Università di Pisa

Dipartimento di Matematica
Corso di Laurea in Matematica

Tesi di Laurea Triennale

Nested decomposition approaches for Delay-Constrained Routing problems

Relatore:
Dott. Antonio Frangioni

Candidato:
Enrico Sorbera

Anno Accademico 2020/2021

Contents

1 Introduction

In many real life situations we have to face the problem of routing a flow on a network, having under control the maximum end-to-end delay that it accumulates, while crossing it (e.g., industrial control systems, remote sensing and surveillance systems, live Internet Protocol Television and IP Telephony).
The purpose of the following paper is to build effective lower bounds for this general problem.
For this to be done we are going to take advantage from a nested decomposition approaches.
As we can check in §2, where a mathematical model is presented, delay constraint has a quadratic formulation, that makes it hard to handle.
Hence, we want to perform a Lagrangian relaxation on that constraint, that is a dual approach that allows to find lower bounds to the optimal value of the DCR problem.
This approach brings a troublesome variable, that in the following we name r_{min}, in the objective function of relaxation and we noted that the problem reduces to a simple *shortest path problem* by fixing that wild card.
To take advantage of this property, we want to perform a Benders' decomposition on that variable, in order to find the optimum for relaxation, studying its function point by point.
For this to be done, we need a dual vector of solution for fixed values in r_{min}, in this context the Lagrangian approach, that is in fact a dual approach, can make sense.
Anyway, if we want to find dual vectors we need the problem to have the integrality property, which is not granted by some previous models for this problem (see [1] and [4]).
Thus, we present a different formulation, with a non-linear constraint involving the cumbersome variable r_{min}, which grants integrality instead.
This makes the approximating function from Benders' decomposition non convex, however, this function being mono dimensional, we can manage this situation via splitting the domain in two halves, in which we build two convex approximation.
Hence, those remarks bring to a nested decomposition method, with a Benders' decomposition non-convex, thus highly non-standard, at first level, taking advantage from a Lagrangian relaxation to project the problem onto a SP structure.
This would eventually end in empirically accurate lower bounds for the general DCR problem.
We are now handling those problems by stages, in the following paragraph (§2) we present mathematical model underlining the constraint that lends itself to both non-linear and linear reformulations; then in paragraph §3 we show the Lagrangian relaxation on delay constraint and how to find primal and dual solution.

We take advantage of those in both paragraphs §4 and §5: in the first we use primal information to find the optimal value for the multiplier λ, meanwhile, in the latter we use mostly dual information in order to perform Benders' decomposition.
Finally, paragraphs §5 and §6 are dedicated to the numerical results of comparison between our method and a general-purpose solver and some conclusive remarks.

First of all, let's start describing the key elements of the graph $G = (N, A)$, that we assume directed, with a single source $s \in N$ and destination $d \in N \backslash \{s\}$.
Each node $i \in N$ is characterized by a fixed *node delay* n_i, while an arc $(i, j) \in A$ is described by its fixed *link delay* l_{ij}, *physical link speed* w_{ij}, *reservable capacity* c_{ij}, and its cost f_{ij}.
We measure the amount of bits of a flow expected to enter the origin (and so be dispatched across the network) in the time frame τ with an *arrival curve* $A(\tau) : \mathbb{R}_+ \to \mathbb{R}_+$.
This is formulated as the affine function $A(\tau) = \sigma + \rho \cdot \tau$, with its parameters called respectively *burst* and *rate*, this formulation is known in literature as the *leaky-bucket arrival curve*.
Moreover let L be the fixed and constant *maximum transmit unit* (MTU), and δ the *deadline* that marks maximum acceptable delay when routing a flow.

In order to keep finite the delay along a path p, we need that

$$r_{ij} \geq \rho \quad \forall (i, j) \in p \tag{1}$$

where r_{ij} stands for the reserved rate for a flow on the arc (i, j).
This property allows us to describe the WCD (*worst-case end-to-end delay*) of a path p:

$$\frac{\sigma}{\min\{r_{ij} \;:\; (i, j) \in p\}} + \sum_{(i,j)\in p} (\theta_{ij} + l_{ij} + n_i) \tag{2}$$

in which θ_{ij} is another core element of the graph: the *latency* encountered on the arc (i, j), that could be expressed in many ways depending on the scheduling algorithm at nodes (e.g. *Group Based*, *Weakly Rate-Proportional*, and so on).
We are going to work with a *Strictly Rate-Proportional* (SRP) scheduling; its name is due to its inverse relationship to the reserved rate except for an additive term:

$$\theta_{ij} = \frac{L}{w_{ij}} + \frac{L}{r_{ij}} \tag{3}$$

this scheduling enables us to ignore the *global admission control*, that is we can neglect the checks needed to guarantee the feasibility of the flows previously routed, for this formulation doesn't depend on them.

We are now ready to start with the description of this problem as a mixed-integer nonlinear model.

2 Mathematical model

We now present our MISOCP model for the DCR problem.
For this, we first introduce arc-flow binary variables $x_{ij} \in \{0,1\}$ indicating whether or not arc (i,j) belongs to the chosen path P, so that we can use the standard flow conservation constraints:

$$\sum_{(j,i)\in BS(i)} x_{ji} - \sum_{(i,j)\in FS(i)} x_{ij} = \begin{cases} -1 & \text{if } i = s \\ 1 & \text{if } i = d \\ 0 & \text{otherwise} \end{cases} \qquad i \in N \tag{4}$$

those are meant to model the s-d-path requirements.
We also introduce arc reserve variable r_{ij}, and a variable r_{min} that represents least reserved capacity along the chosen path.
With c_{max} standing for the maximum capacity of any arc in the graph, we can add the constraints

$$r_{min} x_{ij} \leq r_{ij} \leq c_{ij} x_{ij} \quad (i,j) \in A \tag{5}$$

$$\rho \leq r_{min} \leq c_{max} \tag{6}$$

to ensure that $r_{ij} = 0$ whenever $x_{ij} = 0$, or that the reserved capacity stays within the right set, if $x_{ij} = 1$.
Then we introduce latency constraint

$$\frac{\sigma}{r_{min}} + \sum_{(i,j)\in A} (\theta_{ij} + (l_{ij} + n_i) x_{ij}) \leq \delta \tag{7}$$

Note that the nonlinear term $\frac{\sigma}{r_{min}}$ can be expressed via the rotated SOCP constraint, adding an auxiliary variable t:

$$t + \sum_{(i,j)\in A} (\theta_{ij} + (l_{ij} + n_i) x_{ij}) \leq \delta \tag{8}$$

$$t \; r_{min} \geq \sigma \quad t \geq 0 \tag{9}$$

A further issue now is to represent the fact that θ_{ij} is zero if $x_{ij} = 0$, while it is given by an appropriate (convex) nonlinear expression otherwise, that is

$$\theta_{ij} = \begin{cases} \frac{L}{w_{ij}} + \frac{L}{r_{ij}} & \text{if } x_{ij} = 1 \\ 0 & \text{if } x_{ij} = 0 \end{cases} \tag{10}$$

It is shown in [3] that one can represent this disjunctive set in a convex way via *Perspective Reformulation technique*, that bring us to the following:

$$\theta_{ij} = L s_{ij} + \frac{L x_{ij}}{w_{ij}} \qquad \forall (i,j) \in A \tag{11}$$

$$s_{ij} r_{ij} \geq x_{ij}^2 \qquad \forall (i,j) \in A \tag{12}$$

trading off the simplicity of this form, with a larger number of auxiliary variables s_{ij}.
We now have to define the objective function for our problem, that depends on reserved rates along the path and their costs:

$$\sum_{(i,j)\in A} f_{ij} r_{ij} \tag{13}$$

this completes the presented model for DCR problem.
Anyway, taking a closer look to (5) constraint, we can notice that this could be formulated in this linear form:

$$0 \le r_{ij} \le c_{ij} x_{ij} \qquad (i,j) \in A \tag{14}$$

$$r_{min} \le r_{ij} + c_{max}(1 - x_{ij}) \qquad (i,j) \in A \tag{15}$$

as we heralded in introduction and we are going to explain just before dual solution computing, this form is not fitting our purpose of solving this problem by nested decompositions, because it doesn't allow effective ways to find those dual solution we need.
Hence we are going on ignoring this alternative formulation, and eventually using it for providing the problem to the solver we are facing.
In the next paragraph we are going to set up some of the strategies for an efficient resolution of this problem.

3 Lagrangian relaxation

We now give a brief explanation of this procedure reporting the main theorems that grant the properties that are involved in this work.
Let (P) be the general integer linear programming problem

$$\min cx \quad s.t. \tag{16}$$
$$Ax \geq b \tag{17}$$
$$Bx \geq d \tag{18}$$
$$x_j \in \{0,1\}, \quad j \in I \tag{19}$$

where A, B are matrix and b, c and d vectors of conformable dimension, and the index set I denotes the variables required to be integer.
We distinguished two sets of constraints, because we assume that $Bx \geq d$ has a special structure that makes the resolution of the problem easy.
We define the Lagrangian relaxation of (P) relative to $Ax \geq b$ to be (PL_λ):

$$\min cx + \lambda(b - Ax) \quad s.t. \tag{20}$$
$$Bx \geq d \tag{21}$$
$$x_j \in \{0,1\}, \quad j \in I \tag{22}$$

which represents a set of problems, depending on λ, each of these is a relaxation, that is its set of feasible solutions $F(PL_\lambda) \supseteq F(P) \quad \forall \lambda \geq 0$, and its objective function is less then or equal to that of (P) in $F(P)$.
The ideal choice in λ, to make this relaxation as tighter as possible to (P), is to take the optimal value λ^* of this program that we name (D):

$$\max_{\lambda \geq 0} \; v(PR_\lambda) \tag{23}$$

This is intimately linked to the problem (P^*) obtained from (P) replacing the convex hull of $\{Bx \geq d, \quad x_j \in \{0,1\}, \quad j \in I\}$ to these constraints.
The sense of this linking is explained by these following two theorems, which underline also the relationship between those problems and the continuous relaxation $(\bar{P})$ of (P).

Theorem 1

(a) $F(\bar{P}) \supseteq F(P^*) \supseteq F(P), \quad F(PR_\lambda) \supseteq F(P)$

$$v(\bar{P}) \leq v(P^*) \leq v(P), \quad v(PR_\lambda) \leq v(P) \quad \forall \lambda \geq 0$$

(b) *if* $(\bar{P})$ *is feasible, then* $v(\bar{P}) \leq v(PR_{\bar{\lambda}})$

(c) *if for a given* λ, x *satisfies the three conditions*

(i) x *is optimal in* (PR_λ),

(ii) $Ax \geq b$,

(iii) $\lambda(b - Ax) = 0$

then x *is an optimal solution of* (P)

(d) *if* (P^*) *is feasible then*

$$v(D) = \max_{\lambda \geq 0} \; v(PR_\lambda) = v(PR_{\lambda^*}) = v(P^*)$$

Where $\bar{\lambda}$ is a dual optimal solution associated with $Ax \geq b$ for the linear program $(\bar{P})$.
Moreover, if we define the integrality property to be the fact that the optimal value of (PR_λ) is not altered by dropping the integrality conditions on its variable (i.e. $v(PR_\lambda) = v(\overline{PR_\lambda})$), this second theorem holds:

Theorem 2 *Let* $(\bar{P})$ *be feasible and* (PR_λ) *have the integrality property, then* (P^*) *is feasible and*

$$v(\bar{P}) = v(PR_{\bar{\lambda}}) = v(D) = v(PR_{\lambda^*}) = v(P^*)$$

the proofs of these theorems can be found in [11].
In this setting the fact that we will end finding a lower bound for DCR problem is now clear.
Let's see how we calculate λ^* and primal and dual solution for $(\bar{P})$, given r_{min}, and what a Lagrangian relaxation on the delay constraint looks like in our environment.

3.1 Relaxing delay constraint

First of all, here's a short recap of our formulation, that makes it easier to handle:

$$\min \sum_{(i,j)\in A} f_{ij} r_{ij} \tag{24}$$

$$\sum_{(j,i)\in BS(i)} x_{ji} - \sum_{(i,j)\in FS(i)} x_{ij} = \begin{cases} -1 & \text{if } i = s \\ 1 & \text{if } i = d \\ 0 & \text{otherwise} \end{cases} \qquad i \in N \tag{25}$$

$$\frac{\sigma}{r_{min}} + \sum_{(i,j)\in A} (\frac{L{x_{ij}}^2}{r_{ij}} + (\frac{L}{w_{ij}} + l_{ij} + n_i) x_{ij}) \leq \delta \tag{26}$$

$$r_{min} x_{ij} \leq r_{ij} \leq c_{ij} x_{ij} \qquad (i,j) \in A \tag{27}$$

$$r_{min} \in [\rho, c_{max}],\ x_{ij} \in \{0,1\} \qquad (i,j) \in A \tag{28}$$

Note that, despite the reformulation, still the delay constraint it's quite troublesome, for it's non linear.
We want to multiply it for a Lagrangian multiplier λ, adding it to the objective function to get around this issue, that turns in:

$$\min \sum_{(i,j)\in A} f_{ij} r_{ij} + \lambda(\frac{\sigma}{r_{min}} + \sum_{(i,j)\in A} (\frac{L{x_{ij}}^2}{r_{ij}} + (\frac{L}{w_{ij}} + l_{ij} + n_i) x_{ij}) - \delta) \tag{29}$$

and let in the following $\bar{l}_{ij} = \frac{L}{w_{ij}} + l_{ij} + n_i$.

Now, in order to solve this Lagrangian relaxation, we have to find the maximum in λ of the minimum function in (29); to do this we need both the value of the minimum and the value of a gradient (at least an ipergradient) of the function point by point.
Let's handle these problems in stages, starting with finding a solution for (29) for fixed value of λ and r_{min}.

3.2 Projection on Shortest Path

A very interesting remark about the latest formulation (29) is based on the possibility of projection on a *Shortest Path problem* by fixing x_{ij} variables.
In fact, $x_{ij} = 0$, as already noted, forces us to put $r_{ij} = 0$ too (see (27)), while fixing $x_{ij} = 1$ brings to:

$$\min\{ \sum_{(i,j)\in A} f_{ij} r_{ij} + \frac{\lambda L}{r_{ij}} + \lambda \bar{l}_{ij} \} + \frac{\lambda\sigma}{r_{min}} - \lambda\delta \tag{30}$$

Calling $\phi(r_{ij}) = f_{ij}r_{ij} + \frac{\lambda L}{r_{ij}}$, we have

$$\phi'(r_{ij}) = f_{ij} - \frac{\lambda L}{r_{ij}^2} \tag{31}$$

so the function decreases until it meets its minimum in $\bar{v} = \sqrt{\frac{\lambda L}{f_{ij}}}$ and keeps growing after that point.
Hence, if $x_{ij} = 1$, the optimal value for r_{ij} is

$$r_{ij}^* = \begin{cases} c_{ij} & \text{if } f_{ij} \leq 0 \\ \min\{c_{ij}, \max\{r_{min}, \bar{v}\}\} & \text{otherwise} \end{cases} \tag{32}$$

Now we can show an equivalent formulation for (29)

$$\frac{\lambda\sigma}{\bar{r}_{min}} + \min\{ \sum_{(i,j)\in A} (\lambda\bar{l}_{ij} + \phi(r_{ij}^*))x_{ij} : \ (25) \ \} - \lambda\delta \tag{33}$$

that has a visible SP structure, with arc costs fixed to

$$\tilde{c}_{ij} = \lambda\bar{l}_{ij} + \phi(r_{ij}^*) \tag{34}$$

Solving it by any applicable algorithm eventually gives us the optimal value of the Lagrangian function for fixed λ and r_{min}.

3.2.1 SP feasibility

It can happen, and it empirically does indeed, that SP it's infeasible for some r_{min}, that is that there's no value in λ for that r_{min} that grants feasibility for SP.
That means that we can't find any s-d-path with all reservable rates larger than r_{min}.
If we imagine to delete all arcs with reservable capacity smaller than r_{min}, unfeasibility means a disconnection in the graph between source and destination.
In that case we have to pick a smaller value for r_{min}, or in case $r_{min} = \rho$ we conclude that DCR problem was empty.
We can find the larger value in r_{min} that grants feasibility sorting arc capacities vector and running a binary search on it, with a graph visit at each step.
That procedure has a non-trivial complexity, anyway it can be considered a preprocessing procedure, that takes place only once.
In the following r_μ stands for this value.

3.2.2 Computing dual solution

It takes more effort to find a dual solution for the settled Lagrangian relaxation of DCR problem.

In this regard, note that (27) would lend itself to a linear reformulation:

$$0 \leq r_{ij} \leq c_{ij} x_{ij} \qquad (i,j) \in A \tag{35}$$

$$r_{min} \leq r_{ij} + c_{max}(1 - x_{ij}) \qquad (i,j) \in A \tag{36}$$

we chose the nonlinear one, because carrying on dual optimum calculation we found out that they acquire a more complicated shape in linear case; meanwhile, as we're about to see, nonlinear form allows us to take advantage of SP dual solution.
This depends on the fact that non-linear reformulation matches the convex hull of the set of feasible solutions for (29), for we're about to show a dual solution for the problem (37) (in which we ignored integrality on variables x_{ij}) related to the primal integer we found via solving SPT.
That's a desirable property, in fact almost any reasonable algorithm for SP gives a vector of dual solutions back as a subproduct.

Before we dive in solution seeking, let's write down again (29) in an even more compact shape

$$\begin{gathered} \min\{h(x,r) : \begin{bmatrix} E & 0 \end{bmatrix} \begin{pmatrix} x \\ r \end{pmatrix} = b;\ D^- \begin{pmatrix} x \\ r \end{pmatrix} \leq 0;\ D^+ \begin{pmatrix} x \\ r \end{pmatrix} \leq 0; \\ \begin{bmatrix} -I_{m\text{x}m} & 0 \end{bmatrix} \begin{pmatrix} x \\ r \end{pmatrix} \leq 0\} + g(r_{min}) \end{gathered} \tag{37}$$

where

$$h(x,r) = \sum_{(i,j)\in A} f_{ij} r_{ij} + \lambda L \frac{x_{ij}^2}{r_{ij}} + \lambda \bar{l}_{ij} x_{ij} \tag{38}$$

$$g(r_{min}) = \frac{\lambda \sigma}{r_{min}} - \lambda \delta \tag{39}$$

while $\begin{bmatrix} E & 0 \end{bmatrix} \in \mathbf{M}(n, 2m, \mathbb{R})$ with E representing transition matrix for the graph.
Moreover

$$D^+ = \begin{bmatrix} -diag(c_{ij}) & I_{m\text{x}m} \end{bmatrix} \qquad D^- = \begin{bmatrix} diag(r_{min}) & -I_{m\text{x}m} \end{bmatrix}$$

standing for (25) and (27) constraints.

A primal-dual solution pair $((x^*, r^*), (\pi, \mu^+, \mu^-, \eta))$ is optimal if KKT (Karush-

Kuhn-Tucker) conditions hold:

$$\nabla h(x^*, r^*) + \begin{bmatrix} E & 0 \end{bmatrix}^T \pi + D^{-T} \mu^- + D^{+T} \mu^+ + \begin{bmatrix} -I & 0 \end{bmatrix}^T \eta = 0 \tag{40a}$$

$$\begin{cases} \mu^+ \geq 0 \\ \mu^- \geq 0 \\ \eta \geq 0 \end{cases} \tag{40b}$$

$$\begin{cases} \mu^{+T} (D^+ \binom{x}{r}) = 0 \\ \mu^{-T} (D^- \binom{x}{r}) = 0 \\ \eta^T \begin{bmatrix} -I & 0 \end{bmatrix} \binom{x}{r} = 0. \end{cases} \tag{40c}$$

In which we ignored primal feasibility conditions, already granted by SP solution.
Expanding (40a) and (40c) we get these two linear system of equations:

$$\begin{cases} 2\lambda L \frac{x^*_{ij}}{r^*_{ij}} + \lambda \bar{l}_{ij} - \pi_i + \pi_j - c_{ij}\mu^+_{ij} + r_{min}\mu^-_{ij} - \eta_{ij} = 0 & (i,j) \in A \\ f_{ij} - \lambda L (\frac{x^*_{ij}}{r^*_{ij}})^2 + \mu^+_{ij} - \mu^-_{ij} = 0 & (i,j) \in A \end{cases} \tag{41}$$

$$\begin{cases} \mu^+_{ij}(r_{ij} - c_{ij}x_{ij}) = 0 & (i,j) \in A \\ \mu^-_{ij}(x_{ij}r_{min} - r_{ij}) = 0 & (i,j) \in A \\ -\eta_{ij}x_{ij} = 0 & (i,j) \in A \end{cases} \tag{42}$$

the latter it's been simplified by imposing the same condition arc by arc.
Anyway it's an equivalent form since that was a sum of positive terms.
Note that we assume to have access to SP dual solution d, for which holds the useful condition

$$\begin{cases} \tilde{c}_{ij} + d_i - d_j \geq 0 \\ x_{ij}(\tilde{c}_{ij} + d_i - d_j) = 0 \end{cases} \tag{43}$$

where $\tilde{c}_{ij}$ are the costs of arcs in SP, defined in (34).

3.2.3 Optimal dual solution for $x_{ij} = 1$

We start our research trying to find dual optimum when r^*_{ij} doesn't reach neither of (27) margins, that can happen iff $r_{min} < \bar{v} < c_{ij}$, that, with regard to (32), implies $r^*_{ij} = \bar{v}$.
In this case (42) forces us to choose $\mu^+_{ij} = \mu^-_{ij} = \eta_{ij} = 0$.
Thus, eliminating those multipliers in (41) we get for the arc (i,j)

$$f_{ij} - \lambda L (\frac{x^*_{ij}}{r^*_{ij}})^2 = f_{ij} - \lambda L (\frac{f_{ij}}{\lambda L}) = 0 \tag{44}$$

in the second equation, and

$$\lambda \bar{l}_{ij} + 2\sqrt{\lambda L f_{ij}} - \pi_i + \pi_j = 0 \tag{45}$$

for the first one.
Note that, for this value of r^*_{ij}, cost of arc (i,j) in SP is $\tilde{c}_{ij} = \lambda\bar{l}_{ij} + 2\sqrt{\lambda L f_{ij}}$, as we can check in its definition (see (34)).
In addition, SP dual second condition tells us that in this case

$$\tilde{c}_{ij} = d_j - d_i \tag{46}$$

it follows that we can take $\pi_i = -d_i$ and $\pi_j = -d_j$ with regards to KKT.
We just have to watch out for going on choosing this solution for each arc, in fact more than one arc could start in or ends to the same node, that is we have more conditions that nodes, so we will try to find a suitable solution for all arcs, choosing $\pi = -d$ in any situation.

Whenever $r^*_{ij} = c_{ij} \neq r_{min}$, from (42) we have $\mu^-_{ij} = \eta_{ij} = 0$.
That implies in (40a)

$$f_{ij} - \frac{\lambda L}{c^2_{ij}} + \mu^+_{ij} = 0 \Leftrightarrow \mu^+_{ij} = \frac{\lambda L}{c^2_{ij}} - f_{ij} \tag{47}$$

and

$$2\frac{\lambda L}{c_{ij}} + \lambda\bar{l}_{ij} - \pi_i + \pi_j + c_{ij} f_{ij} - \frac{\lambda L}{c_{ij}} = 0 \tag{48}$$

Again, $\tilde{c}_{ij} = \frac{\lambda L}{c_{ij}} + \lambda\bar{l}_{ij} + c_{ij} f_{ij}$ hence $\pi = -d$ is still a workable choice.
We just have to check out (40b) condition, that is (47) is acceptable only if has a positive value

$$\frac{\lambda L}{c^2_{ij}} - f_{ij} \geq 0 \tag{49}$$

For this purpose note that $r^*_{ij} = c_{ij}$ either if $f_{ij} \leq 0$, that turns (49) in a sum of positive terms, or if $c_{ij} \leq \max\{r_{min}, \bar{v}\} = \bar{v}$ that clearly fulfils (49).
The latter relation holds because if it wouldn't we should have $r^*_{ij} = \min\{c_{ij}, r_{min}\} = r_{min}$, and this is not the case.

Finally, if $r^*_{ij} = r_{min}$ we have $\mu^+_{ij} = \eta_{ij} = 0$ from (42) and this brings us to

$$\mu^-_{ij} = f_{ij} - \frac{\lambda L}{r^2_{min}} \tag{50}$$

for the first equation in (41), while the second becomes

$$\frac{2\lambda L}{r_{min}} + \lambda\bar{l}_{ij} + \pi_i - \pi_j + r_{min} f_{ij} - \frac{\lambda L}{r_{min}} = 0 \tag{51}$$

First, let's check out $\mu^-_{ij} \geq 0$, expanding this condition:

$$f_{ij} - \frac{\lambda L}{r^2_{min}} \geq 0 \;\equiv\; r_{min} \geq \sqrt{\frac{\lambda L}{f_{ij}}} = \bar{v} \tag{52}$$

$r^*_{ij} = r_{min}$ couldn't hold unless $r_{min} \geq \bar{v}$, as we can check in (32), so this condition turns out to be true.
Once again

$$\tilde{c}_{ij} = \frac{\lambda L}{r_{min}} + \lambda \bar{l}_{ij} r_{min} f_{ij} \tag{53}$$

and so we pick $\pi = -d$ once again.
It could be easily verified that these conditions hold in case $r_{min} = c_{ij}$ for some $(i, j) \in A$ too.

3.2.4 Optimal dual solution for $x_{ij} = r_{ij} = 0$

The function $h(x, r)$ is no longer differentiable in $(0, 0)$.
We can assume that $h(0, 0) = 0$.
Let's now see how to generalize KKT conditions, to make them suitable in non differentiable situation.
Restraining h to one arc, ignoring the sum

$$\tilde{h}(x_{ij}, r_{ij}) = f_{ij} r_{ij} + \lambda L \frac{x_{ij}^2}{r_{ij}} + \lambda \bar{l}_{ij} x_{ij} \tag{54}$$

we can note that only the term $\lambda L \frac{{x_{ij}}^2}{r_{ij}}$ is non differentiable.
Thanks to linearity of the derivative, we can focus just on this term and look to the form of his subdifferential ($\partial \tilde{h}$), in which we are going to pick a gradient to generalize (40a).
We could have done this generalization from the start, replacing (40a) with

$$\exists v \in \partial h(x^*, r^*) \; : \; v + \begin{bmatrix} E & 0 \end{bmatrix}^T \pi + {D^+}^T \mu^+ + {D^-}^T \mu^- + \begin{bmatrix} -I & 0 \end{bmatrix}^T \eta = 0 \tag{55}$$

anyway, that would eventually bring us to the same conclusion.
In fact, whenever convexity holds, subgradient equates gradient singlet.
Thus, we restrain h even more, minding just the non differentiable term:

$$\tilde{h}(x, r) = \frac{\lambda L x^2}{r} \tag{56}$$

Setting $f(x) = \lambda L x^2$, note that $\tilde{h}(x, r) = r f(\frac{x}{r})$ [1] that's linear for fixed values of $\frac{x}{r}$.
It can be shown (see [3]) that for any couple $(\bar{x}, \bar{r})$ observing (27) and $x \in [0, 1]^m$, we have that

$$[\nabla f \left(\frac{\bar{x}}{\bar{r}}\right), f\left(\frac{\bar{x}}{\bar{r}}\right) - \nabla f \left(\frac{\bar{x}}{\bar{r}}\right) \frac{\bar{x}}{\bar{r}}] = [2\lambda L \frac{\bar{x}}{\bar{r}}, -\lambda L \left(\frac{\bar{x}}{\bar{r}}\right)^2] \in \partial \tilde{h}(0, 0) \tag{57}$$

Finally, we write down (41) in the light of the above:

$$\begin{cases} 2\lambda L \frac{\bar{x}_{ij}}{\bar{r}_{ij}} + \lambda \bar{l}_{ij} - \pi_i + \pi_j + r_{min} \mu^-_{ij} - c_{ij} \mu^+_{ij} - \eta_{ij} = 0 \\ f_{ij} - \lambda L (\frac{\bar{x}_{ij}}{\bar{r}_{ij}})^2 + \mu^+_{ij} - \mu^-_{ij} = 0 \end{cases} \tag{58}$$

[1] that is $\tilde{h}$ is the *perspective function* of f

Now, we have to distinguish two cases, subjecting to whether $r_{min} \leq c_{ij}$ or not. In the first of which picking

$$\bar{x}_{ij} = 1 \tag{59}$$

$$\bar{r}_{ij} = \begin{cases} c_{ij} & \text{if } f_{ij} \leq 0 \\ \min\{c_{ij}, \max\{\bar{r}_{min}, \bar{v}\}\} & \text{otherwise} \end{cases} \tag{60}$$

renews §2.2.1 environment, ignoring additional dual moltipliers (that is fixing their value at zero).
In this case, we just act as before, while, whether $r_{min} > c_{ij}$, (27) constrains us to $(\bar{x}, \bar{r}) = (0, 0)$, that puts us in a worse situation.
In fact, in this case, those arcs didn't belong to SP graph, opening up the possibility of some nodes disconnection.
That would cause SP dual value for those nodes to be $+\infty$.
Anyway, let's write (58) for $(\bar{x}, \bar{r}) = (0, 0)$

$$\begin{cases} -\pi_i + \pi_j + r_{min}\mu_{ij}^- - c_{ij}\mu_{ij}^+ - \eta_{ij} = 0 & \forall (i,j) \in A \\ \mu_{ij}^+ - \mu_{ij}^- = 0 & \forall (i,j) \in A \end{cases} \tag{61}$$

this second relationship establishes $\mu_{ij}^+ = \mu_{ij}^-$, which we refer to as simply μ_{ij}.
Hence, let's replace it in the first equation:

$$-\pi_i + \pi_j + (r_{min} - c_{ij})\mu_{ij} - \eta_{ij} = 0 \tag{62}$$

in which we remind that $r_{min} > c_{ij}$.

What to do when $d_i = d_j = \infty$

That could seem the easiest setting, for we're free to choose the values of our wild cards, but we have to watch out for the usual trouble: π_i definition has to be common for more arcs, so we have to find a common suitable definition.
Let's try to choose $\pi_i = 0$ whenever $d_i = +\infty$.
In this particular case that can, for instance, drive us to $\mu_{ij} = \eta_{ij} = 0$, despite there's plenty of affordable possibilities.

What to do when we defined just one SP dual label

This case forces to annihilate the other label.
Hence, if $d_i = +\infty$, we choose $\pi_i = 0$, while $\pi_j = -d_j$, obtaining in (62)

$$-d_j + (r_{min} - c_{ij})\mu_{ij} - \eta_{ij} = 0 \tag{63}$$

note that $d_j \geq 0$, so $\eta_{ij} = 0$ and $\mu_{ij} = \frac{d_j}{r_{min} - c_{ij}}$ works, for they're both non negative.

Instead, when $d_j = +\infty$ ($\Rightarrow \pi_i = -d_i$), we have

$$d_i + (r_{min} - c_{ij})\mu_{ij} - \eta_{ij} = 0 \tag{64}$$

and $\mu_{ij} = 0$ and $\eta_{ij} = d_i$ is still a reasonable and correct choice.

What to do when both SP dual solution are defined

This situation could happen indeed, if those nodes are linked by some other SP-arcs to the remaining part of the graph.
Hence (62) becomes

$$d_i - d_j + (r_{min} - c_{ij})\mu_{ij} - \eta_{ij} = 0 \tag{65}$$

but now we have no information on the number $d_i - d_j$, for this is no longer the potential difference of the edges of (i, j).
Hence, we have to find a solution according both their possible signs:
when $d_i - d_j \geq 0$, we choose $\eta_{ij} = d_i - d_j$ and $\mu_{ij} = 0$.
Meanwhile, if $d_i - d_j < 0$ a feasible choice would be $\eta_{ij} = 0$ and $\mu_{ij} = \frac{d_j - d_i}{r_{min} - c_{ij}}$.

All we've stated by now fulfils dual solution seeking.
Now we want to shed light on what happens when λ and r_{min} are free to change their values.
We are going to handle this variation one wild cards from these two by one, starting from λ.
Note that this choice isn't mandatory, we could also reverse this order, but these way seemed much more auspicious.
That's because, as we're about to see in further paragraphs, this keeps the more unusual behaviour of r_{min} variation in the more external loop of the discussed algorithm.

4 Inner loop: finding an optimal multiplier λ

With the knowledge of primal and dual solution (for fixed λ, r_{min}) for

$$p(\lambda) = \min_{(x,r)} \{h(x,r) + g(\bar{r}_{min}) : (25), (27), (28)\} \tag{66}$$

we're going to show how to solve

$$\max\{p(\lambda) \,:\, \lambda \geq 0\} \tag{67}$$

- that is Lagrangian dual of DCR problem due to relaxation of (26) constraint
- via gradient method.

That knowledge obviously grants us the point-by-point value for (66), we reformulate p definition so that we can exploit SP solutions, in order to find one of its ipergradient in that point too.
For this purpose, we define

$$\beta(x,r) = \frac{\sigma}{\bar{r}_{min}} - \delta + \sum_{(i,j)\in A} \frac{Lx_{ij}^2}{r_{ij}} + \bar{l}_{ij} x_{ij} \tag{68}$$

$$\alpha(x,r) = \sum_{(i,j)\in A} f_{ij} r_{ij} \tag{69}$$

Note that β equals LHS of (26), while α is the objective function for DCR.
However, with those terms we can write

$$p(\lambda) = \min_{(x,r)} \{\alpha(x,r) + \lambda\beta(x,r) : (25), (27), (28)\} = \alpha(x(\lambda), r(\lambda)) + \lambda\beta(x(\lambda), r(\lambda)) \tag{70}$$

in which for fixed $\lambda = \bar{\lambda}$, $(x(\bar{\lambda}), r(\bar{\lambda}))$ stands for primal solution of SP, with usual definition of arc's costs (34) for that value in λ.
This reformulation underlines point-by-point value for $p(\bar{\lambda})$ derivative, that, from the latter relation, we can easily check to be $\beta(x(\bar{\lambda}), r(\bar{\lambda}))$.

For algorithm to start, now we just need two points: one with a positive value for the ipergradient, and another with ipergradient of opposite sign, from which line search could eventually start.
For this, first we check $\beta(x(0), r(0))$, provided this turns out to be non positive, we've already found our optimal solution: $\lambda^* = 0$.
Otherwise we look for the point with negative ipergradient on the sequence

$$\begin{cases} \lambda_0 = 1 \\ \lambda_{i+1} = 2\lambda_i \end{cases} \tag{71}$$

for this to be achievable, we need a stop condition, that is we have to be aware of the existence of a point where we can stop our seeking, figuring out dual to

be unlimited.
An upper bound for that point can be easily built by computing an upper bound for worst DCR solution, check this formulation out:

$$\max\{ \sum_{(i,j)\in A} f_{ij} r_{ij} \ : \ (25),(26),(27),(28)\} \tag{72}$$

This is DCR thought as a maximum cost routing problem, its value in optimum has the property to be larger of any feasible solution for our actual problem.
Moreover, for it's a maximum cost problem, the property to be a valid upper bound holds for each relaxation of (72).
Thus, we can ignore all constraints except box ones:

$$\max \sum_{(i,j)\in A} f_{ij} r_{ij} \tag{73}$$

$$\bar{r}_{min} x_{ij} \leq r_{ij} \leq c_{ij} x_{ij} \qquad (i,j) \in A \tag{74}$$

$$x_{ij} \in [0,1] \qquad (i,j) \in A \tag{75}$$

which striking solution is the one that saturates positive cost arcs, and ignores other ones:

$$(\hat{x}_{ij}, \hat{r}_{ij}) = \begin{cases} (0,0) & \text{if } \ f_{ij} \leq 0 \text{ o } \bar{r}_{min} > c_{ij} \\ (1, c_{ij}) & \text{otherwise} \end{cases} \tag{76}$$

its cost

$$z = \sum_{\substack{f_{ij}>0 \\ c_{ij}\geq \bar{r}_{min}}} f_{ij} c_{ij} \tag{77}$$

is the upper bound we needed, in fact, at any point λ_i, $p(\lambda_i)$ is a lower bound for DCR optimal value.
Hence, whenever for some i $p(\lambda_i) > z$ holds, we can state primal problem to be empty, and we have to change r_{min} value, which wasn't within feasibility interval.

4.1 Procedure for finding optimal multiplier λ^*

We found all the ingredients for gradient method, its procedure is quite simple at this point.
We start with two ipergradient, in each iteration we get their interception and calculate value and ipergradient for dual function, unless in $\lambda = 0$ we discovered a negative slope yet, which makes us stop, as already noted, with $\lambda^* = 0$.
This new ipergradient replaces one of the old ones, consistently with its sign (the one with same sign for the slope).
For new interceptions must stay within the interval between the two previous ones, this method's clearly convergent.
We are shutting its running down when relative distance from the value of the

approximation given by gradient interception to the actual dual value in that point becomes lesser that a set accuracy ϵ.

In paragraph ending we suggest some ways to reuse information resulted by running this method.

4.2 Reoptimizing

We are going to discuss in further paragraph r_{min} variation, hence, discover whether some cuts that we generate in the gradient method still work for a new r_{min} value is really useful.
A sufficient condition for them to be still effective ipergradient of our function p is that the primal solution related to them is still feasible.
Note that the only constraint that could violate this feasibility is (27).
Hence, the only check we need is that $r_{ij}(\lambda) \geq \hat{r}_{min}$ for all arcs in optimal path, where $\hat{r}_{min}$ is its new value.
Checking this condition for all optimal solution of all the visited cuts, will ends up in finding a subset $I' \subset I$ of the set I of all feasible solution for our dual problem.
If there's a couple of those solution that pass that check with discordant slopes, we can use them to start our gradient method, skipping some preprocessing.
Moreover, every time we find a negative slope cut we gain an even more interesting condition.
In fact, as we noted, β represent both slope of ipergradients and the LHS of (27) constraint, that means that for each negative β we found a primal feasible solution, that is one that verifies

$$\frac{\sigma}{\bar{r}_{min}} + \sum_{(i,j)\in A} \frac{Lx_{ij}^2}{r_{ij}} + \bar{l}_{ij}x_{ij} \leq \delta \tag{78}$$

This on one hand grants that primal problem's not empty, on the other it allows us to set up an heuristic method, with no more computing needed.
For each new r_{min} we run this method, and we want to save the best (the cheaper one) primal feasible solution found concurrently with point in λ with negative slope, that is making us discover empirically good upper bound to DCR optimal value, with no adding effort needed.

Now that we're able to find optimal solution in λ for fixed values in r_{min}, we discuss this wild card's variation.
To do this, we present a highly non-standard Benders' decomposition, that takes advantage from dual optimal solution we calculated in §3.2, to build up partially convex approximations of Lagrangian objective function.
Let's proceed step by step, showing how to calculate those approximations and how to use them to eventually find an optimal solution for our Lagrangian relaxation.

5 Benders' decomposition

As we did in §3, we start this paragraph giving a brief explanation of the Benders' decomposition.
This is a useful method whenever we have to face a problem of the form

$$\max_{x,y} \; f(x,y) \; s.t. \; G(x,y) \geq 0, \quad x \in X, \quad y \in Y \tag{79}$$

where y stands for a set of variable complicating the solution, G is a vector of constraint functions defined on $X \times Y$.
What we want to do is a projection onto y-space:

$$\max_{y} \; v(y) \; s.t. \; y \in Y \cap V \tag{80}$$

where

$$v(y) = \sup_{x} \; f(x,y) \; s.t. \; G(x,y) \geq 0, \quad x \in X \tag{81}$$

and

$$V = \{y \; : \; G(x,y) \geq 0 \text{ for some } x \in X\} \tag{82}$$

Given y, computing $v(y)$ should be easy, and in our setting we already found a solution for this problem, for $y = r_{min}$ is our only complicating wild card.
Let (1-y) be the problem

$$\max_{x} \; f(x,y) \; s.t. \; G(x,y) \geq 0 \tag{83}$$

the following theorems underline some interesting properties of this projection, for their proofs and further explanation see [9]-[10].

Theorem 3 (Projection) *Problem* (79) *is infeasible or has unbounded optimal value iff the same is true of* (80). *If* (x^*, y^*) *is optimal in* (79), *then* y^* *must be optimal in* (80). *If* y^* *is optimal in* (80) *and* x^* *achieves the sup in* (81) *with* $y = y^*$, *then* (x^*, y^*) *is optimal in* (79).

Theorem 4 (v-Representation) *Assume that* X *is a nonempty convex set and that* f *and* G *are concave on* X *for each fixed* $y \in Y$. *Assume further that, for each fixed* $\bar{y} \in Y \cap V$, *at least one of the following three conditions holds:*
(a) $v(\bar{y})$ *is finite and* (1-$\bar{y}$) *possesses an optimal multiplier vector;*
(b) $v(\bar{y})$ *is finite,* $G(x,\bar{y})$ *and* $f(x,\bar{y})$ *are continuous on* X, X *is closed, and the* ϵ*-optimal solution set of* (1-$\bar{y}$) *is nonempty and bounded for some* $\epsilon \geq 0$;
(c) $v(\bar{y}) = +\infty$.
Then, the optimal value of (1-$\bar{y}$) *equals that of its dual on* $Y \cap V$, *that is,*

$$v(y) = \inf_{u \geq 0} \{ \sup_{x \in X} \; f(x,y) + u^t G(x,y) \}, \quad \forall y \in Y \cap V$$

Hence, under these assumptions we can write the master problem of Benders' decomposition, that is an equivalent formulation for (79):

$$\max_{y\in Y\cap V}\{\inf_{u\geq 0}\{\sup_{x\in X}\ f(x,y)+u^tG(x,y)\}\} \tag{84}$$

We want to apply this formulation to our Lagrangian relaxation in order to handle the complicating variable r_{min}.

5.1 Decomposing DCR problem

We start presenting a Lagrangian relaxation of (37) on each constraint:

$$\begin{aligned}\max_{(\pi,\mu^+,\mu^-,\eta)}&\{\min_{(x,r)}\{h(x,r)+{\mu^+}^T(D^+\begin{pmatrix}x\\r\end{pmatrix})+{\mu^-}^T(D^-\begin{pmatrix}x\\r\end{pmatrix})\\&+\eta^T\begin{bmatrix}-I & 0\end{bmatrix}\begin{pmatrix}x\\r\end{pmatrix}+\pi^T(\begin{bmatrix}E & 0\end{bmatrix}\begin{pmatrix}x\\r\end{pmatrix}-b)\}\}\end{aligned} \tag{85}$$

$$\mu_{ij}^-\geq 0,\ \mu_{ij}^+\geq 0,\eta_{ij}\geq 0\quad\forall(i,j)\in A. \tag{86}$$

where we multiplied each of them for a Lagrangian multiplier and we added it to objective function.
In order to lighten this hard notation, we define $\theta=(\mu^+,\mu^-,\pi,\eta)$ and $q(\theta,x,r,r_{min})$ as the function inside the minimum on the couple (x,r).
We are ready for our first formulation of master problem of Benders' decomposition, where we think to r_{min} as our only complicated variable:

$$\min_{r_{min}}\{g(r_{min})+\max_{\theta\geq 0}\{\min_{(x,r)}\{q(\theta,x,r,r_{min})\}\}\} \tag{87}$$

$$r_{min}\in[\rho,r_\mu] \tag{88}$$

where we recall g function's definition to be $g(r_{min})=\frac{\lambda\sigma}{r_{min}}-\lambda\delta$.
We can already reformulate (87)-(88) taking advantage from maximum function properties:

$$\min_{(r_0,r_{min})} r_0 \tag{89}$$

$$r_0\geq g(r_{min})+\min_{(x,r)}\{q(\theta,x,r,r_{min})\}\quad\forall\theta\geq 0 \tag{90}$$

$$r_{min}\in[\rho,r_\mu]=V \tag{91}$$

renaming $r_{min}=z$, to underline the difference between r_{min} wild card on which we're finding minimum in (87) and fixed value $\bar{r}_{min}$, in which we computed latter iteration.
That's helping us, because what we want to do now is calculate dual vector of solutions $\bar{\theta}$ for some fixed $\bar{r}_{min}$ and use it for computing $\min_{(x,r)}\{q(\bar{\theta},x,r,z)\}$.

The point of this is that if we pretend that a certain $\bar{\theta} \geq 0$ wouldn't change according with z, computing RHS of (90) for that fixed $\bar{\theta}$'s returning us a function that has the property to stay within Lagrangian objective function ipograph.
Building these functions for different points - hence different dual vectors - and taking the maximum among them, shapes an approximation of the function of Lagrangian dual problem, that improves as the amount of points grows.
The minimum value of this approximation is a valid lower bound for the minimum of our Lagrangian problem, so we keep on improving this bound until whether we reach the set accuracy or we realize that we can improve it no more.
As usual, let's proceed gradually, figuring out, first of all, how to build that function up for some fixed $\bar{r}_{min}$ and $\bar{\theta}$.

5.2 Computing $\min_{(x,r)}\{q(\bar{\theta}, x, r, z)\}$

This is how it looks q expanded form:

$$\begin{aligned} q(x,r) = \sum_{(i,j)\in A} (f_{ij}r_{ij} + \frac{\lambda L x_{ij}^2}{r_{ij}} + \lambda \bar{l}_{ij}x_{ij} + \mu_{ij}^{+}(r_{ij} - c_{ij}x_{ij}) + \\ + \mu_{ij}^{-}(zx_{ij} - r_{ij}) - \eta_{ij}x_{ij} - \pi_i x_{ij} + \pi_j x_{ij}) - \sum_{i\in N} \pi_i b_i \end{aligned} \tag{92}$$

where we ignored the fact that q depends on θ and z, because we're going to think to the first as a constant term, and to the second as a parameter.
The following is the gradient in (x,r) of our function (quite familiar as one could note, see (41)):

$$\begin{cases} 2\lambda L\frac{x_{ij}}{r_{ij}} + \lambda\bar{l}_{ij} - \mu_{ij}^{+}c_{ij} + z\mu_{ij}^{-} - \eta_{ij} - \pi_i + \pi_j = 0 \quad (i,j)\in A \\ f_{ij} - \lambda L(\frac{x_{ij}}{r_{ij}})^2 + \mu_{ij}^{+} - \mu_{ij}^{-} = 0 \quad (i,j)\in A \end{cases} \tag{93}$$

The term $\frac{x_{ij}}{r_{ij}}$ prevents us to minimize this function separately on x and r, anyway we can surely work arc by arc, because arc contribution doesn't depend on the one from other arcs.
We now review each potential behaviour of θ, according with what we saw in §3.2.

Whenever we choose $\mu_{ij}^{-} = 0$, each term depending on z annihilates in (93), so its value can't influence where the minimum of the function is located.
We already knew that KKT grants optimality for (x^*, r^*) whether $z = \bar{r}_{min}$, hence, in this case, this'd hold when z is moving too.
Thus, the only case that we didn't solve yet is the one with $\mu_{ij}^{-} \neq 0$, that is when

$$\bar{v} \leq \bar{r}_{min} \leq c_{ij} \tag{94}$$

where we have that $\mu_{ij}^{+} = \eta_{ij} = 0$, while $\pi_i = -d_i$, $\pi_j = -d_j$ (or zero if SP labels equal $+\infty$) and $\mu_{ij}^{-} = f_{ij} - \frac{\lambda L}{\bar{r}_{min}^2}$.
If we replace their values in (92), we find

$$\lambda L \frac{x_{ij}^2}{r_{ij}} - \lambda L \frac{z}{\bar{r}_{min}^2} x_{ij} + \lambda L \frac{r_{ij}}{\bar{r}_{min}^2} + f_{ij} z x_{ij} + \lambda \bar{l}_{ij} x_{ij} + (d_i - d_j) x_{ij} \tag{95}$$

while, if we put them in (93), we get

$$\begin{cases} 2\lambda L \frac{x_{ij}}{r_{ij}} + \lambda \bar{l}_{ij} + z(f_{ij} - \frac{\lambda L}{\bar{r}_{min}^2}) + d_i - d_j = 0 \\ -\lambda L (\frac{x_{ij}}{r_{ij}})^2 + \frac{\lambda L}{\bar{r}_{min}^2} = 0 \end{cases} \tag{96}$$

this second condition gives us

$$-\lambda L (\frac{x_{ij}}{r_{ij}})^2 + \frac{\lambda L}{\bar{r}_{min}^2} = 0 \Leftrightarrow (\frac{x_{ij}}{r_{ij}})^2 = \frac{1}{\bar{r}_{min}^2} \Leftrightarrow \frac{x_{ij}}{r_{ij}} = \frac{1}{\bar{r}_{min}} \tag{97}$$

that we place in the first one, obtaining

$$2\frac{\lambda L}{\bar{r}_{min}} + \lambda \bar{l}_{ij} + z(f_{ij} - \frac{\lambda L}{\bar{r}_{min}^2}) + (d_i - d_j)\frac{r_{ij}}{\bar{r}_{min}} = 0 \tag{98}$$

it seems clear that this latter equation equals zero just for some values in z, while we wanted this to hold all the way through its variation.
Another issue is that, as one may note, some (x_{ij}^n, r_{ij}^n) sequences do exist along which q tends to $-\infty$.
This bothersome situation is due to the complete unboundedness of the formulation of the Benders' master problem in (87).
In order to avoid this issue, we try to do the same steps in a midterm Lagrangian relaxation, in which we make some of the terms in objective function go back to their constraint role.

5.3 Midterm Lagrangian relaxation

In our attempt to restore some constraint, we choose the ones that ensure the set within r can change to be bounded, that is box constraints (27)-(28)

$$\min_{(x_{ij}, r_{ij})} \sum_{(i,j)\in A} (f_{ij} r_{ij} + \lambda L \frac{x_{ij}^2}{r_{ij}} + \lambda \bar{l}_{ij} x_{ij} + d_i x_{ij} - d_j x_{ij}) - \sum_{i \in N} \pi_i b_i \tag{99}$$

$$x_{ij} z \leq r_{ij} \leq c_{ij} x_{ij} \tag{100}$$

$$x_{ij} \in \{0, 1\} \tag{101}$$

from now on, q refers to this function, and Benders' master problem too shifts to:

$$\min_{(r_0, r_{min})} r_0 \tag{102}$$

$$r_0 \geq g(r_{min}) + \min_{(x,r)}\{q(\theta, x, r, r_{min}) : \ (100),\ (101)\} \quad \forall \theta \geq 0 \tag{103}$$

$$r_{min} \in [\rho, r_\mu] = V \tag{104}$$

note that the sum on nodes is a constant shifting term, thus it's irrelevant to the minimum of the function, we explicit its value and put it away by now.
Here it's a recap of b definition:

$$b = \begin{cases} -1 & \text{if } \ i = s \\ 1 & \text{if } \ i = d \\ 0 & \text{otherwise} \end{cases} \tag{105}$$

while π can be either the opposite of dual SP labels, whenever they're finite, or 0, so that sum equates

$$-\sum_{i \in N} \pi_i b_i = d_d - d_s = d_d \tag{106}$$

for s is the source of the graph, and its label's automatically zero.

After this, we can start searching the minimum in (99)-(101).
Note that we can still work on the minimization arc by arc, moreover if $x_{ij} = 0$ it implies $r_{ij} = 0$, and $q(0,0) = 0$.
Hence, we want to figure out what happens when $x_{ij} = 1$, and this would eventually be the optimal choice in x_{ij} if the minimum turns out to be a negative value.
Working arc by arc lets us ignore indices, thus that's how our problem looks like

$$\phi(z) = \min_{(x,r)}\{fr + \frac{\lambda L x^2}{r} + \lambda l x + dx \mid xz \leq r \leq cx, \ x \in \{0,1\}\} \tag{107}$$

That annihilates for $x = 0$, as we noted.
When $x = 1$ instead, this is the function that we have to explicit

$$\phi(z) = \min_{r}\{fr + \frac{\lambda L}{r} + \lambda l + d \mid z \leq r \leq c\} \tag{108}$$

5.3.1 Minimizing within particular-case condition $\lambda = 0$

With this extra condition on λ, the problem becomes quite simple

$$\phi(z) = \min_{r}\{fr + d \mid z \leq r \leq c\} \tag{109}$$

in fact, minimizing function is a line that meets x-axis in the point $p = -\frac{d}{f}$.

Line is clearly strictly increasing, as such we pick lower value in r with regard to the only constraint, that is: $r = \min\{z, c\}$.
Thus, $\phi(z)$, when $\lambda = 0$, equals

$$\phi(z) = \begin{cases} fz + d & \text{se} \quad z \leq \min\{p, c\} \\ 0 & \text{se} \quad z > \min\{p, c\} \end{cases} \tag{110}$$

That definition preserves continuity if $\min\{p, c\} = p$, while it doesn't when that minimum equals the capacity of the arc, that would end up in a jump from negative value $fc + d$ to zero in that point.
In both cases this is a piecewise linear function, that isn't differentiable where it changes definition.
This function is both convex and concave along its two pieces, it's globally concave in case $\min\{p, c\} = p$.

5.3.2 General case

We now give a similar formulation for $\phi(z)$, assuming $\lambda > 0$.
First of all, we want to figure out where is located the minimum in r when $x = 1$, and if it returns a negative value for the function.
If $h(r)$ stands for the minimizing function, its gradient is

$$h'(r) = f - \frac{\lambda L}{r^2} \tag{111}$$

that vanishes in $\pm\bar{v} = \pm\sqrt{\frac{\lambda L}{f_{ij}}}$, thus it has a minimum in a positive point, and a negative point of maximum.
Subjecting our function to its constraint, we find that the optimal value for r is $r = \min\{c, \max\{\bar{v}, z\}\}$.

Now we want to discover whether in this points $h(r)$ equals a positive value, finding its roots and comparing them with that point.
This brought us to this equation

$$fr + \frac{\lambda L}{r} + \lambda l + d = 0 \tag{112}$$

it seems clear that we can assume $r > 0$, and multiply that relationship for its value, finding a quadratic equation

$$fr^2 + (\lambda l + d)r + \lambda L = 0 \tag{113}$$

Hence, h roots are

$$\zeta_{\pm} = \frac{-\lambda l - d \pm \sqrt{(\lambda l + d)^2 - 4f\lambda L}}{2f} \tag{114}$$

when the topic of the square root is non negative, or none.
If that function doesn't annihilates, computing it's limits in 0 and $+\infty$, that are both $+\infty$, assures us that the function is non negative in its minimum point, and we pick $\phi(z) = 0$.
Otherwise, if those roots exist and their value is positive, we have that

$$\zeta_- \leq \bar{v} \leq \zeta_+ \tag{115}$$

We are ready to define general-case ϕ function:

(i) If $h(r)$ doesn't vanish $\Rightarrow \phi(z) \equiv 0$

(ii) If $\zeta_\pm$ exist positive and $\bar{v} \leq c$

$$\phi(z) = \begin{cases} f\bar{v} + \frac{\lambda L}{\bar{v}} + \lambda l + d & \text{if } \ z \leq \bar{v} \\ fz + \frac{\lambda L}{z} + \lambda l + d & \text{if } \ \bar{v} \leq z \leq \min\{\zeta_+, c\} \\ 0 & \text{if } \ z > \min\{\zeta_+, c\} \end{cases} \tag{116}$$

(iii) If $\zeta_\pm$ exist positive and $\bar{v} > c$ and $\zeta_- < c$

$$\phi(z) = \begin{cases} fc + \frac{\lambda L}{c} + \lambda l + d & \text{if } \ z \leq c \\ 0 & \text{if } \ z > c \end{cases} \tag{117}$$

(iv) If $\zeta_\pm$ exist positive and $\bar{v} > c$ and $\zeta_- \geq c \Rightarrow \phi(z) \equiv 0$

let's take a closer look to (ii) and (iii) cases.
In the first one the function starts with a constant negative value and null value in derivative, until it reaches the point $\bar{v}$.
After that point the dependency on z becomes explicit, and the gradient of the function is now positive, as we can check in (111).
In that point both the function and its gradient are continuous and, despite it switches definition, that point preserves convexity, that's because the function changes definition exactly in its minimum.
After that it changes definition again in $\min\{\zeta_+, c\}$.
If that point equals ζ_+ the function is still continuous, but this time its gradient becomes null after a positive stretch, hence our function is no more convex.
When $\min\{\zeta_+, c\} = c$ both continuity and convexity can't hold, in fact the function behavior is similar to $\lambda = 0$ case: it grows until the point c where it jumps to null value both for function and its gradient (therefore it decreases).

Finally in case (iii) the function starts with negative constant value to jump to zero when it reaches the right edge of the constraint.
In this case we have a two-pieces linear function, both convex and concave.

This concludes the analysis of approximating function for fixed $\bar{r}_{min}$ and $\bar{\theta}$.
We end this paragraph noting that at $\bar{r}_{min}$ the function has always a null value.

That's due to SP dual conditions, that states that potential difference at the edges of an arc is smaller then the arc's cost:

$$\tilde{c}_{ij} \geq d_j - d_i = -d \;\equiv\; d \geq -\tilde{c}_{ij} \tag{118}$$

where $\tilde{c}_{ij}$ depends on r^* (see (32) and (34)), that depends on $\bar{r}_{min}$.
We now define

$$\psi(t) = ft + \frac{\lambda L}{t} + \lambda l = h(t) - d \tag{119}$$

note arc cost to be exactly equal to $\psi(r^*)$.
If we define $r(z) = \min\{c, \max\{\bar{v}, z\}\}$ the optimum found in previous analysis, we can read dual SP condition as

$$h(r(z)) = \psi(r(z)) + d \geq \psi(r(z)) - \psi(r^*) \tag{120}$$

Note that $\psi(t)$ increases in $[\bar{v}, +\infty)$, where both $r(z)$ and r^* lay, thus, for $r(z) \geq r^* \; \forall z \geq \bar{r}_{min}$, in (120) we have

$$h(r(z)) \geq \psi(r(z)) - \psi(r^*) \geq 0 \tag{121}$$

that easily implies $\phi(\bar{r}_{min}) = 0$.
This is an useful property, because from $\bar{r}_{min}$ on our approximation is certainly convex.
Now, we want to find a convex approximation for the first part of the function, in order to have an at worst two-pieces convex global approximation for Benders' master problem.
That allows us to perform a double line search to handle the maximum function in (87)-(88), in the subintervals $[\rho, \bar{r}_{min}]$ and $(\bar{r}_{min}, r_\mu]$ of $V = [\rho, r_\mu]$.

5.3.3 Convex approximations for ϕ

In this paragraph we face the issue of finding a convex approximation, shaping ϕ in a certain interval $[\underline{z}, \bar{z}]$.
For our purpose we can assume $\underline{z} > 0$, and $\bar{z} = \bar{r}_{min}$, that's based on latter ending paragraph remark.
This may bring us to quite loose approximation, anyway in this way we found a point that's common to all arcs in graph.
If we want a more strict approximating function, we should section V in c_{ij} and $\zeta_\pm$ too, but we have that those points depends on the features of arcs, thus we could need a 3-points per arc partition of our feasibility set V, that quickly becomes too difficult to handle.

Finding a convex approximation means to review all ϕ forms according with the value of λ.
We start finding one for $\lambda = 0$, we recap that this implies ϕ to be a piecewise linear function.

In this case we intuitively cut off non convexity point, taking in $(\bar{v}, \phi(\bar{v}))$ the linear shortcut to $(\bar{r}_{min}, \phi(\bar{r}_{min}))$.
It could happen that this shortcut crosses our function, in that situation we want to increase its slope, picking the $(c, \phi(c))$-$(\bar{r}_{min}, \phi(\bar{r}_{min}))$ linear worse approximation.
This concludes this easier case, let's have a look to the more general $\lambda > 0$, that is cases (ii) and (iii) in §5.2.2.
Linear shortcut still tempts us, but we have to grant its feasibility.
In case (iii) this is quite simple: the function is a two null slope pieces function, so it seems obvious to make it convex by linking $(c, \phi(c))$ to $(\bar{r}_{min}, \phi(\bar{r}_{min}))$ with a line.
Case (ii) is more troublesome, ϕ is convex in $[0, \min\{\zeta_+, c\}]$, but shortcut is surely no longer consistent, for, in $\bar{v}$, ϕ changes definition and starts increasing with punctual null slope.
Hence, we have to lower our shortcut in order to force it to remain in ipograph of ϕ.
A way to do this is finding a suitable ϵ that makes affordable the linking between $(\bar{v} + \epsilon, \phi(\bar{v}))$ and $(\bar{r}_{min}, \phi(\bar{r}_{min}))$, we define it r.
This could be done by picking the lines passing trough $(\bar{r}_{min}, \phi(\bar{r}_{min}))$ and forcing them to have just one intersection with the curve part of ϕ in $[\bar{v}, \min\{\zeta_+, c\}]$.
The r linking is characterized by its slope and its y-axis intersection:

$$m = \frac{\phi(\bar{v}) - \phi(\bar{z})}{\bar{v} + \epsilon - \bar{z}} \tag{122}$$

$$q = \phi(\bar{z}) - \bar{z}m \tag{123}$$

that we ask for equaling ϕ:

$$fz + \frac{\lambda L}{z} + \lambda l + d = mz + q \tag{124}$$

multiplying for z, that we can assume to be positive, we get

$$(f - m)z^2 + (\lambda l + d - q)z + \lambda L = 0 \tag{125}$$

We now impose that its discriminant vanishes

$$(\lambda l + d - q)^2 - 4(f - m)\lambda L = 0 \tag{126}$$

and we replace here q, to underline the fact it depends on m

$$(\lambda l + d + \bar{z}m)^2 - 4f\lambda L + 4fm\lambda L = 0 \tag{127}$$

defining $\beta = \lambda l + d$ and expanding quadratic term we have

$$\bar{z}^2 m^2 + (2\beta\bar{z} + 4f\lambda L)m - 4f\lambda L + \beta^2 \tag{128}$$

solving it in m eventually brings us to

$$\begin{aligned} m_\pm = \frac{-2\beta\bar{z} - 4f\lambda L \pm \sqrt{(2\beta\bar{z} + 4f\lambda L)^2 - 4\bar{z}^2(\beta^2 - 4f\lambda L)}}{2\bar{z}^2} = \\ \frac{-\beta\bar{z} - 2f\lambda L \pm \sqrt{2f\lambda L(2f\lambda L + \beta\bar{z} - 2\bar{z}^2)}}{\bar{z}^2} \end{aligned} \tag{129}$$

That's not an handling solution for sure, in addition to its potential issues in definition, and the unexpected two solutions ending.
Thus, we want to try another way to build our convex bound.

5.3.4 Alternative consistent bounds for ϵ

We present two supplementary ideas to build convex approximation.
The first is based on trying to force the shortcut to become valid, via lowering it down until it reaches the tangent point with ϕ.
Alternatively we can check the slope that ϕ reaches in $\min\{\zeta_+, c, \bar{z}\}$ and choose among the lines passing through $\bar{z}$ the one with that slope.
Those suggestions seem to decrease the accuracy of our bound, anyway we carry out calculation in both ways, to see if we can find an acceptable compromise between accuracy and ease.

We start with the first idea; this is the formulation for shortcut:

$$s(t) = m_s t + q_s = \frac{\phi(\bar{v})}{\bar{v} - \bar{z}} t - \bar{z} m_s \tag{130}$$

while ϕ reaches its slope in the point that grants

$$\phi'(z) = h'(z) = f - \frac{\lambda L}{z^2} = m_s \tag{131}$$

that is

$$\hat{z} = \sqrt{\frac{\lambda L}{f - m_s}} \tag{132}$$

Thus we can write the formulation for our linear convex approximation:

$$\hat{s}(t) = m_s t + \phi(\hat{z}) - m_s \hat{z} \tag{133}$$

Note that, for this to be defined, we need $m_s < f$.
This doesn't hold when shortcut has already a larger slope than ϕ one at each point.
In this situation we want to simply lower down shortcut until the point that makes ϕ change definition: $\min\{\zeta_+, c, \bar{z}\}$; that turns out to be a consistent bound again.

As regards the second idea, the slope that ϕ has in $\min\{\zeta_+, c, \bar{z}\}$ equals

$$h'(\min\{\zeta_+, c, \bar{z}\}) = f - \frac{\lambda L}{(\min\{\zeta_+, c, \bar{z}\})^2} \tag{134}$$

thus, our approximation is the line passing through $(\bar{z}, \phi(\bar{z})) = (\bar{r}_{min}, 0)$ with that slope:

$$r(t) = h'(\min\{\zeta_+, c, \bar{z}\})t - \bar{z}h'(\min\{\zeta_+, c, \bar{z}\}) \tag{135}$$

In view of our calculation the first idea in this latter paragraph seems to be quite up-and-coming, being careful not to let discontinuity that it produces in $\bar{r}_{min}$ take over.

It follows a brief look to other possibilities for the choice of our midterm relaxation, before we explore what happens to our approximation adding each arc's contribution and the terms $g(z)$ and d_d, that we ignored so far.

5.3.5 Further alternative versions for midterm Lagrangian relaxation

The idea on which our midterm relaxation is based is that the set, where the variables of the problem can move, becomes too easily unbounded.
There are different formulation to avoid this issue, each relying on confining r in a bounded set.
Some of these could eventually bring to easier approximation, with different properties, thus we sort through some of them, hoping we could find a global convex approximation for Benders' master problem.
One interesting idea, for example, is to keep total relaxation formulation, such as in (92), and finally add redundant constraint $\rho x_{ij} \leq r_{ij}$.
We move on those calculation in order to avoiding burden this treatment, especially because this turns out to be a non convex formulation.
We decided to underline this possible way for it frees constraint's depending on z and reduced the approximation to a piecewise linear form, such as $\lambda = 0$ previous case.
Anyway, we discarded this path for it doesn't guarantee an important property: the fact that in $\bar{r}_{min}$ all arcs certainly end their contributions, allowing us to build up a two pieces convex function.
That implies that we need to partition the V interval in a whole bunch of subinterval in which function is convex, or at least exploring all non convexity points to build a global convex bound.
This end in a worse issue that is that the function could be no longer tangent to Benders' master problem one, significantly threatening our bound's accuracy.

We cross over this analysis by showing some possible shapes of DCR problem's objective function.
Those are built by interpolation on its values computed on a dense partition of the feasibility set V.
Those in particular refer to graphs with more than 4000 arcs (up to 9500).
Note that we record lack of global convexity and potential discontinuity in objective function yet.
That remark made us desist from further investigation in potential global convex bound.

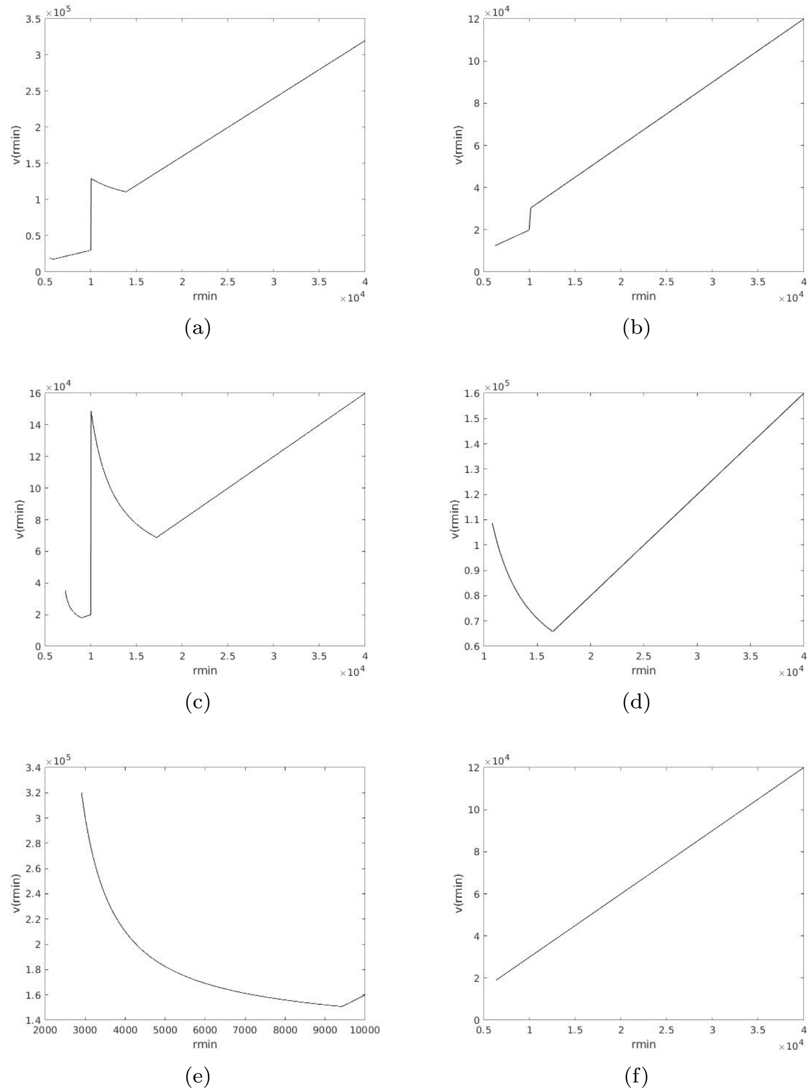

Figure 1: Changing forms for objective function of DCR problem

5.4 Complete formulation for approximating function

What we've done so far was studying single arc contribution to our approximation to Benders' master problem, with $\bar{r}_{min}$, thus $\bar{\theta}$, fixed, thinking at z as a parameter.
Then we simplified it by taking a further piecewise linear and convex lower bound for it; hence now we are ready to put those contributions together, adding constant terms and $g(z)$, to discover what our complete approximation looks like.
In this purpose, let Φ be RHS function in (102) for fixed θ, that is

$$\Phi(z, \bar{r}_{min}) = g(z) + \min_{(x,r)}\{q(\theta, x, r, z) : \ (100), \ (101)\} \tag{136}$$

We now add indices to our previous study, calling ϕ_{ij} the approximating function we found in §5.2 for the arc (i, j).
Thus, what we've done so far can be summed up as

$$\Phi(z, \bar{r}_{min}) = g(z) + d_d + \sum_{(i,j)\in A} \phi_{ij}(z, \bar{r}_{min}) \tag{137}$$

This is a valid lower bound for general objective function in master problem, which has, as we noted, bunches of points were convexity and even continuity don't hold, up to three per arc.
By evaluating that function in $z = \bar{r}_{min}$, it equals Lagrangian optimal value, thus in that point it matches the actual master problem's function, which allows θ to change with regard to r_{min} variation instead.
In order to avoid issues that non convexity could arise, we already built a further lower bound, that is the one we discussed in §5.2.3.
Those functions, which we refer to as $\tilde{\phi}_{ij}$, keep on matching the master problem function in $z = \bar{r}_{min}$, in addition they're convex in the two intervals $[0, \bar{r}_{min}]$ and $(\bar{r}_{min}, +\infty]$, which is a seriously auspicious property.
Thus, following property holds for those functions:

$$\sum_{(i,j)\in A} \phi_{ij}(z) \geq \sum_{(i,j)\in A} \tilde{\phi}_{ij}(z) \quad \forall z \in V \tag{138}$$

hence we can replace them in Φ definition, building a quite more loose approximation, yet it grants the useful properties we just underlined.
This put an end of issues due to the contributions of arcs.
What we have to do now is analyse remaining term:

$$g(z) + d_d = \frac{\lambda\sigma}{z} - \delta + d_d \tag{139}$$

where d_d is constant shifting that we actually can ignore, for we are now focusing on convexity.
Our last term $g(z) = \frac{\lambda\sigma}{z} - \delta$ is strictly decreasing for $z > 0$, and it's convex.

Thus, what's happening here is that the derivative of Φ starts with the value $g(\underline{z})$ plus the non-negative contributions by arcs, that is constant for their objective function is piecewise linear.
Those arcs who are already contributing in $\underline{z}$, are going to keep acting in this way, until z reaches $\bar{r}_{min}$, where all contributions from arcs vanish.
Meanwhile the ones that start with null contribution could pass all the way through in z without changing their derivative contribution, or activate in some point before they reach $\bar{r}_{min}$.
If it's so, they are remaining constant, after their sudden turning on, until they annihilate in $\bar{r}_{min}$.
Thus arcs can produce sudden variation of Φ derivative, but whenever they do change it, they increase its value by a fixed non negative amount, hence their contribution preserves convexity as long as $z \leq \bar{r}_{min}$.
After that point the derivative of Φ equals $g(z) + d_d$, that is a convex decreasing function, thus we eventually built our two-pieces convex approximation.
This can be summed up in

$$\Phi(z) = \begin{cases} g(z) + d_d + \sum_{(i,j)\in A} \tilde{\phi}_{ij}(z) & \text{se} \quad z \in V_1 \\ g(z) + d_d & \text{se} \quad z \in V_2 \end{cases} \tag{140}$$

where $V_1 = [\underline{z}, \bar{r}_{min}]$ and $V_2 = (\bar{r}_{min}, \bar{z}]$.

Now we can finally have a closer look to Benders' procedure.
Before that, though, we take advantage from convexity for last simplification step, that will hopefully clarify why two-pieces convexity was such a compelling necessity too.
In fact, for convex function, we have that local gradient is a global lower bound for the function itself.
Thus, what we want to do is replacing Φ with two lines which slope's given by its right and left derivative in $\bar{r}_{min}$:

$$\Phi'_+(\bar{r}_{min}) = \lim_{z \to \bar{r}^+_{min}} \Phi'(z) = -\frac{\lambda\sigma}{\bar{r}^2_{min}} \tag{141}$$

$$\Phi'_-(\bar{r}_{min}) = \lim_{z \to \bar{r}^-_{min}} \Phi'(z) = -\frac{\lambda\sigma}{\bar{r}^2_{min}} + \sum_{(i,j)\in A} \tilde{\phi}'_{ij}(\bar{r}_{min}) \tag{142}$$

that we're using in its place in the following procedure, given their ease.

5.5 Procedure for Benders' decomposition

First of all, a core issue for this decomposition to be performed is finding the set V of r_{min} values that grants feasibility, which we briefly discussed in §3.1.1.
Just to remind what we said then, we want to sort arc capacities vector in seeking for minimum value for which feasibility doesn't hold and that would be

the right edge of V.
Empirically one can become aware that ρ is actually a lower bound for the left edge in many realistic situation.
We can improve that bound noting that, for the problem to be feasible, we need to find a negative slope while looking for λ optimal value, that is, as we can check in §4, finding a λ and a related (x, r) solution for

$$\beta = \frac{\sigma}{r_{min}} - \delta + \sum_{(i,j)\in A} \frac{Lx_{ij}^2}{r_{ij}} + \bar{l}_{ij}x_{ij} \tag{143}$$

to be negative.
Sum being a positive term one, for β to become negative, we need that

$$\frac{\sigma}{r_{min}} - \delta \leq 0 \tag{144}$$

That is

$$r_{min} > \frac{\sigma}{\delta} \tag{145}$$

Thus we can lift up V left edge by asking r_{min} for being greater than $\max\{\rho, \frac{\sigma}{\delta}\}$. This is actually the point we picked in implementation, adjusting it by convex combination between lower known feasibility point and that edge, all the time we come across any non feasibility.

In case-study iteration we start building our piecewise linear function Φ for a certain fixed value $\bar{r}_{min}$.
Then we split V in two part V_1, V_2, separated by $\bar{r}_{min}$.
We save each linear part of Φ in its natural interval, where it represents a lower bound for Benders' function.
Thus we just have to manage how to run a line search on each subinterval already set at this time, finding the minimum among the maximum function formed by those linear cuts that belong to that certain interval.
We're taking the minimum value point among the minimal value points line search brought to light.
This would eventually be our new point, in which we start again building Φ function.
This method is clearly convergent, for it is so on single subintervals.
We have two stop conditions: one is optimality, that is when the gap between new point value on approximation and its value on actual objective function stays within a chosen accuracy ϵ.
The other one is satisfied whenever we already solved the problem, and then built Φ, for a new value, that is we're visiting again one point, that implies that we can no longer improve our lower bound.

We implemented and tested this whole algorithm and we're about to emphasize its behaviour by comparing it with the one of a general purpose solver, in the next paragraph.

As we're seeing this method finds often a solution that is not just Lagrangian relaxation one, but that matches general DCR problem optimality.
Before that we take a moment to talk about the heuristic and the ideas we analysed in order to reuse some computed information and optimize this latter procedure.

5.6 Heuristic and re-optimization

In running the so far outlined procedure, we want to save best primal feasible solution, as we remarked in §4.2.
Anyway, if we're just concerned about an upper heuristic bound, we can strongly simplify the approximation we proposed in §5.2.3.
In fact, in this case, we can let the approximating function passing through the general objective function, trading off potential looseness of our bound, with more ease in computing it.
For example, this situation enables to always chose the shortcut we talk about in §5.2.3 (that is the line from $(\bar{v}, \phi(\bar{v}))$ to $(\bar{r}_{min}, \phi((\bar{r}_{min})))$, which ends in a clear facilitation, preserving both convexity and continuity for approximating function.

Anyway, whether we're just concerned about upper bound or both upper and lower bounds, there exists some method to quickly cut off unpromising branches of V.
For instance, anytime the line search returns the new value, from which we start the next iteration, this can be used for further branching.
In fact the minimum value of the maximum function among linear cuts is crescent with respect to the operation of adding more cuts.
This guarantees that anytime we find a solution for line search in a subinterval that is lower than this bound, that subinterval can be ignored henceforth, because it could provide just larger valuated points, thus the minimum of Benders' function cannot be located in that branch.

A subtler an more promising condition is due to both empirical findings and theoretical remarks: it happens in many situations that minimum value is located before a descending stretch in which Lagrangian multiplier λ assumes continually zero as optimal value.
Let's have a closer look to what happens in this cases.
First of all our right cut, as one can check in (141), has null slope whenever $\lambda = 0$.
Left cut assumes a non-negative value instead, made up by arcs contributions.
That's enough yet in order to ignore from now on what is situated after this point for which $\lambda = 0$, but we can gain something more from this information.

Checking SP arc costs turns out in this condition for null value in λ

$$\tilde{c}_{ij} = f_{ij} r^*_{ij} \tag{146}$$

and, for $\bar{v} = 0$, $r^*_{ij} = \min\{r_{min}, c_{ij}\}$ (see (32)).
In particular, we're concerned about what happens when $r_{min} \leq c_{ij}$ for all arcs. That's because in c_{ij} we have jumps upwards in approximation function, and sometimes in actual function too, as we can check in figures in §5.2.5.
Hence, following this greedy impulse, we end up in $r^*_{ij} = r_{min}$ for all arcs, thus costs in SP decrease uniformly as r_{min} goes down.
In this situation SP solution cannot do without being constant, as long as $\lambda = 0$ is still optimal.
Thus we already know that this branch cannot contain the minimum for our problem, and we want to rapidly descend along this stretch to find us in a different situation, which empirically means quite often to find ourselves in the actual optimal point.
We just have to work this descending out.
A first remark in that direction is that, as long as the slope of dual function in (67) stays negative, $\lambda = 0$ is still optimal.
That can be thought of as

$$\beta(x, r) = \frac{\sigma}{r_{min}} - \delta + \sum_{(i,j)\in A} \frac{L x^2_{ij}}{r_{ij}} + \bar{l}_{ij} x_{ij} \leq 0 \tag{147}$$

defining P_0 as the constant optimal path for SP we can reformulate this:

$$\frac{\sigma}{r_{min}} - \delta + \sum_{(i,j)\in P_0} \frac{L}{r_{min}} + \bar{l}_{ij} \leq 0 \tag{148}$$

that is decreasing in $\bar{r}_{min}$ and vanishes in

$$r^0_{min} = \frac{\sigma + |P_0| L}{\delta - \sum_{(i,j)\in P_0} \bar{l}_{ij}} \tag{149}$$

Anyway, this condition doesn't grant that the descendent stretch ends here: $\lambda = 0$ can still be optimal solution even when β is positive (it's just an ipergradient of our function), and empirically it does indeed.
Thus, what we want to do is to perform a line search between this point and a smaller one for which we already know that $\lambda = 0$ is non optimal, in order to find the minimum point in which this relation holds.
This may seem so inefficient, because we're doing something similar to our method, but without been helped by the cuts in doing it.
That would be true if we should solve the entire Lagrangian problem to clue whether or not zero is its solution.
Actually this could be done simply evaluating SP optimum in $\lambda = 0$ and $\lambda = \epsilon$, by comparing those value we're capable of stating whether or not zero is our solution.

We conclude this paragraph opening up the possibility of similar properties to be found in other Lagrangian reformulation, that can eventually bring to an even more elaborate and efficient solution.

6 Numerical results

All the experiments have been performed on a (currently, rather low-end) PC with 3.7Gb RAM, running a 64 bits Linux operating system (Ubuntu 20.04.2 LTS).
All the codes were compiled with gcc 9.3.0 and -O3 optimizations.
The MISOCP model wes solved by the state-of-the-art, off-the-shelf, commercial solver Cplex 20.1.
The solver was ran without time limit and with default parameters.
It allowed us to solve both the whole DCR problem and its continuous relaxation, which we solved twice: the second time with a larger amount of cuts.
Constructing a set of significant DCR instances is a nontrivial exercise; fortunately, the FNSS tool [5] provides a number of expert-tuned options to help devising realistic models of current telecommunications networks.
The generation process starts by selecting a network topology.
For this, we considered two sets of real-world IP network topologies: the Internet Topology Zoo [6] ones, and the SNDlib ones [7], which can be downloaded in gml format.
Furthermore, in order to test our models on larger instances we used also random topologies generated according to the Waxman model [8].
This can be done by FNSS, which allows to generate random Waxman topologies simply by specifying the number of nodes n and the probability parameter $\alpha \in (0, 1]$, representing the link density.
In a first tabular some information is shown about graphs we used in testing and the number of flows we routing among them.
Afterwards in second tabular accuracy and speed of our method are reported, one can find there Cplex results for optimal solution and both continuous relaxations, there the rate of optimal solved flows by our method lies.
That tabular is divided into two halves, the first one refers to the method represented in this paper; we can find here the rate of optimal solved cases, medium running time for both optimal and suboptimal cases, and some information related to suboptimal cases performance: the "subopt dgap" column reports the quality of our lower bound in non optimal cases, while "subopt pgap" represents the quality of the heuristic, that is the quality of our upper bound in term of accuracy with respect to the value of Cplex integer solution.
The second half of this table is dedicated to Cplex performance: it reports relaxation and improved relaxation medium running time and the quality of their bound, it ends with medium running time for solving Cplex in integer case.

As we can see, this method and its optimizations enable a very efficient solution for the problem, that results in optimal solution with an high rate.
Moreover when this solution ends in just an upper and lower bound that don't meet at optimum, we still have an accurate bound, and just few branching moves

are required to find DCR optimum.
Note that in this tabular we pointed Waxman100 improved relaxation dual gap, because from this size for the graph on it's no more reliable.
In fact, it starts to return lower bounds that are larger than optimal value, or, even worse, it concludes that problem is infeasible while it is not, and that's why we ignored second relaxation dual gap value from this point on.
Anyway, we tracked its time, to underline its ineffectiveness.
Moreover note that for small graphs our performance is close to cplex one, that is due to preprocessing and reoptimization weight in time, and this could be improved by reducing reoptimization in those cases, especially one suggested in the end of §5.5.
Meanwhile, as soon as dimension starts to grow, we can appreciate how more effective our method is, and how this increases in effectiveness according to the growth in number of arcs.

Topology	m	n	n. flows	m/n
Cogentco	486	197	600	2.47
Colt	354	153	600	2.31
DialTelecomCz	302	193	600	1.56
Pern	258	127	600	2.03
Waxman100	414	100	600	4.14
Waxman200	1550	200	400	7.75
Waxman300	3630	300	200	12.1
Waxman500	9738	500	150	19.46

Table 1: Topologies information and routed flows

	Lagr/Bend				Cplex				
					Rel.		Impr. Rel.		Int.
Topology	opt rate	time	subopt dgap	subopt pgap	time	dgap	time	dgap	time
Cogentco	98.5%	3.35e-1	2.84e-2	9.09e-2	1.72e+0	6.09e-1	3.17e+0	2.71e-2	1.77e+0
Colt	99.5%	2.98e-1	1.17e-1	2.01e-1	2.82e-1	2.26e-1	2.69e-1	2.37e-1	3.01e-1
DialTelecomCz	91.4%	1.33e-1	1.58e-1	1.52e-1	1.26e+0	1.81e-1	1.48e+0	1.38e-1	1.24e+0
Pern	99.8%	2.81e-1	1.48e-2	1.12e-1	6.33e-1	1.32e-3	1.23e+0	1.89e-1	6.65e-1
Waxman100	95.7%	1.75e-1	6.11e-2	8.51e-2	2.66e+0	1.63e-1	3.03e+0	7.86e-1*	3.89e+0
Waxman200	97.3%	1.02e+0	7.62e-2	1.85e-1	9.13e+0	1.97e-1	1.32e+1		1.67e+1
Waxman300	96.5%	2.56e+0	1.51e-2	5.88e-2	4.57e+1	2.32e-1	5.58e+1		8.57e+1
Waxman500	96.7%	8.34e+0	8.81e-2	2.29e-1	1.72e+2	6.11e-1	2.31e+2		3.31e+2

Table 2: Numerical results

7 Conclusions

Routing under QoS constraint is clearly an interesting field for research, for we have to face this problem in a lot of situation, for instance the ones we mentioned in introduction.
The MISOCP formulation that is presented in [1] has the important property to be handled by general-purpose solvers.
Anyway, as we can check in the previous chapter, the numerical results emphasize their limits, when the size of the graph grows.
In this situation those solvers can manage to find a solution in a rather large amount of time, hence the urge of a specialized algorithm becomes more compelling.
In this environment our method finds its natural location, for it solves the problem quickly and at optimality with an high rate.
Whenever this doesn't happen, our bound is still tighter than the one we could hope to find via continuous relaxation, and the time gap allows to adjust those bound performing a B&B procedure.
The difference in time being so relevant, this will hopefully turn out to be still a more effective procedure in those suboptimal cases, as it is in most of the encountered ones.
Moreover this paper shows a even more general form for Benders' decomposition, that could be performed whenever a problem could be simplified by fixing one or more variables, while some nonlinear constraints occurs.
If we are interested in those theoretical aspect, we got to watch computational issues due to excessive partitioning of the domain of those complicated variable, speeding up this procedure by taking advantage of internal and specific properties of the problem, as we did.

Thus, those theoretical aspects, the fact that this problem can be encountered in a such wide range of situations, the complexity of the solution we presented and the specificity of the remarks we found, in order to speed it up, highlight what interested us and prompted us to face this problem.

References

[1] Frangioni, A., and Galli, L and Stea, G, *QoS Routing with worst-case delay constraints: models, algorithms and performance analysis*, technical report, dipartimento di Informatica, università di Pisa, 2015.

[2] Frangioni, A., *Unconstrained optimization I Gradient-type methods*, lezioni, dipartimento di Computer Science, università di Pisa.

[3] Frangioni, A. and Gentile, C., *Perspective cuts for a class of convex 0-1 mixed integer programs*, technical report, Istituto di Analisi dei Sistemi ed Informatica, consiglio nazionale delle ricerche, 2002.

[4] The PRIN Collective, *Lagrangian approaches for the delay-constrained shortest path problem*, technical report.

[5] Saino, L. and Cocora, C. and Pavlou, C. (2013) A toolchain for sim- plifying network simulation setup. *Proceedings of SIMUTOOLS '13*, Cannes, France 06–08 March. ICST, Brussels, Belgium.

[6] The internet topology zoo: http://www.topology-zoo.org/

[7] Sndlib: http://www.sndlib.zib.de

[8] Waxman, B.: Routing of multipoint connections. IEEE J. Sel. Areas Commun. 6(9), 1617–1622 (1988)

[9] Benders, J. F., *Partitioning procedures for solving mixed-variables programming problems* (1962)

[10] Geoffrion, A. M., *Generalized Benders Decomposition*, journal of optimization theory and applications: Vol. 10, No. 4, 1972

[11] Geoffrion, A.M. (1974). *Lagrangian Relaxation and its Uses in Integer Programming.* Math. Prog. Study 2, 82–114.