Short-term Mobility: Scientific report

Claudio Gentile

August 4, 2011

Abstract

During my visit at the Department of Statistics and Operations Research of the Politechnic University of Catalonia in Barcelona from June 13 to July 04, 2011, I studied the Controlled Tabular Adjustment problem (CTA) of statistical tables by means of the Euclidean distance.

Any institution that disseminates data in aggregated form has to guarantee that individual confidential information is not disclosed. This is specially relevant for tabular data released by national statistical agencies (NSAs). Protection techniques for tabular data are divided into perturbative (they modify information) and non-perturbative (they hide information). Among the perturbative approaches we find controlled tabular adjustment, an emerging technique being considered by some NSAs (such as Eurostat). Briefly, given some unsafe tabular data, CTA attempts to find the protected closest one according to some distance. CTA results in a challenging integer optimization problem. Previous research on CTA focused on L_1 distances. This is the first attempt to efficiently solve CTA with Euclidean distances, formulated as a mixed integer quadratic problems. These are approached by perspective reformulations that have recently given good performances. We compare different algorithms to solve perspective reformulations on CTA, reporting the computational results. This report will be the base for a paper that will be sent in a major international journal.

1 Introduction

The most important mission of National Statistical Agencies (NSAs), and a significant mission of several other institutions, is to provide high-quality statistical data. These data are disseminated either in disaggregated (i.e., microdata or microfiles) or aggregated (i.e., tabular data) form. A microdata file is a matrix of individuals by variables, where each cell provides the information of a particular individual for some particular variable. Crossing two or more categorical variables of the microdata file produces tabular

	$\cdots$	S_i	$\cdots$	S_j	$\cdots$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
AS_k	$\cdots$	450M$	$\cdots$	35M$	$\cdots$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
AS_l	$\cdots$	625M$	$\cdots$	770M$	$\cdots$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	...

(a)

	$\cdots$	S_i	$\cdots$	S_j	$\cdots$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
AS_k	$\cdots$	22	$\cdots$	1 / 2	$\cdots$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
AS_l	$\cdots$	27	$\cdots$	33	$\cdots$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	...

(b)

Figure 1: Example of disclosure in tabular data: (a) turnover and (b) number of companies per activity sector and state.

data, either a single multiway or multidimensional table, or a set of related tables. There are stringent requirements that no confidential or sensitive information of any individual can be disclosed from the released data; not only this is dictated by law, but also respondents (e.g., of a census) may be tempted to hide or change information if they suspect that their confidential information may be released. This justifies the interest in *statistical disclosure control*, i.e., the set of techniques that can be deployed to protect sensitive information. In particular, the focus of this work is on *tabular data protection*; seminal work on this field can be found in Bacharach (1966), and the current state-of-the-art is described in the recent surveys of Salazar-González (2008) and Castro (2011), as well as in the monographs Willenborg and de Waal (2000), Hundepool et al. (2010).

Although tabular data provide aggregated information, the publication of some cells may jeopardize individual information. Consider the small example of Figure 1: if there is only one company with activity sector AS_k in state S_j, then any attacker knows the turnover of this company. For two companies, any of them can deduce the other's turnover, becoming an internal attacker. Clearly, the risk in the example is due to a small number of respondents in cell (AS_k,S_j). However, even if the number of respondents was larger, there could be a disclosure risk if some companies can obtain a tight estimator of another's turnover (for instance by subtracting its own contribution from the cell value). Unsafe or sensitive cells are a priori determined before the application of any tabular data protection method, by applying some "sensitivity rules". These rules are out of the scope of this work; e.g., see Domingo-Ferrer and Torra (2002), Hundepool et al. (2010) for details.

Disclosure limitation techniques for tabular data are classified as perturbative if one is allowed to add small perturbations or adjustments to released data, and as nonperturbative if released cell values must be exact, and therefore one is only allowed to entirely eliminate cells. Clearly, nonperturbative approaches are therefore more rigid than perturbative ones. Furthermore, the most widely used nonperturbative approach, *cell suppression* (Kelly et al. 1992, Fischetti and Salazar-González 2001, Castro 2007), requires the solution of large-scale optimization problems to identify the optimal set of cells to be suppressed. It is perhaps not surprising, therefore, that perturbative approaches are being considered as emerging technologies for tabular data protection. In particular, *Controlled Tabular Adjustment* (CTA) is gaining recognition and acceptance among NSAs (Zayatz 2009), as testified by the recent handbook Hundepool et al. (2010) and by the fact that it is currently used by Eurostat (Statistical Office of the European Communities) within a wider protection scheme for tabular data (Giessing et al. 2009). Figure 1 can be used to illustrate CTA. If cell (AS_k,S_j) of table (a) is considered sensitive, with *lower* and *upper protection levels* of 5, then the published value of this cell must be in the range $(-\infty, 30] \cup [40, \infty)$. We say that the *protection sense* is "lower" or "upper" if the published value is, respectively, in $(-\infty, 30]$ or in $[40, \infty)$. The remaining cells in the same column and row of the sensitive cell have to be accordingly adjusted to preserve the marginal values, while minimizing the distance between the original and the released values. Since each sensitive cells introduces a disjunctive constraint, which can be formulated by adding one binary variable, when the number sensitive cells is large CTA is a difficult combinatorial optimization problem.

It is worth remarking that, while the tables of Figure 1 are two-way (two-dimensional) ones, in general the situation can be much more complex. Tables can be classified in (i) k-dimensional tables, which are obtained by crossing k categorical variables; (ii) hierarchical tables, or set of tables that share some variables with hierarchical structure (e.g., "country", "state/province", "city"); (iii) linked tables, the most general situation, which is a set of tables that are obtained from the same microdata. A particularly interesting case for NSAs, which will be tested in this work, are two-dimensional hierarchical tables that share one hierarchical variable (e.g., tables that show the turnover crossing "activity sector" by "country", "activity sector" by "state/province", and "activity sector" by "city"). These are named *one-hierarchical two-dimensional* tables (or 1H2D for short), and their relations can be represented as a tree of tables. However, table relations for any type of table are represented by linear constraints, where the sum of the inner cells is equal to the marginal cell; thus, the techniques developed in this paper are applicable to the most general case (linked tables") as well.

In all previous works on CTA, the L_1 or Manhattan norm has been used to measure the distance between the original and the protected published data (Dandekar and Cox 2002, Castro 2006). This has the advantage that CTA can then be formulated as a Mixed Integer Linear Problem (MILP) with a number of variables and constraints that is linear in the size of the table, and whose solution can therefore be attempted with generic state-of-the-art MILP solvers. By contrast, formulations of the cell suppression problem are much larger and require the application of specialized approaches such as Benders decomposition. This is not to say that CTA, even with the L_1 distance, is an easy problem; for large (1H2D) tables MILP solvers may require a long time even to provide a first feasible solution, and therefore heuristic approaches (González and Castro 2011) are required to provide practical solutions in a reasonable time. It can be expected that CTA with L_2 (Euclidean) distance, which results in a Mixed Integer Quadratic Problem (MIQP), is even more difficult to solve; this is likely the reason why this work is, to the best of our knowledge, the first one where such a feat is attempted. Yet, protecting a table using L_2 in CTA has several benefits:

- Weighting the distance between the original and the published cell value by the inverse of the original cell value, the objective function of CTA minimizes the well-known χ^2 distance between the original and the released table, which is useful for the statistical evaluation of the results.

- The L_2 distance more evenly distributes the deviations induced by sensitive cells to other cells. This avoids concentration of deviations in few cells, which improves the overall utility of the published data, as measured, e.g., by the number of non-sensitive cells whose published value is "significantly" different from the original data.

- From a computational point of view, once the binary variables are fixed (i.e., the protection sense is decided), the solution of the resulting continuous problem can be more efficient for L_2 than for L_1 if specialized interior-point methods are used, which can be orders of magnitude faster than state-of-the-art general-purpose solvers (Castro and Cuesta 2010).

On the other hand, the protected values provided by CTA with the L_2 distance will likely be more fractional than those provided by the L_1 distance, which has been often observed in practice to provide integer values even without imposing integrality constraints. Yet, this is not a significant drawback since CTA is mainly used for "magnitude" tables which do not provide frequencies but information about a third continuous variable (salary, net profit, turnover, ...) which is most often fractional.

Fortunately, as we will see, the main structural characteristic of MIQP formulations of CTA with the L_2 distance (from now on, simply "CTA") is very closely related to *convex separable quadratic-cost models with semicontinuous variables*, which are naturally formulated as in the following (fragment of) MIQP

$$\min \left\{ wz^2 + cy \ : \ yl \leq z \leq yu \ , \ y \in \{0,1\} \right\} \tag{1}$$

where $w > 0$ and $l < u$. This is useful because (1) admits the *Perspective Reformulation* (PR)

$$\min \left\{ wz^2/y + cy \ : \ yl \leq z \leq yu \ , \ y \in \{0,1\} \right\} \ . \tag{2}$$

Despite the weird look and the apparent ill-definiteness at $y = 0$, the objective function in (2) is convex, and it actually is the *convex envelope* of an appropriately re-defined version of the objective function in (1), i.e., the best possible objective function to have when the integrality constraints $y \in \{0,1\}$ are relaxed to $y \in [0,1]$. Indeed, (2) has at least two possible further reformulations which avoid the fractional term in the objective function with the associated difficulties (nondifferentiability, possible numerical problems) at $y = 0$: one is the Mixed Integer Second-Order Cone Program

$$\min\left\{v + cy : yl \leq z \leq yu, \sqrt{wz^2 + (v-y)^2/4} \leq (v+y)/2, y \in \{0,1\}\right\} \tag{3}$$

(Tawarmalani and Sahinidis 2002, Aktürk et al. 2009, Günlük and Linderoth 2008), and another is the Semi-Infinite MILP

$$\min\left\{v + cy : yl \leq z \leq yu, v \geq w(2\gamma z - \gamma^2 y),\ \gamma \in [l,u], y \in \{0,1\}\right\} \tag{4}$$

where γ is the index of the infinitely many linear constraints (called *Perspective Cuts* in Frangioni and Gentile (2006)) whose pointwise supremum completely describes the objective function in (2). Either (3) or any finite approximation to (4)—typically, to be iteratively refined—can be used as models of (2), whose continuous relaxation is significantly stronger than that of (1) and that therefore is a more convenient starting point to develop exact and approximate solution algorithms (Frangioni and Gentile 2006, 2007, Günlük and Linderoth 2008, Aktürk et al. 2009, Frangioni et al. 2009). Somewhat surprisingly, the potentially very large and approximated (4) appears to be most often preferable to the compact and exact (3) in the context of exact or approximate enumerative solution approaches (Frangioni and Gentile 2009), likely due to the better reoptimization capabilities of simplex methods for linear and programs w.r.t. these of interior point methods for conic programs. Furthermore, a different approach has been recently proposed in Frangioni et al. (2011) that can be applied under some restrictive hypotheses, some (but not all) of which satisfied in our application. The idea is to recast the continuous relaxation of (2) as the minimization over $z \in [0, u]$ of the function

$$\phi(z) = \min_y \left\{ \ wz^2/y + cy \ : \ ly \leq z \leq uy \ , \ y \in [0, 1] \ \right\} \tag{5}$$

which effectively eliminates the y variable from the problem. The function ϕ is convex (partial minimization of a convex function), and its closed form can be computed by simple algebraic steps revealing a piecewise-quadratic function with at most two pieces, at most one of them actually quadratic (and the other linear). When the underlying problem has a useful structure (e.g., network flow or knapsack), the continuos relaxation of (2) obtained in this way retains that structure, which allows to use specialized algorithms to solve it and therefore to outperform both (3) and (4). Yet, direct application of that approach is only possible under rather restrictive assumptions that are not satisfied in our case.

In this paper we discuss the application of Perspective Reformulation techniques to the CTA problem. In particular, other than the standard approaches (3) and (4) we develop and test a new reformulation partly inspired by the results of Frangioni et al. (2011). However, since our problem is different and somewhat more complex, the "projected" version of the PR we obtain is substantially different and trickier to use. Thus, instead of insisting in keeping the equivalence with the original formulation we "drop the nastier pieces" and end up with an *approximated* reformulation, which is only as tight as the PR in some special cases, and looser otherwise. However, this reformulation results in a simpler (although non-separable) MIQP to be solved, and therefore it is most often preferable to the standard approaches (3) and (4); furthermore, it suggests a simple modification to the latter which invariably improves its performances. Armed with these results we show on a large experimental set that CTA for 1H2D of realistic sizes can often be solved effectively enough.

We remark that the Perspective Reformulation approach is much more widely applicable than the simple quadratic case we consider here: it not only applies to the objective function but also to constraints $f(z) \leq 0$ that are "activated" if and only if a binary variable y is 1, f can be any closed convex (possibly, SOCP-representable) function, z can be a vector whose feasible region can by any bounded polyhedron; see (Ceria and Soares 1999, Tawarmalani and Sahinidis 2002, Grossmann and Lee 2003, Frangioni and Gentile 2006, Hijazi et al. 2011) and the recent survey Günlük and Linderoth (2011). Thus, some of the ideas developed here could be extendable to more complex situations.

2 Formulations of the CTA problem

Any CTA problem instance, either with one table or a number of tables, can be represented by the following elements:

- a set of n cells a_i, $i \in \mathcal{N} = \{1, \ldots, n\}$, that satisfy m linear relations $Aa = b$ ($a = [a_i]_{i \in \mathcal{N}}$); in the general case, if $\mathcal{I}_j$ is the set of *inner* cells of relation $j \in \{1, \ldots, m\}$, and t_j is the index of the *total* or *marginal* cell of relation j, the constraint associated to this relation is $\sum_{i \in \mathcal{I}_j} a_i - a_{t_j} = 0$;
- the subset $\mathcal{S} \subseteq \mathcal{N}$ of indices of sensitive cells, and hence its complement $\mathcal{U} = \mathcal{N} \setminus \mathcal{S}$;
- a vector of nonnegative cell weights $w = [w_i]_{i \in \mathcal{N}}$;
- *finite* lower and upper bounds $\bar{l}^a \leq a \leq \bar{u}^a$ for each cell reasonably known by any attacker;
- nonnegative *lower and upper protection levels* for each confidential cell $i \in \mathcal{S}$, l_i and u_i respectively, such that the released values $x = [x_i]_{i \in \mathcal{N}}$ are considered to be safe if they satisfy

$$\text{either } x_i \geq a_i + u_i \text{ or } x_i \leq a_i - l_i \qquad \text{for all } i \in \mathcal{S} \ . \tag{6}$$

Given any weighted distance $\| \cdot \|_w$, CTA can then be formulated as

$$\min \left\{ \ \|x - a\|_w \ : \ Ax = b \ , \ \bar{l}^a \leq x \leq \bar{u}^a \ , \ (6) \ \right\} \tag{7}$$

since one seeks for the released values x that are closest (in the given norm) to the true values a, compatible with the relationships that a is known to have to satisfy, and protected according to (6). Of course, the *disjunctive constraints* (6) are the difficult part of the problem, their feasible region being nonconvex. Formulating it hence requires some nonconvex element, the simplest one being a vector of binary variables $y = [y_i]_{i \in \mathcal{S}} \in \{0,1\}^{|\mathcal{S}|}$. It is also convenient to restate the problem in terms of the *deviations* $z = x - a$ from the true cell values, which therefore have to satisfy $\bar{l}^a - a = \bar{l} \leq z \leq \bar{u} = \bar{u}^a - a$; this gives the formulation

$$\begin{array}{lll} \min & \|z\|_w & \\ & Az = 0 & \\ & \bar{l} \leq z \leq \bar{u} & \\ & \bar{l}_i(1-y_i) + u_i y_i \leq z_i \leq \bar{u}_i y_i - l_i(1-y_i) & i \in \mathcal{S} \\ & y_i \in \{0,1\} & i \in \mathcal{S} \end{array} \tag{8}$$

with "natural big-M constraints". Indeed, when $y_i = 1$ one has $z_i \geq u_i$ and thus the *protection sense* is "upper", while when $y_i = 0$ one rather gets $z_i \leq -l_i$ and thus the protection sense is "lower". While this formulation is correct, it would provide rather weak bounds when its *continuous relaxation* is formed by replacing the (nonconvex) integrality constraints $y_i \in \{0,1\}$ with (the convex) $y_i \in [0,1]$. The simple example with $n = 1$, "empty" A, $l_1 = u_1 = 10$ and $-\bar{l}_1 = \bar{u}_1 = 100$ shows that for $y_1 = 1/2$ the solution $z_1 = 0$ is feasible to the relaxation, whose optimal value is therefore 0 while the optimal value of the integer problem is $\|10\|_w$. Since weak bounds are very detrimental for the solution of the problem via exact or approximate approaches, we aim at constructing "better" formulations of the problem.

A first step in this direction is to introduce vectors of *positive and negative deviations* $z^+ \in \mathbb{R}^n$ and $z^- \in \mathbb{R}^n$, respectively, thereby redefining $z = z^+ - z^-$; this allows to reformulate the disjunctive constraints in (8) as

$$\begin{array}{rcll} u_i y_i \leq & z_i^+ & \leq \bar{u}_i y_i & \\ l_i(1-y_i) \leq & z_i^- & \leq -\bar{l}_i(1-y_i) & i \in \mathcal{S} \\ y_i & \in & \{0,1\} & \end{array} \tag{9}$$

As before, when $y_i = 1$, the constraints force $u_i \leq z_i^+ \leq \bar{u}_i$ and $z_i^- = 0$, thus the protection sense is "upper"; conversely, when $y_i = 0$ we get $z_i^+ = 0$ and $l_i \leq z_i^- \leq -\bar{l}_i$, thus the protection sense is "lower". This alone is not enough to improve on the bounds, as in the above example we now have $z_1^+ = z_1^- = 5$ as a feasible solution for $y_1 = 1/2$, which still leads to a null bound. However the advantage of this formulation is that we now have *two semicontinuous variables*, to which we can hope to apply Perspective Reformulation techniques. This is not straightforward: the two semicontinuous variables are governed by the *same integer variable*, and unlike in standard cases—where this is possible, provided that all variables are "active" or "inactive" at the same time—one of them is "active" if and only of the other is not. Furthermore, the objective function is *nonseparable* in z^+ and z^-, and the convex envelope of multilinear functions, even if with only two variables as here, is notoriously a complex object (cf. Luedtke et al. (2010) and the references therein) so that "dirty tricks" have to be used (Frangioni and Gentile 2007) in order to apply PR techniques. Thus, the next paragraph will be devoted to the study of the convex envelope for our particular case.

3 Perspective Reformulations of the CTA problem

In the following we will most often concentrate on a single cell $i \in \mathcal{S}$; thus, to simplify the notation we will consider the index i as fixed and drop it. In order to improve the lower bound provided by the continuous relaxation, one possibility is to compute the convex envelope of the nonconvex function

$$f(z^+, z^-, y) = \begin{cases} w(z^+ - z^-)^2 & \text{if } u \le z^+ \le \bar{u}, z^- = 0 \text{ and } y = 1 \\ w(z^+ - z^-)^2 & \text{if } l \le z^- \le -\bar{l}, z^+ = 0 \text{ and } y = 0 \\ +\infty & \text{otherwise} \end{cases} \tag{10}$$

This can be accomplished by considering two arbitrary points $u \le \bar{z}^+ \le \bar{u}$ and $l \le \bar{z}^- \le -\bar{l}$ and computing the convex combinations of the two tuples in the epigraphical space

$$(\ \bar{z}^+, \ 0, \ 1, \ w(\bar{z}^+)^2 \ ) \qquad\qquad (\ 0, \ \bar{z}^-, \ 0, \ w(\bar{z}^-)^2 \ ) \ .$$

In other words, taking any arbitrary convex combinator $\theta \in [0, 1]$ and using the shorthand $f(z) = wz^2$ (which also suggests how the approach can be generalized), we have

$$\begin{aligned} &\theta(\ \bar{z}^+ \ , \ 0 \ , \ 1 \ , \ f(\bar{z}^+) \ ) + (1-\theta)(\ 0 \ , \ \bar{z}^- \ , \ 0 \ , \ f(\bar{z}^-) \ ) \quad = \\ &(\ \theta\bar{z}^+ \ , \ (1-\theta)\bar{z}^- \ , \ \theta \ , \ \theta f(\bar{z}^+) + (1-\theta) f(\bar{z}^-) \ ) \end{aligned}$$

Now, identifying $\theta \equiv y$, $z^+ \equiv \theta\bar{z}^+$ and $z^- \equiv (1-\theta)\bar{z}^-$ we can rewrite the above as

$$\left(\ z^+ \ , \ z^- \ , \ y \ , \ yf\left(\frac{z^+}{y}\right) + (1-y)f\left(\frac{z^-}{1-y}\right) \ \right)$$

which finally leads to

$$\overline{co}f(z^+,z^-,y) = \begin{cases} w\left(\frac{(z^+)^2}{y} + \frac{(z^-)^2}{1-y}\right) & \text{if } \begin{array}{rcl} uy \leq & z^+ & \leq \bar{u}y \\ l(1-y) \leq & z^- & \leq -\bar{l}(1-y) \\ & y \in [0,1] & \end{array} \\ +\infty & \text{otherwise} \end{cases}$$

and therefore to the following PR of (8):

$$\min \quad \textstyle\sum_{i\in\mathcal{U}} w_i(z_i^+ - z_i^-)^2 + \sum_{i\in\mathcal{S}} w_i\big(\ (z_i^+)^2/y_i + (z_i^-)^2/(1-y_i)\ \big) \quad (11)$$

$$A(z^+ - z^-) = 0 \quad , \quad 0 \leq z^+ \leq \bar{u} \quad , \quad 0 \leq z^- \leq -\bar{l} \quad , \quad (9) \qquad (12)$$

In other words, the PR can be seen as being obtained as follows:

1. substitute $(z^+ - z^-)^2$ in the objective function with $(z^+)^2 + (z^-)^2$, which is correct since in each *integer* solution $z^+z^- = 0$;

2. treat z^+ and z^- as two distinct semicontinuous variables with two distinct binary variables, say y^+ and y^-, and apply the standard PR (2);

3. now exploit the fact that $y^+ + y^- = 1$ to replace $y^+ = y$ and $y^- = 1-y$.

While this sequence of reformulation steps could have been devised independently (but, to the best of our knowledge, has never had), our analysis has suggested them, as well as proved that this is in fact the convex envelope of the fragment. Actually, the analysis suggests that one can further improve the PR even regarding the non-sensitive cells $i \in \mathcal{U}$. In fact, these can be considered as sensitive cells with $l = u = 0$ and $\bar{l} = \bar{u} = +\infty$, and therefore it is clear that one could have been taken

$$\text{(MIQP)} \qquad \min \left\{ \ \textstyle\sum_{i\in\mathcal{N}} w_i\big(\ (z_i^+)^2 + (z_i^-)^2\ \big) \ : \ (12) \right\}$$

as the *original* MIQP formulation of CTA, to which then directly apply steps 2. and 3. above, thus obtaining

$$\text{(PR)} \ \min\left\{ \textstyle\sum_{i\in\mathcal{U}} w_i\big((z_i^+)^2 + (z_i^-)^2\big) + \sum_{i\in\mathcal{S}} w_i\big((z_i^+)^2/y_i + (z_i^-)^2/(1-y_i)\big) : (12) \right\}.$$

Note how (MIQP) have already improved the lower bound: for our example of the previous paragraph (with $w_1 = 1$), $z_1^+ = z_1^- = 5$ and $y_1 = 1/2$, (MIQP) gives a bound of 50 instead of 0. Yet, (PR) is even better: for the same solution it gives a bound of 100, which (as expected) is the optimal solution to the problem.

Then by applying standard SOCP and SI reformulation tricks to (PR), i.e., formulae (3) and (4), to express the objective function in terms of one conic constraint/infinitely many linear constraints respectively, we obtain two reformulations of CTA that we denote as (SOCP) and (P/C), respectively.

Conversely, applying the projection approach of Frangioni et al. (2011) following the same guidelines is not possible. The reason is that the main condition required for that to work is that *the binary variable corresponding to one semicontinuous variable only appear in the corresponding constraints (9) and nowhere else*, or, in other words, that there are no constraints directly linking the binary variables to one another. This is clearly *not* the case here, as the constraint $y^+ + y^- = 1$ is crucial.

In order to extend the projection approach of Frangioni et al. (2011) to CTA we then have to explicitly carry out the analysis for our case. This is done by considering the (clearly, convex) function

$$g(z^+, z^-) = \min_y \; \{ \; \overline{co} f(\; z^+ \; , \; z^- \; , \; y \;) \; : \; y \in [0, 1] \; \} \tag{13}$$

and carrying out a case-by-case analysis of its shape. This is significantly more complex and rather tedious, so the details are best relegated to the Appendix. These can be summarized by the following Theorem.

Theorem 1 *The function $g(z^+, z^-)$ is piecewise-conic-quadratic with at most three pieces. If cell i is* reasonably balanced*, i.e.,* $\max\{ \; l_i \; , \; u_i \; \} < \min\{ \; \bar{u}_i \; , \; -\bar{l}_i \; \}$*, then $g(z^+, z^-)$ has exactly three pieces, the "central" one of which is*

$$(z_i^+ + z_i^-)^2 \tag{14}$$

that is also the lower approximation *to $g(z^+, z^-)$ corresponding to the relaxation of the bounds constraints. If, furthermore, cell i is* totally symmetric*, i.e., $\bar{u}_i = -\bar{l}_i$ and $l_i = u_i$, then (14) actually* coincides *with $g(z^+, z^-)$.*

It would be then possible to derive a projected model analogous to these of Frangioni et al. (2011) for CTA, but the prospects of doing so are not particularly encouraging. First of all, the corresponding model would be a SOCP with up to three conic constraints for each sensitive cells; the standard formulation (SOCP), which already has only two of them, is typically not competitive with (P/C) (Frangioni and Gentile 2009), a fact that we directly verified to be true for CTA also. Furthermore, the rationale of Frangioni et al. (2011) is to exploit structural properties in the original problem, which are notably absent here since the matrix A lacks exploitable characteristics for general tabular data.

Yet, the analysis readily suggests a workable alternative: use the model

$$\text{(MIQP+)} \qquad \min \left\{ \sum_{i \in \mathcal{N}} w_i \left(z_i^+ + z_i^- \right)^2 \ : \ (12) \right\}$$

instead of (MIQP), (SOCP) or (P/C). This is possible since (14) is a lower approximation to (13); furthermore, the two objective function clearly coincide on integer solutions. The model is clearly stronger than (MIQP). The (MIQP+) model is somewhat simpler than (SOCP), not requiring conic constraints; however, it has a nonseparable (albeit only slightly so) objective function. It is also more compact than (P/C), which however is a separable quadratic model.

Note that, as in the previous case, there is no need to distinguish between sensitive and non-sensitive cells: the reformulation of the objective function can be applied to either, and this actually has—as it can be expected—positive results. Indeed, *since non-sensitive cells are equivalent to totally symmetric sensitive ones*, as previously seen the analysis suggests to rather consider

$$\text{(PR+)} \quad \min \left\{ \sum_{i \in \mathcal{U}} w_i (z_i^+ + z_i^-)^2 + \sum_{i \in \mathcal{S}} w_i \big((z_i^+)^2 / y_i + (z_i^-)^2 / (1 - y_i) \big) : (12) \right\}$$

as the "starting" Perspective Relaxation. Thus, other than (MIQP), (SOCP), (P/C) and (MIQP+), there are two further possible models: (SOCP+) and (P/C+), obtained from (PR+) exactly as (SOCP) and (P/C), respectively, are obtained from (MIQP). Compared to (SOCP) and (P/C), these new models have (slightly) nonseparable objective function but may provide better results. The relative strengths and weaknesses of these six models can only be gauged computationally, which is done in the next paragraph.

4 Computational Tests

We performed a large computational experience to compare the six models (MIQP), (P/C), (SOCP), (MIQP+), (P/C+), and (SOCP+). All tests have been done by using CPLEX 12.1 solver and setting `CPX_PARAM_NUMERICALEMPHASIS` to 1. (SOCP) and (SOCP+) have been tested but were regularly worse and therefore not reported.

instance	cells	sens. cells	tab. cons	vars.	cons.	%purebinvar
10-20-3	2877	81	452	5835	777	1.39
10-20-5	3163	150	466	6475	1064	2.31
10-20-10	2772	262	447	5806	1495	4.51
10-30-3	4569	131	612	9270	1137	1.42
10-30-5	4185	201	600	8571	1403	2.34
10-30-10	4706	452	617	9864	2426	4.59
20-20-3	6607	188	630	13401	1381	1.40
20-20-5	6426	305	621	13157	1841	2.32
20-20-10	6212	590	611	13013	2969	4.53
20-30-3	9145	264	760	18554	1816	1.42
20-30-5	8947	431	754	18324	2478	2.35
20-30-10	9164	884	761	19211	4296	4.60

Table 1: size and properties of symmetric instances

4.1 Test instances

We generated a number of instances of type 1H2D by using a generator developed at UPC and validated by an NSA during a European project involving UPC. We considered two types of instances: symmetrical instances and asymmetrical instances. The former ones have the property that $u_i = l_i$, but in general $\bar{u}_i \neq -\bar{l}_i$, because in many cases we have to ensure nonnegativity of the perturbed values. Asymmetrical instances have also $u_i \neq l_i$. For each parameter set we generated 5 instances.

In tables 1 and 2 we specify the size properties of the instances that have been used: the number of cells, the number of sensitive cells, the number of tabular constraints, the number of variables and constraints in the resulting (MIQP) or (MIQP+) models, and % of pure binary variables (that are in one-to-one correspondence with sensitive cells). These data are the average over the 5 instances for each combination of the generator parameters . Note that P/C and SOCP models present more variables and constraints due to the reformulation tricks (3) and (4).

4.2 Computational Results

In tables 3 and 4 we present the results obtained by considering models (MIQP+), (P/C+), (MIQP), (P/C). We used a time limit of 10000 seconds. For each algorithm the first column of the table is the gap at which the instances have been solved

$$gap = (UB - LB)/LB,$$

where UB and LB are the values of the feasible solution and lower bound provided the algorithm, relatively; the second column is the primal gap (pgap) that is obtaind by the following formula:

$$pgap = \frac{UB - bestLB}{bestLB},$$

instance	cells	sens. cells	tab. cons	vars.	cons.	%purebinvar
10-20-3-2	2877	81	452	5835	777	1.39
10-20-3-5	3163	89	466	6414	822	1.39
10-20-3-10	2919	82	454	5920	784	1.39
10-20-5-2	3095	146	462	6337	1048	2.31
10-20-5-5	2835	134	450	5804	986	2.31
10-20-5-10	3188	151	467	6526	1070	2.31
10-20-10-2	3230	306	469	6765	1691	4.52
10-20-10-5	3146	298	465	6589	1655	4.52
10-20-10-10	3024	286	459	6334	1603	4.52
10-30-3-2	4476	129	609	9081	1124	1.42
10-30-3-5	4383	126	606	8893	1110	1.41
10-30-3-10	4452	128	609	9031	1121	1.42
10-30-5-2	4439	213	608	9091	1460	2.34
10-30-5-5	4427	212	608	9066	1457	2.34
10-30-5-10	3999	192	594	8190	1360	2.34
10-30-10-2	4334	416	605	9084	2270	4.58
10-30-10-5	4204	404	601	8811	2216	4.58
10-30-10-10	4545	437	612	9526	2359	4.59
20-20-3-2	5985	170	600	12140	1280	1.40
20-20-3-5	6556	186	627	13299	1372	1.40
20-20-3-10	6737	192	636	13665	1402	1.40
20-20-5-2	5905	280	596	12091	1717	2.32
20-20-5-5	6573	312	628	13458	1876	2.32
20-20-5-10	6409	304	620	13123	1837	2.32
20-20-10-2	6082	577	605	12740	2913	4.53
20-20-10-5	6094	578	605	12767	2919	4.53
20-20-10-10	6577	624	628	13779	3126	4.53
20-30-3-2	8804	254	749	17862	1767	1.42
20-30-3-5	9219	266	762	18705	1828	1.42
20-30-3-10	9176	265	761	18617	1822	1.42
20-30-5-2	9126	440	759	18693	2519	2.35
20-30-5-5	8661	417	744	17740	2414	2.35
20-30-5-10	8996	434	755	18426	2490	2.35
20-30-10-2	9170	884	761	19224	4298	4.60
20-30-10-5	9151	883	760	19185	4291	4.60
20-30-10-10	9033	871	756	18938	4241	4.60

Table 2: size and properties of asymmetric instances

	MIQP+				P/C+				MIQP				P/C			
instance	gap	pgap	time	nodes	gap	pgap	time	n odes	gap	pgap	time	nodes	gap	pgap	time	nodes
10-20-3	0.01	0.00	442	474	0.00	0.00	486	357	6.49	0.01	9686	10365	0.00	0.00	1331	1973
10-20-5	0.01	0.00	765	690	0.01	0.00	1016	611	67.62	0.05	10000	2649	0.16	0.00	6695	8675
10-20-10	0.01	0.01	3852	10507	2.21	0.07	7660	2676	72.75	0.14	10000	5536	12.39	0.14	10000	3230
10-30-3	0.01	0.00	1470	760	0.01	0.00	1749	457	127.03	0.02	10000	778	0.98	0.01	9070	3022
10-30-5	0.01	0.01	4850	4003	0.07	0.01	7102	4769	118.53	0.12	10000	1422	15.80	0.03	10000	1853
10-30-10	2.44	2.44	10000	3512	8.26	2.53	10000	889	128.67	2.62	10000	1619	35.30	2.54	10000	643
20-20-3	0.00	0.00	1710	260	0.00	0.00	1874	291	158.64	0.01	10000	636	17.84	0.04	8559	596
20-20-5	0.01	0.01	3543	1507	1.27	0.01	7237	1185	138.59	0.12	10000	625	12.33	0.01	8808	481
20-20-10	7.10	7.10	10000	1968	24.51	7.21	10000	504	142.82	7.60	10000	777	38.22	7.39	10000	262
20-30-3	0.40	0.40	6113	738	3.60	0.41	6800	458	138.85	0.47	10000	726	27.17	0.45	10000	379
20-30-5	7.39	7.39	8791	751	15.19	7.46	8885	379	156.73	9.37	10000	801	32.83	8.02	10000	406
20-30-10	19.92	19.92	10000	674	32.04	21.13	10000	102	153.79	23.08	10000	496	44.06	21.20	10000	56

Table 3: Results for symmetric instances and MIQP+, P/C+, MIQP, P/C

where bestLB is the best lower obtained obtained with one of the four models The other two columns are the time in seconds and the number of nodes of the B&C tree. All data are averaged over the 5 instances associated with the same generator parameters.

We recall that from our theoretical results we derived that (MIQP+) and (P/C+) provide the same lower bound on fully symmetrical instances. Although this is not the case as data must be nonnegative, (MIQP+) and (P/C+) present similar behaviour on the lower bound results because upper bound constraints are seldom active. From the computational tests we can see that (MIQP+) generally performs better than (P/C+) on symmetrical instances. Indeed (MIQP+) solves the instances in fewer seconds and when it does not succeed to obtain the optimal solution within 10000 seconds it provides a solution with a better gap. This can be explained as (MIQP+) does not require constraint generation to compute the Perspective Reformulation bound. On the same instances (P/C) and (MIQP) perform worse than the other two models. They sport theoretically worse lower bounds and in particular (MIQP) is much worse than the other three models.

When considering asymmetrical instances we have two different cases. If the asymmetric parameter is small we have a behaviour similar to the one observed for symmetrical instances. If the asymmetric parameter is large (see instances 10_30_10_5, 10_30_10_10, 20_30_10_5, and 20_30_10_10) the (MIQP+) model presents worse results than (P/C+) model. This respects the property that the (P/C+) model provides better lower bounds than (MIQP+). Moreover, also the separability of the objective function gives a contribution to the performance. Indeed, (P/C) is even better than (P/C+) on these very asymmetrical instances.

	MIQP+				P/C+				MIQP				P/C			
instance	gap	pgap	time	nodes	gap	pgap	time	nodes	gap	pgap	time	nodes	gap	pgap	time	nodes
10-20-3-2	0.00	0.00	23	9	0.00	0.00	58	1	0.01	0.00	1218	7823	0.00	0.00	106	17
10-20-3-5	0.01	0.00	19	1	0.00	0.00	82	1	0.01	0.00	322	197	0.00	0.00	111	1
10-20-3-10	0.00	0.00	15	7	0.00	0.00	55	1	0.01	0.00	270	124	0.00	0.00	78	1
10-20-5-2	0.01	0.00	58	30	0.00	0.00	119	9	0.04	0.00	10000	113601	0.00	0.00	152	32
10-20-5-5	0.01	0.00	21	15	0.00	0.00	79	1	0.01	0.00	1293	2332	0.00	0.00	81	1
10-20-5-10	0.00	0.00	20	2	0.00	0.00	106	1	0.01	0.00	1483	660	0.00	0.00	111	1
10-20-10-2	0.01	0.00	438	556	0.00	0.00	637	181	0.04	0.00	10000	67541	1.49	0.00	2904	370
10-20-10-5	0.01	0.00	4315	31344	0.00	0.00	142	1	0.08	0.00	10000	102641	0.00	0.00	142	1
10-20-10-10	0.01	0.00	416	2135	0.00	0.00	120	1	0.04	0.00	5044	26508	0.00	0.00	109	1
10-30-3-2	0.00	0.00	115	28	0.00	0.00	271	5	0.02	0.00	10000	55266	0.00	0.00	391	35
10-30-3-5	0.00	0.00	40	4	0.00	0.00	220	1	0.01	0.00	2447	1333	0.00	0.00	237	1
10-30-3-10	0.00	0.00	31	1	0.00	0.00	232	1	0.01	0.00	1468	565	0.00	0.00	258	1
10-30-5-2	0.00	0.00	193	103	0.00	0.00	377	19	0.05	0.00	10000	28721	0.00	0.00	455	72
10-30-5-5	0.01	0.00	119	39	0.00	0.00	333	1	0.01	0.00	4055	24181	0.00	0.00	258	1
10-30-5-10	0.01	0.00	63	46	0.00	0.00	207	1	0.01	0.00	1855	1104	0.00	0.00	216	1
10-30-10-2	0.01	0.00	1158	1035	0.00	0.00	1905	230	7.03	0.00	10000	27461	0.82	0.00	3066	986
10-30-10-5	0.01	0.00	6489	38818	0.00	0.00	401	1	8.53	0.00	10000	60347	0.00	0.00	311	1
10-30-10-10	0.01	0.00	4806	22519	0.00	0.00	522	1	0.09	0.00	10000	52141	0.00	0.00	372	1
20-20-3-2	0.00	0.00	136	25	0.00	0.00	393	1	0.03	0.00	10000	13721	0.00	0.00	502	9
20-20-3-5	0.01	0.00	72	1	0.00	0.00	625	1	0.01	0.00	4074	1207	0.00	0.00	691	1
20-20-3-10	0.00	0.00	76	1	0.00	0.00	574	1	2.18	0.00	5356	465	0.00	0.00	644	1
20-20-5-2	0.00	0.00	257	47	0.00	0.00	601	4	1.40	0.00	10000	14362	0.00	0.00	598	24
20-20-5-5	0.01	0.00	117	10	0.00	0.00	690	1	1.19	0.00	10000	15635	0.00	0.00	638	1
20-20-5-10	0.01	0.00	128	54	0.00	0.00	736	1	0.52	0.00	6434	2076	0.00	0.00	623	1
20-20-10-2	0.01	0.00	1448	212	0.00	0.00	2802	138	63.41	0.04	10000	1006	0.00	0.00	2525	228
20-20-10-5	0.02	0.00	9203	22462	0.00	0.00	943	1	3.40	0.00	10000	9950	0.00	0.00	634	1
20-20-10-10	0.03	0.00	7910	19421	0.00	0.00	1327	1	7.33	0.00	10000	9801	0.00	0.00	801	1
20-30-3-2	0.01	0.00	439	28	0.00	0.00	1477	1	13.94	0.00	10000	1203	0.00	0.00	1649	16
20-30-3-5	0.01	0.00	140	1	0.00	0.00	1597	1	5.39	0.00	8400	1767	0.00	0.00	1510	1
20-30-3-10	0.00	0.00	157	8	0.00	-0.00	1601	1	8.34	0.00	9321	691	0.00	-0.00	1547	1
20-30-5-2	0.00	0.00	777	65	0.00	0.00	2160	17	48.34	0.01	10000	612	0.00	0.00	2111	34
20-30-5-5	0.01	0.00	618	462	0.00	0.00	1800	1	19.74	0.01	10000	1692	0.00	0.00	1622	1
20-30-5-10	0.01	0.00	622	243	0.00	0.00	1988	1	2.14	0.00	9815	2623	0.00	0.00	1625	1
20-30-10-2	1.23	1.23	7575	1454	3.67	1.24	8407	297	79.80	1.39	10000	422	4.16	1.23	7705	262
20-30-10-5	0.52	0.00	10000	12890	0.00	0.00	2784	1	36.91	0.03	10000	718	0.00	0.00	1915	1
20-30-10-10	0.04	0.00	10000	17526	0.00	0.00	2619	1	27.08	0.03	10000	1441	0.00	0.00	1817	1

Table 4: Results for asymmetric instances and MIQP+, P/C+, MIQP, P/C

5 Conclusions and future research

During my visit in Barcelona, I also found that the structure of the problem may be used to find new valid inequalities for CTA problems with both L_1 and L_2 distances. This could be the topic of a future collaboration.

Appendix: Proof of Theorem 1.

As in Section 3 we will concentrate on a fixed cell $i \in \mathcal{S}$ and therefore drop the index i. Also, in the development we assume w.l.o.g. $w = 1$, because it is a multiplicative factor which just goes untouched through the derivation. It is easy to see that the constraint

$$\min\{\ l\ ,\ u\ \} \leq z^+ + z^- \leq \max\{\ \bar{u}\ ,\ -\bar{l}\ \} \tag{15}$$

is implied by (9): in all integral solutions one has either $z^+ \leq \bar{u}$ and $z^- = 0$, or $z^- \leq -\bar{l}$ and $z^+ = 0$, and, analogously, either $z^+ \geq u$ and $z^- = 0$ or $z^- \geq l$ and $z^+ = 0$. Therefore, we can consider (15) as explicitly added to the formulation if we need it. Furthermore, the constraints $0 \leq z^+ \leq \bar{u}$ and $0 \leq z^- \leq -\bar{l}$ are always valid.

From (9) we immediately obtain

$$\begin{aligned} 0 \leq z^+/\bar{u} \quad &\leq y \leq \quad z^+/u \\ (l - z^-)/l \quad &\leq y \leq \quad (z^- + \bar{l})/\bar{l} \leq 1 \end{aligned}$$

which yields

$$\delta(z^+, z^-) = \max\left\{\frac{z^+}{\bar{u}}, 1 - \frac{z^-}{l}\right\} \leq y \leq \min\left\{\frac{z^+}{u}, 1 + \frac{z^-}{\bar{l}}\right\} = \Delta(z^+, z^-)\ . \tag{16}$$

We now want to develop a closed-form formula for the optimal solution $y(z^+, z^-)$ of (13). We therefore need to find the value of y such that

$$\frac{\partial h(z^+, z^-, y)}{\partial y} = -\frac{(z^+)^2}{y^2} + \frac{(z^-)^2}{(1-y)^2} = 0$$

which leads to

$$\begin{aligned} (1-y)^2(z^+)^2 = y^2(z^-)^2 \quad &\Leftrightarrow \quad (1 - 2y + y^2)(z^+)^2 = y^2(z^-)^2 \\ y^2((z^+)^2 - (z^-)^2) - 2y(z^+)^2 + (z^+)^2 = 0 \quad &\Leftrightarrow \quad y = z^+/(z^+ + z^-) = \tilde{y} \end{aligned}$$

as $y \geq 0$, $z^+ \geq 0$ and $z^- \geq 0$. In fact, the other root of the quadratic equation, $z^+/(z^+ - z^-)$, coincides with $\tilde{y}$ when $z^- = 0$, is > 1 when $z^+ >$

$z^- > 0$, is indefinite when $z^+ = z^-$ and is < 0 when $z^- > z^+$, and therefore is never relevant. Moreover, the second derivative

$$\frac{\partial^2 h(z^+, z^-, y)}{\partial y^2} = 2\frac{(z^+)^2}{y^3} + 2\frac{(z^-)^2}{(1-y)^3}$$

is greater then zero in $y = \tilde{y}$ when $0 < \tilde{y} < 1$. Me must now distinguish three cases:

1) $\tilde{y} \leq \delta(z^+, z^-) \quad \Rightarrow \quad y(z^+, z^-) = \delta(z^+, z^-)$;

2) $\delta(z^+, z^-) \leq \tilde{y} \leq \Delta(z^+, z^-) \quad \Rightarrow \quad y(z^+, z^-) = \tilde{y}$;

3) $\Delta(z^+, z^-) \leq \tilde{y} \quad \Rightarrow \quad y(z^+, z^-) = \Delta(z^+, z^-)$.

For case 2), plugging $y = \tilde{y} = z^+/(z^+ + z^-)$ into (9) gives

$$u \leq z^+ + z^- \leq \bar{u} \quad \text{and} \quad l \leq z^+ + z^- \leq -\bar{l} \ . \tag{17}$$

Therefore, under these conditions, the optimal objective function value $f^*(z^+, z^-) = f(z^+, z^-, \tilde{y})$ takes the particularly simple form

$$f^*(\ z^+\ ,\ z^-\ ) = f(\ z^+\ ,\ z^-\ ,\ z^+/(z^+ + z^-)\ ) = (z^+ + z^-)^2 \ ,$$

i.e., (14). Hence, in the totally symmetric case $\bar{u} = -\bar{l}$, $l = u$ one has $\max\{\ \bar{u}\ ,\ -\bar{l}\ \} = \min\{\ \bar{u}\ ,\ -\bar{l}\ \}$ and $\max\{\ u\ ,\ l\ \} = \min\{\ u\ ,\ l\ \}$, only case 2) can happen: $g(z^+, z^-) = f^*(z^+, z^-)$. Note that, as claimed in the Theorem, (14) $\equiv f^*(z^+, z^-) \leq g(z^+, z^-)$ as it corresponds to *unconstrained* minimization over y.

With non-symmetric data, cases 1) and 3) has to be taken into account. The analysis has to be divided into several sub-cases.

1) $\tilde{y} \leq \delta(z^+, z^-)$. Because $\delta(z^+, z^-) = \max\{z^+/\bar{u}, 1 - z^-/l\}$, two sub-cases have to be separately considered:

   1.1) $z^+/\bar{u} \geq 1 - z^-/l$ and $\tilde{y} \leq z^+/\bar{u}$; by simple algebraic manipulations, these two conditions boil down to

$$lz^+ + \bar{u}z^- \geq \bar{u}l \tag{18}$$

$$z^+ + z^- \geq \bar{u} \tag{19}$$

   By rewriting (18) in the equivalent form

$$z^+ + z^-(\bar{u}/l) \geq \bar{u}$$

it is immediately evident that one among (18) and (19) is redundant when the other is imposed; this depends on which of the two conditions

$$\bar{u} \leq l \qquad (20)$$
$$l \leq \bar{u} \qquad (21)$$

holds. In particular,

* $*$ (20) $\Rightarrow$ (18) dominates (19);
* $*$ (21) $\Rightarrow$ (19) dominates (18).

In either case we have $y(z^+, z^-) = z^+/\bar{u}$, which finally leads to

$$f^*(\ z^+\ ,\ z^-\ ) = f(\ z^+\ ,\ z^-\ ,\ z^+/\bar{u}\ ) = \bar{u}((z^-)^2/(\bar{u} - z^+) + z^+)\ . \qquad (22)$$

Note that the objective function value is always positive, as $z^+ \leq \bar{u}$.

1.2) $z^+/\bar{u} \leq 1 - z^-/l$ and $\tilde{y} \leq 1 - z^-/l$; this gives

$$lz^+ + \bar{u}z^- \leq \bar{u}l \qquad (23)$$
$$z^+ + z^- \leq l \qquad (24)$$

Again, by rewriting (23) in the equivalent form

$$z^+(l/\bar{u}) + z^- \leq l$$

we see that one of these is redundant when the other is imposed, depending on *the same* conditions (20)/(21); that is,

* $*$ (20) $\Rightarrow$ (23) dominates (24);
* $*$ (21) $\Rightarrow$ (24) dominates (23).

In either case we have $y(z^+, z^-) = 1 - z^-/l$, which finally leads to

$$f^*(\ z^+\ ,\ z^-\ ) = f(\ z^+\ ,\ z^-\ ,\ 1 - z^-/l\ ) = l((z^+)^2/(l - z^-) + z^-)\ . \qquad (25)$$

Note that the objective function value is always positive, as $z^- \leq z^+ + z^- \leq l$.

3) $\Delta(z^+, z^-) \leq \tilde{y}$. Because $\Delta(z^+, z^-) = \min\{z^+/u, 1 + z^-/\bar{l}\}$, again this can happen in two different ways:

3.1) $z^+/u \leq 1 + z^-/\bar{l}$ and $\tilde{y} \geq z^+/u$; this is equivalent to

$$\begin{aligned} -\bar{l}z^+ + uz^- &\leq -\bar{l}u & (26) \\ z^+ + z^- &\leq u & (27) \end{aligned}$$

where as usual (26) can be rewritten as $z^+ + z^-(u/ - \bar{l}) \leq u$. Thus, according to which among

$$\begin{aligned} -\bar{l} &\leq u & (28) \\ u &\leq -\bar{l} & (29) \end{aligned}$$

holds, one of the constraints is useless; indeed,

* (28) $\Rightarrow$ (26) dominates (27);
* (29) $\Rightarrow$ (27) dominates (26).

In either case we have $y(z^+, z^-) = z^+/u$, which finally leads to

$$f^*(\ z^+\ ,\ z^-\ ) = f(\ z^+\ ,\ z^-\ ,\ z^+/u\ ) = u((z^-)^2/(u - z^+) + z^+)\ . \quad (30)$$

Note that the objective function value is always positive, as $z^+ \leq z^+ + z^- \leq u$.

3.2) $z^+/u \geq 1 + z^-/\bar{l}$ and $\tilde{y} \geq 1 + z^-/\bar{l}$; one has

$$\begin{aligned} -\bar{l}z^+ + uz^- &\geq -\bar{l}u & (31) \\ z^+ + z^- &\geq -\bar{l} & (32) \end{aligned}$$

According to which among (28)/(29) holds, one of the above (considering that (31) can be rewritten as $z^+(-\bar{l}/u) + z^- \geq -\bar{l}$) is irrelevant; that is,

* (28) $\Rightarrow$ (31) dominates (32);
* (29) $\Rightarrow$ (32) dominates (31).

In either case we have $y(z^+, z^-) = 1 + z^-/\bar{l}$, which finally leads to

$$f^*(\ z^+\ ,\ z^-\ ) = f(\ z^+\ ,\ z^-\ ,\ 1{+}z^-/\bar{l}\ ) = (-\bar{l})((z^+)^2/(-\bar{l}{-}z^-){+}z^-)\ . \quad (33)$$

Again, the objective function value is always positive, as $z^- \leq -\bar{l}$.

From the above discussion we conclude, remembering that $0 \leq u \leq \bar{u}$ and $0 \leq l \leq -\bar{l}$, that the (z^+, z^-) space can be partitioned into several subsets, in each of which the objective function is uniquely determined. Again this requires a case-by-case discussion:

- If $\bar{u} \leq l$ (cf. (20)), then $\max\{l, u\} = l \geq \min\{\bar{u}, -\bar{l}\} = \bar{u}$; therefore, case 2) is not significant. Because (18) dominates (19) and (23) dominates (24), we have that for all $u \leq z^+ + z^- \leq -\bar{l}$

$$g(z^+, z^-) = \begin{cases} \bar{u}((z^-)^2/(\bar{u} - z^+) + z^+) & \text{if } lz^+ + \bar{u}z^- \geq \bar{u}l \\ l((z^+)^2/(l - z^-) + z^-) & \text{if } lz^+ + \bar{u}z^- \leq \bar{u}l \end{cases} .$$

- Analogously, if $-\bar{l} \leq u$ (cf. (28)), then $\max\{l, u\} = u \geq \min\{\bar{u}, -\bar{l}\} = -\bar{l}$; therefore, case 2) does not happen. Because (26) dominates (27) and (31) dominates (32), we have that for all $l \leq z^+ + z^- \leq \bar{u}$

$$g(z^+, z^-) = \begin{cases} u((z^-)^2/(u - z^+) + z^+) & \text{if } -\bar{l}z^+ + uz^- \leq -\bar{l}u \\ (-\bar{l})((z^+)^2/(-\bar{l} - z^-) + z^-) & \text{if } -\bar{l}z^+ + uz^- \geq -\bar{l}u \end{cases} .$$

If none of the above two "extreme" cases occur, then the "simple" inequalities (19), (24), (27) and (32) all dominate their "complex" companions (18), (23), (26) and (31), respectively. We can thus continue the discussion listing all other possible ways in which l, u, $-\bar{l}$ and $\bar{u}$ can be arranged along the line:

- If $l \leq u \leq \bar{u} \leq -\bar{l}$, then $\max\{l, u\} = u$ and $\min\{\bar{u}, -\bar{l}\} = \bar{u}$. Thus,

$$g(z^+, z^-) = \begin{cases} u((z^-)^2/(u - z^+) + z^+) & \text{if } l \leq z^+ + z^- \leq u \\ (z^+ + z^-)^2 & \text{if } u \leq z^+ + z^- \leq \bar{u} \\ \bar{u}((z^-)^2/(\bar{u} - z^+) + z^+) & \text{if } \bar{u} \leq z^+ + z^- \leq -\bar{l} \end{cases}$$

- If $l \leq u \leq -\bar{l} \leq \bar{u}$, then $\max\{l, u\} = u$ and $\min\{\bar{u}, -\bar{l}\} = -\bar{l}$. Thus,

$$g(z^+, z^-) = \begin{cases} u((z^-)^2/(u - z^+) + z^+) & \text{if } l \leq z^+ + z^- \leq u \\ (z^+ + z^-)^2 & \text{if } u \leq z^+ + z^- \leq -\bar{l} \\ (-\bar{l})((z^+)^2/(-\bar{l} - z^-) + z^-) & \text{if } -\bar{l} \leq z^+ + z^- \leq -\bar{u} \end{cases}$$

- If $u \leq l \leq -\bar{l} \leq \bar{u}$, then $\max\{l, u\} = l$ and $\min\{\bar{u}, -\bar{l}\} = -\bar{l}$. Thus,

$$g(z^+, z^-) = \begin{cases} l((z^+)^2/(l - z^-) + z^-) & \text{if } u \leq z^+ + z^- \leq l \\ (z^+ + z^-)^2 & \text{if } l \leq z^+ + z^- \leq -\bar{l} \\ (-\bar{l})((z^+)^2/(-\bar{l} - z^-) + z^-) & \text{if } -\bar{l} \leq z^+ + z^- \leq -\bar{u} \end{cases}$$

- If $u \leq l \leq \bar{u} \leq \bar{l}$, then $\max\{l, u\} = l$ and $\min\{\bar{u}, -\bar{l}\} = \bar{u}$. Thus,

$$g(z^+, z^-) = \begin{cases} l((z^+)^2/(l - z^-) + z^-) & \text{if } u \leq z^+ + z^- \leq l \\ (z^+ + z^-)^2 & \text{if } l \leq z^+ + z^- \leq \bar{u} \\ \bar{u}((z^-)^2/(\bar{u} - z^+) + z^+) & \text{if } \bar{u} \leq z^+ + z^- \leq -\bar{l} \end{cases}$$

Thus, we have a total of 6 possible cases; in 4 of them the function has three pieces, two conic ones and a quadratic one, while in the remaining 2 the function has two pieces, all of them being conic. We have therefore completed the proof of Theorem 1.

References

Aktürk, S., A. Atamtürk, S. Gürel. 2009. A strong conic quadratic reformulation for machine-job assignment with controllable processing times. *Operations Research Letters* **37**(3) 187–191. pages 5, 6

Bacharach, M. 1966. Matrix rounding problems. *Management Sci.*, **9** 732–742. pages 2

Castro, J. 2006. Minimum-distance controlled perturbation methods for large-scale tabular data protection. *Eur. J. Oper. Res.* **171** 39–52. pages 4

Castro, J. 2007. A shortest paths heuristic for statistical disclosure control in positive tables. *INFORMS J.Comput.* **19** 520–533. pages 3

Castro, J., J. Cuesta. 2010. Quadratic regularizations in an interior-point method for primal block-angular problems. *Math. Prog.* doi:10.1007/s10107-010-0341-2, in press pages 4

Castro, J. 2011. Recent advances in optimization techniques for statistical tabular data protection. *Eur. J. Oper. Res.* doi:10.1016/j.ejor.2011.03.050, in press. pages 2

Ceria, S., J. Soares. 1999. Convex programming for disjunctive convex optimization. *Mathematical Programming* **86** 595–614. pages 7

Dandekar, R.A., L.H. Cox. 2002. Synthetic tabular Data: an alternative to complementary cell suppression. Manuscript, Energy Information Administration, US. pages 4

Domingo-Ferrer J., V. Torra. 2002. A Critique of the Sensitivity Rules Usually Employed for Statistical Table Protection. *Internat. J. of Uncertainty Fuzziness and Knowledge-Based Systems*, **10(5)** 545–556. pages 2

Fischetti, M., J.J. Salazar-González. 2001. Solving the cell suppression problem on tabular data with linear constraints. *Management Sci.* **47** 1008–1026. pages 3

Frangioni, A., C. Gentile. 2006. Perspective Cuts for 0-1 Mixed Integer Programs. *Math. Prog.* **106(2)** 225–236. pages 5, 6, 7

Frangioni, A., C. Gentile. 2007. SDP Diagonalizations and Perspective Cuts for a Class of Nonseparable MIQP. *Oper. Res. Lett.* **35(2)** 181–185. pages 6, 9

Frangioni, A., C. Gentile. 2009. A Computational Comparison of Reformulations of the Perspective Relaxation: SOCP vs. Cutting Planes. *Operations Research Letters* **37**(3) 206–210. pages 6, 11

Frangioni A., C. Gentile, E. Grande, A. Pacifici. 2011. Projected Perspective Reformulations With Applications in Design Problems. *Operations Research*, to appear. pages 6, 11

Frangioni, A., C. Gentile, F. Lacalandra. 2009. Tighter Approximated MILP Formulations for Unit Commitment Problems. *IEEE Transactions on Power Systems* **24**(1) 105–113. pages 6

Giessing, S., A. Hundepool, J. Castro. 2009. Rounding methods for protecting EU-aggregates. *Worksession on statistical data confidentiality. Eurostat methodologies and working papers*, Eurostat-Office for Official Publications of the European Communities, Luxembourg, 255–264. pages 3

González, J.A., J. Castro. 2011. A heuristic block coordinate descent approach for controlled tabular adjustment. *Comput. Oper. Res.* **38** 1826-1835. pages 4

Grossmann, I. and S. Lee. 2003. Generalized convex disjunctive programming: Nonlinear convex hull relaxation. *Computational Optimization and Applications*, **26** 83-100. pages 7

Günlük, O., J. Linderoth. 2008. Perspective relaxation of MINLPs with indicator variables. A. Lodi, A. Panconesi, G. Rinaldi, eds., *Proceedings 13th IPCO*, *Lecture Notes in Computer Science*, vol. 5035. 1–16. pages 5, 6

Günlük, O., J. Linderoth. 2011. Perspective reformulation and applications. *IMA Volumes*, to appear. pages 7

Hijazi, H., P. Bonami, G. Cornuejols and A. Ouorou, Mixed Integer NonLinear Programs featuring "On/Off" Constraints: Convex Analysis and Applications. `http://integer.tepper.cmu.edu/webpub/ON_OFF_MINLPS.pdf`, revised July 2011. pages 7

Hundepool, A., J. Domingo-Ferrer, L. Franconi, S. Giessing, R. Lenz, J. Naylor, E. Schulte-Nordholt, G. Seri and P.P. de Wolf. 2010. *Handbook on Statistical Disclosure Control (v. 1.2)*, Network of Excellence in the European Statistical System in th field of Statistical Disclosure Control. Available on-line at `http://neon.vb.cbs.nl/casc/SDC_Handbook.pdf`. pages 2, 3

Kelly, J.P., B.L. Golden, A.A. Assad. 1992. Cell suppression: disclosure protection for sensitive tabular data. *Networks* **22** 28–55. pages 3

Luedtke, J., M. Namazifar, J.T. Linderoth. 2010. Some Results on the Strength of Relaxations of Multilinear Functions *Technical Report* **#1678**, Computer Sciences Department, University of Wisconsin-Madison. pages 9

Salazar-González, J.J. 2008. Statistical confidentiality: Optimization techniques to protect tables, *Comput. Oper. Res.* **35** 1638–1651. pages 2

Tawarmalani, M., N. Sahinidis. 2002. Convex extensions and envelopes of lower semi-continuous functions. *Math. Prog.* **93** 515–532. pages 5, 7

Willenborg, L., T. de Waal. 2000. *Lecture Notes in Statistics. Elements of Statistical Disclosure Control* 155, Springer, New York. pages 2

Zayatz, L. 2009. U.S. Census Bureau. Communication at Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Bilbao (Basque Country, Spain). pages 3