \in \Gamma(\{\mu_{ij}:(i,j)\in A\})} \mathbb{E}_{\tilde{u}\sim\theta}[Z(\tilde{u})]. \tag{*}$$

Towards this goal, observe that $Z(\tilde{u})$ can be written as a discrete minimization problem (the *min cut* problem) over the set $\mathcal{C}$ of incidence vectors to (s-t) cut-sets in the graph. Further, $conv(\mathcal{C})$ is a 0-1, integral polyhedron.

The first main result is a linear program formulation.

Theorem 1.1. *The expected value of the max flow under the worst-case correlation, defined as* $\inf_{\theta \in \Gamma(\{\mu_{ij}:(i,j)\in A\})} \mathbb{E}_{\tilde{u}\sim\theta}[Z(\tilde{u})]$*, is equal to the optimal objective value of the following problem.*

$$\max_{x,v,w} \left(v - \sum_{(i,j)\in A} \int_{\tilde{u}_{ij}} \max\{w_{ij} - \tilde{u}_{ij}, 0\} d\mu_{ij} \right) \tag{MaxFlow-2}$$

$$s.t. \sum_{j:(i,j)\in A} x_{ij} - \sum_{j:(j,i)\in A} x_{ji} = \begin{cases} v, & i = s \\ 0, & i \in N \setminus \{s,t\} \\ -v, & i = t \end{cases} \quad \forall i \in N$$

$$0 \le x_{ij} \le w_{ij},\ (i,j) \in A$$

Problem (MaxFlow-2) resembles the original LP (MaxFlow-1), except now the arc capacities, instead of being fixed realizations $\tilde{u}_{ij}$, are decision variables w_{ij}. For each arc (i,j) there is a "penalty" term $\int_{\tilde{u}_{ij}} \max\{w_{ij} - \tilde{u}_{ij}, 0\} d\mu_{ij}$ in the objective function, dependent on the marginal distribution μ_{ij}, which dissuades w_{ij} from being as large as possible.

In general, the worst-case joint distribution θ between the random arc capacities will be high-dimensional and intractable to compute. Nonetheless, Corollary 1.1 shows that we can still evaluate the worst-case expected value of the max flow. Indeed, (MaxFlow-2) is tractable since the penalty terms are convex—in fact, if each μ_{ij} is given as a discrete distribution, then (MaxFlow-2) can be reformulated as a polynomial-sized LP.

2 Recovering the Worst-Case Coupling of Marginals with Finite Support

In Theorem 1.1, we showed that the optimization problem (*) can essentially be written in the form of a max-cost flow problem. When solved, we find the worst-case expected value under the worst-case coupling of the random arc capacities, without actually even having the worst-case coupling on-hand. In this subsection, we demonstrate that when the random arc capacities are discrete random variables with finite support, we can in fact find the worst-case coupling.

2.1 A Lagrangian

Let $supp(\tilde{u}_{ij}) \coloneqq \{u_{ij}^1, \ldots, u_{ij}^{m_{ij}}\}$ (values written in increasing order), with respective probabilities $\{p_{ij}^1, \ldots, p_{ij}^{m_{ij}}\}$, and let $\mathcal{X}_{cut}$ denote the set consisting of any vector χ that is the 0-1 characteristic vector to some (s-t) cut-set in the digraph G. For $w \in \mathbb{R}^A$ and $\nu \in \mathcal{P}(\mathcal{X}_{cut})$, let us define the following Lagrangian function

$$L(\nu, w) \coloneqq \left(\sum_{(i,j)\in A} w_{ij} \cdot \Pi_{ij}\nu(1) - \sum_{(i,j)\in A} \sum_{k=1}^{m_{ij}} \max\{w_{ij} - u_{ij}^k, 0\} \cdot p_{ij}^k \right).$$

We can establish that problem (*) is equivalent to the max-min/min-max pair of problems that form a primal-dual pair.

$$\max_{w} \min_{\nu \in \mathcal{P}(\mathcal{X}_{cut})} L(\nu, w) = \text{(Problem MaxFlow-2)} \qquad \text{(Lagrangian Primal)}$$

and

$$\min_{\nu \in \mathcal{P}(\mathcal{X}_{cut})} \{q(\nu) := \max_{w} L(\nu, w) = \sum_{(i,j)\in A} \sum_{k=1}^{k_{ij}-1} u_{ij}^{k} \cdot p_{ij}^{k} + u_{ij}^{k_{ij}} \left(\Pi_{ij}\nu(1) - \sum_{\tau=1}^{k_{ij}-1} p_{ij}^{\tau} \right) \}, \quad \text{(Lagrangian Dual)}$$

where $k_{ij} := \max\{k \in [m_{ij}] : \Pi_{ij}\nu(1) - \sum_{\tau=1}^{k-1} p_{ij}^{\tau} \geq 0\}$.

Remark 2.1. At this juncture, we note that the objective function q in Lagrangian Dual is an expectation with respect to a very particular coupling. We describe the coupling at an intutiive level as follows: If $\chi \in \mathcal{X}_{cut}$, let χ_{ij} denote whether or not arc (i,j) is in the cut that χ represents. Now let $\tilde{\chi} \sim \nu^* \in \mathcal{P}(\mathcal{X}_{cut})$, and let $\Pi_{ij}\nu^*(1)$ denote the probability that $\tilde{\chi}_{ij} = 1$. And we'll let $\Pi_{ij}\nu^*(0)$ denote the probability that $\tilde{\chi}_{ij} = 0$. For each arc (i,j) define the conditional distribution $\tilde{u}_{ij}|_{\tilde{\chi}_{ij}=1}$ as the $\tilde{\chi}_{ij}$ distribution conditioned on the "bottom $\Pi_{ij}\nu^*(1)$ -values". For each arc (i,j) define the conditional distribution $\tilde{u}_{ij}|_{\tilde{\chi}_{ij}=0}$ as the $\tilde{u}_{ij}$ distribution conditioned on the "top $\Pi_{ij}\nu^*(0)$ -values". Then, the worst-case joint distribution $\tilde{u}$ can be described as follows: Draw a cut $\tilde{\chi}$ according to ν^*, and denote the realization by χ; then, for each (i,j), draw a $\tilde{u}_{ij}$ according to $\tilde{u}_{ij}|_{\tilde{\chi}_{ij}=\chi_{ij}}$. Hence, $q(\nu)$ can alternatively be written as an expectation wrt this coupling as:

$$q(\nu) = E_{\tilde{\chi}\sim\nu}\left[E\left[\sum_{(i,j)\in A} \tilde{u}_{ij} \cdot \tilde{\chi}_{ij} || \tilde{\chi} \right]\right]$$

As we are given that $\inf_{\theta \in \Gamma(\{\mu_{ij}:(i,j)\in A\})} \mathbb{E}_{\tilde{u}\sim\theta}[\min_{\chi\in\mathcal{X}_{cut}} \sum_{(i,j)\in A} \tilde{u}_{ij} \cdot \chi_{ij}]$ equals $\min_{\nu\in\mathcal{P}(\mathcal{X}_{cut})} q(\nu)$, it then suffices to solve for a distribution over cut-sets in order to find the worst-case coupling. △

2.2 An Alternative Network Formulation Dual

Via the Lagrangian object introduced in the last section, we were able to establish a dual problem to Lagrangian Primal, equivalently, problem (*). In this section, we show that the network structure affords us yet another problem that is dual to Lagrangian Primal, whose connections to Lagrangian Dual prove instrumental in our analysis. To obtain this dual problem that we term LP Dual, observe first that Lagrangian Primal can be simplified to:

$$\max_{w,w_{ts}} \quad w_{ts} - \sum_{(i,j)\in A}\sum_{k=1}^{m_{ij}} \max\{w_{ij} - u_{ij}^k, 0\}\cdot p_{ij}^k$$

$$\text{subject to} \quad w, w_{ts} \text{ satisfy flow balance at all nodes}$$

$$w \geq 0$$

(Add arc (t,s) with cost +1, and upper capacity = $+\infty$)

At this point, we have an Uncapacitated max-cost flow problem w/ piecewise linear concave arc profits. From here, we can perform the standard transformation to a capacitated max-cost flow problem with linear arc costs by replacing each arc $(i,j) \in A$ with m_{ij} parallel arcs. More precisely, for any $(i,j) \in A$, in place of arc (i,j) we now have m_{ij} parallel arcs directed from i to j, where for any $k \in \{1,\ldots,m_{ij}\}$, the k-th parallel arc has profit equal to $-\sum_{\tau=1}^{k-1} p_{ij}^{\tau}$ and upper capacity equal to $u_{ij}^k - u_{ij}^{k-1}$. **Note:** $u_{ij}^0 := 0$ for any $(i,j) \in A$. And we leave the arc (t,s) as is. In addition, for the sake of analysis, we will also add the following arcs wherein it will never be profitable to have nonzero flow:

- For any node $i \notin \{s,t\}$, add an arc from i pointed to t, with cost = -1, capacity = $+\infty$
- For any node $i \notin \{s,t\}$, add an arc from t pointed to i, with cost = 0, capacity = $+\infty$

With this equivalent capacitated max-cost flow problem just designed, the interest turns towards its linear programming dual:

$$\begin{aligned}
\min_{\pi,\lambda} \quad & \sum_{(i,j)\in A}\sum_{k=1}^{m_{ij}} (u_{ij}^k - u_{ij}^{k-1})\cdot \lambda_{ij}^k \\
\text{subject to} \quad & 1 + \pi_t - \pi_s = 0 \\
& -\sum_{\tau=1}^{k-1} p_{ij}^{\tau} + \pi_i - \pi_j \leq \lambda_{ij}^k \qquad (i,j)\in A, k = 1,\ldots,m_{ij} \\
& \pi_t - \pi_i \geq -1; \quad \forall i \in N \setminus \{s,t\} \\
& \pi_i - \pi_t \geq 0; \quad \forall i \in N \setminus \{s,t\} \\
& \lambda_{ij}^k \geq 0; \quad (i,j)\in A, k = 1,\ldots,m_{ij} \\
& \pi_i \text{ free} \qquad i \in N
\end{aligned}$$

(Label: **LP Dual**)

Let us call the feasible region $LPDual_{feas}$, and note the following. If we further constrain

$\pi_t = 0$, the optimal value is not changed. Indeed, we can do this because for any fixed λ, $Proj_\pi(LPDual_{feas}) := \{\pi : (\pi, \lambda) \in LPDual_{feas}\}$ is translation-invariant in the "all-ones" direction. What's more, for any $\pi \in Proj_\pi(LPDual_{feas})$ that satisfies $\pi_t = 0$, necessarily $\pi_s = 1$ and $\pi_i \in [0,1]$ for all $i \in N$. So let us define

$$\Pi_{feasible} := \{\pi \in \mathbb{R}^N : \pi_s = 1, \pi_t = 0, \pi_i \in [0,1] \forall i \in N\},$$

and observe that for any $\pi \in \Pi_{feasible}$,

$$\min_{\lambda \in Proj_\lambda(LPDual_{feas})} \sum_{(i,j)\in A} \sum_{k=1}^{m_{ij}} (u_{ij}^k - u_{ij}^{k-1}) \cdot \lambda_{ij}^k = \sum_{(i,j)\in A} \sum_{k=1}^{m_{ij}} (u_{ij}^k - u_{ij}^{k-1}) \cdot \max(\pi_i - \pi_j - \sum_{\tau=1}^{k-1} p_{ij}^\tau, 0).$$

This expression, as a function of π, resembles Lagrangian Dual Objective and motivates the definition of an LP Dual Objective q', i.e, LP Dual becomes

$$\min_{\pi \in \Pi_{feasible}} \{q'(\pi) := \sum_{(i,j)\in A} \sum_{k=1}^{m_{ij}} (u_{ij}^k - u_{ij}^{k-1}) \cdot \max(\pi_i - \pi_j - \sum_{\tau=1}^{k-1} p_{ij}^\tau, 0)\} \qquad \textbf{(LP Dual)}$$

Written another way, q' takes the form

$$q'(\pi) = \sum_{(i,j)\in A} \left(\sum_{k=1}^{k'_{ij}-1} u_{ij}^k \cdot p_{ij}^k + u_{ij}^{k'_{ij}} \left(\pi_i - \pi_j - \sum_{\tau=1}^{k'_{ij}-1} p_{ij}^\tau \right) \right) \mathbb{1}_{\pi_i - \pi_j > 0} \qquad \text{(LP Dual Objective)}$$

where $k'_{ij} := \max\{k \in [m_{ij}] : \pi_i - \pi_j - \sum_{\tau=1}^{k-1} p_{ij}^\tau \geq 0\}$.

2.3 Connecting Lagrangian Dual and LP Dual: Recovering the Worst-Case Coupling

The optimization problems LP Dual and Lagrangian Dual are equivalent in the sense that $\min_{\pi \in \Pi_{feasible}} q'(\pi) = \min_{\nu \in \mathcal{P}(\mathcal{X}_{cut})} q(\nu)$. LP Dual is a linear program and hence more tractable than the optimization problem over $\mathcal{P}(\mathcal{X}_{cut})$ in Lagrangian Dual. But, since it is of interest to find an optimal solution to Lagrangian Dual so that we may construct the worst-case correlation of arc capacities, merely solving LP Dual does not help. The following key observation reveals how indeed an optimal solution to LP Dual yields an optimal solution to Lagrangian Dual, and hence, the desired worst-case coupling of arc capacities.

Comparing the expressions $q(\nu)$ and $q'(\pi)$, we note that for $\nu \in \mathcal{P}(\mathcal{X}_{cut})$ and $\pi \in \Pi_{feasible}$,

$$\Pi_{ij}\nu(1) = (\pi_i - \pi_j) \cdot \mathbb{1}_{(\pi_i - \pi_j) > 0}$$

is a sufficient condition to obtain $q(\nu) = q'(\pi)$. The fact that for $\pi \in \Pi_{feasible}$, $\pi_i - \pi_j \leq 1$ for all $(i,j) \in A$ is encouraging and suggests there may exist a mapping $\pi \mapsto \nu_\pi$ that takes $\pi \in \Pi_{feasible}$ into some probability distribution ν_π over the set of cut-sets. With the goal of ensuring that π and ν_π satisfy the sufficient condition, let us consider the following mapping. Given $\pi \in \Pi_{feasible}$, order the range of π as $0 = \pi^{(0)} < \pi^{(1)} < \ldots < \pi^{(K)} = 1$, where K denotes the number of different values in the range of π. Then, for each $k \in \{0, 1, \ldots, K-1\}$, let $T_k := \{i \in N : \pi_i \leq \pi^{(k)}\}$, and $S_k := N \setminus T_k$. Consequently, each T_k contains node t and S_k contains node s, so that $C_k := (S_k, T_k)$ is an (s-t) cut with corresponding cut-set $A_k := \{(i,j) \in A : i \in S_k, j \in T_k\}$. Furthermore, $T_0 \subset T_1 \subset \ldots \subset T_{K-1}$, while $S_0 \supset S_1 \supset \ldots \supset S_{K-1}$; in other words, the collection of (s-t) cuts C_k is "nested". Let us define a probability distribution over this collection of (s-t) cuts via

$$C_k \ \ w.p. \ \ \pi^{(k+1)} - \pi^{(k)}, \qquad \forall k = 0, \ldots, K-1.$$

This distribution on (s-t) cuts induces a distribution ν_π over the set of cut-sets $\mathcal{X}_{cut}$; namely, if $C^\pi = (S^\pi, T^\pi)$ is a random (s-t) cut, distributed according to the above over $\{C_k\}_{k=1}^{K-1}$, then we correspondingly have a random cut-set A^π whose distribution over $\{A_k\}_{k=1}^{K-1}$ we will call ν_π.

Finally, given any arc $(i,j) \in A$,

- If $\pi_i - \pi_j > 0$, then

  $$j \in T^\pi, i \in S^\pi \iff [T^\pi = T_{k(j)}] \vee [T^\pi = T_{k(j)+1}] \vee \ldots \vee [T^\pi = T_{k(i)-1}],$$

  where k(j) is defined by $\pi_j = \pi^{(k(j))}$. Noting this, then the event $[(i,j) \in A]$ happens with probability $\Pi_{ij}\nu_\pi(1) = \sum_{\tau=k(j)}^{k(i)-1} \pi^{(\tau+1)} - \pi^{(\tau)} = \pi^{(k(i))} - \pi^{(k(j))} = \pi_i - \pi_j$.

- If $\pi_i - \pi_j < 0$, then $(i,j) \notin A_k$ for all $k \in \{0, 1, \ldots, K-1\}$. Hence, the event $[(i,j) \in A]$ happens with probability $\Pi_{ij}\nu_\pi(1) = 0$.

Thus, we have confirmed that this mapping $\pi \longmapsto \nu_\pi$ satisfies the sufficient condition, so that

$$q'(\pi) = q(\nu_\pi) \quad \forall \pi \in \Pi_{feasible}.$$

Thus, it suffices to solve LP Dual for an optimal solution π^* and perform the mapping above to obtain a probability distribution ν_{π^*} over (s-t) cut-sets that yields the worst-case coupling of arc capacities. Interestingly, along the way, the structure of our mapping revealed that it suffices to restrict the set of distributions over (s-t) cut-sets to the subset of distributions whose support satisfies the "nested" property described above.

References

S. Agrawal, Y. Ding, A. Saberi and Yinyu Ye. (2012). *Price of correlations in stochastic optimization.* Operations Research, 60(1), 150-162.

R. K. Ahuja, T. L. Magnanti and J. B. Orlin. (1993). *Network flows: theory, algorithms, and applications.* Prentice Hall.

F. Bach. (2018). *Submodular Functions: from Discrete to Continous Domains.* Mathematical Programming.

A. Ben-Tal, D. den Hertog, A. De Waegenaere, B. Melenberg and G. Rennen (2013). *Robust solutions of optimization problems affected by uncertain probabilities.* Management Science, 59, 341-357.

D. Bertsimas, X. V. Doan, K. Natarajan and C-P. Teo. (2010). *Models for minimax stochastic linear optimization problems with risk aversion.* Mathematics for Operations Research, 35(3), 580-602.

D. Bertsimas, K. Natarajan and C-P. Teo. (2004). *Probabilistic combinatorial optimization: Moments, semidefinite programming and asymptotic bounds.* SIAM Journal on Optimization, 15(1), 185-209.

D. Bertsimas, K. Natarajan and C-P. Teo. (2006). *Persistence in discrete optimization under data uncertainty.* Mathematical Programming, Series B 108(2-3), 251-274.

D. Bertsimas and I. Popescu. (2005). *Optimal inequalities in probability theory: A convex optimization approach.* SIAM Journal on Optimization, 15(3), 780-804.

J. R. Birge and M. J. Maddox. (1995). *Bounds on expected project tardiness.* Operations Research, 43(5), 838-850.

J. Blanchet, Y. Kang, F. Zhang and K. Murthy. (2017). *Data-driven optimal transport cost selection for distributionally robust optimization.*

H. L. Bodlaender, P. Gritzmann, V. Klee and J. Van Leeuwen. (1990). *Computational complexity of norm-maximization.* Combinatorica, 10(2), 203-225.

L. Chen, W. Ma, J. Orlin, D. Simchi-Levi, and Z. Yan. (2018). *Distributionally Robust Linear and Discrete Optimization with Marginals.* Submitted to Operations Research.

G. Carlier and I. Ekeland. (2010). *Matching for teams.* Economic Theory, 42(2), 397-418.

E. Delage and Y. Ye. (2010). *Distributionally robust optimization under moment uncertainty with application to data-driven problems.* Operations Research, 58(3), 595-612.

I. Ekeland and R. Temam. (1999). *Convex analysis and variational problems.* SIAM Classics in Applied Mathematics, Philadelphia.

P. M. Esfahani and D. Kuhn. (2017). *Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations.* Mathematical Programming, 1-52.

A. Galichon. (2016). *Optimal transport methods in economics.* Princeton University Press.

R. Gao and A. J. Kleywegt. (2016). *Distributionally robust stochastic optimization with Wasserstein distance.*

D. Gupta and B. Denton. (2008). *Appointment scheduling in health care: Challenges and opportunities.* IIE Transactions, 40(9), 800-819.

A. N. Kolmogorov and S. V. Fomin. (1970). *Introductory real analysis.* Translated and edited by R. A. Silverman. Englewood Cliffs, N.J: Prentice-Hall.

Q. Kong, C-Y. Lee, C-P. Teo and Z. Zheng. (2013). *Scheduling arrivals to a stochastic service delivery system using copositive cones.* Operations Research, 61(3), 711-726.

T. L. Lai and H. Robbins. (1976). *Maximally dependent random variables.* Proceedings of the National Academy of Sciences of the United States of America, 73(2), 286-288.

H-Y. Mak, Y. Rong, and J. Zhang. (2015). *Appointment scheduling with limited distributional information.* Management Science, 61(2), 316-334.

O. L. Mangasarian and T. H. Shiau. (1986). *A variable-complexity norm maximization problem.* SIAM Journal on Algebraic and Discrete Methods, 7(3), 455-461.

I. Meilijson (1991). *Sharp bounds on the largest of some linear combinations of random variables with given marginal distributions.* Probability in the Engineering and Informational Sciences, 5, 1-14.

I. Meilijson and A. Nadas (1979). *Convex majorization with an application to the length of critical paths.* Journal of Applied Probability, 16(3), 671-677.

V. K. Mishra, K. Natarajan, D. Padmanabhan, C-P. Teo and X. Li. (2014). *On theoretical and empirical aspects of marginal distribution choice models.* Management Science, 60(6), 1511-1531.

R. H. Möhring, A. S. Schulz, F. Stork and M. Uetz. (2001). *On project scheduling with irregular starting time costs.* Operations Research Letters, 28, 149154.

Murota. (2003). *Discrete Convex Analysis.* SIAM, Philadelphia.

A. Nadas (1979). *Probabilistic PERT.* IBM Journal of Research and Development, 23(3), 339-347.

K. Natarajan, D. Shi and K-C. Toh. (2017). *Bounds for random binary quadratic programs.* To appear in SIAM Journal on Optimization.

K. Natarajan, M. Song and C-P. Teo. (2009). *Persistency model and its applications in choice modeling.* Management Science, 55(3), 453-469.

B. Pass. (2015). *Multi-marginal optimal transport: Theory and applications.* ESAIM: Mathematical Modelling and Numerical Analysis, 49, 1771-1790.

B. Pass. (2012). *On the local structure of optimal measures in the multi-marginal optimal transportation problem.* Calculus of Variations, 43: 529-536.

S. T. Rachev and L. Rüschendorf. (1998). *Mass transportation problems: Volume I: Theory* Springer.

R. T. Rockafellar. (1997). *Convex analysis.* Princeton University Press.

H. L. Royden and P. M. Fitzpatrick. (2010). *Real analysis.* Fourth Edition, Prentice Hall.

F. Santambrogio. (2015). *Optimal transport for applied mathematicians.* Birkhäuser Basel.

H. Scarf. (1958). *A min-max solution of an inventory problem*, Studies in The Mathematical Theory of Inventory and Production. (K. Arrow, S. Karlin and H. Scarf, Eds.) 201-209, Stanford University Press, California.

A. Shapiro, D. Dentcheva and A. Ruszczyński. (2014). *Lectures on stochastic programming: Modeling and theory.* Second edition. MOS-SIAM Series on Optimization.

D. Simchi-Levi, X. Chen, J. Bramel. (2014). *The Logic of Logistics.* Third Edition, Springer.

G. Weiss. (1986). *Stochastic bounds on distributions of optimal value functions with applications to PERT, network flows and reliability.* Operations Research, 34(4), 595-605.

W. Wiesemann, D. Kuhn and M. Sim. (2014). *Distributionally robust convex optimization.* Operations Research, 62(6), 1358-1376.

C. Villani. (2003). *Topics in optimal transportation.* American Mathematical Society.

C. Villani. (2009). *Optimal transport: old and new.* Grundlehren der Mathematischen Wissenschaften, Fundamental Principles of Mathematical Sciences, 338, Springer-Verlag, Berlin.

W. I. Zangwill. (1966). *A deterministic multi-period production scheduling model with backlogging.* Management Science, 13(1), 105-119.

W. I. Zangwill. (1969). *A backlogging model and a multi-echelon model of a dynamic economic lot size production system - a network approach.* Management Science, 15(9), 506-527.