# New reformulations of distributionally robust stochastic shortest path problem

Abdel Lisser

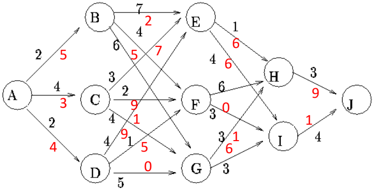
Laboratoire de Recherche en Informatique (LRI)
Université Paris Sud

Summer School Grenoble
Optimization under uncertainty
July 4-7, 2016

# Outline of the talk

- **Introduction** — problem formulation
- Stochastic shortest path problem
- Distributionally robust shortest path problem
  - New approximations
- Numerical results
- Copositive reformulation

## Shortest Path Problem




- $G = (V, A)$ is simple and acyclic digraph
- $V$ is the set of vertices
- $A$ is the set of arcs with cost (black) and delay (red)

## Introduction

- $G = (V, A)$: simple and acyclic digraph
- $s, t$: two given vertices (source/sink)
- $c(a)$: cost (deterministic) for each arc $a \in A$
- $\delta(a)$: random variables for the delay of the arc $a \in A$
- $D$: maximum delay allowed without penalty
- $d$: penalty for exceeding D (penalty unit time)
- Objective: **Find a path from s to t which minimizes the expectation of total costs**

## Introduction

- $G = (V, A)$: simple and acyclic digraph
- $s, t$: two given vertices (source/sink)
- $c(a)$: cost (deterministic) for each arc $a \in A$
- $\delta(a)$: random variables for the delay of the arc $a \in A$
- $D$: maximum delay allowed without penalty
- $d$: penalty for exceeding D (penalty unit time)
- Objective: **Find a path from s to t which minimizes the expectation of total costs**

# Outline of the talk

- Introduction — problem formulation
- **Stochastic shortest path problem**
- Distributionally robust shortest path problem
  - New approximations
- Numerical results
- Copositive reformulation

## Stochastic Shortest Path with Delay Excess Penalty (SSPP)

$$\min_{x \in \{0,1\}^{|A|}} \quad \mathbb{E}[\mathcal{J}(x,\delta)] := \sum_{a \in A} c(a)x_a + d \cdot \mathbb{E}[[\sum_{a \in A} \delta(a)x_a - D]^+]$$

$$\text{s.t.} \quad Mx = b$$

- $[x]^+ := max(0, x)$, $\mathbb{E}[X]$ expectation of a random variable $X$
- $M \in \mathbb{R}^{n \times |A|}$ is the *node-arc incidence matrix*
- $b \in \mathbb{R}^n$, all elements are 0 except the *s-th* and *t-th* element, which are 1 and -1, respectively.

## Assumptions

- $\delta(a)$ is normally distributed
- for two distinct arcs $a$ and $a'$, $\delta(a)$ and $\delta(a')$ are independent
- mean $\mu(a)$ and variance $\sigma^2(a)$ are known

### Assumptions

- $\delta(a)$ is normally distributed
- for two distinct arcs $a$ and $a'$, $\delta(a)$ and $\delta(a')$ are independent
- mean $\mu(a)$ and variance $\sigma^2(a)$ are known

## Deterministic Equivalent Reformulation of SSPD

$$\min_{x\in\{0,1\}^{|A|}} \sum_{a\in A} c(a)x_a + d\cdot[\hat{\sigma}\cdot f(\frac{D-\hat{\mu}}{\hat{\sigma}}) + (\hat{\mu}-D)[1-F(\frac{D-\hat{\mu}}{\hat{\sigma}})]]$$
$$\text{s.t. } Mx = b$$

- $\hat{\mu} := \sum_{a\in A}\mu(a)x_a$
- $\hat{\sigma} := \sqrt{\sum_{a\in A}\sigma^2(a)x_a^2}$
- $f(\cdot)$ : probability density function of the standard normal distribution.
- $F(\cdot)$ : cumulative distribution function of the standard normal distribution.

## Deterministic Equivalent Reformulation of SSPD

$$\min_{x\in\{0,1\}^{|A|}} \sum_{a\in A} c(a)x_a + d\cdot[\hat{\sigma}\cdot f(\frac{D-\hat{\mu}}{\hat{\sigma}}) + (\hat{\mu}-D)[1-F(\frac{D-\hat{\mu}}{\hat{\sigma}})]]$$
$$\text{s.t. } Mx = b$$

- $\hat{\mu} := \sum_{a\in A} \mu(a)x_a$
- $\hat{\sigma} := \sqrt{\sum_{a\in A} \sigma^2(a)x_a^2}$
- $f(\cdot)$ : probability density function of the standard normal distribution.
- $F(\cdot)$ : cumulative distribution function of the standard normal distribution.

## Problem Solving Method

- Solve the problem by applying a **branch-and-bound** algorithm
- Get a lower bound by solving the **linear relaxation** of *SSPD*
- Use a **Stochastic Gradient Projection Method** to solve the relaxed *SSPD* associated with **Active set Methods**

## Solving the relaxed SSPD

The linear relaxation of SSPD, where $x \in \{0,1\}^{|A|}$ is replaced by $x \in [0,1]^{|A|}$), is a convex problem. We can solve it by using Gradient Projection Method.

## Problem Solving Method

- Solve the problem by applying a **branch-and-bound** algorithm
- Get a lower bound by solving the **linear relaxation** of *SSPD*
- Use a **Stochastic Gradient Projection Method** to solve the relaxed *SSPD* associated with **Active set Methods**

## Solving the relaxed SSPD

The linear relaxation of SSPD, where $x \in \{0,1\}^{|A|}$ is replaced by $x \in [0,1]^{|A|}$), is a convex problem. We can solve it by using Gradient Projection Method.

## Problem Solving Method

- Solve the problem by applying a **branch-and-bound** algorithm
- Get a lower bound by solving the **linear relaxation** of *SSPD*
- Use a **Stochastic Gradient Projection Method** to solve the relaxed *SSPD* associated with **Active set Methods**

## Solving the relaxed SSPD

The linear relaxation of SSPD, where $x \in \{0,1\}^{|A|}$ is replaced by $x \in [0,1]^{|A|}$), is a convex problem. We can solve it by using Gradient Projection Method.

## Problem Solving Method

- Solve the problem by applying a **branch-and-bound** algorithm
- Get a lower bound by solving the **linear relaxation** of *SSPD*
- Use a **Stochastic Gradient Projection Method** to solve the relaxed *SSPD* associated with **Active set Methods**

## Solving the relaxed SSPD

The linear relaxation of SSPD, where $x \in \{0,1\}^{|A|}$ is replaced by $x \in [0,1]^{|A|}$), is a convex problem. We can solve it by using Gradient Projection Method.

### References

J. Cheng, S. Kosuch, A. Lisser (2012):

- NP-hard
- Assume $\delta(a)$ is independently normally distributed
- Mean $\mu(a)$ and variance $\sigma^2(a)$ are known

### What happens if the probability distribution is unknown ?

- An answer can be given by **Distributionally robust optimization**

### References

J. Cheng, S. Kosuch, A. Lisser (2012):

- NP-hard
- Assume $\delta(a)$ is independently normally distributed
- Mean $\mu(a)$ and variance $\sigma^2(a)$ are known

### What happens if the probability distribution is unknown ?

- An answer can be given by **Distributionally robust optimization**

## References

J. Cheng, S. Kosuch, A. Lisser (2012):

- NP-hard
- Assume $\delta(a)$ is independently normally distributed
- Mean $\mu(a)$ and variance $\sigma^2(a)$ are known

What happens if the probability distribution is unknown ?

- An answer can be given by **Distributionally robust optimization**

## References

J. Cheng, S. Kosuch, A. Lisser (2012):

- NP-hard
- Assume $\delta(a)$ is independently normally distributed
- Mean $\mu(a)$ and variance $\sigma^2(a)$ are known

## What happens if the probability distribution is unknown ?

- An answer can be given by **Distributionally robust optimization**

## References

J. Cheng, S. Kosuch, A. Lisser (2012):

- NP-hard
- Assume $\delta(a)$ is independently normally distributed
- Mean $\mu(a)$ and variance $\sigma^2(a)$ are known

## What happens if the probability distribution is unknown ?

- An answer can be given by **Distributionally robust optimization**

## References

J. Cheng, S. Kosuch, A. Lisser (2012):

- NP-hard
- Assume $\delta(a)$ is independently normally distributed
- Mean $\mu(a)$ and variance $\sigma^2(a)$ are known

## What happens if the probability distribution is unknown ?

- An answer can be given by **Distributionally robust optimization**

## References

J. Cheng, S. Kosuch, A. Lisser (2012):

- NP-hard
- Assume $\delta(a)$ is independently normally distributed
- Mean $\mu(a)$ and variance $\sigma^2(a)$ are known

## What happens if the probability distribution is unknown ?

- An answer can be given by **Distributionally robust optimization**

# Outline of the talk

- Introduction — problem formulation
- Stochastic shortest path problem
- **Distributionally** robust shortest path problem
    - New approximations
- Numerical results
- Copositive reformulation

# Distributionally Robust SSPP

- SSPP requires a strong assumption which is the knowledge of the exact information about the distribution $\tilde{F}$
- This is not often the case especially in practice.
- Distributionally robust optimization requires only a mild assumption on the probability distribution such as known supports, means and covariances.

We model the SSPP as distributionally robust SSPP as follows:

$$\text{(DRSSPP)} \quad \min_{x \in \{0,1\}^m} \quad c^T x + d \cdot \max_{F \in \mathcal{D}} \mathbb{E}_F [\tilde{\delta}^T x - D]^+ \tag{1a}$$

$$\text{s.t.} \;\; Mx = b \tag{1b}$$

where $\mathcal{D}$ is the collection of probability distributions of interest.

# Deterministic equivalent problem of a moment problem

We show how to come up with a deterministic equivalent problem of a simple example moment problem.

**Assumptions**

- (A1) There exists a deterministic Matrix $B \in \mathbb{R}^{m \times K}$, a known vector $\mu \in \mathbb{R}^m$ and random vector $\delta \in \mathbb{R}^K$ such that

$$\tilde{\delta} = B\delta + \mu$$

- (A2) The distributional uncertainty set accounts for information about the support $\mathcal{S}$, mean 0, and an upper bound $\Sigma$ on the covariance matrix of the random vector $\delta$

$$\mathcal{D}(\mathcal{S}, 0, \Sigma) = \left\{ F \in \mathcal{M} \;\middle|\; \begin{array}{l} Prob(x \in \mathcal{S}) = 1 \\ \mathbb{E}_F[\delta] = 0 \\ \mathbb{E}_F[\delta\delta^T] \preceq \Sigma \end{array} \right\}.$$

where $\mathcal{M}$ is the set of all probability measures on the measurable space $(\mathbb{R}^K, \mathcal{B})$, with $\mathcal{B}$ the Borel $\sigma$-algebra on $\mathbb{R}^K$.

# A deterministic equivalent problem

Under assumptions (A1) and (A2),

$$\max_{\tilde{F}\in\tilde{\mathcal{S}}} \mathbb{E}_{\tilde{F}} \quad [\sum_{a\in A} \tilde{\delta}(a)x_a - D]^+ \tag{2}$$

is equivalent to the problem

$$
\begin{aligned}
\text{(DR-PART)} \quad & \max_F \int_{\mathcal{S}} [(B\delta+\mu)^T x - D]^+ dF(\delta) \\
\text{s.t.} \quad & \int_{\mathcal{S}} dF(\delta) = 1 \\
& \int_{\mathcal{S}} \delta dF(\delta) = 0 \qquad \text{(3a)} \\
& \int_{\mathcal{S}} \delta\delta^T dF(\delta) \preceq \Sigma \\
& F \in \mathcal{M}
\end{aligned}
$$

# A deterministic equivalent problem

**Lemma 1**

*Under assumptions (A1) and (A2), together with $\mathcal{S} = \mathbb{R}^K$, the problem (3) is equivalent to the following deterministic problem*

$$\begin{aligned} \min \quad & \Sigma \bullet \mathbf{Q} + t \\ s.t. \quad & \begin{bmatrix} t + D - \mu^T x & \frac{(\mathbf{q} - B^T x)^T}{2} \\ \frac{\mathbf{q} - B^T x}{2} & \mathbf{Q} \end{bmatrix} \succeq 0 \\ & \begin{bmatrix} t & \frac{\mathbf{q}^T}{2} \\ \frac{\mathbf{q}}{2} & \mathbf{Q} \end{bmatrix} \succeq 0, \quad \mathbf{Q} \succeq 0, \quad \mathbf{q} \in \mathbb{R}^K, t \in \mathbb{R} \end{aligned} \tag{4a}$$

# A deterministic equivalent problem

**Proof**

- The main idea of the proof consists in transforming the distributionally robust objective function into a deterministic equivalent problem [Delage&Ye 2010].
- The proof is twofold :

  We establish the primal-dual relationship between Problem (3) and Problem (4).

  We show that the strong duality holds.

**Primal-dual relationship:** The Lagrangian function of Problem (3) is as follows:

$$\begin{aligned}\mathfrak{L}\{r,q,Q\} &= \int_{\mathcal{S}} [(B\delta+\mu)^T x - D]^+ dF(\delta) + r(1 - \int_{\mathcal{S}} dF(\delta)) \\ &\quad -\mathbf{q}^T \int_{\mathcal{S}} \delta dF(\delta) + \mathbf{Q} \bullet (\Sigma - \int_{\mathcal{S}} \delta\delta^T dF(\delta))\end{aligned}$$

where $\mathbf{Q} \succeq 0, \mathbf{q} \in \mathbb{R}^K, r \in \mathbb{R}$.

# A deterministic equivalent problem

**Primal-dual relationship**

The Lagrangian function of Problem (3) is as follows:

$$\begin{aligned}\mathfrak{L}\{r, q, Q\} &= \int_{\mathcal{S}} [(B\delta + \mu)^T x - D]^+ dF(\delta) + r(1 - \int_{\mathcal{S}} dF(\delta)) \\ &\quad - \mathbf{q}^T \int_{\mathcal{S}} \delta dF(\delta) + \mathbf{Q} \bullet (\Sigma - \int_{\mathcal{S}} \delta\delta^T dF(\delta))\end{aligned}$$

where $\mathbf{Q} \succeq 0, \mathbf{q} \in \mathbb{R}^K, r \in \mathbb{R}$.

# A deterministic equivalent problem

**Proof**

$$\mathfrak{L}\{r,q,Q\} = r + \mathbf{Q}\bullet\Sigma + \int_{\mathcal{S}}\{[(B\delta+\mu)^T x - D]^+ - r - \mathbf{q}^T\delta - \mathbf{Q}\bullet\delta\delta^T\}dF(\delta)$$

To maximize $\mathfrak{L}$ with $F \in \mathcal{M}$, we have

$$\max_{F\in\mathcal{M}} \mathfrak{L}\{r,q,Q\} = \min \quad r + \mathbf{Q}\bullet\Sigma + t$$
$$\text{s.t.} \quad t \geq [(B\delta+\mu)^T x - D]^+ - r - \mathbf{q}^T\delta - \mathbf{Q}\bullet\delta\delta^T, \forall\, \delta \in \mathcal{S}.$$

Precisely, the optimal $F$ is the dirac distribution with probability 1 on the point $\delta \in \mathcal{S}$ which maximizes $[(B\delta+\mu)^T x - D]^+ - r - \mathbf{q}^T\delta - \mathbf{Q}\bullet\delta\delta^T$.

# A deterministic equivalent problem

**Proof**

The dual of Problem (3) is as follows:

$$\min_{\mathbf{Q}\succeq 0,\mathbf{q}\in\mathbb{R}^K,r\in\mathbb{R}} \max_{F\in\mathcal{M}} \mathfrak{L} \quad = \quad \min \quad r + \mathbf{Q}\bullet\Sigma + t$$

$$\text{s.t.} \quad t \geq [(B\delta+\mu)^T x - D]^+ - r - \mathbf{q}^T\delta - \mathbf{Q}\bullet\delta\delta^T, \forall\,\delta\in\mathcal{S}$$

$$\mathbf{Q}\succeq 0, \mathbf{q}\in\mathbb{R}^K, r\in\mathbb{R}, t\in\mathbb{R}$$

Furthermore, it is equivalent to

$$\begin{aligned} \min \quad & \mathbf{Q}\bullet\Sigma + t \\ \text{s.t.} \quad & t \geq (B\delta+\mu)^T x - D - \mathbf{q}^T\delta - \mathbf{Q}\bullet\delta\delta^T, \forall\,\delta\in\mathcal{S} \\ & t \geq -\mathbf{q}^T\delta - \mathbf{Q}\bullet\delta\delta^T, \forall\,\delta\in\mathcal{S} \\ & \mathbf{Q}\succeq 0, \mathbf{q}\in\mathbb{R}^K, t\in\mathbb{R} \end{aligned}$$

# A deterministic equivalent problem

Equivalently,

$$\begin{array}{ll} \min & \mathbf{Q} \bullet \Sigma + t \\ \text{s.t.} & (1;\delta)^T \begin{bmatrix} t + D - \mu^T x & \frac{(\mathbf{q} - B^T x)^T}{2} \\ \frac{\mathbf{q} - B^T x}{2} & \mathbf{Q} \end{bmatrix} (1;\delta) \geq 0 \\ & (1;\delta)^T \begin{bmatrix} t & \frac{\mathbf{q}^T}{2} \\ \frac{\mathbf{q}}{2} & \mathbf{Q} \end{bmatrix} (1;\delta) \geq 0, \quad \mathbf{Q} \succeq 0, \quad \mathbf{q} \in \mathbb{R}^K, t \in \mathbb{R} \end{array}$$

As $\mathcal{S} = \mathbb{R}^K$, thus the dual is equivalent to

$$\begin{array}{ll} \min & \Sigma \bullet \mathbf{Q} + t \\ \text{s.t.} & \begin{bmatrix} t + D - \mu^T x & \frac{(\mathbf{q} - B^T x)^T}{2} \\ \frac{\mathbf{q} - B^T x}{2} & \mathbf{Q} \end{bmatrix} \succeq 0 \\ & \begin{bmatrix} t & \frac{\mathbf{q}^T}{2} \\ \frac{\mathbf{q}}{2} & \mathbf{Q} \end{bmatrix} \succeq 0, \quad \mathbf{Q} \succeq 0, \quad \mathbf{q} \in \mathbb{R}^K, t \in \mathbb{R} \end{array}$$

# A deterministic equivalent problem

**Strong duality**

- We can easily show that the conditions on $\Sigma$ to ensure that the dirac distribution lies in the relative interior set of Problem (3).
- Furthermore, based on the weaker version of Proposition 3.4 in Shaprio [2001], the strong duality exists.

### Assumptions

In the following, DRSSPP is considered under the following key assumption:

*Assumption* (A1): The distributional uncertainty set accounts for information about the support $\mathcal{S}$, mean $\mu$, and an upper bound $\Sigma$ on the covariance matrix of the random vector $\tilde{\delta}$

$$\mathcal{D}(\mathcal{S}, \mu, \Sigma) = \left\{ F \in \mathcal{M} \middle| \begin{array}{l} \mathbb{P}(\tilde{\delta} \in \mathcal{S}) = 1 \\ \mathbb{E}_F[\tilde{\delta}] = \mu \\ \mathbb{E}_F[(\tilde{\delta} - \mu)(\tilde{\delta} - \mu)^T] \preceq \Sigma \end{array} \right\}.$$

where $\mathcal{M}$ is the set of all probability distributions on the measurable space $(\mathbb{R}^m, \mathcal{B})$, with $\mathcal{B}$ the Borel $\sigma$-algebra on $\mathbb{R}^m$.

## Theorem 2

*Under assumption (A1), together with $\mathcal{S} = \mathbb{R}^m$, problem (1) is equivalent to the following deterministic problem.*

$$\text{(DRSSPP1)}: \min \quad c^T x + d \cdot ((\Sigma + \mu\mu^T) \bullet \mathbf{Q} + \mu^T \mathbf{q} + t) \tag{5a}$$

$$\begin{bmatrix} t + D & \frac{(\mathbf{q}-x)^T}{2} \\ \frac{\mathbf{q}-x}{2} & \mathbf{Q} \end{bmatrix} \succeq 0 \tag{5b}$$

$$\begin{bmatrix} t & \frac{\mathbf{q}^T}{2} \\ \frac{\mathbf{q}}{2} & \mathbf{Q} \end{bmatrix} \succeq 0, \tag{5c}$$

$$Mx = b \tag{5d}$$

$$x \in \{0,1\}^m, t \in \mathbb{R}, \mathbf{q} \in \mathbb{R}^m, \mathbf{Q} \in \mathbb{R}^{m \times m} \tag{5e}$$

*where $\bullet$ is the inner product defined by $A \bullet B = \sum_{i,j} A_{ij} B_{ij}$.*

- Apart from the binary constraint, problem (5) is an SDP problem which is theoretically solvable in polynomial time.
- However, in practice, solving the SDP problem is very time-consuming.
- In order to overcome this drawback, we propose a new efficient formulation without imposing any additional assumption.

## Theorem 3

*Under assumption (A1) and $\mathcal{S} = \mathbb{R}^m$, problem (1) is equivalent to*

$$
\begin{aligned}
\textit{(DRSSPP2):} \min \quad & c^T x + d \cdot ((x^T \Sigma x + (\mu^T x)^2) \cdot p_0 + \mu^T x \cdot q_0 + t) && \text{(6a)}\\
& \begin{bmatrix} t + D & \frac{q_0 - 1}{2} \\ \frac{q_0 - 1}{2} & p_0 \end{bmatrix} \succeq 0 && \text{(6b)}\\
& \begin{bmatrix} t & \frac{q_0}{2} \\ \frac{q_0}{2} & p_0 \end{bmatrix} \succeq 0, && \text{(6c)}\\
& Mx = b && \text{(6d)}\\
& x \in \{0,1\}^m, p_0, q_0, t \in \mathbb{R} && \text{(6e)}
\end{aligned}
$$

## Remark 1

**:** In problem (6), the dimension of the linear matrix inequality is $2 \times 2$, compared to $(m+1) \times (m+1)$ for the linear matrix inequality of problem (5).

**Proof.**

The proof consists of two parts: first, we establish the primal-dual relationship between problem (1) and problem (6). Second, we show that the strong duality holds.

□

- In the case where the mean $\mu$ and the covariance matrix $\Sigma$ are positively proportional to the cost of the arc and to $\mu\mu^T$ respectively, DRSSPP can be significantly simplified.
- There assumptions are realistic for some real world problems, e.g., in transportation networks the traffic delay is proportional to the length of the roads.

### Theorem 4

*If there exists $K \geq 0$ such that $\Sigma = K\mu\mu^T$. Then problem (1) is equivalent to*

$$\min_{x\in\{0,1\}^m, t,p,q\in\mathbb{R}} \quad c^T x + d\cdot((K+1)\cdot p + q + t) \tag{7a}$$

$$\begin{bmatrix} t+D & \frac{q-\mu^T x}{2} \\ \frac{q-\mu^T x}{2} & p \end{bmatrix} \succeq 0 \tag{7b}$$

$$\begin{bmatrix} t & \frac{q}{2} \\ \frac{q}{2} & p \end{bmatrix} \succeq 0 \tag{7c}$$

$$Mx = b \tag{7d}$$

$$x \in \{0,1\}^m, t, p, q \in \mathbb{R} \tag{7e}$$

### Remark 2

**:** In problem (7), the dimension of its linear matrix inequality is $2\times 2$, but also its objective function is linear.

# Relaxed approximation

- As DRSSPP is NP-hard, special interest is given to its relaxations.
- We present the relaxed approximations of the two deterministic formulations of problem (1) when the support of $\tilde{\delta}$ is the whole space, i.e., $\mathbb{R}^m$.
- For the first deterministic formulation DRSSPP1, there is a natural linear relaxation on binary variables $x$ which is as follows:

# DRSSPP1

$$\min \quad c^T x + d \cdot ((\Sigma + \mu\mu^T) \bullet \mathbf{Q} + \mu^T \mathbf{q} + t) \tag{8a}$$

$$\begin{bmatrix} t + D & \frac{(\mathbf{q}-x)^T}{2} \\ \frac{\mathbf{q}-x}{2} & \mathbf{Q} \end{bmatrix} \succeq 0 \tag{8b}$$

$$\begin{bmatrix} t & \frac{\mathbf{q}^T}{2} \\ \frac{\mathbf{q}}{2} & \mathbf{Q} \end{bmatrix} \succeq 0 \tag{8c}$$

$$Mx = b \tag{8d}$$

$$x \geq 0, t \in \mathbb{R}, \mathbf{q} \in \mathbb{R}^m, \mathbf{Q} \in \mathbb{R}^{m \times m} \tag{8e}$$

In this case, problem(8) is an SDP problem. It is easy to check that the optimal objective value of $DRSSPP1 - SDP$ is a lower bound of DRSSPP.

- If we take variables $p_0$, $q_0$ and $t$ as parameters, the second deterministic formulation of DRSSPP2 becomes a quadratic problem with binary constraints.
- Thus, we solve the problem by applying SDP relaxation methods.
- By introducing redundant constraints, we get the following SDP approximation of DRSSPP2 as follows:

# DRSSPP2-SDP

$$\min \quad c^T x + d \cdot (p_0(\Sigma + \mu\mu^T) \bullet \mathbf{X} + \mu^T x \cdot q_0 + t) \tag{9a}$$

$$\begin{bmatrix} t + D & \frac{q_0 - 1}{2} \\ \frac{q_0 - 1}{2} & p_0 \end{bmatrix} \succeq 0 \tag{9b}$$

$$\begin{bmatrix} t & \frac{q_0}{2} \\ \frac{q_0}{2} & p_0 \end{bmatrix} \succeq 0, \tag{9c}$$

$$M_i x = b_i, \ i = 1, \ldots, n \tag{9d}$$

$$M_i^T \mathbf{X} M_i = b_i^2, \quad X_{ii} = x_i, i = 1, \ldots, n \tag{9e}$$

$$\begin{bmatrix} 1 & x^T \\ x & \mathbf{X} \end{bmatrix} \succeq 0 \tag{9f}$$

$$t, p_0, q_0 \in \mathbb{R}, x \geq 0 \tag{9g}$$

where $M_i$ is the $i$-th row vector of the matrix $M$. As the binary quadratic terms are replaced by an SDP relaxation, then the optimal objective value of DRSSPP2-SDP is a lower bound of DRSSPP as well.

- When the variables $p_0$, $q_0$ and $t$ are fixed in DRSSPP2-SDP, we obtain an SDP problem which can be solved in polynomial time.
- When $x$ and $\mathbf{X}$ are fixed, DRSSPP2-SDP gives rise to another SDP problem.
- Thus, we can apply the alternating direction method which provides in this case a conservative approximation of DRSSPP2-SDP.

# Alternating Direction Method

Let $p_0 = \bar{p_0}, q_0 = \bar{q_0}$ and $t = \bar{t}$ such that constraints (9b) and (9c) are feasible, then DRSSPP2-SDP can be written as

$$(P(\bar{p_0}, \bar{q_0}, \bar{t})) : \min_{x,\mathbf{X}} \quad c^T x + d \cdot (\bar{p_0}(\Sigma + \mu\mu^T) \bullet \mathbf{X} + \mu^T x \cdot \bar{q_0} + \bar{t}) \tag{10a}$$

$$M_i x = b_i, \ i = 1, \dots, n \tag{10b}$$

$$M_i^T \mathbf{X} M_i = b_i^2, \quad X_{ii} = x_i, i = 1, \dots, n \tag{10c}$$

$$\begin{bmatrix} 1 & x^T \\ x & \mathbf{X} \end{bmatrix} \succeq 0 \tag{10d}$$

If we consider $x = \bar{x}$ and $\mathbf{X} = \bar{X}$ such that constraints (9d), (9e) and (9f) are feasible, then the second SDP problem formulation of DRSSPP2-SDP is

$$(P(\bar{x}, \bar{X})) : \min_{t, p_0, q_0 \in \mathbb{R}} \quad c^T \bar{x} + d \cdot (p_0(\Sigma + \mu\mu^T) \bullet \bar{X} + \mu^T \bar{x} \cdot q_0 + t) \tag{11a}$$

$$\begin{bmatrix} t + D & \frac{\mathbf{q}_0 - 1}{2} \\ \frac{\mathbf{q}_0 - 1}{2} & p_0 \end{bmatrix} \succeq 0 \tag{11b}$$

$$\begin{bmatrix} t & \frac{q_0}{2} \\ \frac{q_0}{2} & p_0 \end{bmatrix} \succeq 0 \tag{11c}$$

**Algorithm 1:** *Alternating Direction Method*

**Step 0.** Let $\epsilon \geq 0$ be a given numerical precision parameter and choose initial parameters $p_0 = p^0, q_0 = q^0$ and $t = t^0$ such that constraints (9b) and (9c) are satisfied. Set the iteration counter $k = 0$ and $f^0 = -\inf$.
**Step 1.** Solve the subproblem $P(p^k, q^k, t^k)$ and let $(x^{k+1}, X^{k+1})$ be the obtained optimal solutions while the optimal objective value is denoted by $f^{k+1}$ .
**Step 2.** If $f^k - f^{k+1} \leq \epsilon$, return $(x^{k+1}, X^{k+1}, p^k, q^k, t^k, f^{k+1})$ and stop.
**Step 3.** Solve the subproblem $P(x^{k+1}, X^{k+1})$ to obtain an optimal solution $(p^{k+1}, q^{k+1}, t^{k+1})$.
**Step 4.** Set $k := k + 1$ and go to **Step 1**.

**Theorem 5**

*If the problem DRSSPP2-SDP is bounded and has a feasible solution for the initial values of $p_0$,$q_0$ and $t$, then the sequence of the objective values $\{f^k\}$ generated by Algorithm 1 is nonincreasing. Moreover, the sequence $\{f^k\}$ converges to a finite limit and $f^k$ is an upper bound of DRSSPP2-SDP.*

- We consider three directed graphs for our numerical tests with $(|V|, |A|)$ equal to $(21, 39)$, $(30, 68)$ and $(40, 112)$ respectively.
- The input data for the models are randomly generated as follows.
- The cost $c$ is uniformly generated from $[0, 10]$.
- The mean $\mu$ is uniformly generated from the interval $[5, 10]$ and the covariance matrix $\Sigma$ is generated by the MATLAB function "gallery('randcorr',n)*2".
- The penalty $d$ is set to 5 and $D$ is set to the mean of the delay of the shortest path.

- We set the initial parameters for Algorithm 1 as follows: $\epsilon = 0.1$, $p_0 = \frac{1}{4D}$, $q_0 = 0$ and $t = 0$.
- For the sake of simplicity, the problems DRSSPP1-SDP and DRSSPP2-SDP are called hereafter original and modified approximations respectively.
- In order to compare the quality of our two relaxations, we use the branch-and-bound method to come up with the integer optimal solutions.
- The bound used in the branch-and-bound method corresponds to the original SDP relaxations.
- We denote the optimal values of the two SDP relaxations and the optimal solution obtained with the branch-and-bound method by $V^{SDP1}$, $V^{SDP2}$ and $V^{OPT}$, respectively.

# Outline of the talk

- Introduction — problem formulation
- Stochastic shortest path problem
- Distributionally robust shortest path problem
  - New approximations
- **Numerical** results
- Copositive reformulation

## Computational results of DRSSPP

The gap is defined by Gap $= \frac{V^{OPT} - V^{SDP}}{V^{OPT}} 100\%$.

| DATA | | | B& B | | Original | | | Modified | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Name | n | m | $V^{OPT}$ | CPU (s) | $V^{SDP1}$ | CPU (s) | Gap(%) | $V^{SDP2}$ | CPU (s) | Gap(%) |
| Inst1 | 21 | 39 | 38.17 | 37.1 | 36.15 | 6.3 | 5.29 | 38.17 | 3.0 | 0.00 |
| Inst2 | 21 | 39 | 32.03 | 36.7 | 31.40 | 4.7 | 1.97 | 32.03 | 1.7 | 0.00 |
| Inst3 | 21 | 39 | 38.38 | 35.7 | 36.06 | 4.7 | 6.04 | 38.28 | 2.7 | 0.26 |
| Inst4 | 21 | 39 | 35.37 | 36.9 | 33.67 | 4.8 | 4.81 | 35.37 | 2.2 | 0.00 |
| Inst5 | 21 | 39 | 31.24 | 31.6 | 30.73 | 4.6 | 1.63 | 31.24 | 3.3 | 0.00 |
| Inst6 | 30 | 68 | 137.63 | 1055.2 | 137.63 | 198.6 | 0.00 | 137.63 | 5.0 | 0.00 |
| Inst7 | 30 | 68 | 133.06 | 1159.7 | 133.06 | 238.8 | 0.00 | 133.06 | 2.8 | 0.00 |
| Inst8 | 30 | 68 | 140.81 | 1004.4 | 140.81 | 187.5 | 0.00 | 140.81 | 2.9 | 0.00 |
| Inst9 | 30 | 68 | 132.25 | 1148.3 | 132.25 | 167.0 | 0.00 | 132.25 | 2.9 | 0.00 |
| Inst10 | 30 | 68 | 131.79 | 1066.9 | 131.79 | 167.8 | 0.00 | 131.79 | 2.9 | 0.00 |
| Inst11 | 40 | 112 | 167.33 | 41383.4 | 166.65 | 4230.9 | 0.41 | 167.26 | 23.0 | 0.04 |
| Inst12 | 40 | 112 | 170.73 | 42696.1 | 168.18 | 3821.8 | 1.49 | 169.70 | 25.3 | 0.60 |
| Inst13 | 40 | 112 | 170.75 | 42284.9 | 170.75 | 4620.6 | 0.00 | 170.75 | 7.6 | 0.00 |
| Inst14 | 40 | 112 | 170.41 | 42326.7 | 170.41 | 4617.5 | 0.00 | 170.41 | 7.7 | 0.00 |
| Inst15 | 40 | 112 | 170.50 | 42631.2 | 167.84 | 4025.8 | 1.56 | 169.27 | 33.1 | 0.72 |

Table: DRSSPP Computational results

## Computational results of DRSSPP for large size graphs

The gap is defined by Gap $= \frac{V^{SDP2} - V^{SDP1}}{V^{SDP1}} 100\%$.

| DATA | | | Original | | Modified | | |
|---|---|---|---|---|---|---|---|
| Name | n | m | $V^{SDP1}$ | CPU (s) | $V^{SDP2}$ | CPU (s) | Gap(%) |
| Inst1 | 30 | 177 | 5.90 | 15510 | 5.90 | 7.7 | 0.00 |
| Inst2 | 45 | 190 | 10.96 | 22496 | 10.96 | 10.3 | 0.00 |
| Inst3 | 65 | 199 | 194.16 | 25118 | 194.16 | 14.9 | 0.00 |
| Inst4 | 65 | 206 | 214.23 | 29487 | 214.23 | 15.8 | 0.00 |
| Inst5 | 100 | 223 | 770.98 | 49513 | 770.98 | 47.5 | 0.00 |
| Inst6 | 100 | 481 | – | – | 144.43 | 127.3 | – |
| Inst7 | 100 | 753 | – | – | 17.09 | 365.0 | – |
| Inst8 | 100 | 999 | – | – | 10.85 | 1319.4 | |

Table: DRSSPP Computational results; "–" indicates that no solution was found because of lack of memory

### Stochastic optimization vs Distributionally robust optimization

Recall the Stochastic Shortest Path Problem (SSPP) as follows:

$$\text{(SSPP)} \quad \min_{x\in\{0,1\}^m} \quad c^T x + d \cdot \mathbb{E}_F[\tilde{\delta}^T x - D]^+ \tag{12a}$$

$$\text{s.t.} \;\; Mx = b \tag{12b}$$

where $F$ has a known mean and covariance matrix structure. We consider either a normal distribution or a log-normal distribution.

# SAA to evaluate the expectation function

- For SSPP, we propose to apply the Sample Average Approximation (SAA) method to solve it.
- The SAA problem is:

$$OPT_N = \min_{x \in \{0,1\}^m} \quad c^T x + d \cdot \frac{\sum_{k=1}^{N} [{\delta^k}^T x - D]^+}{N} \tag{13a}$$

$$\text{s.t. } Mx = b \tag{13b}$$

which is equivalent to the following mixed integer linear programming problem.

$$
\begin{aligned}
OPT_N = \min_{x \in \{0,1\}^m} \quad & c^T x + d \cdot \frac{\sum_{k=1}^N s_k}{N} && \text{(14a)}\\
\text{s.t. } \; & s_k \geq {\delta^k}^T x - D, \; s_k \geq 0, \; k = 1, \ldots, N && \text{(14b)}\\
& Mx = b && \text{(14c)}
\end{aligned}
$$

where the scenarios $\delta^1, \ldots, \delta^N$ are independent and sampled from the distribution $F$.

- We use the graph of size $(n, m) = (21, 39)$.
- In our numerical tests, we set the number of scenarios $N$ to 1000 for the SAA method.

# Comparison bw stochastic optimization and distributionally robust optimization

| | | Stochastic solutions | | Robust solutions |
|---|---|---|---|---|
| | | Normal | Log-Normal | Robust |
| Current distribution | Normal dist. | 29.32(0%) | 35.02(19%) | 32.80(12%) |
| | Log-normal dist. | 43.17(11%) | 38.73(0%) | 40.70(5%) |

Table: Comparison between stochastic and distributionally robust solutions.

# Outline of the talk

- Introduction — problem formulation
- Stochastic shortest path problem
- Distributionally robust shortest path problem
  - New approximations
- Numerical results
- **Copositive** reformulation

# Conic programs

## Conic Programs

- Let $\mathbb{K}$ be some closed convex cone.
- Primal Problem

$$(P)\begin{cases} \min & \langle C, X\rangle \\ \text{subject to} & \langle A_i, X\rangle = b_i, i = 1, \dots, m \\ & X \in \mathbb{K} \end{cases} \tag{15}$$

- Dual Problem

$$(D)\begin{cases} \max & \sum_{i=1}^{m} b_i y_i \\ \text{subject to} & C - \sum_{i=1}^{m} y_i A_i \in \mathbb{K}^* \end{cases} \tag{16}$$

- $\mathbb{K}^*$ is the dual cone: $\mathbb{K}^* := \{A \in S, \langle A_i, X\rangle \geq 0, \forall B \in \mathbb{K}\}$

# Well known cones

## Cones

- $\mathbb{K} = \mathbb{R}^n_+$
- $\mathbb{K} = \mathbb{L}$, the second order cone or Lorentz cone.
- $\mathbb{K} = \mathbb{S}^+$, the semidefinite cone.
- All are selfdual, i.e., $(\mathbb{R}^n_+)^* = \mathbb{R}^n_+$, $(\mathbb{L})^* = \mathbb{L}$, $(\mathbb{S}^+)^* = \mathbb{S}^+$
- Strong duality holds provided CQ is fulfilled.
- Efficient polynomial algorithms.
- What about the copositive cone ?

# Prelude

## Completely Positive Matrices

- Let $A = (a1, \ldots, ak)$ be a nonnegative $n \times k$ matrix, then $X = a_1 a_1^T + \ldots + a_k a_k^T = AA^T$ is called completely positive.
- $COP = \{X : X$ completely positive $\}$ is closed, convex cone.
- From the definition we get $COP = conv\{aa^T : a \geq 0\}$.

## Copositive Matrices

- Dual cone $COP^*$ of $COP$ in $S_n$ (sym. matrices):
  $Y \in COP^* \iff \mathrm{Tr}(XY) \geq 0 \quad \forall X \in COP$
  $\iff a^T Y a \geq 0 \quad \forall$ vectors $a \geq 0$.
- By definition, this means $Y$ is copositive.
  $CP = \{Y : a^T Y a \geq 0, \quad \forall a \geq 0\}$
- $X \notin CP$ is NP-complete decision problem.

### Semidefinite Programs vs Completely Positive Programs

- Semidefinite matrices : $PSD = \{X : a^T X a \geq 0, \quad \forall a\}$
- Copositive matrices : $CP = \{Y : a^T Y a \geq 0, \quad \forall a \geq 0\}$

### Semidefinite Programs vs Completely Positive Programs

- Semidefinite Programs : $\max\langle C, X\rangle$ s.t. $A(X) = b, X$ÂĄ $\in PSD$
- Copositive Programs : $\max\langle C, X\rangle$ s.t. $A(X) = b, X$ÂĄ $\in CP$

  $\max\langle C, X\rangle$ s.t. $A(X) = b, X \in COP$

We consider the case where the support is nonnegative, i.e., $\mathcal{S} = \mathbb{R}^m_+$. Before presenting the deterministic reformulation, we introduce the following lemma:

**Lemma 6**

*Given* $\mathbf{Q} \in \mathbb{R}^{m \times m}, \mathbf{q} \in \mathbb{R}^m, t \in \mathbb{R}$. *If matrix* $\mathbf{Q}$ *is positive semidefinite, then* $P_1 := \begin{bmatrix} t & \mathbf{q}^T \\ \mathbf{q} & \mathbf{Q} \end{bmatrix} \in \mathcal{CP}^{m+1}$ *is equivalent to*

$$\exists \mathbf{p} \in \mathbb{R}^m_+, P_2 := \begin{bmatrix} t & (\mathbf{q}-\mathbf{p})^T \\ \mathbf{q}-\mathbf{p} & \mathbf{Q} \end{bmatrix} \succeq 0,$$

*where* $\mathcal{CP}^m$ *is the cone of copositive matrices:*

$$\mathcal{CP}^m = \{M \in \mathcal{S}^m : x^T M x \geq 0 \text{ for all } x \in \mathbb{R}^m_+\}.$$

### Theorem 7

*Under assumption (A1), together with $\mathcal{S} = \mathbb{R}^m_+$, problem (1) is equivalent to the following deterministic problem*

$$\min_{\mathbf{Q}\succeq 0, p, q \in \mathbb{R}^m, t\in\mathbb{R}, x\in\{0,1\}^m} \quad c^T x + d\cdot((\Sigma+\mu\mu^T)\bullet\mathbf{Q} + \mu^T\mathbf{q} + t) \tag{17a}$$

$$\begin{bmatrix} t+D & \frac{(\mathbf{q}-x-\mathbf{p})^T}{2} \\ \frac{\mathbf{q}-x-\mathbf{p}}{2} & \mathbf{Q} \end{bmatrix} \succeq 0, \tag{17b}$$

$$\begin{bmatrix} t & \frac{\mathbf{q}-\lambda}{2}^T \\ \frac{\mathbf{q}-\lambda}{2} & \mathbf{Q} \end{bmatrix} \succeq 0, \quad \lambda \in \mathbb{R}^m_+ \tag{17c}$$

$$Mx = b \tag{17d}$$

# Conclusion

- Comparison between stochastic shortest path problem and distributionally robust shortest path problem.
- Numerical results for the distributionally robust SSP.
- Comparison with DRSSP with other SSP with different probability distributions.